Optimal Taxation with Multiple Incomes and Types*

Kevin SPIRITUS† Etienne LEHMANN‡ Sander RENES§

Floris T. ZOUTMAN¶

January 19, 2022

Abstract

We analyze the optimal nonlinear income tax schedule when taxpayers earn multiple incomes and differ along many unobserved dimensions. We derive the necessary conditions for the government's optimum using both a tax perturbation and a mechanism design approach, and show that 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

*We wish to thank for their valuable comments and 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. Furthermore, this paper benefited from comments and suggestions made by participants at the 2011 Nake Conference, Utrecht, the 2013 CESifo Area Conference on Public Economics, Munich, the annual meeting of the European Economic Association 2015, Mannheim, the annual meeting of the Italian Society of Public Economics 2016, Lecce, the FIBE conference 2017, Bergen, the IIPF 2017 in Tokyo, the 2017 NTA in Philadelphia, the 2017 Tax workshop in Rotterdam, the 2019 IIPF conference in Glasgow; and seminar participants at the Erasmus School of Economics, the Norwegian University of Science and Technology, the Norwegian School of Economics, the University of Cologne the University of Konstanz, the University of Mannheim, Umeå Universitet, the Paris Tax Workshop at Paris II and the Center for European Economic Research (ZEW).

†Erasmus School of Economics, Erasmus Universiteit Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands. Kevin SPIRITUS is also a member of Department of Economics, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium. Email: spiritus@ese.eur.nl. Kevin SPIRITUS acknowledges the financial support of the Belgian Federal Science Policy Office (BELSPO) via the BRAIN.be project BR/121/A5/CRESUS.

‡CRED, Université Panthéon-Assas Paris II, 12 Place du Panthéon, 75 234 Paris Cedex 05, France. Email: elehmann@u-paris2.fr. Etienne LEHMANN is also member of CEPR, CESifo, IZA, IDEP and TEPP.

§Erasmus Universiteit Rotterdam, Department of Business Economics, P.O. box 1738, 3000 DR Rotterdam, The Netherlands. Email srenes@ese.eur.nl.

¶NHH Norwegian School of Economics, Department of Business and Management Science, Helleveien 30, 5045 Bergen, Norway. Email: floris.zoutman@nhh.no. Floris T. ZOUTMAN is also a member of the Norwegian Center for Taxation (NOCET) and CESifo.

Although the one-dimensional population is an extremely useful model for computations and examination of particular issues, it is not, in that respect, an accurate representation of reality.

– Mirrlees (1986)

The bi-variate composition of income more generally introduces the question as to how the optimal design of taxation depends on the degree of correlation of different income sources.

– Atkinson and Stiglitz (2015)

I Introduction

Households typically earn incomes from multiple sources, such as salaries, business incomes, dividends, interests or capital gains. Moreover, households often consist of multiple individuals, each of whom have their own earning abilities. The tax treatment of these different sources of income differs widely between countries. Optimal tax theory generally neglects the fact that households earn multiple incomes and that households differ in multiple dimensions of unobserved heterogeneity. While Mirrlees (1976, Section 4) derives an optimal tax formula in such a context, it offers virtually no guidance for policymakers. Very few results have been derived since. 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] 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 decomposed into 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 an ABC-formula (see Equation (19a)) that is 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

[1]The one-dimensional model has been 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.

