DISCUSSION PAPER SERIES

DP16797
(v. 3)

OPTIMAL TAXATION WITH MULTIPLE INCOMES AND TYPES

Kevin Spiritus, Etienne Lehmann, Sander Renes and Floris Zoutman

PUBLIC ECONOMICS

CEPR

ISSN 0265-8003

OPTIMAL TAXATION WITH MULTIPLE INCOMES AND TYPES

Kevin Spiritus, Etienne Lehmann, Sander Renes and Floris Zoutman

Discussion Paper DP16797
First Published 19 January 2022
This Revision 31 October 2022

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

OPTIMAL TAXATION WITH MULTIPLE INCOMES AND TYPES

Abstract

We analyze the optimal nonlinear income tax schedule when taxpayers earn multiple incomes and differ along many unobserved dimensions. We refine the tax perturbation and mechanism design approaches to derive the necessary conditions for the government's optimum and derive conditions under which both methods produce the same results. Our main contribution is to propose a numerical method to find the optimal tax schedule. Applied to the optimal taxation of couples, we find that optimal isotax curves are very close to linear and parallel. The slope of isotax curves is strongly affected by the relative tax-elasticity of male and female income. We make several additional contributions, including a test for Pareto efficiency and a condition on primitives that ensures the government's necessary conditions are sufficient and the solution to the problem is unique.

JEL Classification: H21, H23, H24, D82

Keywords: Nonlinear optimal taxation, Multidimensional screening, Household income taxation

Kevin Spiritus - spiritus@ese.eur.nl
Erasmus Universiteit Rotterdam

Etienne Lehmann - etienne.lehmann@gmail.com
CRED (TEPP), University Paris II Pantheon-Assas and CEPR

Sander Renes - srenes@ese.eur.nl
Delft University of Technology

Floris Zoutman - floris.zoutman@nhh.no
Norwegian School of Economics and Business Administration (NHH)

Acknowledgements
This paper merges earlier projects of Renes and Zoutman (2017) and Spiritus (2017). We thank for valuable suggestions Spencer Bastani, Felix Bierbrauer, Pierre Boyer, Robin Boadway, Katherine Cuff, André Decoster, Eva Gavrilova, Aart Gerritsen, Tom Gresik, Nathan Hendren, Yasusi Iwamoto, Bas Jacobs, Laurence Jacquet, Roman Kozlof, Jonas Loebbing, Luca Micheletto, John Morgan, Nicola Pavoni, Emmanuel Saez, Dominik Sachs, Leif Sandal, Dirk Schindler, Guttorm Schjelderup, Erik Schokkaert, Matti Tuomala, Casper de Vries, Bauke Visser, Hendrik Vrijburg and NicolasWerquin. This paper benefited from suggestions by participants at numerous conferences.

Optimal Taxation with Multiple Incomes and Types*

Kevin SPIRITUS† Etienne LEHMANN‡ Sander RENES§
Floris T. ZOUTMAN¶

October 21, 2022

Abstract

We analyze the optimal nonlinear income tax schedule when taxpayers earn multiple incomes and differ along many unobserved dimensions. We refine the tax perturbation and mechanism design approaches to derive the necessary conditions for the government's optimum and derive conditions under which both methods produce the same results. Our main contribution is to propose a numerical method to find the optimal tax schedule. Applied to the optimal taxation of couples, we find that optimal isotax curves are very close to linear and parallel. The slope of isotax curves is strongly affected by the relative tax-elasticity of male and female income. We make several additional contributions, including a test for Pareto efficiency and a condition on primitives that ensures the government's necessary conditions are sufficient and the solution to the problem is unique.

Keywords: Nonlinear Optimal Taxation, Multidimensional Screening, Household Income Taxation

JEL codes: H21, H23, H24 and D82

*This paper merges earlier projects of Renes and Zoutman (2017) and Spiritus (2017). We thank for valuable suggestions Spencer Bastani, Felix Bierbrauer, Pierre Boyer, Robin Boadway, Katherine Cuff, André Decoster, Eva Gavrilova, Aart Gerritsen, Tom Gresik, Nathan Hendren, Yasusi Iwamoto, Bas Jacobs, Laurence Jacquet, Roman Kozlof, Jonas Loebbing, Luca Micheletto, John Morgan, Nicola Pavoni, Emmanuel Saez, Dominik Sachs, Leif Sandal, Dirk Schindler, Guttorm Schjelderup, Erik Schokkaert, Matti Tuomala, Casper de Vries, Bauke Visser, Hendrik Vrijburg and Nicolas Werquin. This paper benefited from suggestions by participants at numerous conferences.

†Erasmus School of Economics, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands. Email: spiritus@ese.eur.nl. Kevin SPIRITUS acknowledges financial support of BELSPO via BRAIN.be project BR/121/A5/CRESUS.

‡CRED, University Pantheon-Assas Paris II, 12 Place du Panthéon, 75 234 Paris Cedex 05, France. Email: elehmann@u-paris2.fr.

§Delft University of Technology, Faculty of Technology Policy and management, Jaffalaan 5, 2628 BX Delft, The Netherlands. Email s.renes-1@tudelft.nl.

¶NHH Norwegian School of Economics, Department of Business and Management Science, Helleveien 30, 5045 Bergen, Norway. Email: floris.zoutman@nhh.no.

I Introduction

Optimal tax theory generally neglects the fact that households earn multiple incomes and that households differ in multiple dimensions of unobserved heterogeneity. Most papers that allow for a multidimensional tax base, either assume that taxpayers differ in one dimension only, or they impose restrictions on the tax schedule to simplify the problem.[1] While Mirrlees (1976, Section 4) does derive a general optimal tax formula in a context with multiple characteristics and multiple incomes, he offers virtually no guidance for policymakers. The aim of this paper is to investigate the properties of the optimal tax schedule when both the tax base and the unobserved heterogeneity are multidimensional.

In this paper, we show that the multidimensional optimal tax problem can be seen as consisting of two steps. For this purpose, we introduce the concept of an isotax curve, i.e. a set of income bundles that are associated with the same tax liability. The first step in solving the multidimensional optimal tax problem is then to determine the shape of these isotax curves. The second step concerns the assignment of a tax liability to each isotax curve. We show that the assignment of tax liabilities to the isotax curves satisfies a generalization of the ABC-formulas derived by Diamond (1998) and Saez (2001). This ABC-formula shows how the distributional benefit of a marginal tax increase along an isotax curve, is to be balanced against the efficiency costs of doing so. The key challenge to multi-dimensional taxation is then to understand the optimal shape of the isotax curves.

Finding the optimal shape of the isotax curves requires solving a Partial Differential Equation, which is much more challenging than solving the Ordinary Differential Equation implied by the optimal tax formula for a single tax base. To understand the difficulty, note that in the one-dimensional case one can study the effects of perturbing the marginal tax rate at one income level. The optimal marginal tax rate at that income level is then expressed as the ratio of mechanical and income effects at all incomes above, to compensated effects at the income level under consideration. In the multidimensional case, one cannot study the effects of a change in the tax gradient at one combination of incomes

[1] The one-dimensional model was pioneered by Mirrlees (1971) and further developed by Diamond (1998). Saez (2001), Scheuer and Werning (2016) and Jacquet and Lehmann (2021b) discuss to what extent the one-dimensional model extends to the case with one income but many dimensions of unobserved heterogeneity. Atkinson and Stiglitz (1976) study a multidimensional tax base with one dimension of unobserved heterogeneity.

without causing additional changes in the tax gradients at other combinations of incomes. Figure 1 on page 17 illustrates the problem. Perturbing the tax liabilities in a subset of income bundles (the shaded area in Figure 1), affects the marginal tax rates along the boundary of that subset.

To the best of our knowledge, we are the first to develop a numerical algorithm that addresses this geometric difficulty and that can solve the optimal multidimensional tax problem in its general form. We apply our algorithm to the taxation of couples. In our application we make some simplifying assumptions, similar to Kleven et al. (2006, 2007). We assume quasilinear and additively separable household preferences. Moreover, in line with the empirical literature, we assume that the labor supply of wives is more elastic (0.43) than that of husbands (0.11) (Bargain and Peichl, 2016). Finally, we non-parametrically calibrate the joint distribution of skills starting from the joint distribution of incomes in the Current Population Survey (CPS) of the US census.

We find that the optimal isotax curves are almost linear and parallel, with positive marginal tax rates for both spouses. A joint income tax that discounts female income by approximately 53 % closely approximates the fully optimized schedule in terms of social welfare. Furthermore, we investigate the desirability of negative jointness, i.e. the requirement that the optimal marginal tax rates of males decrease with female income (and vice versa). Kleven et al. (2006, 2007) show analytically that negative jointness is desirable when the productivities of both spouses are assumed uncorrelated. We numerically find that this result is not robust to a more realistic joint distribution of productivities.

We perform additional sensitivity analyses to investigate the determinants of the optimal isotax curves. In each case, we first conjecture the effect that changing some primitive has on the solution and then numerically check our prediction. Varying the government's aversion to inequality, or jointly varying the labor supply elasticity of both males and females has virtually no effect on the shape of optimal isotax curves. This only narrows or widens the gap between isotax curves depending on whether optimal marginal tax rates increase (when aversion to inequality increases) or decrease (when both elasticities increase). Conversely, only changing the labor supply elasticity of one spouse changes the slope of the isotax curves. For instance, when the female labor supply elasticity increases, the optimal marginal tax rate on female income decreases, whereas the optimal marginal tax rate on male income increases. These changes in marginal taxes shift the burden of taxation to the less elastic tax base.

Besides our numerical algorithm, we make several theoretical contributions.

First, we derive a test for Pareto efficiency. If welfare weights revealed by the optimal tax formula are negative for some income bundles, then decreasing tax liabilities at these income bundles is a self-financed Pareto improvement. We thus extend the revealed social preference approach of Bourguignon and Spadaro (2012), Bargain et al. (2014a), Bargain et al. (2014b), Jacobs et al. (2017), Bierbrauer et al. (2020) and Hendren (2020) to the multidimensional context.

Second, we use the mechanism design approach pioneered by Mirrlees (1976) to derive conditions under which the first-order conditions are sufficient to characterize the optimal allocation. This is the case when the government's Lagrangian is concave with respect to the taxpayers' utilities and to the gradient of the mapping between the taxpayers' type and utility. We analytically verify that the specification we use in our numerical exercise satisfies these sufficiency conditions. Hence, once we have obtained a numerical solution that verifies the government's necessary conditions, we can be sure that it is the unique solution. It is not necessary to conduct sensitivity analyses with respect to the initial conditions of our algorithm.

Third, we show that the tax perturbation approach and the mechanism design approach lead to the same "hybrid" optimal tax formula expressed in terms of welfare weights, behavioral elasticities and type densities, thereby ensuring that the two approaches are mutually consistent. Moreover, this hybrid formula turns out to be the most suitable to implement numerically.

Fourth, we address a concern with the tax perturbation approach that both Saez (2001) and Golosov et al. (2014) assume that incomes respond smoothly to the size of tax perturbations. We contribute by making explicit which assumptions on the tax schedule ensure smooth responses of taxpayers to tax reforms. Our assumptions rule out kinks in the tax schedule, and the existence of multiple global maxima, preventing incremental tax perturbations from causing jumps in the taxpayers' choices. Furthermore, we make explicit the underlying single-crossing assumptions that enable the derivation of the optimality conditions in the tax perturbation approach.

Fifth, we develop a new approach to derive the optimal mechanism. Mirrlees (1976, 1986) and Kleven et al. (2006, 2007) derive necessary conditions for the optimal allocation of utilities and incomes. There are many different income allocations that fulfill the necessary conditions for the optimum. For each such income allocation, the first-order incentive constraints imply the partial derivatives of the attained utilities with respect to the types. Nothing at this stage ensures that the obtained partial derivatives of the attained utilities are mutually consistent, i.e. that they imply symmetric second-order partial derivatives. Mirrlees (1976, p. 342) and Kleven et al. (2007, p. 18) acknowledge this difficulty by stating that among the different solutions of the partial differential equation, only the one that implies symmetric second-order cross derivatives should be considered. We prevent these difficulties by directly choosing the utility profile and deriving the incomes as functions of the utility profile and its partial derivatives.

Lastly, we derive optimal tax schedules when the numbers of types and incomes differ. When there are more types than incomes, the tax perturbation approach is the most natural. In that case, the same optimal tax formulas are obtained as before by averaging sufficient statics among the different taxpayers with the same income bundles. We thus extend the results obtained by Saez (2001) and Jacquet and Lehmann (2021b) to the case where taxpayers earn more than one income. When there are more incomes than types, we show that the government's problem consists of two steps. It starts with a subprogram that finds the most efficient way of distributing income choices to generate a given mapping from types to utility levels. The solution to this subprogram does not depend on the preferences of the government, but only on the resource costs of providing these utility levels. In a second step, the government selects the optimal mapping of types to utilities from the set of possible mappings. Our result helps to clarify the presence of similar subprograms that are implicitly found in the settings of Atkinson and Stiglitz (1976), Golosov et al. (2003), Golosov et al. (2007), Gerritsen et al. (2020) and Ferey et al. (2021).

Related Literature

Our paper is related to the multidimensional screening problem that was studied in the context of monopoly pricing by Armstrong (1996), Rochet and Choné (1998) and Basov (2005). Rochet and Choné (1998) show that bunching is a problem in this setting because of the interplay between the participation and the incentive constraints. Kleven et al. (2007) show that bunching is not an issue in the optimal tax problem if taxpayers do not face a participation constraint, provided that aversion to inequality is not too high. They stress that there is a wide range of redistributive preferences where bunching does not occur in the optimum.

Our paper also relates to the literature which studies multi-dimensional heterogeneity in the context where the government can only observe and tax a single income (e.g., Choné and Laroque, 2010; Rothschild and Scheuer, 2013; Rothschild and Scheuer, 2016; Lockwood and Weinzierl, 2015; Jacquet and Lehmann, 2021b; Bergstrom and Dodds, 2021). We rely on the insights in this literature in formulating our expressions in terms of sufficient statistics. Specifically, in the context of multi-dimensional heterogeneity, sufficient statistics can be strongly endogenous to the tax schedule. We use the approach of Jacquet and Lehmann (2021b) to overcome this issue by expressing our optimal-tax formulas in terms of *total elasticities* that incorporate this endogeneity. We expand on this literature by allowing for multi-dimensional incomes in addition to multi-dimensional heterogeneity.

Scheuer (2014) and Gomes et al. (2018) study a setting with multi-dimensional heterogeneity in which agents choose to earn income in one of two different sectors, and the government can tax the income of each sector according to a separate tax schedule. The main difference with our approach is that in our model agents can earn multiple incomes at the same time.

Our paper is also related to Jacquet and Lehmann (2021a), who also consider the optimal taxation of multiple incomes, additionally allowing for general equilibrium effects. However, they derive their optimal tax expressions by restricting the overall tax schedule to be the sum of separate schedules of single tax bases, a restriction that we do not impose.

Like in our application, Frankel (2014) studies the optimal taxation of couples in a setting with multi-dimensional heterogeneity and taxation of both male and female income. The main contrast between the approaches is that we allow for a continuous-type distribution, whereas Frankel (2014) studies a discrete 2×2 distribution of married couples. Cremer et al. (2001, 2003) also consider multidimensional settings. However, they only allow labor income to be taxed non-linearly, whereas taxes on commodity/capital are constrained to be linear.

Another important paper on taxation with multiple dimensions of labor through mechanism design tools is Boerma et al. (2022).[2] Their paper differs from ours on several dimensions. First, they solve the government's problem using Legendre transformations. This method requires one to restrict individual utility functions to be additively separable with isoelastic cost of effort. Conversely, while our numerical analysis assumes isoelastic cost of effort, neither our numerical algorithm nor our analytical results rely on such restrictions on individual preferences. Second, Boerma et al. (2022) introduce a participation

[2] Boerma et al. (2022) describe Golosov and Krasikov (n.d.) as closely related to the present paper. As of today, and after contacting the authors, we are unable to obtain a copy of the latter paper.

constraint in the form of an outside option. We know from Rochet and Choné (1998) that the interplay between participation and incentive constraints generates bunching in multidimensional screening. Conversely, there is no participation constraint in our optimal tax problem with only intensive margins.

The paper is organized as follows. We describe the problem of multidimensional optimal taxation in Section II. Section III is devoted to the tax perturbation approach, and Section IV is devoted to the mechanism design approach. We present our numerical algorithm and results in Section V.

II The model

II.1 Taxpayers

The economy consists of a unit mass of taxpayers who differ in a p-dimensional vector of characteristics denoted $\mathbf{w} \stackrel{\text{def}}{\equiv} (w_1, ..., w_p)$. We refer to the complete vector of characteristics of a taxpayer as her type. Types are drawn from the type space, which is denoted $\mathcal{W} \subset \mathbb{R}^p$ and is assumed to be closed and convex. Types are distributed according to a twice continuously differentiable density denoted by $f(\cdot)$, which is positive over $\mathcal{W}$.

Taxpayers make $n \geq 2$ choices. This implies the existence of n observable tax bases, $\mathbf{x} \stackrel{\text{def}}{\equiv} (x_1, ..., x_n) \in \mathbb{R}^n_+$. We call these tax bases incomes for brevity.[3] Taxpayers pay a tax $T(\mathbf{x})$ that can depend on all incomes in a nonlinear way. Taxpayers who earn incomes $\mathbf{x}$ consume after-tax income $c = \sum_{i=1}^n x_i - T(x_1, ..., x_n)$.

The preferences of taxpayers of type $\mathbf{w}$ over consumption c and income choices $\mathbf{x}$ are described by a thrice continuously differentiable utility function $\mathcal{U}(c, \mathbf{x}; \mathbf{w})$ defined over $\mathbb{R}^{n+1}_+ \times \mathcal{W}$. Taxpayers enjoy utility from consumption but endure disutility to obtain income, so $\mathcal{U}_c > 0$ and $\mathcal{U}_{x_i} < 0$. Let $\mathcal{C}(\cdot, \mathbf{x}; \mathbf{w})$ be the inverse of $\mathcal{U}(\cdot, \mathbf{x}; \mathbf{w})$. That is, a taxpayer of type $\mathbf{w}$ earning incomes $\mathbf{x}$ should consume $\mathcal{C}(u, \mathbf{x}; \mathbf{w})$ to enjoy utility level u. It follows from the implicit function theorem that $\mathcal{C}_u = 1/\mathcal{U}_c$ and $\mathcal{C}_{x_i} = -\mathcal{U}_{x_i}/\mathcal{U}_c$. We assume the utility function $\mathcal{U}(\cdot, \cdot; \mathbf{w})$ is weakly concave in $(c, \mathbf{x})$ and indifference sets defined by $c = \mathcal{C}(u, \mathbf{x}; \mathbf{w})$ are strictly convex in $(c, \mathbf{x})$ for all utility levels u and all types $\mathbf{w}$.

We assume taxpayers maximize utility subject to their budget constraints.

[3]Our model could be extended to include observable actions like private expenditures in education, which correspond to negative cash-flows for the households. This extension would not affect the validity of our results.

Therefore, a taxpayer of type $\mathbf{w}$ solves:

$$U(\mathbf{w}) \stackrel{\text{def}}{\equiv} \max_{x_1,...,x_n} \quad \mathcal{U}\left(\sum_{i=1}^{n} x_i - T(x_1,...,x_n), x_1,...,x_n; \mathbf{w}\right). \tag{1}$$

Let $\mathbf{X}(\mathbf{w}) \stackrel{\text{def}}{\equiv} (X_1(\mathbf{w}),...,X_n(\mathbf{w}))$ denote the solution to this program and let $C(\mathbf{w}) \stackrel{\text{def}}{\equiv} \sum_{i=1}^{n} X_i(\mathbf{w}) - T(\mathbf{X}(\mathbf{w}))$ denote the corresponding consumption. In addition, we denote the marginal rate of substitution between the i^{th} income and consumption as:

$$\mathcal{S}^i(c,\mathbf{x};\mathbf{w}) \stackrel{\text{def}}{\equiv} -\frac{\mathcal{U}_{x_i}(c,\mathbf{x};\mathbf{w})}{\mathcal{U}_c(c,\mathbf{x};\mathbf{w})} = \mathcal{C}_{x_i}\left(U(c,\mathbf{x};\mathbf{w}),\mathbf{x};\mathbf{w}\right) > 0. \tag{2}$$

The first-order conditions for taxpayers of type $\mathbf{w}$ are:

$$\forall j \in \{1,...,n\}: \qquad \mathcal{S}^j\left(C(\mathbf{w}),\mathbf{X}(\mathbf{w});\mathbf{w}\right) = 1 - T_{x_j}\left(\mathbf{X}(\mathbf{w})\right). \tag{3}$$

II.2 Government

The government's budget constraint is given by:

$$\mathcal{B} \stackrel{\text{def}}{\equiv} \iint_{\mathcal{W}} T(\mathbf{X}(\mathbf{w}))\, f(\mathbf{w})\mathrm{d}\mathbf{w} - E \geq 0, \tag{4}$$

where $E \geq 0$ is an exogenous amount of public expenditure. The government's objective is a social welfare function $\mathcal{O}$ which aggregates the utility of the households in the economy:

$$\mathcal{O} \stackrel{\text{def}}{\equiv} \iint_{\mathcal{W}} \Phi\left(U(\mathbf{w});\mathbf{w}\right)\, f(\mathbf{w})\mathrm{d}\mathbf{w}, \tag{5}$$

where the transformation $(u;\mathbf{w}) \mapsto \Phi(u;\mathbf{w})$ is twice continuously differentiable in $(u,\mathbf{w})$, increasing and weakly concave in u and potentially type-dependent. The government's problem consists of finding the tax function $T(\cdot)$ that maximizes the social welfare function (5) subject to revenue constraint (4), considering the households' optimization in (1).