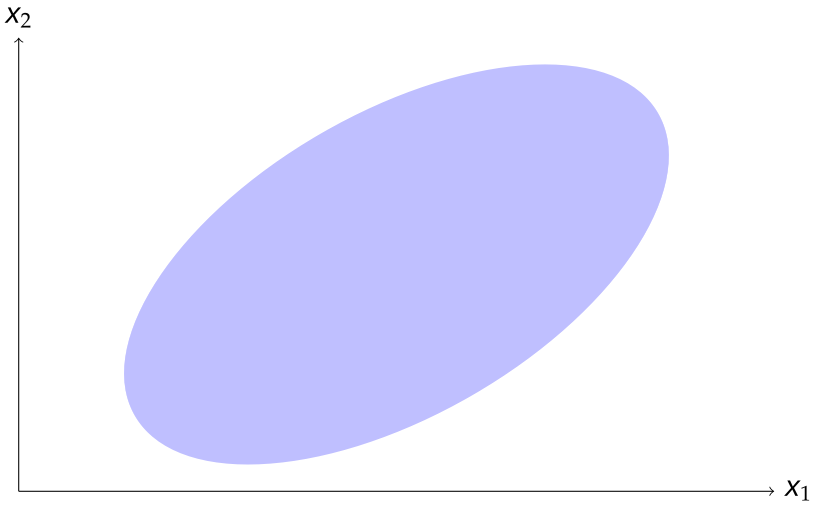


Figure 1: A perturbation in two dimensions. x_1 and x_2 are incomes. Perturbing the tax liabilities within the shaded area, affects the marginal tax rates along the boundary of the shaded area. It is not possible to change the marginal tax rates at one combination of incomes without affecting the marginal tax rates at other combinations of incomes.

by the optimal tax formula for a single tax base. In the one-dimensional case, the optimal marginal tax rate at one income level is 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 can only relate mechanical and income effects within a subset of income bundles (for example, the shaded area in Figure 1) to compensated responses all around this subset (the border of the shaded area in Figure 1). Hence, one cannot study the effects of a change in the tax gradient at one combination of incomes, without causing additional changes in the tax gradients at other combinations of incomes.

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 quasi-linear 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. We show that 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 use a tax perturbation approach similar to Golosov et al. (2014) to 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 Pareto improving tax reform. We hence extend the revealed social preference approach of Bourguignon and Spadaro (2012), Bargain et al. (2014a,b), 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 unique and 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 this solution is unique. Therefore, 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, some scholars may doubt the tax perturbation approach rigorously derives the optimal tax formulas. We address their concerns. The tax perturbation approach pioneered by Saez (2001) in the one-dimensional case states that the optimal tax schedule should be such that some incremental tax reforms do not lead to first-order improvements. Golosov et al. (2014) extend the approach of Saez (2001) to the case with multiple incomes. Furthermore, they address the concern that Saez (2001) only considers reforms that consist in a uniform change in the marginal tax rate in a small interval of the income distribution and a uniform change in tax liability at all incomes above. While these tax reforms are very convenient to convey the economic intuitions behind the optimal tax formula, these reforms generate kinks whose effects are neglected. Golosov et al. (2014) consider the responses to all types of smooth tax reforms. We address a remaining concern, namely that both Saez (2001) and Golosov et al. (2014) *assume* that incomes respond smoothly to the size of tax perturbations. We instead apply the implicit function theorem to *demonstrate* how incomes respond smoothly to the size of tax perturbations. For that purpose, we make explicit the assumptions on the tax schedule that 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. With these incomes, it is possible to use the first-order incentive constraints to deduce the partial derivatives of the attained utilities with respect to the types. However, many different income allocations fulfill the necessary conditions for the optimum, and 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. This mapping 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. Similar subprograms are implicitly found in the settings of, for example, Atkinson and Stiglitz (1976), Golosov et al. (2003, 2007) and Gerritsen et al. (2020).

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. 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.

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

In this section we introduce the set-up of our model. First, we introduce the preferences and the budget constraints of the households, and we derive the first-order conditions for their optimization problem. Next, we turn to the preferences and the budget constraint of the government.

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 $W \subset \mathbb{R}^p$ and is assumed to be closed and convex. Types are distributed over the type space according to a twice continuously differentiable density denoted by $f(\cdot)$, which is positive over W.

Taxpayers make $n \geq 2$ different choices. This implies the existence of n different 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.[2] 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 $U(c, \mathbf{x}; \mathbf{w})$ defined over $\mathbb{R}_+^{n+1} \times W$. Taxpayers enjoy utility from consumption but endure disutility to obtain income, so $U_c > 0$ and $U_{x_i} < 0$. Let $C(\cdot, \mathbf{x}; \mathbf{w})$ be the inverse of $U(\cdot, \mathbf{x}; \mathbf{w})$. That is, a taxpayer of type $\mathbf{w}$ earning incomes $\mathbf{x}$ should consume $C(u, \mathbf{x}; \mathbf{w})$ to enjoy utility level u. It follows from the implicit function theorem that $C_u = 1/U_c$ and $C_{x_i} = -U_{x_i}/U_c$. We assume the utility function $U(\cdot, \cdot; \mathbf{w})$ is weakly

[2] 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.