The Lagrangian for the government's optimization problem is defined in monetary terms as:

$$\mathcal{L} \stackrel{\text{def}}{\equiv} \mathcal{B} + \frac{\mathcal{O}}{\lambda} = \iint_{\mathcal{W}} \left(T(\mathbf{X}(\mathbf{w})) + \frac{\Phi\left(U(\mathbf{w});\mathbf{w}\right)}{\lambda}\right) f(\mathbf{w})\mathrm{d}\mathbf{w} - E, \tag{6}$$

where λ is the shadow price of tax revenue and coincides with the Lagrange multiplier of the government's budget constraint at the optimum. Following Saez (2001), we define the welfare weights of taxpayers of type $\mathbf{w}$ as the social marginal utility of consumption expressed in monetary terms:

$$g(\mathbf{w}) \stackrel{\text{def}}{\equiv} \frac{\Phi_u\left(U(\mathbf{w});\mathbf{w}\right)\ \mathcal{U}_c\left(C(\mathbf{w}),\mathbf{X}(\mathbf{w});\mathbf{w}\right)}{\lambda} \geq 0. \tag{7}$$

III The Tax Perturbation Approach

III.1 Effects of tax perturbations

A necessary condition for a tax schedule to be optimal is that small perturbations of the schedule do not change social welfare. Golosov et al. (2014) are the first to systematically apply this logic to the case with multiple types and incomes. They argue that the effects of a tax perturbation on social welfare consist of mechanical effects on the government budget, effects on household utilities through the altered tax liabilities, and effects on the government budget through behavioral responses of the taxpayers. In the optimum, the sum of these effects should be zero.

In what follows, we formally introduce general perturbations to the tax schedule, and we discuss the behavioral responses to such perturbations. Contrary to earlier contributions in the literature, we do not assume from the outset that individuals respond smoothly to the perturbations. We rather reveal the underlying assumptions on the tax schedule that induce taxpayers' responses to tax perturbations to be smooth. Taking as given the assumptions that lead to smooth behavioral responses, we then provide a condition under which a given tax reform is socially desirable. This condition allows us to characterize the optimal tax schedule: if no tax reform exists that is socially desirable, then we are in the optimum.

We first formally introduce the perturbations to the tax schedule. Perturbing the tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ in the *direction* $R(\cdot)$ by *magnitude* $t \lesseqgtr 0$ leads to the perturbed tax schedule $\mathbf{x} \mapsto T(\mathbf{x}) - t\,R(\mathbf{x})$. If, for example, $R(\mathbf{x}) > 0$ and $t > 0$ or if $R(\mathbf{x}) < 0$ and $t < 0$, the perturbation decreases the tax liabilities at incomes $\mathbf{x}$. Given a tax perturbation in the direction $R(\cdot)$, the utility of taxpayers of type $\mathbf{w}$ becomes a function of magnitude t through:

$$\widetilde{U}^R(\mathbf{w},t) \stackrel{\text{def}}{\equiv} \max_{x_1,\ldots,x_n} \quad \mathcal{U}\left(\sum_{i=1}^{n} x_i - T(x_1,\ldots,x_n) + t\,R(x_1,\ldots,x_n), x_1,\ldots,x_n;\mathbf{w}\right). \tag{8}$$

By definition, we know that: $\widetilde{U}^R(\mathbf{w},0) = U(\mathbf{w})$. The first-order conditions associated with (8) are:

$$\forall j \in \{1, ..., n\}: \quad \mathcal{S}^j\left(\sum_{i=1}^{n} x_i - T(\mathbf{x}) + t\,R(\mathbf{x}), \mathbf{x}; \mathbf{w}\right) = 1 - T_{x_j}(\mathbf{x}) + t\,R_{x_j}(\mathbf{x}). \tag{9}$$

If we perturb the tax schedule or any of the characteristics of the households, then the households will update their choices $\widetilde{\mathbf{X}}^R(\mathbf{w}, t)$ such that first-order conditions (9) remain satisfied. We now introduce assumptions on the unperturbed tax schedule that allow applying the implicit function theorem to (9) in order to study these behavioral responses. If the conditions for the implicit function theorem are satisfied, then the function $\widetilde{\mathbf{X}}^R(\mathbf{w}, t)$ that solves (8) is continuously differentiable for t close to 0, i.e. the behavioral responses to tax reforms are smooth.

Assumption 1. *The tax schedule $T(\cdot)$ verifies the following assumptions:*

i) The tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ is twice continuously differentiable.

ii) For each type $\mathbf{w} \in \mathcal{W}$, the second-order conditions associated with (1) are strictly verified, i.e. the matrix $[\mathcal{S}^i_{x_j} + \mathcal{S}^j \mathcal{S}^i_c + T_{x_i x_j}]_{i,j}$ is positive definite at $c = C(\mathbf{w})$ and $\mathbf{x} = \mathbf{X}(\mathbf{w})$.[4]

iii) For each type $\mathbf{w} \in \mathcal{W}$, the function $\mathbf{x} \mapsto \mathcal{U}\left(\sum_{i=1}^{n} x_i - T(\mathbf{x}), \mathbf{x}; \mathbf{w}\right)$ admits a single global maximum.

Assumption 1.*i*) rules out kinks like those in piecewise linear tax schedules. It ensures that the first-order conditions (9) are continuously differentiable in t, $\mathbf{w}$ and $\mathbf{x}$, provided that the direction $R(\cdot)$ is twice continuously differentiable. Assumption 1.*ii*) ensures that the first-order conditions (9) are associated with a local maximum of the taxpayers' program (8). Parts *i*) and *ii*) of Assumption 1 together enable one to apply the implicit function theorem to determine how a local maximum of (8) is affected by a small tax perturbation or a small change in types. Assumption 1 *iii*) rules out the existence of multiple global maxima. This prevents an incremental tax perturbation from causing a "jump" in the taxpayers' choices from one maximum to another. At such jumps, the derivative of $\widetilde{\mathbf{X}}^R(\mathbf{w}, t)$ with respect to the size t of the perturbation tends to infinity, so the perturbation approach cannot be used.

[4] We let $[a(k)]_k$ denote a column vector whose k^{th} row is $a(k)$, $[A(k,\ell)]_{k,\ell}$ denotes a rectangular matrix whose k^{th} row and ℓ^{th} column is $A(k,\ell)$, and $\cdot$ stands for the matrix product. The transpose operator is denoted with superscript T, and the inverse operator is denoted with superscript -1.

Geometrically, Assumption 1 implies that for each type $\mathbf{w}$, the indifference set defined by $c = \mathcal{C}(U(\mathbf{w}), \mathbf{x}; \mathbf{w})$ admits a single tangency point with the budget set defined by $c = \sum_{i=1}^{n} x_i - T(\mathbf{x})$ and lies strictly above the budget set elsewhere. Given that we assume that the indifference sets defined by $c = \mathcal{C}(u, \mathbf{x}; \mathbf{w})$ are strictly convex, Assumption 1 is automatically verified if the tax schedule is linear (see Appendix A.1).

In the simulations, we characterize the optimal tax schedule under the presumption that Assumption 1 holds, and we verify ex post that this is the case. This is similar to the standard first-order mechanism design approach which presumes the second-order incentive constraints do not bind in the optimum, and verifies ex post that this is the case (Mirrlees, 1971, p. 188).

We now show that the effects of any tax perturbation can be decomposed into the effects of two types of prototypical tax reforms. The first is the *lump sum* perturbation which decreases the tax liability by a uniform amount:

$$\mathbf{x} \mapsto T(\mathbf{x}) - \rho \qquad \text{such that :} \qquad R(\mathbf{x}) = 1, \tag{10a}$$

where we use ρ to denote the magnitude of this specific perturbation. Second, there are *compensated perturbations of the j^{th} marginal tax rate for taxpayers of type* $\mathbf{w}$ which are defined as:

$$\mathbf{x} \mapsto T(\mathbf{x}) - \tau_j \left(x_j - X_j(\mathbf{w})\right) \qquad \text{such that :} \qquad R(\mathbf{x}) = x_j - X_j(\mathbf{w}), \tag{10b}$$

where we use τ_j to denote the magnitude of these specific perturbations. These perturbations are said to be "compensated for taxpayers of type $\mathbf{w}$" because they change the marginal tax rate of type $\mathbf{w}$ but leave the tax liability at incomes $\mathbf{x} = \mathbf{X}(\mathbf{w})$ unchanged.

Let us denote by $\partial X_i(\mathbf{w})/\partial \rho$ and $\partial X_i(\mathbf{w})/\partial \tau_j$ the responses of taxpayers of type $\mathbf{w}$ of their i^{th} income to, respectively, the lump sum perturbation (10a) and to the compensated perturbation (10b) of the j^{th} marginal tax rate.[5] A variation in t affects the first-order conditions (9) through the changes in the marginal tax rates on the right-hand side of (9). In addition, a variation in t affects the first-order conditions (9) through the changes in the tax liabilities that determine the

[5]Strictly speaking, these responses do not just depend on the type $\mathbf{w}$, but also on the Hessian of the tax function. When the tax function is nonlinear, the responses to a tax reform generate changes in the marginal tax rates, which further induce compensated responses to these changes in marginal tax rates, etc. (Jacquet and Lehmann, 2021b). By applying the implicit function theorem, the behavioral responses $\partial X_i(\mathbf{w})/\partial \rho$ and $\partial X_i(\mathbf{w})/\partial \tau_j$ encapsulate this "circular process" through the endogeneity of the marginal tax rates. We refer to these responses as *total* responses. We discuss the relation between direct and total responses in the supplementary materials.

marginal rates of substitution on the left-hand side of (9). Consequently, for each type $\mathbf{w}$, a variation $\mathrm{d}t$ induces the same responses as a lump-sum perturbation (10a) of size $R\left(\mathbf{X}(\mathbf{w})\right)\,\mathrm{d}t$, combined with compensated perturbations of each of the n marginal tax rates (10b) of respective sizes $R_{x_j}\left(\mathbf{X}(\mathbf{w})\right)\,\mathrm{d}t$. We thus get (see Appendix A.2):

$$\left.\frac{\partial \widetilde{X}_i^R(\mathbf{w},t)}{\partial t}\right|_{t=0} = \underbrace{\frac{\partial X_i(\mathbf{w})}{\partial \rho}\,R(\mathbf{X}(\mathbf{w}))}_{\text{Income responses}} + \underbrace{\sum_{j=1}^{n}\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\,R_{x_j}\left(\mathbf{X}(\mathbf{w})\right).}_{\text{Compensated responses}} \tag{11}$$

Note that we do not explicitly assume that responses to tax perturbations are smooth. We rather show that if the unperturbed tax schedule verifies Assumption 1, the function $t \mapsto \widetilde{\mathbf{X}}^R(\mathbf{w},t)$ is continuously differentiable at $t = 0$, and that Eq. (11) holds in that case.

We now investigate whether, starting from a tax schedule $T(\cdot)$ that is not necessarily optimal, a perturbation in a direction $R(\cdot)$ is socially desirable. We evaluate the social desirability of the tax reform by investigating its effects on the following perturbed "Lagrangian":

$$\widetilde{\mathcal{L}}^R(t,\lambda) \overset{\text{def}}{\equiv} \iint_{\mathcal{W}} \left\{ T(\widetilde{\mathbf{X}}^R(\mathbf{w},t)) - t\,R(\widetilde{\mathbf{X}}^R(\mathbf{w},t)) + \frac{\Phi\left(\widetilde{U}^R(\mathbf{w},t);\mathbf{w}\right)}{\lambda} \right\} f(\mathbf{w})\mathrm{d}\mathbf{w}, \tag{12}$$

where $\lambda > 0$ denotes the shadow price of tax revenue. We evaluate the effects of a tax reform on the perturbed Lagrangian by computing its effects, first, on the governments' revenue (4), and second, on the social objective (5).

We compute the response of the tax liabilities $T(\widetilde{\mathbf{X}}^R(\mathbf{w},t)) - t\,R(\widetilde{\mathbf{X}}^R(\mathbf{w},t))$ to a change in the magnitude t of the tax perturbation and evaluate at $t = 0$. For each taxpayer, the tax liabilities are modified in two ways. First, independently of any behavioral change, the tax revenue is directly affected by the *mechanical* effect: $-R(\mathbf{X}(\mathbf{w}))$. Second, taxpayers of type $\mathbf{w}$ respond to the tax perturbation by changing their incomes through the behavioral responses $(\partial \widetilde{X}_i^R(\mathbf{w},t)/\partial t)|_{t=0}$, for $i = 1,\ldots,n$. The total change in the tax liability due to the perturbation equals:

$$\begin{aligned}\left.\frac{\partial T(\widetilde{\mathbf{X}}^R(\mathbf{w},t)) - t\,R(\widetilde{\mathbf{X}}^R(\mathbf{w},t))}{\partial t}\right|_{t=0} &= \left[-1 + \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w}))\,\frac{\partial X_i(\mathbf{w})}{\partial \rho}\right] R(\mathbf{X}(\mathbf{w})) \\ &+ \sum_{1\le i,j\le n} T_{x_i}(\mathbf{X}(\mathbf{w}))\,\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\,R_{x_j}\left(\mathbf{X}(\mathbf{w})\right),\end{aligned} \tag{13}$$

where we use (11) to decompose the behavioral effects.

Next, we evaluate the effect of the tax perturbation on the social objective. Applying the envelope theorem to social welfare $\Phi(U)$ after inserting (8) and using (7) leads to:

$$\frac{1}{\lambda}\left.\frac{\partial\Phi\left(\widetilde{U}^R(\mathbf{w},t);\mathbf{w}\right)}{\partial t}\right|_{t=0} = g(\mathbf{w})\; R\left(\mathbf{X}(\mathbf{w})\right). \tag{14}$$

For any perturbation in direction $R(\cdot)$ and with magnitude t, there exists a lump-sum transfer denoted $\ell^R(t)$ such that the combination of the two perturbations is budget-balanced, i.e. $\mathbf{x} \mapsto T(\mathbf{x}) - t\, R(\mathbf{x}) + \ell^R(t)$ is a budget-balanced perturbation. Given a direction $R(\cdot)$, it is not easy to compute the lump sum transfer $\ell^R(\cdot)$ that makes the overall combination budget-balanced. However, if we normalize λ such that:

$$0 = \iint\limits_{\mathcal{W}} \left[1 - g(\mathbf{w}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w}))\,\frac{\partial X_i(\mathbf{w})}{\partial\rho}\right] f(\mathbf{w})\mathrm{d}\mathbf{w}, \tag{15}$$

then a lump-perturbation (10a) has no impact on the Lagrangian (6). Consequently, one only needs to evaluate the effect of the perturbation in the direction $R(\cdot)$ on the Lagrangian (6) to get the sign of the effect of the budget neutral perturbation on social welfare $\mathcal{O}$. This finding, which is valid even if the tax schedule is not optimal, is expressed in the following proposition (proven in Appendix A.3).

Proposition 1. *Under Assumption 1 and if λ is such that (15) holds, a tax perturbation in the direction $R(\cdot)$ with $t > 0$ $\left(\text{respectively } t < 0\right)$ combined with a lump-sum rebate of the net budget surplus generated by the perturbation is welfare improving if and only if $(\partial\widetilde{\mathcal{L}}^R(t,\lambda)/\partial t)|_{t=0} > 0$ $\left(\text{resp. } (\partial\widetilde{\mathcal{L}}^R(t,\lambda)/\partial t)|_{t=0} < 0\right)$, where:*

$$\begin{aligned}\left.\frac{\partial\widetilde{\mathcal{L}}^R(t,\lambda)}{\partial t}\right|_{t=0} \quad = \quad & \iint\limits_{\mathcal{W}} \left\{\left[g(\mathbf{w}) - 1 + \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w}))\,\frac{\partial X_i(\mathbf{w})}{\partial\rho}\right] R\left(\mathbf{X}(\mathbf{w})\right)\right. \\ & \left. + \sum_{1\leq i,j\leq n} T_{x_i}(\mathbf{X}(\mathbf{w}))\,\frac{\partial X_i(\mathbf{w})}{\partial\tau_j}\, R_{x_j}\left(\mathbf{X}(\mathbf{w})\right)\right\} f(\mathbf{w})\mathrm{d}\mathbf{w}.\end{aligned} \tag{16}$$

In subsections III.2 and III.3, we apply this proposition to derive the optimal-tax function under fixed isotax curves and for the general case, respectively.

III.2 Optimal taxation for given isotax curves

Intuitively, the design of the optimal tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ can be decomposed into two steps. The first step concerns the design of the isotax curves, which are the loci of incomes $\mathbf{x}$ that are associated with the same tax liability.[6] The second step concerns the assignment of a specific tax liability to each isotax curve. In this subsection, we apply Proposition 1 to solve the solution to the second step of finding the optimal tax liability for given isotax curves. We show that the assignment of tax liabilities to given isotax curves is characterized by a tax formula reminiscent of the ABC-formula of Saez (2001) that characterizes the optimal schedule with a one-dimensional base.

Isotax curves are implicitly given through the function $\Gamma(\cdot)$ which defines *taxable income* as follows:

$$y = \Gamma(\mathbf{x}) \in \mathbb{R}$$

for each combination of incomes $\mathbf{x}$. Values of $\mathbf{x}$ with the same tax liability $T(\mathbf{x})$ map to the same value of $\Gamma(\mathbf{x})$. Assuming that $\Gamma(\cdot)$ is twice continuously differentiable and that it admits a non-zero gradient everywhere, it follows that combinations of incomes with equal values of taxable income y are on the same isotax curve. We then solve for the optimal assignment of tax liabilities to each taxable income, denoted by $\mathcal{T}$, so that we have $T(\mathbf{x}) = \mathcal{T}(\Gamma(\mathbf{x}))$.[7] We then consider perturbations of the form $\mathbf{x} \mapsto \mathcal{T}(\Gamma(\mathbf{x})) - t\, R(\Gamma(\mathbf{x}))$, where the direction $R(\cdot)$ admits taxable income $\Gamma(\mathbf{x})$ as its single argument. We thus only consider perturbations of the function $\mathcal{T}$, preserving isotax curves $y = \Gamma(\mathbf{x})$. We denote as $Y(\mathbf{w}) = \Gamma(\mathbf{X}(\mathbf{w}))$ the realized taxable income for taxpayers of type $\mathbf{w}$ under the unperturbed tax schedule, and as $\widetilde{Y}^R(\mathbf{w}, t) = \Gamma(\widetilde{\mathbf{X}}^R(\mathbf{w}, t))$ the realized taxable income of taxpayers of type $\mathbf{w}$ under the perturbed tax schedule $\mathbf{x} \mapsto \mathcal{T}(\Gamma(\mathbf{x})) - t\, R(\Gamma(\mathbf{x}))$.

The lump sum perturbation (10a) defines the income response of taxable income as:

$$\frac{\partial Y(\mathbf{w})}{\partial \rho} = \sum_{i=1}^{n} \Gamma_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho}. \tag{17a}$$

We show in Appendix A.4 that the compensated tax perturbation at taxable income $Y(\mathbf{w})$ in the direction $R(y) = y - Y(\mathbf{w})$ of size τ causes the following

[6] Formally, these loci are "curves" only if $n = 2$. If $n = 3$, they are isotax *surfaces*. If $n \geq 4$, they are isotax *hypersurfaces*, etc. We maintain the term "isotax curves" for simplicity.

[7] We call the summary statistic y "taxable income" because this is the most natural interpretation. Mathematically, it is merely a statistic determined by the combination of income choices $\mathbf{x}$.

compensated responses of taxable income for taxpayers of type $\mathbf{w}$:

$$\frac{\partial Y(\mathbf{w})}{\partial \tau} \stackrel{\text{def}}{\equiv} \sum_{1\leq i,j\leq n} \Gamma_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \tau_j} \Gamma_{x_j}(\mathbf{X}(\mathbf{w})). \tag{17b}$$

Let $m(\cdot)$ denote the density of taxable income Y and let $M(\cdot)$ denote the corresponding cumulative density function. In addition, let $\partial \overline{Y}(y)/\partial \tau$, $\partial \overline{Y}(y)/\partial \rho$ and $\overline{g}(y)$ denote the mean values among taxpayers earning $Y(\mathbf{w}) = y$ of the compensated responses $\partial Y(\mathbf{w})/\partial \tau$, the income responses $\partial Y(\mathbf{w})/\partial \rho$ and the welfare weights $g(\mathbf{w})$ respectively. We show in Appendix A.4 that the optimal assignment of tax liabilities to the isotax curves verifies the following Proposition.

Proposition 2. *The optimal assignment of tax liabilities to each isotax curve verifies the optimal income tax formula:*

$$\frac{\mathcal{T}'(y)}{1-\mathcal{T}'(y)} = \frac{1}{\varepsilon(y)} \frac{1-M(y)}{y\, m(y)} \int_{z=y}^{\infty} \left[1 - \overline{g}(z) - \mathcal{T}'(z) \frac{\partial \overline{Y}(z)}{\partial \rho}\right] \frac{m(z)}{1-M(y)} \mathrm{d}z, \tag{18a}$$

together with transversality condition:

$$0 = \int_{z=0}^{\infty} \left[1 - \overline{g}(z) - \mathcal{T}'(z) \frac{\partial \overline{Y}(z)}{\partial \rho}\right] m(z) \mathrm{d}z, \tag{18b}$$

where we define the compensated elasticity at income y:

$$\varepsilon(y) \stackrel{\mathit{def}}{\equiv} \frac{1-\mathcal{T}'(y)}{y} \frac{\partial \overline{Y}(y)}{\partial \tau}. \tag{18c}$$

Formula (18a) is similar to Equation (19) in Saez (2001) with the exception that it is defined over taxable income rather than labor income. The distortions arising from a change in the marginal tax rate in the neighborhood of isotax curve y are proportional to the compensated elasticity $\varepsilon(y)$ and to the size of the tax base $y\, m(y)$. In the optimum, these distortions should be offset by the sum of the mechanical effects, $1 - \overline{g}(z)$, and the income response effects, $\mathcal{T}'(z)\ (\partial \overline{Y}(z)/\partial \rho)$, for all taxable incomes z above y.