concave in $(c, \mathbf{x})$ and indifference sets defined by $c = C(u, \mathbf{x}; \mathbf{w})$ 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. Therefore, a taxpayer of type $\mathbf{w}$ solves:

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

Let $\mathbf{X}(\mathbf{w}) \overset{\text{def}}{\equiv} (X_1(\mathbf{w}),\ldots,X_n(\mathbf{w}))$ denote the solution to this program and let $C(\mathbf{w}) \overset{\text{def}}{\equiv} \sum_{i=1}^{n} X_i(\mathbf{w}) - T(\mathbf{X}(\mathbf{w}))$ denote the corresponding consumption of taxpayers of type $\mathbf{w}$. In addition, we denote the marginal rate of substitution between the i^{th} income and consumption at any bundle $(c, \mathbf{x})$ as:[3]

$$S^i(c, \mathbf{x}; \mathbf{w}) \overset{\text{def}}{\equiv} -\frac{U_{x_i}(c, \mathbf{x}; \mathbf{w})}{U_c(c, \mathbf{x}; \mathbf{w})} = C_{x_i}(U(c, \mathbf{x}; \mathbf{w}), \mathbf{x}; \mathbf{w}) > 0. \tag{2}$$

We can then write the first-order conditions for taxpayers of type $\mathbf{w}$ as:

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

II.2 Government

The government's budget constraint is given by:[4]

$$B \overset{\text{def}}{\equiv} \iint_W T(\mathbf{X}(\mathbf{w})) f(\mathbf{w}) d\mathbf{w} - E, \tag{4}$$

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

$$O \overset{\text{def}}{\equiv} \iint_W \Phi(U(\mathbf{w}); \mathbf{w}) f(\mathbf{w}) d\mathbf{w}, \tag{5}$$

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), taking into account the households' optimization in (1).

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

$$L \overset{\text{def}}{\equiv} B + \frac{O}{\lambda} = \iint_W \left[T(\mathbf{X}(\mathbf{w})) + \frac{\Phi(U(\mathbf{w}); \mathbf{w})}{\lambda}\right] f(\mathbf{w}) d\mathbf{w} - E, \tag{6}$$

where λ is the Lagrange multiplier of the government's budget constraint. The Lagrange multiplier λ is interpreted as the social value of public funds (expressed in social utility units). 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:

$$g(\mathbf{w}) \overset{\text{def}}{\equiv} \frac{\Phi_u(U(\mathbf{w}); \mathbf{w}) \, U_c(C(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w})}{\lambda} \geq 0. \tag{7}$$

[3] While the functions S^i and C_{x_i} both correspond to the marginal rate of substitution, they admit different arguments. For given pre-tax incomes $\mathbf{x}$ and types $\mathbf{w}$, the marginal rate of substitution S^i admits consumption c as an argument, while function C admits utility level u as an argument. Hence, when differentiating with respect to incomes or types, the derivatives of S^i hold for a given consumption income c, while the derivatives of C hold along the indifference surface $c = C(u, \mathbf{x}; \mathbf{w})$.

[4] We use $\iint$ as shorthand notation to denote an integral over several dimensions.

III The Tax Perturbation Approach

We first study the government's optimization problem using the tax perturbation method. The intuition behind the tax perturbation method is well known in the one-dimensional case since Saez (2001). Golosov et al. (2014) extend the method to the case with many incomes and types. We make a number of contributions to this literature. We start by providing sufficient conditions on the tax schedule under which we can *show* that taxpayers' choices respond smoothly to tax reforms. Up to now, the tax perturbation literature *assumed* that taxpayers respond smoothly to small changes in the tax reform. We show how this imposes an implicit requirement on the smoothness of the tax function.

With the given assumptions about the smoothness of the tax schedule, we provide a condition under which a given tax reform is socially desirable (Section III.1). This condition allows us to study the optimal tax schedule: if no tax reform exists that is socially desirable, then we are in the optimum.

We provide new intuitions by splitting the search for the optimal tax schedule into two consecutive steps. In the first step, the shapes of the isotax curves are determined.[5] In the second step, the tax liabilities for each of the isotax curves are determined. We show that once the shapes of the isotax curves are known, the tax liability belonging to each isotax curve can be determined through an ABC-style (Diamond, 1998) optimal tax formula (III.2).

Next, we derive conditions for the fully optimal tax schedule, and interpret them in economic terms. We use this formula for the optimal tax schedule to provide an equation that allows performing an inverse optimum exercise, extracting the revealed welfare weights from a given tax schedule. Using these revealed welfare weights, we explain how to test whether a given tax schedule is Pareto dominated. Finally, we show how to construct Pareto-improving tax reforms when the tax schedule is indeed Pareto dominated (III.3).

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) 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.

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* $t \gtrless 0$ leads to the perturbed tax schedule $\mathbf{x} \mapsto T(\mathbf{x}) - t\,R(\mathbf{x})$. If $R(\mathbf{x}) > 0$ and $t > 0$ or if $R(\mathbf{x}) < 0$ and $t < 0$, the perturbation decreases the tax liabilities at incomes $\mathbf{x}$. The reverse occurs if $R(\mathbf{x}) < 0$ and $t > 0$ or if $R(\mathbf{x}) > 0$ and

[5] An isotax "curve" is defined as the loci of incomes $\mathbf{x}$ that are associated with the same tax liability $T(\mathbf{x})$. 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.

$t < 0$. Given a tax perturbation in the direction $R(\cdot)$, the utility of taxpayers of type $\mathbf{w}$ becomes a function of magnitude t through:

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

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

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

If we perturb the tax schedule or any of the characteristics of the households, then the households will update their choices 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 we can apply the implicit function theorem, then it necessarily follows that the function $\tilde{\mathbf{X}}^R(\mathbf{w},t)$ that solves (8) is continuously differentiable for t close to 0.

Assumption 1. *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 $\mathbf{w} \in W$, the second-order conditions associated to (1) are strictly verified, i.e. the matrix $[S^i_{x_j} + S^j S^i_c + T_{x_i x_j}]_{i,j}$ is positive definite at $c = C(\mathbf{w})$ and $\mathbf{x} = \mathbf{X}(\mathbf{w})$.[6]

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

Assumption 1.*i*) rules out kinks like those in piecewise linear tax schedules.[7] 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 $\tilde{\mathbf{X}}^R(\mathbf{w},t)$ with respect to the size t of the perturbation tends to infinity, so the perturbation approach cannot be used.

Given that we assume that the indifference sets defined by $c = C(u,\mathbf{x};\mathbf{w})$ are strictly convex, Assumption 1 is automatically verified if the tax schedule is linear (see Appendix A.1).

[6] We let $[a(k)]_k$ denote a column vector whose k^{th} row is $a(k)$, $[A(k,\ell)]_{k,\ell}$ denotes a square matrix of size n 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.

[7] In reality, most convex kinks in the tax schedule do not cause significant bunching. It is thus reasonable to assume that taxpayers base their decisions on a smooth approximation of the actual tax schedule. Smoothed approximations of piecewise linear schedules are twice-differentiable as assumed in 1 *i*).

Moreover, Assumption 1 amounts to assuming that the budget mapping $\mathbf{x} \mapsto \sum_{i=1}^{n} x_i - T(\mathbf{x})$ induced by the tax function is everywhere either concave, linear or less convex than the corresponding indifference sets. Geometrically, it implies that for each type $\mathbf{w}$, the indifference set defined by $c = C(U(\mathbf{w}), \mathbf{x}; \mathbf{w})$ 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. 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 actually the case (Mirrlees, 1971, p. 188).

The effects of any tax perturbation can be decomposed into the effects of two types of prototypical tax reforms (see Appendix A.2 for a proof). The first is the *lump sum* perturbation which decreases the tax liability by a uniform amount:

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

where we use r 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:

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

where we use t_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 $\partial X_i(\mathbf{w})/\partial r$ and $\partial X_i(\mathbf{w})/\partial t_j$ the responses for 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.[8] 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 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 $R(\mathbf{X}(\mathbf{w}))\,\mathrm{d}t$, combined with compensated perturbations of each of the n marginal tax rates (10b) of respective sizes $R_{x_j}(\mathbf{X}(\mathbf{w}))\,\mathrm{d}t$. We thus get (see Appendix A.2):

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

[8] Strictly speaking, these responses do not just depend on the type $\mathbf{w}$, but also on the consumption $c = C(\mathbf{w})$ and the incomes $\mathbf{x} = \mathbf{X}(\mathbf{w})$ of the evaluated types, as well as 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. (Saez, 2001). By applying the implicit function theorem, the income responses $\partial X_i(\mathbf{w})/\partial r$ and compensated responses $\partial X_i(\mathbf{w})/\partial t_j$ encapsulate this "circular process" through the endogeneity of the marginal tax rates. We therefore refer to these responses as *total* responses. Conversely, the empirical literature typically estimates *direct* responses by assuming the tax schedule is linear, thus ignoring the circularity of the process. We discuss the relation between direct and total responses in Appendix A.3.