Since we can replicate known results from the one-dimensional problem so readily in assigning tax liabilities to given isotax curves, the difficulty of solving the multidimensional tax problem does not lie in this step. However, the complementing step of designing the optimal shape of isotax curves is novel and causes new difficulties.

III.3 Optimal tax formula

We now apply Proposition 1 to the more general problem of designing the optimal income tax schedule in the income space. We introduce the following notations. Let $\mathcal{X} \stackrel{\text{def}}{\equiv} \{\mathbf{x} | \exists \mathbf{w} \in \mathcal{W} : \mathbf{x} = \mathbf{X}(\mathbf{w})\}$ denote the range of the type set $\mathcal{W}$ under the allocation $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$. Let $h(\mathbf{x})$ denote the joint density of incomes $\mathbf{x}$, which is defined over $\mathcal{X}$. Finally, for each combination of incomes $\mathbf{x} \in \mathcal{X}$, let $\partial \overline{X_i}(\mathbf{x})/\partial \tau_j$, $\partial \overline{X_i}(\mathbf{x})/\partial \rho$ and $\overline{g}(\mathbf{x})$ respectively denote the means of $\partial X_i(\mathbf{w})/\partial \tau_j$, $\partial X_i(\mathbf{w})/\partial \rho$ and $g(\mathbf{w})$ among taxpayers that earn the combination of incomes $\mathbf{X}(\mathbf{w}) = \mathbf{x}$.

At the optimum, there should not exist an infinitesimal perturbation of the tax schedule that would induce a first-order effect on the government's objective. According to Proposition 1, this is equivalent to demanding that the right-hand side of (16) equals zero for any direction $R(\cdot)$. To derive an optimal tax formula from this requirement, we rewrite (16) in the income space, which requires the following assumption about the regularity of the optimal allocation:

Assumption 2. *The sufficient statistics $h(\mathbf{x})$, $\partial \overline{X_i}(\mathbf{x})/\partial \tau_j$, $\partial \overline{X_i}(\mathbf{x})/\partial \rho$ and $\overline{g}(\mathbf{x})$ are continuously differentiable functions of $\mathbf{x}$.*

At the end of this subsection, we provide sufficient microfoundations to illustrate the plausibility of Assumption 2. The following proposition then characterizes the optimal tax schedule (see the proof in Appendix A.5).

Proposition 3. *Under Assumptions 1 and 2, the optimum verifies the Euler-Lagrange equation:*

$$\left[1 - \overline{g}(\mathbf{x}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \frac{\partial \overline{X_i}(\mathbf{x})}{\partial \rho}\right] h(\mathbf{x}) = - \sum_{j=1}^{n} \frac{\partial \left[\sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \frac{\partial \overline{X_i}(\mathbf{x})}{\partial \tau_j} h(\mathbf{x})\right]}{\partial x_j}, \tag{19a}$$

for all $\mathbf{x}$ in $\mathcal{X}$, and the boundary conditions:

$$\forall \mathbf{x} \in \partial \mathcal{X} : \quad \sum_{1 \le i,j \le n} T_{x_i}(\mathbf{x}) \frac{\partial \overline{X_i}(\mathbf{x})}{\partial \tau_j} h(\mathbf{x})\, e_j(\mathbf{x}) = 0, \tag{19b}$$

where $\partial \mathcal{X}$ denotes the boundary of $\mathcal{X}$, and $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), ..., e_n(\mathbf{x}))$ denotes the outward unit vector normal to the boundary at $\mathbf{x}$.

Proposition 3 provides a divergence equation that should hold for any income $\mathbf{x} \in \mathcal{X}$. Euler-Lagrange equation (19a) corresponds to what is derived by

Golosov et al. (2014, p. 49) in the proof of their Proposition 3, while boundary conditions (19b) are first derived in the present paper. A more intuitive formulation can be obtained by integrating the Euler-Lagrange Partial Differential Equation (19a) on any subset $\Omega \subseteq \mathcal{X}$ of the income set. Applying the divergence theorem yields the following corollary:

Corollary 1. *Under Assumptions 1 and 2, the optimum must verify the following integrated Euler-Lagrange equations for any subset of incomes* $\Omega \subseteq \mathcal{X}$ *with smooth boundary* $\partial\Omega$ *and outward unit normal vectors* $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), ..., e_n(\mathbf{x}))$*:*

$$\begin{aligned} & -\oint_{\partial\Omega} \sum_{1\le i,j\le n} T_{x_i}(\mathbf{x}) \frac{\partial \overline{X}_i(\mathbf{x})}{\partial \tau_j} e_j(\mathbf{x})\, h(\mathbf{x}) \mathrm{d}\Sigma(\mathbf{x}) \\ & = \iint_{\Omega} \left[1 - \overline{g}(\mathbf{x}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \frac{\partial \overline{X}_i(\mathbf{x})}{\partial \rho} \right] h(\mathbf{x}) \mathrm{d}\mathbf{x}, \end{aligned} \tag{19c}$$

where the symbol $\oint$ *denotes a (hyper)-surface integral and* $\mathrm{d}\Sigma(\mathbf{x})$ *is the corresponding measure.*

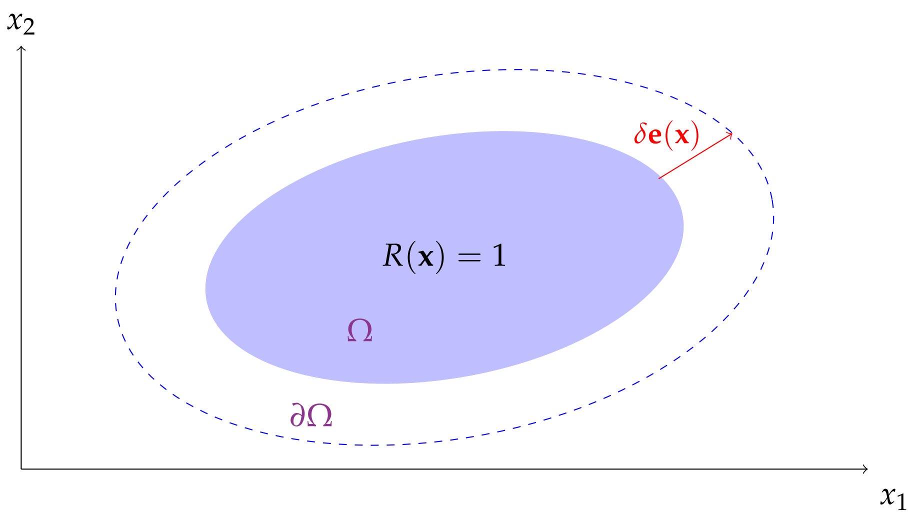


Figure 1: Intuition for Proposition 3. x_1 and x_2 are incomes.

Equation (19c) corresponds to Equation (17) in Golosov et al. (2014). To clarify the economic intuition of Corollary 1, we now provide a heuristic derivation of Eq. (19c) for the case with two incomes ($n = 2$). In doing so, we extend the heuristic derivation of the optimal tax formula provided by Saez (2001) for the one-dimensional case to the multidimensional case. We consider a tax reform, illustrated in Figure 1, that consists of two parts:

i. **Inside the subset of incomes** Ω (shaded area in Figure 1): A lump sum perturbation (10a) that uniformly decreases the tax liability by t for all households with incomes $\mathbf{x} \in \Omega$ before the reform.[8] Using $R(\mathbf{X}(\mathbf{w})) = 1$ and $R_{x_i}(\mathbf{X}(\mathbf{w})) = 0$ inside Ω, only mechanical and income effects matter for types with incomes $\mathbf{X}(\mathbf{w}) \in \Omega$. From (13) and (14), the contributions of the mass $h(\mathbf{x})$ of taxpayers with initial income $\mathbf{x}$ inside Ω to the change in the government's objective $\widetilde{\mathcal{L}}(t)$ is therefore given by:

$$-\left[1 - \overline{g(x)} - \sum_{i=1}^{n} T_{x_i} \frac{\partial \overline{X_i}(\mathbf{x})}{\partial \rho}\right] h(\mathbf{x}).$$

Integrating these effects over all incomes $\mathbf{x}$ inside the shaded area Ω leads to minus the right-hand side of (19c).

ii. **Inside a ring of width δ around Ω** (area between the shaded area and the dashed curve in Figure 1): The tax gradient $(T_{x_1}, ..., T_{x_n})$ must change to ensure tax liabilities uniformly decrease by t inside Ω and are unchanged outside a ring of width δ around Ω. For this purpose, along any radius normal to the boundary $\partial\Omega$, the tax gradient $(T_{x_1}, ..., T_{x_n})$ must be perturbed in a direction such that $R_{x_j}(\mathbf{X}(\mathbf{w})) = -e_j(\mathbf{x})/\delta$ for all $j \in \{1, ..., n\}$, where $(e_1(\mathbf{x}), ..., e_n(\mathbf{x}))$ is the outward unit vector normal to $\partial\Omega$ at income $\mathbf{x}$. If the width δ of the ring around Ω is sufficiently small, then the effects of changes in tax liabilities within the ring are of second-order importance compared to those inside Ω. We therefore approximate the tax perturbation in the ring by the n *compensated* tax perturbations (10b) of sizes $-e_j(\mathbf{x})/\delta$. This allows us to use (13) and $R(\mathbf{X}(\mathbf{w})) \simeq 0$ to approximate the contribution by taxpayers with initial income $\mathbf{x}$ inside the ring to the change in the government's objective $\widetilde{\mathcal{L}}(t)$ as:

$$-\frac{1}{\delta} \sum_{1 \le i,j \le n} T_{x_i} \frac{\partial \overline{X_i}(\mathbf{x})}{\partial \tau_j} e_j(\mathbf{x})\, h(\mathbf{x}).$$

Integrating this expression, first along a radius of width δ normal to $\partial\Omega$, and second along the boundary $\partial\Omega$ of Ω leads to the left-hand side of (19c).

If the initial tax schedule is optimal, the substitution effects inside the ring of width δ around Ω must be exactly offset by the mechanical and income effects inside Ω, which leads to (19c).

Following Bourguignon and Spadaro (2012), Bargain et al. (2014b) and Jacobs et al. (2017), one can use Proposition 3 to reveal the social preferences that

[8] The area Ω in our Figure 1 does not need to be convex.

are consistent with an existing tax schedule. According to (19a), the revealed marginal welfare weights are:

$$\widehat{g}(\mathbf{x}) \stackrel{\text{def}}{\equiv} \left[1 - \sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \frac{\partial \overline{X}_i(\mathbf{x})}{\partial \rho}\right] + \frac{1}{h(\mathbf{x})} \sum_{j=1}^{n} \frac{\partial \left[\sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \frac{\partial \overline{X}_i(\mathbf{X}(\mathbf{w}))}{\partial \tau_j} h(\mathbf{x})\right]}{\partial x_j}. \tag{21}$$

If for some income $\mathbf{x}$ these revealed marginal welfare weights are negative, then there exists a Pareto-improvement to the current tax schedule. We thus get a necessary condition for a given tax schedule to be Pareto efficient (see the proof in Appendix A.6).

Proposition 4. *Under Assumptions 1 and 2:*

i) If for some incomes $\mathbf{x}^\star$ inside $\mathcal{X}$, one has $\widehat{g}(\mathbf{x}^\star) < 0$ then an incremental tax perturbation that decreases the tax liabilities in an interior neighborhood of $\mathbf{x}^\star$ is Pareto improving.

ii) A Pareto efficient tax schedule must lead to $\widehat{g}(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \mathcal{X}$.

Part *ii)* of Proposition 4 provides a necessary condition in terms of observable statistics to test whether the current tax system is Pareto efficient. If the test fails, Part *i)* of Proposition 4 provides a Pareto improving tax reform. This result extends the findings of Werning (2007), Lorenz and Sachs (2016), Hendren (2020), Bierbrauer et al. (2020) and Gaube (2022) to the case where taxpayers earn many incomes.

In the remainder of this section, we discuss a microfoundation under which Assumption 2 holds. We will show that Assumption 2 holds under the following extension of the single crossing condition to the multidimensional context.

Assumption 2'. *The utility function $\mathcal{U}$ and the tax schedule $T(\mathbf{x})$ satisfy the following conditions.*

i) The number of incomes is equal to the number of unobserved characteristics: $n = p$.

ii) The matrix $\left[\mathcal{S}^i_{w_j}\right]_{i,j}$ is invertible.

iii) The mapping $\mathbf{w} \mapsto \left(\mathcal{S}^1(c, \mathbf{x}; \mathbf{w}), ..., \mathcal{S}^n(c, \mathbf{x}; \mathbf{w})\right)$ defined on $\mathcal{W}$ is injective.

iv) The tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ is thrice differentiable.

Part *ii*) of Assumption 2′ is standard (see Mirrlees, 1976, Section 4 and Renes and Zoutman, 2017). For $n = p = 1$, Parts *i*) *ii*) and *iii*) of Assumption 2′ are equivalent to the standard single crossing condition. For $n = p \geq 1$, when the utility function is additively separable:

$$\mathcal{U}(c, \mathbf{x}; \mathbf{w}) = \gamma(c) - \sum_{i=1}^{n} v^i(x_i, w_i) \qquad \text{where :} \qquad \gamma', v^i_{x_i}, v^i_{x_i,x_i} > 0 \neq v^i_{x_i,w_i} \quad (22)$$

then both Part *ii*) and Part *iii*) become equivalent to $v^i_{x_i,w_i} \neq 0$.[9] Part *iv*) of Assumption 2′ is more demanding than Assumption 1. It is necessary to ensure that behavioral responses $\partial X_i(\mathbf{w})/\partial \tau_j$, $\partial X_i(\mathbf{w})/\partial \rho$ and $\partial X_i(\mathbf{w})/\partial w_j$, which are defined along nonlinear income tax schedules, vary in a smooth way in the type space at the optimum (see Appendix A.7). Combining Assumptions 1 and 2′, we obtain the following Lemma, which we prove in Appendix A.7.

Lemma 1. *Under Assumptions 1 and 2′, the mapping* $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ *is a continuously differentiable bijection from* $\mathcal{W}$ *into* $\mathcal{X}$*, and Assumption 2 holds.*

In the case where the dimension p of the type set is larger than the dimension n of the income set, Propositions 3 and 4 remain valid under Assumptions 1 and 2. One can follow Jacquet and Lehmann (2021b) by assuming that Assumption 2 holds with respect to an n-dimensional subset of types, and "pooling" the $p - n$ types of taxpayers who get the same combinations of incomes.

IV The Mechanism Design approach

In this section, we rederive the optimal tax system using the mechanism-design approach instead of the tax-perturbation approach. This exercise serves three purposes. First, it allows us to verify under what conditions the two approaches result in the same optimal allocation. Second, by addressing the same problem from two sides, we arrive at a hybrid formulation for the optimal-tax function that is particularly useful for the purpose of simulation. Third, we use the mechanism-design approach to verify under what conditions the solution to the first-order conditions uniquely describe the maximum.

The mechanism design approach relies on the Taxation Principle (Hammond, 1979; Guesnerie, 1995), according to which it is equivalent for the gov-

[9] When the utility function takes the form (22), we get $\mathcal{S}^i(c, \mathbf{x}; \mathbf{w}) = v^i_{x_i}(x_i, w_i)/\gamma'(c)$. Assumption 2′ then amounts to demanding that the n one-dimensional mappings $w_i \mapsto v^i_{x_i}(x_i, w_i)/\gamma'(c)$ are injective, which is guaranteed by $v^i_{x_i,w_i}$ being either everywhere positive or everywhere negative.

ernment to select a tax function $\mathbf{x} \mapsto T(\mathbf{x})$ taking into account the taxpayers' decisions through (1), or to directly select an allocation $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ that verifies the self-selection (or incentive) constraints:

$$\forall \mathbf{w}, \hat{\mathbf{w}} \in \mathcal{W} : \qquad U(\mathbf{w}) \stackrel{\text{def}}{\equiv} \mathcal{U}\left(C(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right) \geq \mathcal{U}\left(C(\hat{\mathbf{w}}), \mathbf{X}(\hat{\mathbf{w}}); \mathbf{w}\right). \tag{23}$$

Instead of dealing with the double continuum of inequalities in (23), we follow Mirrlees (1971, 1976) by adopting a First Order Mechanism Design approach (henceforth the FOMD). This approach consists of, first, considering only "smooth" allocations. We formalize this in Assumption 3:

Assumption 3. *The allocation $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ is continuously differentiable and verifies* (23).

Second, we consider only the first-order incentive constraints:

$$\forall \mathbf{w} \in \mathcal{W}, \forall i \in \{1, ..., p\} : \qquad U_{w_i}(\mathbf{w}) = \mathcal{U}_{w_i}\left(C(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right), \tag{24}$$

which are obtained by applying the envelope theorem to the maximization of $\mathcal{U}(C(\hat{\mathbf{w}}), \mathbf{X}(\hat{\mathbf{w}}); \mathbf{w})$ with respect to $\hat{\mathbf{w}}$ and demanding that the maximand equals $\mathbf{w}$.

Under Assumption 3, the FOMD consists of finding a continuously differentiable allocation $\mathbf{w} \mapsto (U(\mathbf{w}), \mathbf{X}(\mathbf{w}), C(\mathbf{w}))$ that verifies the first-order incentive constraint (24) and maximizes the government's Lagrangian:

$$\iint_{\mathcal{W}} \left\{ \sum_{i=1}^{n} X_i(\mathbf{w}) - C(\mathbf{w}) + \frac{\Phi(U(\mathbf{w}); \mathbf{w})}{\lambda} \right\} f(\mathbf{w}) \mathrm{d}\mathbf{w} - E. \tag{25}$$

Our approach is to divide the optimization problem in two stages. In the first stage, the government chooses the utility profile $\mathbf{w} \mapsto U(\mathbf{w})$. In the second stage, which we label the subprogram, the government chooses the allocation $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ to maximize the resources extracted from taxpayers conditional on the utility profile chosen in the first stage. Formally, the government chooses the utility profile $\mathbf{w} \mapsto U(\mathbf{w})$ to maximize:

$$\iint_{\mathcal{W}} L\left(U(\mathbf{w}), U_{w_1}, \ldots, U_{w_p}; \mathbf{w}, \lambda\right) \mathrm{d}\mathbf{w}, \tag{26}$$

where the Lagrangian $L(\cdot)$ is defined as:

$$L(u, \mathbf{z}; \mathbf{w}, \lambda) \stackrel{\text{def}}{\equiv} \left(\mathcal{R}\left(u, \mathbf{z}; \mathbf{w}\right) + \frac{\Phi(u; \mathbf{w})}{\lambda} \right) f(\mathbf{w}) - E, \tag{27}$$

and the function $\mathcal{R}(\cdot)$ is defined via the subprogram:

$$\mathcal{R}(u, \mathbf{z}; \mathbf{w}) \overset{\text{def}}{\equiv} \max_{x_1,\dots,x_n} \sum_{i=1}^{n} x_i - \mathcal{C}(u; \mathbf{x}; \mathbf{w}) \tag{28}$$
$$s.t: \quad \forall i \in \{1, \dots, p\} \quad : \quad z_i = \mathcal{U}_{w_i}\left(\mathcal{C}\left(u, \mathbf{x}; \mathbf{w}\right), \mathbf{x}; \mathbf{w}\right).$$

Our approach differs from the traditional approach in Mirrlees (1976), Kleven et al. (2007) and Renes and Zoutman (2017), who directly maximize Lagrangian (25) subject to the incentive constraint (24) with respect to both the utility profile and the allocation. The traditional approach hides a conceptual problem in the multidimensional context. To see this, consider an example in which utility is additively separable as in (22). In that case, for a given candidate allocation $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ the first-order incentive constraints (24) form a system of Partial Differential Equations in $\mathbf{w} \mapsto U(\mathbf{w})$. If there is only one type, $p = 1$, the system simplifies to an Ordinary Differential Equation which can be integrated to provide the corresponding mapping $\mathbf{w} \mapsto U(\mathbf{w})$ (up to a constant). Conversely, when $p \geq 2$, the system of Partial Differential Equations (24) for a given *candidate* mapping $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ yields a *candidate* for the gradient of $\mathbf{w} \mapsto U(\mathbf{w})$ with components $\mathbf{w} \mapsto Z_i(\mathbf{w}) \overset{\text{def}}{\equiv} -v^i_{w_i}(X_i(\mathbf{w}), w_i)$ for all $i \in \{1, \dots, p\}$. However, not every combination of mappings $\mathbf{w} \mapsto Z_i(\mathbf{w})$ can effectively be the gradient of a mapping $\mathbf{w} \mapsto U(\mathbf{w})$. The utility profile $\mathbf{w} \mapsto U(\mathbf{w})$ must exhibit symmetric second-order cross-derivatives, i.e. $U_{w_j,w_k}(\mathbf{w}) = U_{w_k,w_j}(\mathbf{w})$ for all j, k and all $\mathbf{w}$. Hence, only candidate mappings $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ that imply a utility profile that verifies $\partial Z_k(\mathbf{w})/\partial w_j = \partial Z_j(\mathbf{w})/\partial w_k$ for all j, k and for all $\mathbf{w}$, are implementable. These additional *implementability* constraints are irrelevant in one-dimensional optimal tax problems but cannot be ignored in the multidimensional case.

Our approach overcomes this challenge by explicitly choosing the utility profile $U(\mathbf{w})$ in the first stage. Therefore, the solution automatically satisfies the implementation condition $U_{w_i,w_j}(\mathbf{w}) = U_{w_j,w_i}(\mathbf{w})$.

It is important to discuss the FOIC (24) in greater detail, since these form the constraints to the subprogram (28). Consider first the case in which $n = p = 1$, and assume the single-crossing condition is satisfied, i.e. $\mathcal{U}_{xw}(\cdot)$ is either strictly positive or negative for all (u, x, w), such that the right-hand side of (24) is monotonic in x. In that case, for any utility profile the FOIC (24) admits one solution, and hence, the inner optimization problem (28) becomes trivial since only one allocation can implement the utility profile.[10]

[10] Depending on the specification of the utility function, there may exist utility profiles for which (24) does not admit a solution. For instance, consider $n = p = 1$ and assume utility is defined

In a setting with more incomes $n > 1$, the same utility profile can often be offered through multiple allocations, the number of free variables in the system of equations (24) is larger. In that case, the subprogram (28) ensures that the government maximizes the resources extracted from each taxpayer. This division of the optimization problem into two stages explains several efficiency results in the literature, the most famous example being the Atkinson and Stiglitz (1976) theorem. If the type space is one-dimensional, $p = 1$, and $\mathcal{U}_w$ depends only on c and x_1, but not on $(x_2, ..., x_p)$, then marginal tax rates on $(x_2, ..., x_p)$ should be zero. More generally, as argued by Gauthier and Laroque (2009), if one includes externalities or public good provision in $(x_2, ..., x_p)$, then one retrieves first-best rules, such as a Pigouvian tax rule in case of an externality, or a Samuelson rule in case of public good provision.

A final concern is that we have not excluded the possibility that two allocations which implement the same utility profile, also extract exactly the same amount of resources. In addition, a concern is that the solution to (28) is not differentiable. Assumption (4) rules out these possibilities.

Assumption 4. *Subprogram (28) admits a single solution for each $(u, z; \mathbf{w})$. We denote this solution by $\mathbb{X}_1(u, \mathbf{z}; \mathbf{w})$, ..., $\mathbb{X}_n(u, \mathbf{z}; \mathbf{w})$ and assume that it is twice continuously differentiable in $(u, \mathbf{z}; \mathbf{w})$.*

We provide a micro-foundation for (4) in Assumption (4′) below.

Under Assumptions 3 and 4 we can derive the necessary conditions for the FOMD problem (26) using a perturbation approach. In Appendix B.1 we derive these conditions by considering perturbations in the utility profile $\mathbf{w} \mapsto U(\mathbf{w})$ that leave the value of the Lagrangian (26) unchanged. The resulting expressions are summarized in Proposition. 5.

Proposition 5. *Under Assumptions 3 and 4, the optimal utility profile $\mathbf{w} \mapsto U(\mathbf{w})$ must verify for all w in $\mathcal{W}$:*

$$\left(1 - \mathcal{S}^i\right) f(\mathbf{w}) = \mathcal{U}_c \sum_{j=1}^{n} \theta_j(\mathbf{w})\, \mathcal{S}^i_{w_j}, \tag{29a}$$

such that $\mathcal{U}_{\mathbf{w}}$ is strictly positive for all $(u, \mathbf{x}, \mathbf{w})$. In that case a candidate utility profile with $U_w(w) < 0$ for at least one w cannot be implemented. Technically, we can handle this concern by assigning to the subprogram the value $\mathcal{R}(\cdot) = -\infty$ when the constraints to the subprogram do not admit a solution. Assigning negative infinity to the subprogram ensures that such a utility profile cannot maximize the Lagrangian (26). The first stage nevertheless remains well defined, as long as there exists at least one utility profile that satisfies the FOIC. This is ensured because the Laissez-Faire allocation is by definition incentive compatible. In the remainder we will ignore this issue because it is unlikely to be of practical relevance.

$$\sum_{j=1}^{n} \frac{\partial \theta_j}{\partial w_j}(\mathbf{w}) = \left(\frac{1}{\mathcal{U}_c} - \frac{\Phi_u(U(\mathbf{w});\mathbf{w})}{\lambda} \right) f(\mathbf{w}) - \sum_{j=1}^{n} \theta_j(\mathbf{w}) \frac{\mathcal{U}_{cw_j}}{\mathcal{U}_c}, \tag{29b}$$

$$0 = \sum_{j=1}^{n} \theta_j(\mathbf{w}) e_j(\mathbf{w}), \qquad \forall \mathbf{w} \in \partial \mathcal{W}, \tag{29c}$$

where we define:

$$\theta_j(\mathbf{w}) \stackrel{def}{\equiv} -L_{z_j}\left(U(\mathbf{w}), U_{w_1}(\mathbf{w}), ..., U_{w_n}(\mathbf{w}); \mathbf{w}, \lambda\right). \tag{29d}$$

Eq. (29a) characterizes the optimal incomes $\mathbf{X}(\mathbf{w})$. Eq. (29b) is the Euler-Lagrange equation characterizing the cost $\theta(w_j)$ of distorting the j^{th} component of the gradient of $\mathbf{w} \mapsto U(\mathbf{w})$ (see (29d)). Eq. (29c) corresponds to the boundary conditions that must hold along the boundary of the type space $\partial \mathcal{W}$. Equations (29a), (29b), and (29c) respectively correspond to equations (60), (61) and (62) in Mirrlees (1976). Note that $\theta_j(\cdot)$ corresponds to the multiplier of the incentive constraints in Mirrlees (1976), as well as to the multiplier of the incentive constraints in the resource maximization subprogram (28). Our approach of perturbing $\mathbf{w} \mapsto U(\mathbf{w})$ and deducing the implied perturbation of the allocation $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ from the first-order incentive constraints, thus shows that shadow cost on the incentive constraint can be interpreted as the marginal resource cost of providing the utility profile.

We now provide a microfoundation to show the plausibility of Assumption 4.

Assumption 4′. *The number of incomes equals the number of unobserved characteristics, i.e. $n = p$, and for each utility level u and each type $\mathbf{w} \in \mathcal{W}$, the mapping*

$$(u, \mathbf{x}; \mathbf{w}) \mapsto \left(\mathcal{U}_{w_1}\left(\mathcal{C}(u, \mathbf{x}; \mathbf{w}), \mathbf{x}; \mathbf{w}\right), ..., \mathcal{U}_{w_p}\left(\mathcal{C}(u, \mathbf{x}; \mathbf{w}), \mathbf{x}; \mathbf{w}\right) \right)$$

is twice continuously differentiable in $(u, \mathbf{x}, \mathbf{w})$, and bijective in $\mathbf{x}$ with an invertible Jacobian.

When the utility function is of the additively separable form described in (22), Assumption 4′ is equivalent to $v^i_{x_i, w_i} \neq 0$. Hence, Assumption 4′ is a way to extend the single crossing condition in a multidimensional context.[11]

In Appendix B.2, we rewrite the conditions of Proposition 5 in terms of behavioral elasticities, type densities and welfare weights when $n = p$. We show that under Assumptions 3 and 4, the Euler-Lagrange Equation (29b) can

[11] Mirrlees (1976, page 342) adds a similar assumption when interpreting his optimality condition.

be rewritten:

$$\left[1 - g(\mathbf{w}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w}))\frac{\partial X_i(\mathbf{w})}{\partial \rho}\right] f(\mathbf{w}) = \sum_{j=1}^{n} \frac{\partial \sum_{i=1}^{n} \left(T_{x_i}(\mathbf{X}(\mathbf{w}))\, \mathcal{A}_{j,i}(\mathbf{w})\, f(\mathbf{w})\right)}{\partial w_j}, \tag{30a}$$

for all $\mathbf{w}$ in $\mathcal{W}$, while the Boundary conditions become:

$$\sum_{1 \leq i,j \leq n} T_{x_i}(\mathbf{X}(\mathbf{w}))\, \mathcal{A}_{j,i}(\mathbf{w})\, e_j(\mathbf{w}) = 0, \tag{30b}$$

for all $\mathbf{w}$ in $\mathcal{W}$, where the matrix $\mathcal{A}$ is defined by:

$$\left[\mathcal{A}_{i,j}\right]_{i,j} \stackrel{\text{def}}{\equiv} \left[\mathcal{S}^i_{w_j}\right]^{-1}_{i,j} = -\left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]^{-1}_{i,j} \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j}. \tag{30c}$$

In Appendix A.8, we show how optimal tax formulas (30) can also be retrieved by applying the tax perturbation approach using Assumption 1 and only Parts *i*) and *ii*) of Assumption 2′. Hence, we show that the optimality conditions derived by Mirrlees (1976, 1986) through a mechanism design approach are consistent with the optimality conditions in terms of sufficient statistics derived by Golosov et al. (2014) using a tax perturbation approach. We thus confirm that these two approaches are consistent in the multi-dimensional setting, as Saez (2001) shows for the one-dimensional case. The tax perturbation approach and FOMD approach are two faces of the same coin: while the FOMD approach computes the effects of directly perturbing a utility profile, the tax perturbation approach considers the effects of perturbing a tax function that decentralizes an allocation and the corresponding utility profile. The tax perturbation approach thus indirectly deals with perturbed allocations and yields conditions for optimality that are consistent with those that follow from the FOMD approach using direct perturbations.

Finally, we discuss the strengths and weaknesses of the tax perturbation and the mechanism design approaches. First assume the number of incomes equals the number of types, $n = p$, and assume that individual preferences verify Parts *i*), *ii*) and *iii*) of Assumptions 2′ and Assumption 4′.[12] The tax perturbation approach then requires the tax function to be smooth enough in the sense of Assumption 1 and of Part *iv*) of Assumption 2′, while the first-order mechanism approach requires incentive-compatible allocations that are smooth in the sense

[12]Recall when individual preferences are additively separable as in (22), Parts *i*), *ii*) and *iii*) of Assumption 2′ on the one hand and Assumption 4′ on the other hand are both equivalent to $v^i_{w_i,y_i} \neq 0$.

of Assumption 3. We show in Appendix A.2 that Assumptions 1 and 2′ together imply Assumption 3. Conversely, combining Assumption 3 with the left-hand side of (3) implies that marginal tax rates are only once-differentiable functions of type $\mathbf{w}$, while Part *iv*) of Assumption 2′ requires they are twice-differentiable in incomes $\mathbf{x}$. The tax perturbation approach is thus slightly more demanding than the mechanism design approach when $n = p$.

When the number of types is larger than the number of incomes, $p > n$, the tax perturbation method is easier to trace than the mechanism design approach. The reason is that the tax perturbation approach directly optimizes over the tax function $T(\mathbf{x})$, which depends on n inputs whereas the mechanism-design function optimizes over the indirect-utility function $V(\mathbf{w})$ which depends on $p > n$ inputs. Hence, using the tax perturbation approach rather than the mechanism-design approach reduces the dimensionality of the problem. Conversely, and by the same logic the mechanism-design approach is easier to trace when $n > p$.

Proposition 6. *Under Assumptions 3 and 4, if for each type $\mathbf{w} \in \mathcal{W}$ and each $\lambda \in \mathbb{R}_+$ the mapping $(U, \mathbf{z}) \mapsto L(U, \mathbf{z}; \mathbf{w}, \lambda)$ is concave and $\mathbf{w} \mapsto U^\star(\mathbf{w})$ verifies Equations (29), then $\mathbf{w} \mapsto U^\star(\mathbf{w})$ is the unique solution to the relaxed problem.*

This result is especially important for the numerical simulations below because it ensures that, whenever $(U, \mathbf{z}) \mapsto L(U, \mathbf{z}; \mathbf{w}, \lambda)$ is concave, and an allocation is found that verifies the necessary conditions, this allocation is the unique solution to the government's problem.

V Numerical simulations

Because both the type space and the income space are multidimensional, the optimal tax formulas do not take the form of ordinary differential equations, as in Mirrlees (1971), Diamond (1998) and Saez (2001) but they take the form of a second-order partial differential equation, as in Mirrlees (1976) and Golosov et al. (2014). This significantly complicates the process of solving the optimal tax equations. To understand this, it helps to consider the effects of a tax perturbation from a geometric perspective. In the one-dimensional case, the change in the marginal tax rate at a given income level is directly connected to changes in tax liabilities at all higher incomes. In the multidimensional case, the relation is more complicated. To change the gradient of the tax function at a given point, one must change the tax liabilities near that point, causing changes in the tax gradient elsewhere, see for instance Figure 1. To deal with this complexity, we rely on numerical simulations.

We develop a new numerical algorithm and apply it to the optimal taxation of couples. We consider an economy where couples differ in the productivity of females (w_f) and males (w_m), so unobserved heterogeneity is bi-dimensional ($p = 2$). Each couple chooses the labor supply of both spouses, so there are two incomes (i.e. $n = p = 2$). Preferences over the couple's consumption c, female income x_f and male income x_m are quasilinear in consumption, additively separable and isoelastic in each income:

$$\mathcal{U}(c, x_f, x_m; w_f, w_m) = c - \frac{\varepsilon_f}{1+\varepsilon_f}\, x_f^{\frac{1+\varepsilon_f}{\varepsilon_f}}\, w_f^{-\frac{1}{\varepsilon_f}} - \frac{\varepsilon_m}{1+\varepsilon_m}\, x_m^{\frac{1+\varepsilon_m}{\varepsilon_m}}\, w_m^{-\frac{1}{\varepsilon_m}}. \tag{31}$$

Income effects are thus assumed away (i.e. $\partial X_f(\mathbf{w})/\partial\rho = \partial X_m(\mathbf{w})/\partial\rho = 0$). Moreover, if the tax schedule is additively separable, the cross responses are equal to zero (i.e. $\partial X_f(\mathbf{w})/\partial\tau_m = \partial X_m(\mathbf{w})/\partial\tau_f = 0$). Finally, ε_f and ε_m respectively denote the direct elasticities of male and female incomes with respect to their own net-of-marginal-tax rates. Our baseline values are $\varepsilon_f = 0.43$ and $\varepsilon_m = 0.11$, which correspond to the mean labor supply elasticity for married women and for men in the meta-analysis of Bargain and Peichl (2016, Figure 1).

We calibrate the skill density $f(\cdot)$ using the Current Population Survey (CPS) of the US census of March 2016. We focus on married, mixed-gender couples that live together. We only consider income from labor. We drop couples in which either partner earns less than \$1,000 per year or in which either of the partners' incomes is top-coded. We drop same-sex couples because in our simulations we attach labor elasticities based on gender in each couple. From each observed couple, we recover their type (w_f, w_m) from their labor earnings (x_f, x_m) by inverting the first-order conditions (3). For this purpose, we use a rough approximation of the current tax schedule in the US by assuming a constant marginal tax rate of 37%, a figure which is consistent with Barro and Redlick (2011, Table 1). Next, we estimate the type density through a bi-dimensional kernel. We specify the social welfare function to be CARA with $\Phi(u, w_1, w_2) = 1 - \exp(-\beta\, u)/\beta$, where $\beta > 0$ stands for the degree of inequality aversion. For our baseline simulation, we select β such that the assumed 37% tax rate coincides with the optimal linear tax rate. This leads to $\beta = 0.0061$. Throughout the simulations, we assume that the government's revenue requirement equals 15% of GDP, which is close to the observed share of public spending in GDP for the US.

With our functional specifications, the government's Lagrangian (27) becomes:

$$L(u, \mathbf{z}; \mathbf{w}, \lambda) = \left[\sum_{i=f,m} \left((1+\varepsilon_i)^{\frac{\varepsilon_i}{1+\varepsilon_i}} w_i \, z_i^{\frac{\varepsilon_i}{1+\varepsilon_i}} - \varepsilon_i \, z_i \, w_i \right) - u - \frac{\exp\left(-\beta \, u\right)}{\lambda} \right] f(\mathbf{w}),$$

which is concave in (u, z_f, z_m). Since the Lagrangian is concave, Proposition 6 applies, meaning that our optimal tax formulas are both necessary and sufficient for the unique optimum.

We first give an overview of the simulation algorithm, in Subsection V.1. Next, in Subsection V.2, we report the results of the simulations for the baseline calibration. Finally, in Subsection V.3, we consider a number of comparative statics. We conjecture what happens when we vary the labor supply elasticities, the inequality aversion, or the simulation domain, and we verify our conjectures in the simulations.

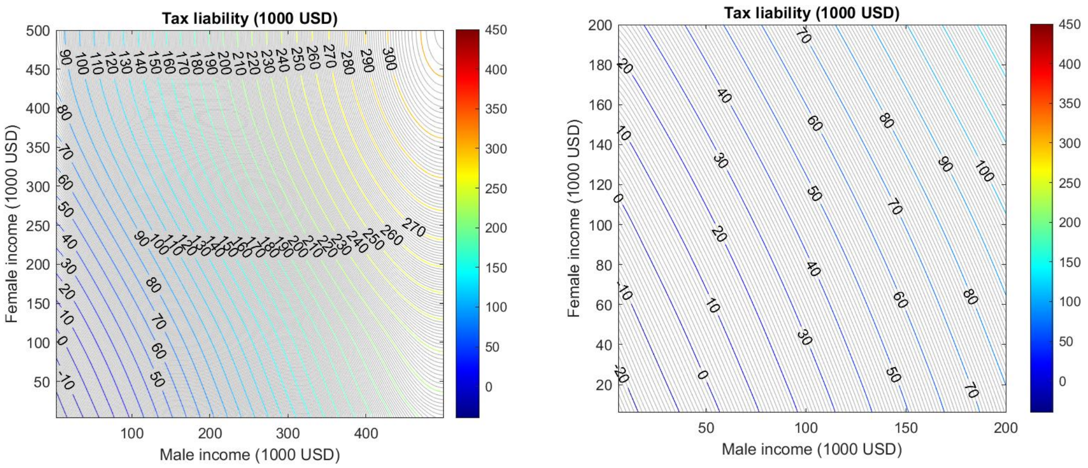
V.1 Simulation algorithm

The idea of our numerical algorithm is to first solve an optimal tax formula for given values of sufficient statistics, then to update the sufficient statistics using the tax schedule derived from the optimal tax formula, and to repeat this procedure until it converges to the optimal tax schedule. To do so, we can a priori use three optimal tax formulas, namely (19), (29) and (30). Let us explain why we choose (30). The optimal formula in (29) takes the form of a second-order *nonlinear* partial differential equation in the type space, which is numerically much more challenging than solving a linear second-order partial differential equation. Conversely, the optimal formula in Equations (19) is a *linear* second-order partial differential equation. However, it is defined in the income set $\mathcal{X}$. Hence, if one solves the optimal tax formula (19a) using the same income set from one iteration to the next, which is required given the boundary conditions (19b), then the corresponding typeset is changing from one iteration to the next. This is problematic when, for instance, comparing the values obtained for the tax revenue or for the social objective from one iteration to the next. Finally, the partial differential equation described in (30) is linear, provided that the sufficient statistics $g(\mathbf{w})$ and $\mathcal{A}(\mathbf{w})$ are taken as given. In addition, it is defined over the fixed type set $\mathcal{W}$.

Here again, there is a difficulty. Equations (30a)-(30b) are defined in the type space, while $(T_{x_1}, \ldots, T_{x_n})$ stands for the gradient of tax liability with respect to incomes. However, one can rewrite (30a)–(30b) in terms of the gradient of tax liability in the skills-space by scaling matrix $\mathcal{A}$ by the matrix $\left[\partial X_j(\mathbf{w})/\partial w_i\right]_{i,j}^{-1}$. We then iterate by *i*) finding the mapping $\mathbf{w} \mapsto T(\mathbf{X}(\mathbf{w}))$ that solves Equations (30a)–(30b) for given matrix $\mathcal{A}$, Jacobian $\left[\partial X_j(\mathbf{w})/\partial w_i\right]_{i,j}$ and type density $f(\mathbf{w})$ and getting a tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ from this solution, and *ii*) updating the matrix $\mathcal{A}$ and the Jacobian $\left[\partial X_j(\mathbf{w})/\partial w_i\right]_{i,j}$ given the new tax schedule. This hybrid approach thus combines the strength of the mechanism design approach (a fixed typeset over which to integrate), with the strength of the tax perturbation approach (a linear PDE). We describe the algorithm in more detail in the supplementary materials.

V.2 Results under the baseline calibration

Figure 2 displays the solution of the optimal tax problem using our baseline calibration. The optimal tax schedule is represented by the isotax curves, which are the loci of incomes for which the tax liability is constant at a given value. Male income is shown on the horizontal axis, while female income is indicated on the vertical axis. The left panel displays the whole domain of the simulations running up to $500,000, while the right panel zooms in at incomes below $200,000, where we find most taxpayers, roughly 97% of males and 99% of females.

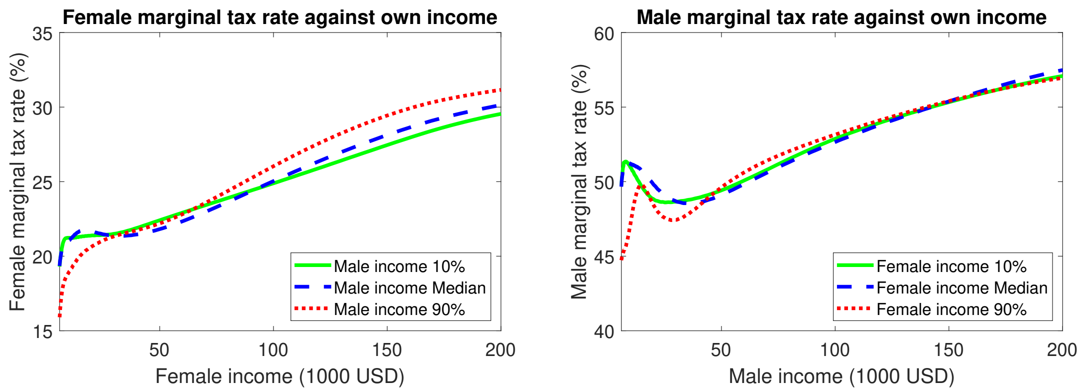


(a) Full simulation set (b) Zoom on $x_f, x_m \leq \$200,000$

Figure 2: Isotax curves in the baseline case

Strikingly, isotax curves are almost linear and parallel, except close to the boundaries. There, isotax curves are curved to satisfy boundary constraints (19b). This curvature pattern is most notable at high income levels where there are very few taxpayers. For lower incomes, the curvature only affects isotax curves very close to the lower bound.