Note that our approach differs from that of Golosov et al. (2014). While they *assume* that the function $t \mapsto \hat{\mathbf{X}}^R(\mathbf{w},t)$ is Lipschitz continuous, we *show* that $t \mapsto \hat{\mathbf{X}}^R(\mathbf{w},t)$ is continuously differentiable at $t = 0$, such that Equation (11) holds whenever the unperturbed tax schedule verifies Assumption 1.

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":

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

where $\lambda > 0$ denotes the social value of public funds indicating how the government trades off social welfare (5) for tax revenue (4). 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).

To analyze the effect of a tax perturbation on the government's budget constraint (4), we compute the response of the tax liabilities $T(\hat{\mathbf{X}}^R(\mathbf{w},t)) - t\,R(\hat{\mathbf{X}}^R(\mathbf{w},t))$ 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 $(\partial \hat{X}_i^R(\mathbf{w},t)/\partial t)|_{t=0}$, for $i = 1, \ldots, n$. The total change in the tax liability due to the perturbation thus equals:

$$\left.\frac{\partial T(\hat{\mathbf{X}}^R(\mathbf{w},t)) - t\,R(\hat{\mathbf{X}}^R(\mathbf{w},t))}{\partial t}\right|_{t=0} = \underbrace{-R(\mathbf{X}(\mathbf{w}))}_{\text{Mechanical effects}} + \underbrace{\sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \left.\frac{\partial \hat{X}_i^R(\mathbf{w},t)}{\partial t}\right|_{t=0}}_{\text{Behavioral effects}}. \tag{13}$$

Combining Equations (11) and (13) leads to:

$$\begin{aligned}\left.\frac{\partial T(\hat{\mathbf{X}}^R(\mathbf{w},t)) - t\,R(\hat{\mathbf{X}}^R(\mathbf{w},t))}{\partial t}\right|_{t=0} = & -\left[1 + \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial r}\right] R(\mathbf{X}(\mathbf{w})) \\ & + \sum_{1 \le i,j \le n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial t_j} R_{x_j}(\mathbf{X}(\mathbf{w})).\end{aligned} \tag{14}$$

Next, we evaluate the effect of the tax perturbation on the social objective. Under Assumption 1, behavioral responses only induce second-order effects on the taxpayers' utilities. Therefore, the tax perturbation only affects the social objective through mechanical effects. Applying the envelope theorem to social welfare $\Phi(U)$ after inserting (8) and using (7) leads to:

$$\frac{1}{\lambda} \left.\frac{\partial \Phi\left(\hat{U}^R(\mathbf{w},t);\mathbf{w}\right)}{\partial t}\right|_{t=0} = g(\mathbf{w})\, R\left(\mathbf{X}(\mathbf{w})\right). \tag{15}$$

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 magnitude $\ell^R(\cdot)$ of the lump sum transfer that makes the overall combination budget-balanced. However, if the social value of public funds λ is normalized such that a lump-perturbation (10a) has no impact on the Lagrangian (6), i.e. if:

$$0 = \iint_{\mathcal{W}} \left[1 - g(\mathbf{w}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho} \right] f(\mathbf{w}) d\mathbf{w}, \tag{16}$$

then 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 combined perturbation on social welfare Ω. This finding is expressed in the following proposition (proven in Appendix A.4):

Proposition 1. *Under Assumption 1, and when λ is such that (16) holds, a tax perturbation in the direction $R(\cdot)$ with $t > 0$ (respectively $t < 0$) combined with a lump-sum rebate of the net budget surplus generated by the perturbation is welfare improving if and only if $(\partial \mathscr{L}^R(t,\lambda)/\partial t)|_{t=0} > 0$ (resp. $(\partial \mathscr{L}^R(t,\lambda)/\partial t)|_{t=0} < 0$), where:*

$$\left.\frac{\partial \mathscr{L}^R(t,\lambda)}{\partial t}\right|_{t=0} = \iint_{\mathcal{W}} \Bigg\{ \left[g(\mathbf{w}) - 1 + \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho} \right] R(\mathbf{X}(\mathbf{w})) \\ + \sum_{1 \le i,j \le n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial t_j} R_{x_j}(\mathbf{X}(\mathbf{w})) \Bigg\} f(\mathbf{w}) d\mathbf{w}. \tag{17}$$

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. However, Proposition 1 also holds outside of the social optimum.