Compared with the current economy, which is approximated by a linear tax rate of 37%, the optimal tax schedule leads to an improvement of the social objective equivalent to 0.82% of GDP in monetary terms. To understand which forces drive this gain, we decompose the welfare gain in different steps. Going from our approximation of the current economy (where we assume linear tax rates) to the optimal *joint* tax $(x_f, x_m) \mapsto T(x_f + x_m)$ captures the welfare gain of allowing the joint income tax schedule to be nonlinear. We find this welfare gain to be only 0.03%. If we now maintain the requirement that the isotax curves are linear and parallel, but remove the requirement that both marginal tax rates are equal, so $(x_f, x_m) \mapsto T(x_f + \alpha\, x_m)$ where α is optimized, we obtain a welfare gain from the current economy equal to 0.81%. The optimal value of α is 2.13, which implies that female income is discounted by 53%. Hence, while the gain of optimizing the slope of the isotax curves (optimizing α) is economically significant, the welfare gain of relaxing the constraint that isotax curves must be linear and parallel appears to be small.



(a) Optimal marginal tax rates of female income against own income

(b) Optimal marginal tax rates of male income against own income

Figure 3: Optimal Jointness

Kleven et al. (2006, 2007) show that under our individual and social preferences, when the abilities of both spouses are not correlated, the optimal marginal tax rates of each partner decrease in the income of the other partner. This is the so-called *negative jointness* of the optimal tax system. In a separate simulation with a population that replicates the moments of male and female incomes, but removes any correlation between the two, we confirm the optimality of the negative jointness of the tax system. In reality, however, the assumption that the skills of both partners are not correlated, is counterfactual. We show in Figure

3 that the optimal negative jointness result is not robust to using more realistic type densities with positive assortative matching. Figure 3a (resp. Figure 3b) displays the marginal tax rate for females (males) as a function of their own income. Each curve graphs this marginal income fixing male (female) income at the 10-th, 50-th and 90-th percentile of the male (female) income distribution. In case of negative jointness, the curve corresponding to male (female) income at the 10-th percentile should be everywhere above the curve corresponding to male (female) income at 50-th and 90-th percentiles of the distribution. Figures 3a and 3b contradict this prediction, thereby rejecting the idea that negative jointness holds at the optimum.

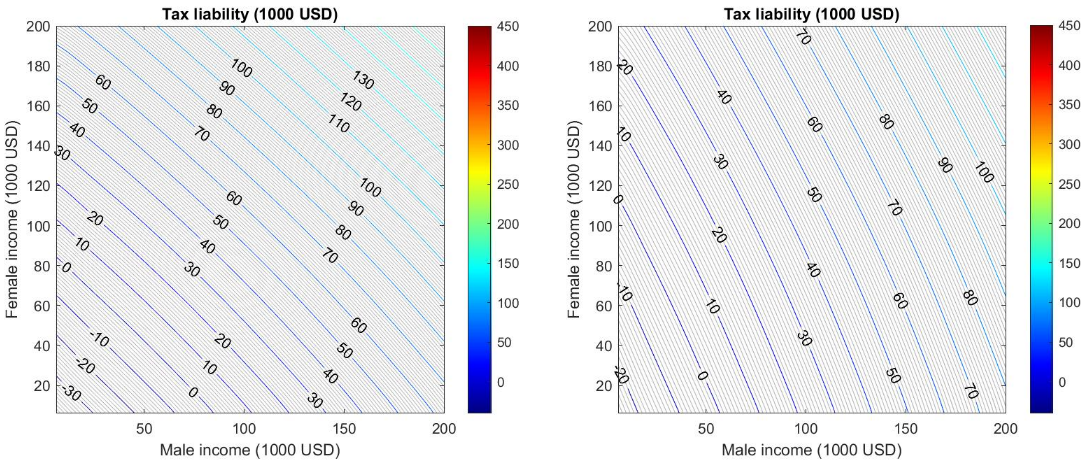
V.3 Comparative statics

To better understand the determinants of the optimal schedule, we examine how the simulated optimal schedule varies when we change the parameters of the simulations. For each parameter that we vary, we first intuitively provide conjectures on how changes in these parameters are going to affect optimal isotax curves before examining whether our simulations confirm or negate our a priori guess.

V.3.a Varying labor supply elasticity

One may conjecture that the slope of the isotax curves is affected by the ratio of the labor supply elasticities of both spouses. Whenever the labor supply elasticities of the two spouses are different, we expect that it is optimal to levy the lowest marginal tax rate on the spouse with the highest elasticity. Doing so shifts the burden of taxation away from the most elastic tax base. Recall that the empirical literature finds that married females have higher labor supply elasticities than married males. With male earnings in the horizontal axis, this amounts to making the isotax curves steeper. We thus conjecture that the larger is the ratio of female to male labor supply elasticity, the steeper are the optimal isotax curves.

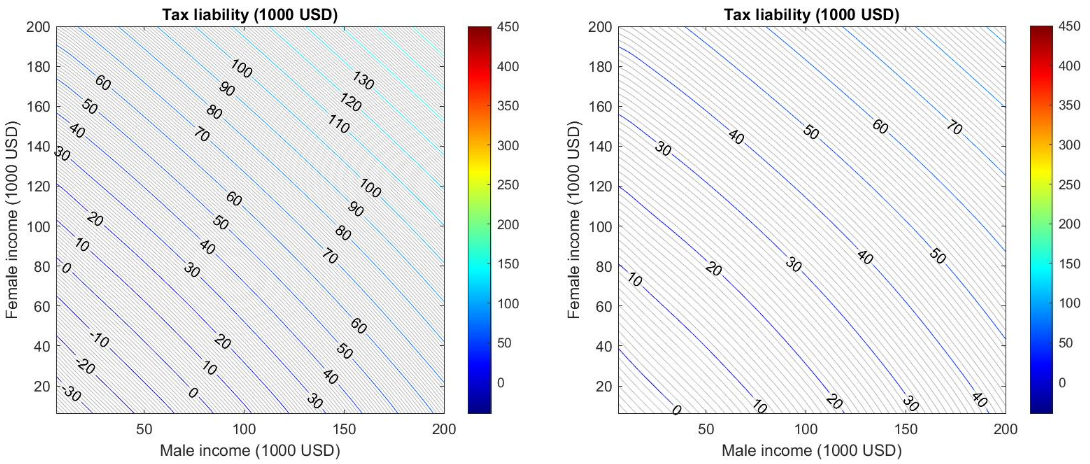
We investigate the validity of this conjecture in Figure 4. The left panel displays isotax curves when the two elasticities are equal $\varepsilon_f = \varepsilon_m = 0.11$, while the right panel shows the benchmark values $\varepsilon_f = 0.43$ for female and $\varepsilon_m = 0.11$ for male income. In both cases, we maintain inequality aversion at its baseline value $\beta = 0.0061$. As we conjectured, isotax curves are steeper when the two elasticities are different.



(a) $\varepsilon_f = \varepsilon_m = 0.11$ (b) $\varepsilon_f = 0.43, \varepsilon_m = 0.11$

Figure 4: Isotax curves with different elasticities.

In Figure 5, we assume $\varepsilon_f = \varepsilon_m = 0.11$ in the left panel and $\varepsilon_f = \varepsilon_m = 0.43$ in the right panel. In this figure, the slope of isotax curves looks the same in both panels. However, all else equal, higher labor supply elasticities decrease the optimal marginal tax rates, similar to what is found in the one-dimensional case. This is visible in the figure in the increasing distance between the isotax curves.

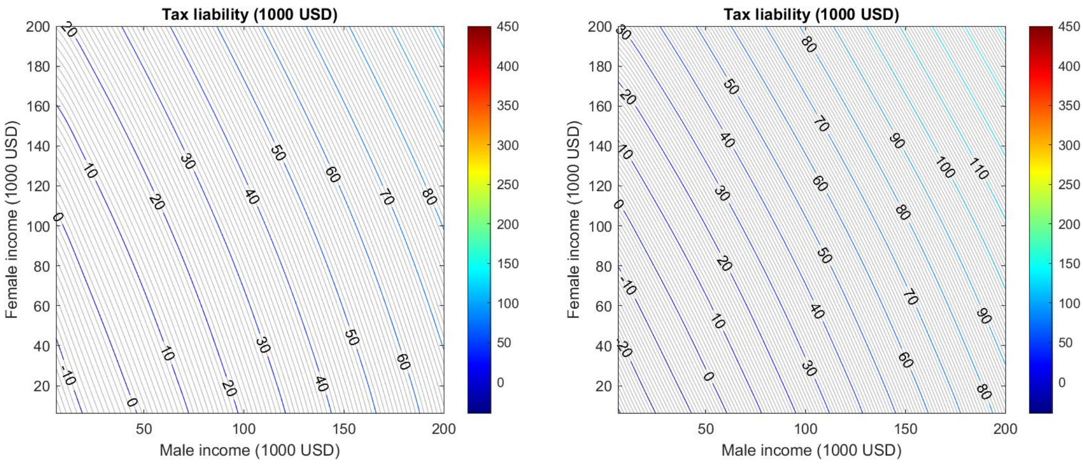


(a) $\varepsilon_f = \varepsilon_m = 0.11$ (b) $\varepsilon_f = \varepsilon_m = 0.43$

Figure 5: Isotax curves with different elasticities.

V.3.b Varying inequality aversion

We now investigate the sensitivity of the optimal tax schedule to the inequality aversion parameter β. In Figure 6, we contrast the case where the inequality aversion parameter is half lower than its baseline value in the left panel to the case where this parameter is half above its baseline value in the right panel. The shapes of isotax curves are virtually unaffected by change in the inequality aversion. However, the isotax curves are much closer together in the right panel when the government is more inequality averse. As in the one-dimensional case, a higher inequality aversion, all else being equal, leads to higher optimal marginal tax rates, causing the isotax curves to be closer together.

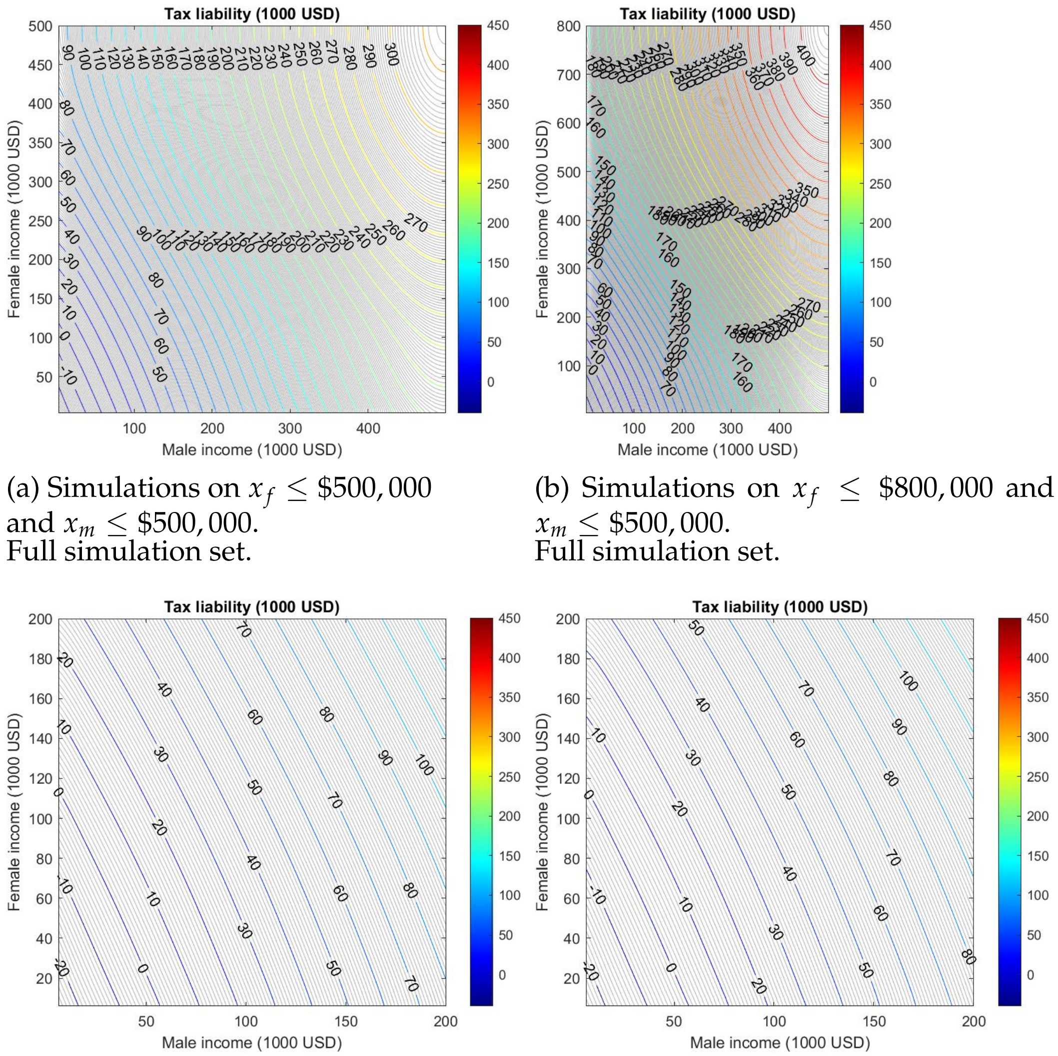


(a) $\beta = 0.003$ (b) $\beta = 0.009$

Figure 6: Isotax curves with different inequality aversion.

V.3.c Simulation domain

So far, it seems that the main departure from parallel and linear isotax curves is the curvature imposed by the boundary conditions. To verify the plausibility of this conjecture, we see what happens if we move the boundaries of the type space. In Figure 7 we compare the simulated isotax curves when both incomes are below $\$500,000$ (Figures 7a and 7c), and when male income is below $\$500,000$ while female income is below $\$800,000$ (Figures 7b and 7d). As expected, changing the simulation domain has virtually no effect for incomes below $\$200,000$ (Figures 7c and 7d). One difference between the simulations is that the larger domain adds some very rich taxpayers, which causes a minor increase in the distributional benefits of higher marginal tax rates, triggering a slight inwards shift of the isotax curves. Simultaneously, the curvatures of the isotax curves near the high-income boundaries adapt to where these boundaries are (Figures 7a and 7b). For instance, when female income is simulated up to $\$800,000$, then the $\$130,000$ isotax curve is concave everywhere (see Figure 7b). In this case, the shape of the isotax curve is affected by the boundary con-



(a) Simulations on $x_f \leq \$500,000$ and $x_m \leq \$500,000$.
Full simulation set.

(b) Simulations on $x_f \leq \$800,000$ and $x_m \leq \$500,000$.
Full simulation set.

(c) Simulations on $x_f \leq \$500,000$ and $x_m \leq \$500,000$.
Zoom on $x_f, x_m \leq \$200,000$.

(d) Simulations on $x_f \leq \$800,000$ and $x_m \leq \$500,000$.
Zoom on $x_f, x_m \leq \$200,000$.

Figure 7: Isotax curves with different domains

dition for a zero male income and a female income just above $\$500,000$. Conversely, when the income space is limited to female incomes below $\$500,000$, the isotax curve for a tax liability of $\$130,000$ exits the domain by crossing the top boundary and therefore becomes convex for low male income. Figure 7 thus confirms our conjecture that boundary conditions are the main explanation for the nonlinearity of the isotax curves at high income levels.

VI Conclusion

We study the optimal tax problem with multiple incomes and multiple dimensions of unobserved heterogeneity. We propose a numerical algorithm that addresses the difficulties inherent to the multidimensional tax problem. We apply this algorithm to the optimal taxation of couples. We find that the optimal isotax curves are close to linear and parallel. Optimal isotax curves are closer together when labor supply elasticities are higher or when inequality aversion is higher. When the labor supply elasticity of one spouse increases, the optimal marginal tax rate for this spouse decreases. We show that the optimal negative jointness of the tax schedules when skills are uncorrelated is not robust to the introduction of a more realistic distribution based on empirical simulations.

Analytically, we find a necessary condition for the tax schedule to be Pareto Efficient. If this condition is not verified, we describe a tax reform that is Pareto-improving. Second, we find conditions that ensure the necessary conditions of the optimal tax problem are unique and sufficient. Third, we derive conditions under which the tax perturbation and mechanism design approaches lead to the same tax formula. Fourth, we improve the tax perturbation approach by proposing conditions under which income bundles respond smoothly to small tax reforms. Fifth, we propose a mechanism design approach that encapsulates not only incentive constraints, but also the implementability constraints embedded in the multidimensional optimal tax problem. Lastly, we consider the cases where the number of incomes differs from the number of types.

Comparing the mechanism design approach to the tax perturbation approach, we find that the latter implies slightly more demanding restrictions on the smoothness of the tax schedule. The tax perturbation approach is thus slightly more demanding than the mechanism design approach. An additional advantage of the mechanism design approach is that it allows identifying a condition under which the necessary optimality conditions are also sufficient. A disadvantage of the mechanism design approach is that it is tractable only when the number of dimensions of unobserved heterogeneity does not exceed the number of incomes. An advantage of the tax perturbation approach is that it is allows providing an intuitive, graphical interpretation for the optimality conditions. We have shown that the tax perturbation approach is not less rigorous than the mechanism design approach.

Our paper can be extended in different ways. First, one could apply our algorithm to cases where the labor supplies of spouses interact through childcare or home production. Second, one could also apply our algorithm to the cases where tax units receive different source of incomes such as labor and capital incomes. Third, one could introduce general equilibrium effects. While our algorithm is sufficiently general to tackle these problems, implementing them would require significant changes to our simulations. We leave these problems for further research.

Appendix

A Tax Perturbation approach

A.1 Convexity of the indifference sets

We verify that assuming convex indifference sets is equivalent to assuming the second-order conditions of the taxpayers' program strictly hold when the tax schedule is linear.

On the one hand, the indifference sets are defined by $c = \mathcal{C}(u, \mathbf{x}; \mathbf{w})$. Applying the implicit function theorem to the definition of $\mathcal{C}(u, \mathbf{x}; \mathbf{w})$, we find the gradient of the indifference sets:

$$\mathcal{C}_{x_i}(u, \mathbf{x}; \mathbf{w}) = -\frac{\mathcal{U}_{x_i}\left(\mathcal{C}(u, \mathbf{x}; \mathbf{w}), \mathbf{x}; \mathbf{w}\right)}{\mathcal{U}_c\left(\mathcal{C}(u, \mathbf{x}; \mathbf{w}), \mathbf{x}; \mathbf{w}\right)}.$$

The Hessian of the indifference surfaces is therefore a matrix with i^{th} row and j^{th} column:

$$\mathcal{C}_{x_i,x_j} = -\frac{\mathcal{U}_{x_i,x_j}\mathcal{U}_c - \mathcal{U}_{c,x_i}\frac{\mathcal{U}_{x_j}}{\mathcal{U}_c}\mathcal{U}_c - \mathcal{U}_{c,x_j}\mathcal{U}_{x_i} + \mathcal{U}_{c,c}\frac{\mathcal{U}_{x_j}}{\mathcal{U}_c}\mathcal{U}_{x_i}}{\mathcal{U}_c^2}.$$

On the other hand, from (2), we get:

$$\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c = -\frac{\mathcal{U}_{x_i,x_j}\mathcal{U}_c - \mathcal{U}_{c,x_j}\mathcal{U}_{x_i}}{\mathcal{U}_c^2} + \frac{\mathcal{U}_{x_j}}{\mathcal{U}_c}\frac{\mathcal{U}_{c,x_i}\mathcal{U}_c - \mathcal{U}_{c,c}\mathcal{U}_{x_i}}{\mathcal{U}_c^2}, \tag{32}$$

The assumption that indifference sets are convex thus implies that the matrix $\left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c\right]_{i,j}$ is symmetric and positive definite. If then taxes are linear, so $T_{x_i}x_j = 0$, Assumption 1 is fulfilled.

A.2 Behavioral Responses

Under Assumption 1, one can differentiate (9) with respect to t, $\mathbf{x}$ and $\mathbf{w}$ to get:

$$\left[\mathcal{C}_{x_j x_i} + T_{x_j x_i}\right]_{j,i} \cdot [\mathrm{d}x_i]_i = \left[R_{x_j}(\mathbf{X}(\mathbf{w}))\right]_j \mathrm{d}t - \left[\mathcal{S}_c^j\right]_j R(\mathbf{X}(\mathbf{w}))\,\mathrm{d}t - \left[\mathcal{S}_{w_k}^j\right]_{j,k} \cdot [\mathrm{d}w_k]_k, \tag{33}$$

where the expressions are evaluated at $t = 0$, $\mathbf{x} = \mathbf{X}(\mathbf{w})$ and $c = C(\mathbf{w})$, and we use (3) and (32).

From (10b), a compensated reform of the j^{th} marginal tax rate is characterized by $R(\mathbf{X}(\mathbf{w})) = 0$, $R_{x_j}(\mathbf{X}(\mathbf{w})) = 1$ and $R_{x_k}(\mathbf{X}(\mathbf{w})) = 0$ for $k \neq j$. Using (33), the matrix of compensated responses for type $\mathbf{w}$ is:

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j} = \left[\frac{\partial X_j(\mathbf{w})}{\partial \tau_i}\right]_{i,j} = \left[\mathcal{C}_{x_j x_i} + T_{x_j x_i}\right]_{j,i}^{-1}. \tag{34a}$$

Since the matrix of compensated responses is the inverse of the symmetric and positive definite matrix $\left[\mathcal{C}_{x_j x_i} + T_{x_j x_i}\right]_{j,i}$, it is also symmetric and positive definite.

From (10a), a lump-sum perturbation of the tax function is characterized by $R(\mathbf{X}(\mathbf{w})) = 1$ and $R_{x_j}(\mathbf{X}(\mathbf{w})) = 0$. Using (33), the vector of income responses of type $\mathbf{w}$ is therefore given by:

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial \rho}\right]_i = -\left[\mathcal{C}_{x_j x_i} + T_{x_j x_i}\right]_{j,i}^{-1} \cdot \left[\mathcal{S}_c^j\right]_j = -\left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j} \cdot \left[\mathcal{S}_c^j\right]_j. \tag{34b}$$

Multiplying both sides of (33) by the Matrix $\left[\mathcal{C}_{x_j x_i} + T_{x_j x_i}\right]_{j,i}^{-1}$ and using (34a)–(34b) leads to (11). Finally, the implicit function theorem ensures that the mapping $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ is differentiable for all $\mathbf{w} \in \mathcal{W}$ with a Jacobian given by:

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial w_k}\right]_{i,k} = -\left[\mathcal{C}_{x_j x_i} + T_{x_j x_i}\right]_{j,i}^{-1} \cdot \left[\mathcal{S}_{w_k}^j\right]_{j,k} = -\left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j} \cdot \left[\mathcal{S}_{w_k}^j\right]_{j,k}. \tag{34c}$$

Eq. (34c) shows that when the tax schedule verifies Assumption 1 and individual preferences verify Assumption 2′, the ensuing allocation $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ verifies Assumption 3.

A.3 Proof of Proposition 1

To find the derivative of (12) with respect to t, we add (13) to (14). We integrate the result over all types $\mathbf{w}$ to obtain (16). To obtain (15), we use the lump-sum perturbation (10a) in (16), i.e. we set $R(\mathbf{X}(\mathbf{w})) = 1$ and $R_{x_j}(\mathbf{X}(\mathbf{w})) = 0$ in (16).

We now show that a tax perturbation in the direction $R(\cdot)$ with $t > 0$ is welfare improving if and only if the effect on the perturbed Lagrangian is positive. For all t, let $\ell^R(t)$ denote the lump-sum transfer that ensures that the following tax perturbation keeps the government's budget balanced: $\mathbf{x} \mapsto T(\mathbf{x}) - t\, R(\mathbf{x}) - \ell^R(t)$. Let $(\partial \widetilde{\mathcal{L}^R}(t)/\partial t)|_{t=0}$ denote the partial derivatives of the government's Lagrangian with respect to size t of the perturbation $\mathbf{w} \mapsto T(\mathbf{x}) - t\, R(\mathbf{x})$. Similarly, let $(\partial \widetilde{\mathcal{O}^{R+\ell}}(t)/\partial t)|_{t=0}$, $(\partial \widetilde{\mathcal{B}^{R+\ell}}(t)/\partial t)|_{t=0}$ and $(\partial \widetilde{\mathcal{L}^{R+\ell}}(t)/\partial t)|_{t=0}$ denote the partial derivatives of, respectively, the social objective, of government's revenue and of government's Lagrangian with respect to size t of the budget-balanced perturbation $\mathbf{w} \mapsto T(\mathbf{x}) - t\, R(\mathbf{x}) - \ell^R(t)$. Let finally $(\partial \widetilde{\mathcal{L}^\rho}(\rho)/\partial \rho)|_{\rho=0}$ denote the partial derivatives of the government's Lagrangian with respect to size ρ of the lump sum perturbation (10a). From (16), one gets that:

$$\left.\frac{\partial \widetilde{\mathcal{L}^{R+\ell}}(t)}{\partial t}\right|_{t=0} = \left.\frac{\partial \widetilde{\mathcal{L}^{R}}(t)}{\partial t}\right|_{t=0} + b'(t)\, \left.\frac{\partial \widetilde{\mathcal{L}^{\rho}}(\rho)}{\partial \rho}\right|_{\rho=0}.$$

Since (15) is equivalent to $(\partial \widetilde{\mathcal{L}^\rho}(\rho)/\partial \rho)|_{\rho=0}$, we thus get:

$$\left.\frac{\partial \widetilde{\mathcal{L}^{R+\ell}}(t)}{\partial t}\right|_{t=0} = \left.\frac{\partial \widetilde{\mathcal{L}^{R}}(t)}{\partial t}\right|_{t=0}.$$

Finally, since the perturbation $\mathbf{w} \mapsto T(\mathbf{x}) - t\, R(\mathbf{x}) - \ell^R(t)$ is budget balanced, one gets that $(\partial \widetilde{\mathcal{B}^{R+\ell}}(t)/\partial t)|_{t=0} = 0$, so that $(1/\lambda)(\partial \widetilde{\mathcal{O}^{R+\ell}}(t)/\partial t)|_{t=0} = (\partial \widetilde{\mathcal{L}^{R+\ell}}(t)/\partial t)|_{t=0}$ and eventually:

$$\frac{1}{\lambda}\left.\frac{\partial \widetilde{\mathcal{O}^{R+\ell}}(t)}{\partial t}\right|_{t=0} = \left.\frac{\partial \widetilde{\mathcal{L}^{R}}}{\partial t}\right|_{t=0}.$$

The above derivations are valid if Assumption 1 holds, regardless of $n \lesseqqgtr p$.

A.4 Optimal Tax for given isotax curves, Proof of Proposition 2

We decompose the tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ in two consecutive mappings: the first mapping defines a *taxable income* $y = \Gamma(\mathbf{x}) \in \mathbb{R}$ for each combination of incomes $\mathbf{x}$; the second mapping denoted $\mathcal{T}$ assigns a tax liability to each taxable income y. The tax liability at incomes $\mathbf{x}$ thus equals $T(\mathbf{x}) = \mathcal{T}(\Gamma(\mathbf{x}))$.

We first consider tax perturbations that preserve the isotax curves. Applying

Eq. (11) to the tax perturbation $\mathbf{x} \mapsto \mathcal{T}(\Gamma(\mathbf{x})) - t\, R(\Gamma(\mathbf{x}))$ leads to:

$$\left.\frac{\partial \widetilde{X}_i^R(\mathbf{w},t)}{\partial t}\right|_{t=0} = \frac{\partial X_i(\mathbf{w})}{\partial \rho} R(\Gamma(\mathbf{X}(\mathbf{w}))) + \sum_{i=1}^{n} \frac{\partial X_i(\mathbf{w})}{\partial \tau_j} \Gamma_{x_j}(\mathbf{w})\, R'(\Gamma(\mathbf{X}(\mathbf{w}))). \tag{35}$$

The definition of perturbed taxable income $\widetilde{Y}^R(\mathbf{w},t) = \Gamma\left(\widetilde{\mathbf{X}}^R(\mathbf{w},t)\right)$ implies that

$$\left.\frac{\partial \widetilde{Y}^R(\mathbf{w},t)}{\partial t}\right|_{t=0} = \sum_{i=1}^{n} \Gamma_{x_i}(\mathbf{X}(\mathbf{w})) \left.\frac{\partial \widetilde{X}_i^R(\mathbf{w},t)}{\partial t}\right|_{t=0}. \tag{36}$$

Applying (35) to the tax liability perturbation $R(y) = 1$ and using (36) leads to (17a). Applying (35) and (36) to the compensated perturbation $R(y) = y - Y(\mathbf{w})$ leads to (17b). Combining (17a), (17b), (35) and (36), the response of taxable income to a generic tax perturbation $R(\cdot)$ is given by:

$$\left.\frac{\partial \widetilde{Y}^R(\mathbf{w},t)}{\partial t}\right|_{t=0} = \frac{\partial Y(\mathbf{w})}{\partial \rho} R(Y(\mathbf{w})) + \frac{\partial Y(\mathbf{w})}{\partial \tau} R'(Y(\mathbf{w})).$$

The response of tax liability to a tax perturbation in the direction $R(\cdot)$ is thus:

$$\begin{aligned}
\left.\frac{\partial\left(\mathcal{T}\left(\widetilde{Y}^R(\mathbf{w},t)\right) - t\, R\left(\widetilde{Y}^R(\mathbf{w},t)\right)\right)}{\partial t}\right|_{t=0} &= -R(Y(\mathbf{w})) + \mathcal{T}'(Y(\mathbf{w})) \left.\frac{\partial \widetilde{Y}^R(\mathbf{w},t)}{\partial t}\right|_{t=0} \\
&= \left[-1 + \mathcal{T}'(Y(\mathbf{w}))\frac{\partial Y(\mathbf{w})}{\partial \rho}\right] R(Y(\mathbf{w})) \\
&\quad + \mathcal{T}'(Y(\mathbf{w})) \frac{\partial Y(\mathbf{w})}{\partial \tau} R'(Y(\mathbf{w})).
\end{aligned}$$

Using (14), the response of the perturbed Lagrangian (12) then is:

$$\begin{aligned}
\left.\frac{\partial \mathcal{L}^R(t)}{\partial t}\right|_{t=0} &= \iint_{\mathbf{w}\in\mathcal{W}} \left\{\left[g(\mathbf{w}) - 1 + \mathcal{T}'(Y(\mathbf{w})) \frac{\partial Y(\mathbf{w})}{\partial \rho}\right] R(Y(\mathbf{w})) \right. \\
&\quad \left. + \mathcal{T}'(Y(\mathbf{w})) \frac{\partial Y(\mathbf{w})}{\partial \tau} R'(Y(\mathbf{w}))\right\} f(\mathbf{w})\mathrm{d}\mathbf{w} \\
&= \int_{y\in\mathbb{R}_+} \left\{\int_{Y(\mathbf{w})=y} \left\{\left[g(\mathbf{w}) - 1 + \mathcal{T}'(y)\frac{\partial Y(\mathbf{w})}{\partial \rho}\right] R(y) \right.\right. \\
&\quad \left.\left. + \mathcal{T}'(y) \frac{\partial Y(\mathbf{w})}{\partial \tau} R'(y)\right\} f\left(\mathbf{w}|Y(\mathbf{w}) = y\right) \mathrm{d}\mathbf{w}\right\} m(y)\mathrm{d}y,
\end{aligned} \tag{37}$$

where $m(\cdot)$ denotes the density of taxable income Y as before. Denote the mean of the compensated responses among taxpayers earning $Y(\mathbf{w}) = y$ as:

$$\frac{\partial \overline{Y}(y)}{\partial \tau} \stackrel{\text{def}}{\equiv} \int_{Y(\mathbf{w})=y} \frac{\partial Y(\mathbf{w})}{\partial \tau} f\left(\mathbf{w}|Y(\mathbf{w}) = y\right) \mathrm{d}\mathbf{w}. \tag{38a}$$

Similarly, denote the mean of the income responses among taxpayers earning

$Y(\mathbf{w}) = y$ as:

$$\frac{\partial \overline{Y}(y)}{\partial \rho}(y) \stackrel{\text{def}}{\equiv} \int_{Y(\mathbf{w})=y} \frac{\partial Y(\mathbf{w})}{\partial \rho} f\left(\mathbf{w} | Y(\mathbf{w}) = y\right) \mathrm{d}\mathbf{w}. \tag{38b}$$

Finally, denote the mean of welfare weights among taxpayers earning $Y(\mathbf{w}) = y$ as:

$$\overline{g}(y) \stackrel{\text{def}}{\equiv} \int_{Y(\mathbf{w})=y} g(\mathbf{w}) f\left(\mathbf{w} | Y(\mathbf{w}) = y\right) \mathrm{d}\mathbf{w}. \tag{38c}$$

Eq. (37) then simplifies to:

$$\left.\frac{\partial \mathcal{L}^R(t)}{\partial t}\right|_{t=0} = \int_{y\in\mathbb{R}_+} \left\{ \left[\overline{g}(y) - 1 + \mathcal{T}'(y) \frac{\partial \overline{Y}(y)}{\partial \rho} \right] R(y) + \mathcal{T}'(y) \frac{\partial \overline{Y}(y)}{\partial \tau} R'(y) \right\} m(y)\mathrm{d}y. \tag{39}$$

Integrating by parts leads to:

$$\begin{aligned} \left.\frac{\partial \mathcal{L}^R(t)}{\partial t}\right|_{t=0} = & \int_{y\in\mathbb{R}_+} \left\{ \int_{z=y}^{\infty} \left[\overline{g}(z) - 1 + \mathcal{T}'(z) \frac{\partial \overline{Y}(z)}{\partial \rho} \right] m(z)\mathrm{d}z + \mathcal{T}'(y) \frac{\partial \overline{Y}(y)}{\partial \tau} m(y) \right\} R'(y)\mathrm{d}y \\ & - R(0) \int_{y\in\mathbb{R}_+} \left[\overline{g}(y) - 1 + \mathcal{T}'(y) \frac{\partial \overline{Y}(y)}{\partial \rho} \right] m(y)\mathrm{d}y. \end{aligned} \tag{40}$$

The effect of the perturbation on the Lagrangian is nil for all directions R if and only if (18b) and the following Equation:

$$\forall y: \qquad \mathcal{T}'(y) \frac{\partial \overline{Y}(y)}{\partial \tau} m(y) = \int_{z=y}^{\infty} \left[1 - \overline{g}(z) - \mathcal{T}'(z) \frac{\partial \overline{Y}(z)}{\partial \rho} \right] m(z)\mathrm{d}z, \tag{41}$$

are valid. Rearranging terms using (18c) leads to (18a) if $\mathcal{T}'(y) < 1$.

The definition of the two mappings $\mathbf{x} \xrightarrow{\Gamma} y \xrightarrow{\mathcal{T}} \mathcal{R}$ is not unique. Let $\alpha(\cdot)$ be a differentiable and increasing mapping, let $\widehat{\Gamma}(\mathbf{x}) \stackrel{\text{def}}{\equiv} \alpha\left(\Gamma(\mathbf{x})\right)$ be an alternative definition of taxable income that we denote $\widehat{y} = \alpha(y)$ and let $\widehat{\mathcal{T}}(\hat{y}) \stackrel{\text{def}}{\equiv} \mathcal{T}\left(\alpha^{-1}(\widehat{y})\right)$ be the associated assignment of tax liability to taxable income. Finally, let $\widehat{m}(\cdot)$ and $\widehat{M}(\cdot)$ be the PDF and CDF of $\widehat{y}$. We get

$$\widehat{\mathcal{T}}'\left(\widehat{y}\right) = \frac{\mathcal{T}'\left(\alpha^{-1}(\widehat{y})\right)}{\alpha'\left(\alpha^{-1}(\widehat{y})\right)} = \frac{\mathcal{T}'\left(y\right)}{\alpha'\left(y\right)}.$$

Differentiating both sides of $\widehat{M}\left(\alpha(y)\right) = M\left(y\right)$ leads to:

$$\widehat{m}(\widehat{y}) = \frac{m(y)}{\alpha'(y)}.$$

Applying respectively (17) and (17b) to $\widehat{Y}(\mathbf{w}) = \alpha\left(Y(\mathbf{w})\right)$ leads to:

$$\frac{\partial \widehat{Y}(\mathbf{w})}{\partial \rho} = \alpha'(Y(\mathbf{w}))\,\frac{\partial Y(\mathbf{w})}{\partial \rho} \qquad \text{and} \qquad \frac{\partial \widehat{Y}(\mathbf{w})}{\partial \tau} = \left(\alpha'(Y(\mathbf{w}))\right)^2 \frac{\partial Y(\mathbf{w})}{\partial \tau}.$$

Hence

$$\widehat{\mathcal{T}}'\left(\widehat{y}\right) \frac{\partial \overline{\widehat{Y}}(\widehat{y})}{\partial \rho} = \mathcal{T}'(y)\,\frac{\partial \overline{Y}(y)}{\partial \rho} \qquad \text{and} \qquad \widehat{\mathcal{T}}'\left(\widehat{y}\right) \frac{\partial \overline{\widehat{Y}}(\widehat{y})}{\partial \tau}\,\widehat{m}(\widehat{y}) = \mathcal{T}'(y)\,\frac{\partial \overline{Y}(y)}{\partial \tau}\,m(y).$$

Therefore (41) and (18b) are equivalent in terms of y or in terms of $\widehat{y}$.

A.5 Proof of Proposition 3

In Appendix A.3, we show that (16) holds under Assumption 1. We rewrite (16) in terms of the income density $h(\cdot)$ (which is well defined under Assumption 2):

$$\begin{aligned} \left.\frac{\partial \widetilde{\mathcal{L}}^R(t)}{\partial t}\right|_{t=0} = & \iint_{\mathcal{X}} \left\{ \left[\overline{g}(\mathbf{X}(\mathbf{w})) - 1 + \sum_{i=1}^{n} T_{x_i}(\mathbf{x})\,\frac{\partial \overline{X}_i(\mathbf{x})}{\partial \rho} \right] R\left(\mathbf{x}\right) \right. \\ & \left. + \sum_{1 \le i,j \le n} T_{x_i}(\mathbf{x})\,\frac{\partial \overline{X}_i(\mathbf{x})}{\partial \tau_j}\,R_{x_j}\left(\mathbf{x}\right) \right\} h(\mathbf{x})\mathrm{d}\mathbf{x}. \end{aligned} \tag{42}$$

Using the divergence theorem to integrate the term on the second line of this equation by parts and rearranging, yields:

$$\begin{aligned} \left.\frac{\partial \widetilde{\mathcal{L}}^R(t)}{\partial t}\right|_{t=0} = & \oint_{\partial\mathcal{X}} \sum_{1 \le i,j \le n} T_{x_i}(\mathbf{x})\,\frac{\partial \overline{X}_i(\mathbf{X}(\mathbf{w}))}{\partial \tau_j}\,h(\mathbf{x})\,e_j(\mathbf{x})\,R\left(\mathbf{x}\right)\mathrm{d}\Sigma(\mathbf{x}) \\ & - \iint_{\mathcal{X}} \left\{ \left[1 - \overline{g}(\mathbf{X}(\mathbf{w})) - \sum_{i=1}^{n} T_{x_i}(\mathbf{x})\,\frac{\partial \overline{X}_i(\mathbf{x})}{\partial \rho} \right] h(\mathbf{x}) + \sum_{j=1}^{n} \frac{\partial \left[\sum_{i=1}^{n} T_{x_i}(\mathbf{x})\,\frac{\partial \overline{X}_i(\mathbf{X}(\mathbf{w}))}{\partial \tau_j}\,h(\mathbf{x}) \right]}{\partial x_j} \right\} R\left(\mathbf{x}\right)\mathrm{d}\mathbf{x}. \end{aligned} \tag{43}$$

If the tax schedule $T(\cdot)$ is optimal, (43) must equal 0 for all possible directions $R(\cdot)$. This is only possible if the Euler-Lagrange Partial Differential Equation (19a) and the boundary conditions (19b) are both satisfied.

A.6 Proof of Proposition 4

Equations (21) and (43) are valid when Assumption 1 and Assumption 2 hold true. According to (5), (6) and (14), removing the term $\overline{g}(\mathbf{X}(\mathbf{w}))$ from (43) provides the effects of a tax perturbation on government's revenue:

$$\left.\frac{\partial \widetilde{\mathcal{B}}^R(t)}{\partial t}\right|_{t=0} = \oint_{\partial \mathcal{X}} \sum_{1 \le i,j \le n} T_{x_i}(\mathbf{x}) \; \frac{\partial \overline{X}_i(\mathbf{X}(\mathbf{w}))}{\partial \tau_j} \; h(\mathbf{x}) \; e_j(\mathbf{x}) \; R\left(\mathbf{x}\right) \; \mathrm{d}\Sigma(\mathbf{x})$$

$$- \iint_{\mathcal{X}} \left\{ \left[1 - \sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \; \frac{\partial \overline{X}_i(\mathbf{x})}{\partial \rho} \right] h(\mathbf{x}) + \sum_{j=1}^{n} \frac{\partial \left[\sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \; \frac{\partial \overline{X}_i(\mathbf{X}(\mathbf{w}))}{\partial \tau_j} \; h(\mathbf{x}) \right]}{\partial x_j} \right\} R\left(\mathbf{x}\right) \; \mathrm{d}\mathbf{x},$$

which, given (21), can be simplified to:

$$\left.\frac{\partial \widetilde{\mathcal{B}}(t)}{\partial t}\right|_{t=0} = \oint_{\partial \mathcal{X}} \sum_{1 \le i,j \le n} T_{x_i}(\mathbf{x}) \frac{\partial \overline{X}_i(\mathbf{X}(\mathbf{w}))}{\partial \tau_j} h(\mathbf{x}) \; e_j(\mathbf{x}) \; R\left(\mathbf{x}\right) \mathrm{d}\Sigma(\mathbf{x}) - \iint_{\mathcal{X}} \widehat{g}(\mathbf{x}) \; R(\mathbf{x}) \; h(\mathbf{x}) \mathrm{d}\mathbf{x}, \quad (44)$$

Note that the right-hand side of (21), and thus also $\widehat{g}(\cdot)$, is continuous with respect to $\mathbf{x}$. Let $\mathbf{x}^\star$ be an income bundle in the interior of $\mathcal{X}$ such that $\widehat{g}(\mathbf{x}^\star) < 0$. By continuity of $\widehat{g}(\cdot)$, there exists $r > 0$ such that the ball of radius r around $\mathbf{x}^\star$ remains in the interior of $\mathcal{X}$ and $\widehat{g}(\cdot)$ remains negative everywhere in this ball. Consider then a tax perturbation $\mathbf{x} \mapsto T(\mathbf{x}) - t \; R(\mathbf{x})$ where $R(\cdot)$ is twice continuously differentiable, positive inside the ball of radius r around $\mathbf{x}^\star$ and nil otherwise. Hence, $\widehat{g}(\mathbf{x}) \; R(\mathbf{x})$ is negative inside the ball of radius r around $\mathbf{x}^\star$ and nil outside.