III.2 Optimal taxation for given isotax curves

It is possible to decompose the design of the optimal tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ 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. The second step concerns the assignment of a specific tax liability to each isotax curve. In this subsection, we apply Proposition 1 to show that the solution to the second step is characterized by a tax formula reminiscent of the ABC-formula of Saez (2001) that characterizes the optimal schedule with a one-dimensional base.

We thus decompose the tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ into two consecutive mappings. The first mapping defines a *taxable income* $y = \mathtt{G}(\mathbf{x}) \in \mathbb{R}$ for each combination of incomes $\mathbf{x}$. Values of $\mathbf{x}$ with the same tax liability $T(\mathbf{x})$ map to equal values of $\mathtt{G}(\mathbf{x})$. Assuming that $\mathtt{G}(\cdot)$ 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. The second mapping, denoted $\mathcal{T}$, assigns a tax liability to each taxable income y so that we have $T(\mathbf{x}) = \widetilde{T}(G(\mathbf{x}))$.[9] We then consider perturbations of the form $\mathbf{x} \mapsto \widetilde{T}(G(\mathbf{x})) - \tau R(G(\mathbf{x}))$, where the direction $R(\cdot)$ admits taxable income $G(\mathbf{x})$ as its single argument. We thus only consider perturbations of the function $\widetilde{T}$, while isotax curves $y = G(\mathbf{x})$ are preserved. We denote as $Y(\mathbf{w}) = G(\mathbf{X}(\mathbf{w}))$ the realized taxable income for taxpayers of type $\mathbf{w}$ under the unperturbed tax schedule, and as $\hat{Y}^R(\mathbf{w},\tau) = G(\hat{\mathbf{X}}^R(\mathbf{w},\tau))$ the realized taxable income of taxpayers of type $\mathbf{w}$ under the perturbed tax schedule $\mathbf{x} \mapsto \widetilde{T}(G(\mathbf{x})) - \tau R(G(\mathbf{x}))$.

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

$$\frac{\partial Y(\mathbf{w})}{\partial \rho} = \sum_{i=1}^{n} G_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho}. \tag{18a}$$

We show in Appendix A.5 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 compensated responses of taxable income for taxpayers of type $\mathbf{w}$:

$$\frac{\partial Y(\mathbf{w})}{\partial \tau} \overset{\text{def}}{=} -\sum_{1 \leq i,j \leq n} G_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial t_j} G_{x_j}(\mathbf{X}(\mathbf{w})). \tag{18b}$$

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.5 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:*

$$\frac{T'(y)}{1 - T'(y)} = \frac{1}{\varepsilon(y)} \frac{1 - M(y)}{y\, m(y)} \int_{z=y}^{\infty} \left[1 - \overline{g}(z) - T'(z) \frac{\partial \overline{Y}(z)}{\partial \rho} \right] \frac{m(z)}{1 - M(y)} \mathrm{d}z, \tag{19a}$$

together with transversality condition:

$$0 = \int_{z=0}^{\infty} \left[1 - \overline{g}(z) - T'(z) \frac{\partial \overline{Y}(z)}{\partial \rho} \right] m(z) \mathrm{d}z, \tag{19b}$$

where we define the compensated elasticity at income y:

$$\varepsilon(y) \overset{\text{def}}{\equiv} \frac{1 - T'(y)}{y} \frac{\partial \overline{Y}(y)}{\partial \tau}. \tag{19c}$$

Formula (19a) 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 $y\, m(y)$. In the optimum, these distortions should be offset by the sum of the mechanical effects, $1 - \overline{g}(z)$, and the income response effects, $T'(z) \cdot (\partial \overline{Y}(z)/\partial \rho)$, for all taxable incomes z above y.