The first term on the right-hand side of (44) is nil, because the tax schedule is unperturbed on the boundary $\partial \mathcal{X}$ of $\mathcal{X}$, while the second term is positive. Implementing this tax perturbation with $t > 0$ therefore generates tax revenue. Moreover, for incomes $\mathbf{x}$ inside the ball of radius r around $\mathbf{x}^\star$, utility increases since there $R(\mathbf{x})$ is positive so perturbed tax liability $T(\mathbf{x}) - t \; R(\mathbf{x}) < T(\mathbf{x})$ decreases. Finally, utility is unchanged outside the ball. Consequently, implementing this tax perturbation and rebating the extra revenue in a lump-sum way strictly increase the welfare for all taxpayers and is thereby Pareto-improving. This ends the proof of Part *i*) of Proposition 4. If a tax schedule is Pareto efficient, then such Pareto improving reform should not exist, which requires $\widehat{g}(\mathbf{x}) \ge 0$ for all $\mathbf{x} \in \mathcal{X}$.

A.7 Proof of Lemma 1

Given that $\mathcal{X}$ is defined as the range of the typeset $\mathcal{W}$ under the allocation $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$, it is sufficient to show that the mapping $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ is injective to establish that it is a bijection. Assume there exists $\mathbf{x} \in \mathcal{X}$ and $\mathbf{w}, \widehat{\mathbf{w}} \in \mathcal{W}$ such that $\mathbf{X}(\mathbf{w}) = \mathbf{X}(\widehat{\mathbf{w}}) = \mathbf{x}$. From Assumption 1, the first-order conditions (3) must be verified both at $(c, \mathbf{x}; \mathbf{w})$ and at $(c, \mathbf{x}; \widehat{\mathbf{w}})$, so we get $\mathcal{S}^i(c, \mathbf{x}, \mathbf{w}) = \mathcal{S}^i(c, \mathbf{x}, \widehat{\mathbf{w}})$ for all $i \in \{1, ..., n\}$. According to Part *iii*) of Assumption 2′, these n equalities imply that $\mathbf{w} = \widehat{\mathbf{w}}$. Differentiability of $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ is ensured under Assumption 1 by the implicit function theorem applied to (3). Part *ii*) of Assumption 2′ then ensures the Jacobian of $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ is invertible (see (34c) in Appendix A.2).

Because the mapping $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ is injective, we get that $\overline{g}(\mathbf{X}(\mathbf{w})) = g(\mathbf{w})$, $\partial \overline{X_i}(\mathbf{X}(\mathbf{w}))/\partial \tau_j = \partial X_i(\mathbf{w})/\partial \tau_j$ and $\partial \overline{X_i}(\mathbf{X}(\mathbf{w}))/\partial \rho = \partial X_i(\mathbf{w})/\partial \rho$. According to Equations (7), (34a) and (34b), $g(\mathbf{w})$, $\partial X_i(\mathbf{w})/\partial \tau_j$ and $\partial X_i(\mathbf{w})/\partial \rho$ are continuously differentiable functions of c, $\mathbf{x}$, $\mathbf{w}$ and, for the latter two, of the terms $T_{x_i x_j}$ in the Hessian of the tax schedule. Hence, because the mapping $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ is continuously differentiable and invertible, and because of Part *iv*) of Assumption 2′, $\partial \overline{X_i}(\mathbf{x})/\partial \tau_j$, $\partial \overline{X_i}(\mathbf{x})/\partial \rho$ and $\overline{g}(\mathbf{x})$ are continuously differentiable in $\mathbf{x}$. Finally, the income density is given by:

$$h(\mathbf{X}(\mathbf{w})) = f(\mathbf{w}) \left| \det \left[\frac{\partial X_i(\mathbf{w})}{\partial w_j} \right]_{i,j} \right|^{-1}, \tag{45}$$

which ensures the income density is also continuously differentiable in income. Hence Assumption 2 holds.

A.8 Optimal tax formula in the type space

To get an optimal tax formula in the type space, we need to rewrite the derivative of the perturbed Lagrangian, (16), in the type space rather than in the income space. To reparametrize the direction of a tax perturbation as a function of types, define $\widehat{R}(\mathbf{w}) \stackrel{\text{def}}{\equiv} R(\mathbf{X}(\mathbf{w}))$. Differentiating both sides with respect to w_j yields:

$$\widehat{R}_{w_j}(\mathbf{w}) = \sum_{i=1}^{n} (\partial X_i(\mathbf{w})/\partial w_j) \; R_{x_i}(\mathbf{X}(\mathbf{w})).$$

In matrix notation, the latter equality becomes:

$$\left[\widehat{R}_{w_j}(\mathbf{w})\right]_j^T = \left[R_{x_i}(\mathbf{X}(\mathbf{w}))\right]_i^T \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]_{i,j} \quad \Leftrightarrow \quad \left[R_{x_i}(\mathbf{X}(\mathbf{w}))\right]_i^T = \left[\widehat{R}_{w_j}(\mathbf{w})\right]_j^T \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]_{i,j}^{-1},$$

where we use Parts *i*) and *ii*) of Assumption 2′ and Eq. (34c) to ensure that matrix $\left[\partial X_i(\mathbf{w})/\partial w_j\right]_{i,j}$ is invertible. Using the symmetry of the matrix of compensated effects $\left[\partial X_i(\mathbf{w})/\partial \tau_j\right]_{i,j}$, we can rewrite the last term of (16):

$$\begin{aligned}
\sum_{1\leq i,j\leq n} T_{x_i}(\mathbf{X}(\mathbf{w}))\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\, R_{x_j}(\mathbf{X}(\mathbf{w})) &= \left[R_{x_j}(\mathbf{X}(\mathbf{w}))\right]_j^T \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j} \cdot \left[T_{x_i}(\mathbf{X}(\mathbf{w}))\right]_i \\
&= \left[\widehat{R}_{w_j}(\mathbf{w})\right]_j^T \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]_{i,j}^{-1} \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j} \cdot \left[T_{x_i}(\mathbf{X}(\mathbf{w}))\right]_i \\
&= -\left[\widehat{R}_{w_j}(\mathbf{w})\right]_j^T \cdot \left[\mathcal{S}^i_{w_j}\right]_{i,j}^{-1} \cdot \left[T_{x_i}(\mathbf{X}(\mathbf{w}))\right]_i,
\end{aligned}$$

where the last Equality follows from (34c). Using the definition of matrix $\mathcal{A}_{i,j}(\mathbf{w})$ in (30c), Eq. (16) can be rewritten as:

$$\begin{aligned}
\left.\frac{\partial \widetilde{\mathcal{L}}^R(t)}{\partial t}\right|_{t=0} = & \iint\limits_{\mathcal{W}} \Bigg\{ \left[g(\mathbf{w}) - 1 + \sum_{i=1}^n T_{x_i}(\mathbf{X}(\mathbf{w}))\frac{\partial X_i(\mathbf{w})}{\partial \rho}\right] \widehat{R}\,(\mathbf{w}) \\
& - \sum_{1\leq i,j\leq n} T_{x_i}(\mathbf{X}(\mathbf{w}))\; \mathcal{A}_{j,i}(\mathbf{w})\; \widehat{R}_{w_j}(\mathbf{w}) \Bigg\}\, f(\mathbf{w})\mathrm{d}\mathbf{w}.
\end{aligned}$$

Using the Divergence theorem to perform integration by parts, we get:

$$\begin{aligned}
\left.\frac{\partial \widetilde{\mathcal{L}}(t)}{\partial t}\right|_{t=0} = & - \oint_{\partial \mathcal{W}} \sum_{1\leq i,j\leq n} T_{x_i}(\mathbf{X}(\mathbf{w}))\; \mathcal{A}_{j,i}(\mathbf{w})\; e_j(\mathbf{w})\; f(\mathbf{w})\; \widehat{R}(\mathbf{w})\; \mathrm{d}\Sigma(\mathbf{w}) \\
& - \iint\limits_{\mathcal{W}} \Bigg\{ \left[1 - g(\mathbf{w}) - \sum_{i=1}^n T_{x_i}(\mathbf{X}(\mathbf{w}))\frac{\partial X_i(\mathbf{w})}{\partial \rho}\right] f(\mathbf{w}) \\
& \quad - \sum_{j=1}^n \frac{\partial \left(\sum_{i=1}^n T_{x_i}(\mathbf{X}(\mathbf{w}))\; \mathcal{A}_{j,i}(\mathbf{w})\; f(\mathbf{w}) \right)}{\partial w_j} \Bigg\}\, \widehat{R}\,(\mathbf{w})\;\; \mathrm{d}\mathbf{w}.
\end{aligned}$$

This partial derivative equals zero for any direction of tax perturbation $\widehat{R}(\cdot)$ if and only if Euler-Lagrange Equation (30a) and Boundary conditions (30b) are verified.

B First-Order Mechanism Design approach (FOMD)

B.1 Proof of Proposition 5

Let R be a twice differentiable function defined over $\mathcal{W}$ into $\mathbb{R}$. We consider the effects of perturbing the utility profile $\mathbf{w} \mapsto U(\mathbf{w})$ in the direction R. Consider the perturbed Lagrangian, with the unperturbed Lagrangian defined by

(27):

$$\widetilde{L}^R(t) \stackrel{\text{def}}{\equiv} \iint_{\mathcal{W}} L\left(U(\mathbf{w}) + t\,R(\mathbf{w}), U_{w_1}(\mathbf{w}) + t\,R_{w_1}(\mathbf{w}), ..., U_{w_p}(\mathbf{w}) + t\,R_{w_p}(\mathbf{w}); \mathbf{w}, \lambda\right) \mathrm{d}\mathbf{w}. \tag{46}$$

Applying the chain rule and denoting $\langle \mathbf{w} \rangle$ as a shortcut to denote that a function is evaluated at $\left(U(\mathbf{w}), U_{w_1}(\mathbf{w}), ..., U_{w_p}(\mathbf{w}); \mathbf{w}, \lambda\right)$, we obtain:

$$\left.\frac{\partial \widetilde{L}^R(t)}{\partial t}\right|_{t=0} = \iint_{\mathcal{W}} \left\{ L_u \langle \mathbf{w} \rangle\, R(\mathbf{w}) + \sum_{j=1}^{p} L_{z_j} \langle \mathbf{w} \rangle\;\, R_{w_j}(\mathbf{w}) \right\} \mathrm{d}\mathbf{w}.$$

Applying integration by parts using the divergence theorem leads to:

$$\left.\frac{\partial \widetilde{L}^R(t)}{\partial t}\right|_{t=0} = \iint_{\mathcal{W}} \left\{ L_u \langle \mathbf{w} \rangle - \sum_{j=1}^{p} \frac{\partial L_{z_j} \langle \mathbf{w} \rangle}{\partial w_j} \right\} R(\mathbf{w}) \mathrm{d}\mathbf{w} + \oint_{\partial\mathcal{W}} \sum_{j=1}^{p} L_{z_j} \langle \mathbf{w} \rangle\;\, e_j(\mathbf{w})\, R(\mathbf{w})\, \mathrm{d}\Sigma(\mathbf{w}).$$

At the optimal allocation, the latter expression is nil for any perturbation R. Using (29d), we find boundary conditions (29c), and the Euler-Lagrange Equation:

$$\forall \mathbf{w} \in \mathcal{W}: \qquad \sum_{j=1}^{p} \frac{\partial \theta_j(\mathbf{w})}{\partial w_j} = -L_u \langle \mathbf{w} \rangle\,. \tag{47}$$

Using incentive compatibility constraint (24), we can rewrite Lagrangian (27):

$$\left[\sum_{i=1}^{p} X_i(\mathbf{w}) - \mathcal{C}(U(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}) + \frac{\Phi(U(\mathbf{w}); \mathbf{w})}{\lambda}\right] f(\mathbf{w}) = \tag{48}$$
$$L\left(U(\mathbf{w}), \mathcal{U}_{w_1}\left(\mathcal{C}(U(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right), ..., \mathcal{U}_{w_n}\left(\mathcal{C}(U(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right); \mathbf{w}, \lambda\right).$$

Differentiating both sides of (48) with respect to $X_i(\mathbf{w})$ and using (2) and (29d):

$$\left(1 - \mathcal{S}^i \langle \mathbf{w} \rangle\right) f(\mathbf{w}) = -\sum_{j=1}^{p} \theta_j(\mathbf{w}) \left[\mathcal{U}_{x_i w_j} \langle \mathbf{w} \rangle + \mathcal{S}^i \langle \mathbf{w} \rangle\, \mathcal{U}_{c w_j} \langle \mathbf{w} \rangle\right],$$

which leads to (29a) given that $\mathcal{S}^i_{w_j} = (\mathcal{U}_{cw_j}\mathcal{U}_{x_i} - \mathcal{U}_{x_i w_j}\mathcal{U}_c)/\mathcal{U}_c^2 = -\left[\mathcal{U}_{x_i w_j} + \mathcal{S}^i\, \mathcal{U}_{cw_j}\right]/\mathcal{U}_c$. Differentiating (48) with respect to $U(\mathbf{w})$ and using $\mathcal{C}_u = 1/\mathcal{U}_c$ and (29d) leads to:

$$\left(-\frac{1}{\mathcal{U}_c \langle \mathbf{w} \rangle} + \frac{\Phi_u(U(\mathbf{w}); \mathbf{w})}{\lambda}\right) f(\mathbf{w}) = L_u \langle \mathbf{w} \rangle - \sum_{j=1}^{p} \theta_j(\mathbf{w}) \frac{\mathcal{U}_{c,w_j} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle}. \tag{49}$$

Substituting (47) into (49) yields (29b).

B.2 Derivation of the optimal tax formula in the type space

Using (3), Eq. (29a) leads to:

$$T_{x_i}(\mathbf{X}(\mathbf{w}))\, f(\mathbf{w}) = \sum_{j=1}^{p} \mu_j(\mathbf{w})\, \mathcal{S}^i_{w_j} \langle \mathbf{w} \rangle\,, \tag{50}$$

where we denote $\mu_j(\mathbf{w}) \stackrel{\text{def}}{\equiv} \theta_j(\mathbf{w})\, \mathcal{U}_c\left(C(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right)$. This can be rewritten $[T_{x_i}(\mathbf{X}(\mathbf{w}))]_i\, f(\mathbf{w}) = \left[\mathcal{S}^i_{w_j}\right]_{i,j} \cdot \left[\mu_j(\mathbf{w})\right]_j$ in matrix notation, which leads to: $\left[\mu_j(\mathbf{w})\right]_j = \left[\mathcal{S}^i_{w_j}\right]^{-1}_{i,j} \cdot [T_{x_i}(\mathbf{X}(\mathbf{w}))]_i\, f(\mathbf{w})$. Using (30c), we therefore get:

$$\forall \mathbf{w} \in \mathcal{W}, \forall i \in \{1, ..., p\} \qquad \mu_i(\mathbf{w}) = \sum_{j=1}^{n} \mathcal{A}_{i,j}(\mathbf{w})\, T_{x_j}(\mathbf{X}(\mathbf{w}))\, f(\mathbf{w}). \tag{51}$$

Combining (29c) with (51) thus leads to (30b). Using (7), Eq. (29b) implies that:

$$\begin{aligned}
\sum_{j=1}^{p} \frac{\partial \mu_j(\mathbf{w})}{\partial w_j} &= (1-g(\mathbf{w}))f(\mathbf{w}) - \sum_{j=1}^{p} \theta_j(\mathbf{w})\, \mathcal{U}_{cw_j} \langle \mathbf{w} \rangle \\
&+ \sum_{j=1}^{p} \theta_j(\mathbf{w}) \left[\mathcal{U}_{cc} \langle \mathbf{w} \rangle \frac{\partial C(\mathbf{w})}{\partial w_j} + \sum_{i=1}^{n} \mathcal{U}_{cx_i} \langle \mathbf{w} \rangle \frac{\partial X_i(\mathbf{w})}{\partial w_j} + \mathcal{U}_{cw_j} \langle \mathbf{w} \rangle\right] \\
&= (1-g(\mathbf{w}))f(\mathbf{w}) + \sum_{j=1}^{p} \theta_j(\mathbf{w}) \left[\mathcal{U}_{cc} \langle \mathbf{w} \rangle \frac{\partial C(\mathbf{w})}{\partial w_j} + \sum_{i=1}^{n} \mathcal{U}_{cx_i} \langle \mathbf{w} \rangle \frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]
\end{aligned} \tag{52}$$

Differentiating $C(\mathbf{w}) = \mathcal{C}(U(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w})$ with respect to w_j and using $\mathcal{C}_u = 1/\mathcal{U}_c$, $\mathcal{C}_{x_i} = -\mathcal{U}_{x_i}/\mathcal{U}_c$, $\mathcal{C}_{w_j} = -\mathcal{U}_{w_j}/\mathcal{U}_c$ and (24) leads to:

$$\frac{\partial C(\mathbf{w})}{\partial w_j} = \frac{\mathcal{U}_{w_j} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle} - \sum_{i=1}^{n} \frac{\mathcal{U}_{x_i} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle} \frac{\partial X_i(\mathbf{w})}{\partial w_j} - \frac{\mathcal{U}_{w_j} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle} = -\sum_{i=1}^{n} \frac{\mathcal{U}_{x_i} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle} \frac{\partial X_i(\mathbf{w})}{\partial w_j}.$$

Plugging this equality into (52) leads to

$$\begin{aligned}
\sum_{j=1}^{p} \frac{\partial \mu_j(\mathbf{w})}{\partial w_j} &= (1-g(\mathbf{w}))f(\mathbf{w}) + \sum_{j=1}^{p} \sum_{i=1}^{n} \theta_j(\mathbf{w}) \left[\mathcal{U}_{cx_i} \langle \mathbf{w} \rangle - \frac{\mathcal{U}_{x_i} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle} \mathcal{U}_{cc} \langle \mathbf{w} \rangle\right] \frac{\partial X_i(\mathbf{w})}{\partial w_j} \\
&= (1-g(\mathbf{w}))f(\mathbf{w}) - \sum_{j=1}^{p} \sum_{i=1}^{n} \mu_j(\mathbf{w})\, \mathcal{S}^i_c \langle \mathbf{w} \rangle\, \frac{\partial X_i(\mathbf{w})}{\partial w_j}.
\end{aligned} \tag{53}$$

Substituting (34c) into (53) yields:

$$\sum_{j=1}^{p} \frac{\partial \mu_j(\mathbf{w})}{\partial w_j} = (1-g(\mathbf{w}))f(\mathbf{w}) + \sum_{j=1}^{p} \sum_{i=1}^{n} \sum_{k=1}^{n} \mu_j(\mathbf{w})\, \mathcal{S}^i_c \langle \mathbf{w} \rangle\, \frac{\partial X_i(\mathbf{w})}{\partial \tau_k}\, \mathcal{S}^k_{w_j}(\mathbf{w}). \tag{54}$$

Substituting (34b) into (54) and using $\frac{\partial X_i(\mathbf{w})}{\partial \tau_k} = \frac{\partial X_k(\mathbf{w})}{\partial \tau_i}$ yields:

$$\sum_{j=1}^{p} \frac{\partial \mu_j(\mathbf{w})}{\partial w_j} = (1 - g(\mathbf{w}))f(\mathbf{w}) - \sum_{j=1}^{p} \sum_{k=1}^{n} \mu_j(\mathbf{w})\, \mathcal{S}_{w_j}^{k}(\mathbf{w}) \frac{\partial X_k(\mathbf{w})}{\partial \rho}. \tag{55}$$

Plugging (50) into (55) leads to:

$$\sum_{j=1}^{p} \frac{\partial \mu_j(\mathbf{w})}{\partial w_j} = \left(1 - g(\mathbf{w}) - \sum_{k=1}^{n} T_{x_k}(\mathbf{X}(\mathbf{w})) \frac{\partial X_k(\mathbf{w})}{\partial \rho} \right) f(\mathbf{w}). \tag{56}$$

Plugging (51) into (56) leads to (30a). The last equality in (30c) follows from (34c).

B.3 Proof of Proposition 6

If $(u, \mathbf{z}) \mapsto L(u, \mathbf{z}; \mathbf{w}, \lambda)$ is concave then for any perturbation p, the function $t \mapsto \widetilde{L}^R(t)$ defined in (46) is concave. Let $\mathbf{w} \mapsto U(\mathbf{w})$ be another utility profile that verifies (29a) and take the perturbation $R(\mathbf{w}) = U(\mathbf{w}) - U^{\star}(\mathbf{w})$. As the utility profile $\mathbf{w} \mapsto U(\mathbf{w})$ verifies (29), we get that function $t \mapsto \widetilde{L}^R(t)$ admits a zero derivative at $t = 0$ and is concave. So $\widetilde{L}^R(0) > \widetilde{L}^R(1)$ and $U^{\star}(\cdot)$ provides a strictly higher welfare than $U(\cdot)$.

If two distinct allocations $\mathbf{w} \mapsto U^{\star}(\mathbf{w})$ and $\mathbf{w} \mapsto U(\mathbf{w})$ verify (29) then, following the reasoning above, $U(\cdot)$ strictly dominates $U^{\star}(\cdot)$ and $U^{\star}(\cdot)$ strictly dominates $U(\cdot)$, a contradiction. So at most one allocation can verify (29).

References

Armstrong, Mark, "Multiproduct Nonlinear Pricing," *Econometrica*, 1996, *64* (1), 51–75.

Atkinson, Anthony B and Joseph E Stiglitz, "The design of tax structure: Direct versus indirect taxation," *Journal of Public Economics*, 1976, *6* (1-2), 55–75.

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

_ **, Mathias Dolls, Dirk Neumann, Andreas Peichl, and Sebastian Siegloch**, "Comparing inequality aversion across countries when labor supply responses differ," *International Tax and Public Finance*, 2014, *21* (5), 845–873.

— , — , — , — , **and** — , "Tax-benefit revealed social preferences in Europe and the US," *Annals of Economics and Statistics*, 2014, *113-114*, 257–289.

Barro, Robert and Charles Redlick, "Macroeconomic Effects From Government Purchases and Taxes," *The Quarterly Journal of Economics*, 2011, *126* (1), 51–102.

Basov, Suren, *Multidimensional Screening* Studies in Economic Theory, Springer-Verlag Berlin Heidelberg, 2005.

Bergstrom, Katy and William Dodds, "Optimal taxation with multiple dimensions of heterogeneity," *Journal of Public Economics*, 2021, *200*, 104442.

Bierbrauer, Felix, Pierre Boyer, and Emanuel Hansen, "Pareto-improving tax reforms and the Earned Income Tax Credit," CEPR Discussion Papers 14 853 2020.

Boerma, Job, Aleh Tsyvinski, and Alexander P Zimin, "Bunching and Taxing Multidimensional Skills," NBER Working Paper 30 015, 2022.

Bourguignon, François and Amadeo Spadaro, "Tax-benefit revealed social preferences," *Journal of Economic Inequality*, 2012, *10* (1), 75–108.

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

Cremer, Helmuth, Pierre Pestieau, and Jean-Charles Rochet, "Direct versus Indirect Taxation: The Design of the Tax Structure Revisited," *International Economic Review*, 2001, *42* (3), 781–799.

— , — , **and** — , "Capital income taxation when inherited wealth is not observable," *Journal of Public Economics*, 2003, *87* (11), 2475–2490.

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

Ferey, Antoine, Benjamin Lockwood, and Dmitry Taubinsky, "Sufficient Statistics for Nonlinear Tax Systems with Preference Heterogeneity," Technical Report, National Bureau of Economic Research 2021.

Frankel, Alexander, "Taxation of couples under assortative mating," *American Economic Journal: Economic Policy*, 2014, *6* (3), 155–77.

Gaube, Thomas, "Pareto Efficient Income Taxation and Self-financing Tax Cuts," *Mimeo*, 2022.

Gauthier, Stéphane and Guy Laroque, "Separability and public finance," *Jour-*

nal of Public Economics, 2009, *93* (11-12), 1168–1174.

Gerritsen, Aart, Bas Jacobs, Alexandra Rusu, and Kevin Spiritus, "Optimal taxation of capital income when capital returns are heterogeneous," CESifo Working Paper 8395 2020.

Golosov, Mikhail, Aleh Tsyvinski, and Iván Werning, "New Dynamic Public Finance: A User's Guide," in "NBER Macroeconomics Annual 2006, Volume 21" 2007, pp. 317–388.

_ , _ , and Nicolas Werquin, "A Variational Approach to the Analysis of Tax Systems," NBER Working Papers 20 780, 2014.

_ and Ilia Krasikov, "Multidimensional screening in public finance: the optimal taxation of couples," Not yet publicly available, n.d.

_ , Narayana Kocherlakota, and Aleh Tsyvinski, "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 2003, *70* (3), 569–587.

Gomes, Renato, Jean-Marie Lozachmeur, and Alessandro Pavan, "Differential taxation and occupational choice," *The Review of Economic Studies*, 2018, *85* (1), 511–557.

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

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

Hendren, Nathaniel, "Measuring economic efficiency using inverse-optimum weights," *Journal of Public Economics*, 2020, *187*, 104 198.

Jacobs, Bas, Egbert Jongen, and Floris Zoutman, "Revealed social preferences of Dutch political parties," *Journal of Public Economics*, 2017, *156* (C), 81–100.

Jacquet, Laurence and Etienne Lehmann, "How to tax different Incomes?," CESifo Discussion Papers 9324 2021.

_ and _ , "Optimal income taxation with composition effects," *Journal of the European Economic Association*, 2021, *19* (2), 1299–1341.

Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez, "The Optimal Income Taxation of Couples," NBER Working Papers 12685 2006.

_ , _ , and _ , "The Optimal Income Taxation of Couples as a Multi-Dimensional Screening Problem," CESifo Working Paper 2092, CESifo 2007.

Lockwood, Benjamin B and Matthew Weinzierl, "De Gustibus non est Taxandum: Heterogeneity in preferences and optimal redistribution," *Journal of*

Public Economics, 2015, *124*, 74–80.

Lorenz, Normann and Dominik Sachs, "Identifying Laffer Bounds: A Sufficient-Statistics Approach with an Application to Germany," *Scandinavian Journal of Economics*, 2016, *118* (4), 646–665.

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

_ , "Optimal tax theory : A synthesis," *Journal of Public Economics*, 1976, *6* (4), 327–358.

_ , "The Theory of Optimal Taxation," in Kenneth Arrow and Michael Intriligator, eds., *Handbook of Mathematical Economics*, Vol. 3, Elsevier, 1986, chapter 24, pp. 1197–1249.

Renes, Sander and Floris Zoutman, "As Easy as ABC? Multidimensional Screening in Public Finance," CESifo Working Paper 6 301 2017.

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

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

_ **and** _ , "Optimal taxation with rent-seeking," *The Review of Economic Studies*, 2016, *83* (3), 1225–1262.

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

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

_ **and Iván Werning**, "Mirrlees meets Diamond-Mirrlees," NBER Working Papers 22076 2016.

Spiritus, Kevin, "Characterizations for Optimal Multidimensional Taxation with Multidimensional Heterogeneity of Agents," Technical Report, mimeo 2017.

Werning, Iván, "Pareto Efficient Taxation," *mimeo*, 2007.

Supplementary materials

C Total versus Direct Responses

We define "direct responses" as the behavioral responses to a tax perturbation or to a change in the taxpayer's type if the tax schedule were linear. Let $\partial X_i^\star(\mathbf{w})/\partial\rho$ and $\partial X_i^\star(\mathbf{w})/\partial\tau_j$ denote the direct income and compensated responses of the incomes and let $\partial X_i^\star(\mathbf{w})/\partial w_j$ denote the direct responses to a change in types.

We now clarify the difference between direct and total responses. Let $\Delta_1\mathbf{x}$ denote the change in income induced by a tax perturbation or a perturbation in types if we assume the tax schedule is linear. This vector is obtained by setting $[T_{x_i x_j}]_{i,j} = 0$ in (33). We thus get *direct responses* ignoring the effects due to the non-linearity of the tax schedule:

$$\Delta_1\mathbf{x} = \left[\mathcal{C}_{x_i x_j}\right]_{i,j}^{-1} \cdot \mathrm{d}B,$$

where $\mathrm{d}B$ is the column vector on the right-hand side of (33).

When the tax function is nonlinear, this "first" change $\Delta_1\mathbf{x}$ in income induces a change $[T_{x_i,x_j}]_{i,j} \cdot \Delta_1\mathbf{x}$ in the vector of marginal tax rates that generates a "second" change in income through compensated responses that are given by:

$$\Delta_2\mathbf{x} = -\left[\frac{\partial X_i^\star}{\partial\tau_j}\right]_{i,j} \cdot [T_{x_i x_j}]_{i,j} \cdot \Delta_1\mathbf{x},$$

which in turn generates a further change in marginal tax rates. Hence, the k^{th} change in income $\Delta_k\mathbf{x}$ is related to $k-1^{\text{th}}$ change in income $\Delta_{k-1}\mathbf{x}$ by:

$$\Delta_k\mathbf{x} = -\left[\frac{\partial X_i^\star}{\partial\tau_j}\right]_{i,j} \cdot [T_{x_i x_j}]_{i,j} \cdot \Delta_{k-1}\mathbf{x},$$

and so:

$$\Delta_k\mathbf{x} = \left(-\left[\frac{\partial X_i^\star}{\partial\tau_j}\right]_{i,j} \cdot [T_{x_i x_j}]_{i,j}\right)^{k-1} \cdot \Delta_1\mathbf{x}.$$

Adding all the effects and assuming convergence leads to a *total* effect:

$$\Delta\mathbf{x} = \sum_{k=1}^{\infty} \Delta_k\mathbf{x} = \sum_{k=1}^{\infty} \left(-\left[\frac{\partial X_i^\star}{\partial\tau_j}\right]_{i,j} \cdot [T_{x_i x_j}]_{i,j}\right)^{k-1} \cdot \Delta_1\mathbf{x} = \left(I_n + \left[\frac{\partial X_i^\star}{\partial\tau_j}\right]_{i,j} \cdot [T_{x_i x_j}]_{i,j}\right)^{-1} \cdot \Delta_1\mathbf{x}$$

$$= \left(I_n + \left[\mathcal{C}_{x_i x_j}\right]_{i,j}^{-1} \cdot [T_{x_i x_j}]_{i,j}\right)^{-1} \cdot \left[\mathcal{C}_{x_i x_j}\right]_{i,j}^{-1} \cdot \mathrm{d}B = \left[\mathcal{C}_{x_i x_j} + T_{x_i x_j}\right]_{i,j}^{-1} \cdot \mathrm{d}B,$$

where I_n denotes the identity matrix of rank n. We thus retrieve (33), which we showed in Appendix A.2 leads to (11), and we thus obtain *total responses* including the effects due to the non-linearity of the tax schedule:

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j} = \left(I_n + \left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i x_j}]_{i,j}\right)^{-1} \cdot \left[\frac{\partial X_i^\star(\mathbf{w})}{\partial \tau_j}\right]_{i,j}, \tag{57a}$$

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial \rho}\right]_{i} = \left(I_n + \left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i x_j}]_{i,j}\right)^{-1} \cdot \left[\frac{\partial X_i^\star(\mathbf{w})}{\partial \rho}\right]_{i}, \tag{57b}$$

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]_{i,j} = \left(I_n + \left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i x_j}]_{i,j}\right)^{-1} \cdot \left[\frac{\partial X_i^\star(\mathbf{w})}{\partial w_j}\right]_{i,j}. \tag{57c}$$

Equations (34a) and (57a) imply that Part ii) of Assumption 1 is equivalent to assuming that the matrix $I_n + \left[\partial X_i^\star / \partial \tau_j\right]_{i,j} \cdot [T_{x_i x_j}]_{i,j}$ is positive definite despite the nonlinearity of the tax schedule.

D Appendix on the Numerical Simulations

In this section, we document the simulations for the unrestricted tax schedule. The simulations for the individual and joint tax schedules are more standard; we include their documentation with the source code of the simulations.

We assume $n = p = 2$. Denote the tax liability assigned to type $\mathbf{w}$ as $\mathcal{T}(\mathbf{w}) \stackrel{\text{def}}{\equiv} T(\mathbf{X}(\mathbf{w}))$. Let $\mathcal{J}(\mathbf{w})$ denote the inverse of the Jacobian matrix associated with the mapping $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$:

$$\mathcal{J}(\mathbf{w}) \stackrel{\text{def}}{\equiv} \begin{pmatrix} \frac{\partial X_1(\mathbf{w})}{\partial w_1} & \frac{\partial X_1(\mathbf{w})}{\partial w_2} \\ \frac{\partial X_2(\mathbf{w})}{\partial w_1} & \frac{\partial X_2(\mathbf{w})}{\partial w_2} \end{pmatrix}^{-1}.$$

Given the mapping $\mathbf{w} \mapsto \mathcal{T}(\mathbf{w})$ and the allocation $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$, we find the marginal tax rates for a type-$\mathbf{w}$ taxpayer:

$$T_{x_i}(\mathbf{X}(\mathbf{w})) = \sum_{k=1}^{n} \mathcal{T}_{w_k}(\mathbf{w}) \mathcal{J}_{k,i}(\mathbf{w}). \tag{58}$$

Considering that individual preferences (31) do not feature income effects, we rewrite optimal tax condition (30a) in the type space:

$$(1 - g(\mathbf{w}))\, f(\mathbf{w}) = \sum_{j=1}^{p} \frac{\partial \left(\sum_{1 \le i,k \le n} \mathcal{T}_{w_k}(\mathbf{w})\, \mathcal{J}_{k,i}(\mathbf{w})\, \mathcal{A}_{j,i}(\mathbf{w})\, f(\mathbf{w}) \right)}{\partial w_j},$$

with boundary conditions:

$$\forall\ \mathbf{w} \in \partial \mathcal{W} \quad : \quad \sum_{1 \le i,j,k \le n} \mathcal{T}_{w_k}(\mathbf{w})\, \frac{\partial X_k^{-1}(\mathbf{X}(\mathbf{w}))}{\partial x_i}\, \mathcal{A}_{j,i}(\mathbf{w})\, e_j(\mathbf{w}) = 0.$$

The simulation algorithm then works as follows. We start from some initial value of the government budget multiplier λ.

1. Start a loop from an initial tax function. Denote the tax function in iteration ℓ by $\mathbf{x} \mapsto T^{(\ell)}(\mathbf{x})$. Starting from the tax function $\mathbf{x} \mapsto T^{(\ell)}(\mathbf{x})$, we use the individual first-order conditions to calculate the corresponding allocation $\mathbf{w} \mapsto \mathbf{X}^{(\ell)}(\mathbf{w})$, and the corresponding inverse Jacobian $\mathbf{w} \mapsto \mathcal{J}^{(\ell)}(\mathbf{w}) \stackrel{\text{def}}{\equiv} \partial \left(X_k^{(\ell)}\right)^{-1} (\mathbf{X}(\mathbf{w}))/\partial x_i$.

2. We use the Partial Differential Equation toolbox 3.5 in MATLAB R2020b to find the mapping $\mathbf{w} \mapsto \mathcal{T}^{(\ell+1)}(\mathbf{w})$ that solves the Partial Differential Equation using the finite element method:

$$(1 - g(\mathbf{w}))\, f(\mathbf{w}) = \sum_{j=1}^{p} \frac{\partial \left(\sum_{1 \le i,k \le n} \mathcal{T}_{w_k}^{(\ell+1)}(\mathbf{w})\, \mathcal{J}_{k,i}^{(\ell)}(\mathbf{w})\, \mathcal{A}_{j,i}(\mathbf{w})\, f(\mathbf{w}) \right)}{\partial w_j}, \tag{59a}$$

   with boundary conditions:

$$\forall\ \mathbf{w} \in \partial \mathcal{W} \quad : \quad \sum_{1 \le i,j,k \le n} \mathcal{T}_{w_k}^{(\ell+1)}(\mathbf{w})\, \mathcal{J}_{k,i}^{(\ell)}(\mathbf{w})\, \mathcal{A}_{j,i}(\mathbf{w})\, e_j(\mathbf{w}) = 0. \tag{59b}$$

   In (59), welfare weights $g(\mathbf{w})$ are computed endogenously through (7), and matrices $\mathcal{A}(\mathbf{w})$ through (30c), both as functions of the allocation $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$. The allocation is computed from marginal tax rates $\mathbf{x} \mapsto T_{x_i}(\mathbf{x})$. Marginal tax rates are deduced from $\mathbf{w} \mapsto \mathcal{T}_{w_k}^{(\ell+1)}(\mathbf{w})$ and from $\mathbf{w} \mapsto \mathcal{J}^{(\ell)}(\mathbf{w})$ using (58). By keeping the Jacobian $\mathbf{w} \mapsto \mathcal{J}^{(\ell)}(\mathbf{w})$ fixed, the Partial Differential Equation remains solvable by MATLAB.

3. We repeat these steps until the process converges to a fixed point $\mathbf{x} \mapsto T(\mathbf{x})$. As convergence criterion, we require that for more than 99.9% of all

points on the simulation mesh, the difference of the tax liability with the previous iteration is smaller than 0.5% or 50 USD, whichever is larger.[13]

We repeat this algorithm for various values of λ until the budget constraint (4) is fulfilled.

While solving the partial differential equation (59a) for $\mathcal{T}^{(\ell+1)}$, MATLAB's solver will inspect different candidate solutions $\mathcal{T}^{(\ell+1)}$ with corresponding partial derivatives $\mathcal{T}_{w_j}^{(\ell+1)}$. Unavoidably, some candidates will correspond through (58) to marginal tax rates $T_{x_l}^{(\ell+1)}$ which are larger than one for at least some taxpayers. Since the individual optimization problem yields no solution when $T_{x_l}^{(\ell+1)} > 1$, the algorithm halts when such a point is reached. Since we cannot control the candidate solutions $\mathcal{T}^{(\ell+1)}$ inspected by MATLAB, we need a way to guide the solver past any points that imply $T_{x_l}^{(\ell+1)} > 1$. Suppose that straightforward application of (58) yields candidate marginal tax rates denoted by $T_{x_j}^{*(\ell+1)}$. We then use instead the following marginal tax rates to solve the individual optimization problem and to compute $g(\mathbf{w})$ and $\mathcal{A}(\mathbf{w})$:

$$\forall \mathbf{w}: \qquad T_{x_j}^{(\ell+1)}(\mathbf{X}(\mathbf{w})) \equiv \begin{cases} \dfrac{T_{x_j}^{*(\ell+1)}(\mathbf{X}(\mathbf{w}))}{T_{x_j}^{*(\ell+1)}(\mathbf{X}(\mathbf{w})) + 1 - T_{x_j}^{(\ell)}(\mathbf{X}(\mathbf{w}))} & \text{if } T_{x_j}^{*(\ell+1)}(\mathbf{X}(\mathbf{w})) \geq 0, \\ T_{x_j}^{*(\ell+1)}(\mathbf{X}(\mathbf{w})) & \text{if } T_{x_j}^{*(\ell+1)}(\mathbf{X}(\mathbf{w})) < 0. \end{cases} \tag{60}$$

Eq. (60) ensures that $T_{x_j}^{(\ell+1)}(\mathbf{X}(\mathbf{w})) < 1$, given that $T_{x_l}^{(\ell)} < 1$. Moreover, $T_{x_j}^{(\ell+1)}(\mathbf{X}(\mathbf{w}))$ is continuous and increasing in $T_{x_j}^{\star(\ell+1)}(\mathbf{X}(\mathbf{w}))$. Finally, if the algorithm converges, it converges to the correct schedule $\mathbf{w} \mapsto \mathcal{T}(\mathbf{w})$, i.e. if $T_{x_j}^{\star(\ell+1)}(\mathbf{X}(\mathbf{w})) = T_{x_j}^{(\ell)}(\mathbf{X}(\mathbf{w}))$, then one has $T_{x_j}^{(\ell+1)}(\mathbf{X}(\mathbf{w})) = T_{x_j}^{\star(\ell+1)}(\mathbf{X}(\mathbf{w})) = T_{x_j}^{(\ell)}(\mathbf{X}(\mathbf{w}))$.

The Partial Differential Equation Toolbox creates an evenly spaced mesh for the skills of the individuals. It is not possible to directly increase the detail of the mesh in certain regions. To have sufficient detail where necessary, for example near the boundaries and where most households are, we use a transformation of the types. We use the following utility function:

$$\mathcal{U}(c, x; w) = c - \sum_{i=m,f} \frac{\varepsilon_i}{1+\varepsilon_i} x_i^{\frac{1+\varepsilon_i}{\varepsilon_i}} \left[W_i(w_i)\right]^{-\frac{1}{\varepsilon_i}},$$

where $W_i(w_i)$ are transformations of the individual abilities w_i. For given ob-

[13] Given that the lower bound for the income space equals 3.000 USD in the empirical baseline, and the upper bound equals 500.000 USD, this is a high level of precision for practical purposes.

servations of the incomes and for given marginal tax rates, we find for an optimizing individual:

$$W_i(w_i) = x_i(1 - T_{x_i})^{-\varepsilon_i}.$$

An appropriate choice of the transformations $W_i(w_i)$ allows increasing the detail of the mesh grid where desired. We use the following transformations:

$$W_i(w_i) = \frac{\int_{\underline{w}_i}^{w_i} \frac{1}{D_i(\hat{w}_i)} \mathrm{d}\hat{w}_i}{\int_{\underline{w}_i}^{\overline{w}_i} \frac{1}{D_i(\hat{w}_i)} \mathrm{d}\hat{w}_i} (\overline{w}_i - \underline{w}_i) + \underline{w}_i,$$

where $D_i(w_i)$ are functions that determine the detail of the mesh grid. Note that $W_i(\underline{w}_i) = \underline{w}_i$ and $W_i(\overline{w}_i) = \overline{w}_i$. The transformations $w_i \mapsto W_i(w_i)$ thus maintain the domain of the types. Furthermore:

$$\frac{\mathrm{d}W_i(w_i)}{\mathrm{d}w_i} = \frac{\frac{1}{D_i(w_i)}}{\int_{\underline{w}_i}^{\overline{w}_i} \frac{1}{D_i(\hat{w}_i)} \mathrm{d}\hat{w}_i} (\overline{w}_i - \underline{w}_i) + \underline{w}_i > 0.$$

With an evenly spaced grid for $\mathbf{w}$, the grid for $\boldsymbol{W}(\mathbf{w})$ will be more detailed where $\mathrm{d}W_i(w_i)/\mathrm{d}w_i$ is smaller, and thus $D_i(w_i)$ is larger. We increase the detail of the simulation grid near the lower bounds, where the income densities are larger, by choosing the detail functions:

$$D_i(w_i) = 5 \frac{(\overline{w}_i - w_i)^8}{\max_{w_i} [(\overline{w}_i - w_i)^8]} + 0.1.$$

We approximate he inverse Jacobian matrices $\mathcal{J}(\mathbf{w})$ of the allocation $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ for distances $\mathrm{d}w_k = 10^{-7}$ in the skill domain. We smooth the resulting inverse Jacobian by interpolating one fourth of the nodes of the mesh in each dimension using a spline method and by extrapolating linearly for the bottom 0.16% of the population.