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

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. The problem of assigning tax liabilities to given isotax curves is reminiscent of solving an optimal tax problem with one observable income and n dimensions of heterogeneity. The solution to this problem has already been described in Saez (2001), Hendren (2020) and Jacquet and Lehmann (2021b) through equations equivalent to (19a) and (19b). However, the complementing step of designing the optimal shape of isotax curves is novel and causes new difficulties.

Two observations are worth emphasizing at this point. First, Equations (19a) and (19b) are also valid if the design of the isotax curves is suboptimal. Appendix A.5 derives both equations without assuming that the isotax curves are set optimally. Second, the decomposition of the tax schedule into a definition of taxable income, $G(\cdot)$, and an assignment of a tax liability to each level of taxable income, $T(\cdot)$, is not unique. If a is a differentiable and increasing function, the same tax schedule can also be decomposed by defining taxable income as $\hat{G}(\mathbf{x}) \overset{\text{def}}{=} a(G(\mathbf{x}))$ and assigning tax liability by $\hat{T}(\hat{y}) \overset{\text{def}}{=} T(a^{-1}(\hat{y}))$. We verify in Appendix A.5 that Equations (19a) and (19b) remain valid provided that taxable income densities, compensated responses and income responses are adequately redefined. Hence, this decomposition is without loss of generality.

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 $X \overset{\text{def}}{=} \{\mathbf{x} \mid \exists \mathbf{w} \in W : \mathbf{x} = \mathbf{X}(\mathbf{w})\}$ denote the range of the type set 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 X. Finally, for each combination of incomes $\mathbf{x} \in X$, let $\partial\overline{X_i}(\mathbf{x})/\partial t_j$, $\partial\overline{X_i}(\mathbf{x})/\partial r$ and $\overline{g}(\mathbf{x})$ respectively denote the means of $\partial X_i(\mathbf{w})/\partial t_j$, $\partial X_i(\mathbf{w})/\partial r$ and $g(\mathbf{w})$ among taxpayers whose type $\mathbf{w}$ is such that they 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 Equation (17) equals zero for any direction $R(\cdot)$. To derive an optimal tax formula from this requirement, we rewrite (17) in the income space, which requires the following assumption about the regularity of the optimal allocation:

Assumption 2. *The sufficient statistics $h(\mathbf{x})$, $\partial\overline{X_i}(\mathbf{x})/\partial t_j$, $\partial\overline{X_i}(\mathbf{x})/\partial r$ 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.6).

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

$$\forall \mathbf{x} \in X: \quad \left[1 - \bar{g}(\mathbf{x}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \frac{\partial X_i(\mathbf{x})}{\partial r}\right] h(\mathbf{x}) = -\sum_{j=1}^{n} \frac{\partial \left(\sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \frac{\partial X_i(\mathbf{x})}{\partial t_j} h(\mathbf{x})\right)}{\partial x_j}, \tag{20a}$$

and the boundary conditions:

$$\forall \mathbf{x} \in \partial X: \quad \sum_{1 \leq i,j \leq n} T_{x_i}(\mathbf{x}) \frac{\partial X_i(\mathbf{x})}{\partial t_j} h(\mathbf{x})\, e_j(\mathbf{x}) = 0, \tag{20b}$$

where ∂X denotes the boundary of 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 $\mathbf{x} \in X$. A more intuitive formulation can be obtained by integrating the Euler-Lagrange Partial Differential Equation (20a) on any subset $\Omega \subset X$ of the income set. Applying the divergence theorem yields the following corollary:

Corollary 1. *Under Assumptions 1 and 2, the optimum has to verify the following integrated Euler-Lagrange equations for any subset of incomes $\Omega \subset X$ with smooth boundary $\partial\Omega$ and outward unit normal vectors $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), ..., e_n(\mathbf{x}))$:*[10]

$$\oint_{\partial\Omega} \sum_{1 \leq i,j \leq n} T_{x_i}(\mathbf{x}) \frac{\partial X_i(\mathbf{x})}{\partial t_j} e_j(\mathbf{x})\, h(\mathbf{x}) dS(\mathbf{x}) = \iint_{\Omega} \left[1 - \bar{g}(\mathbf{x}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \frac{\partial X_i(\mathbf{x})}{\partial r}\right] h(\mathbf{x}) d\mathbf{x}. \tag{20c}$$

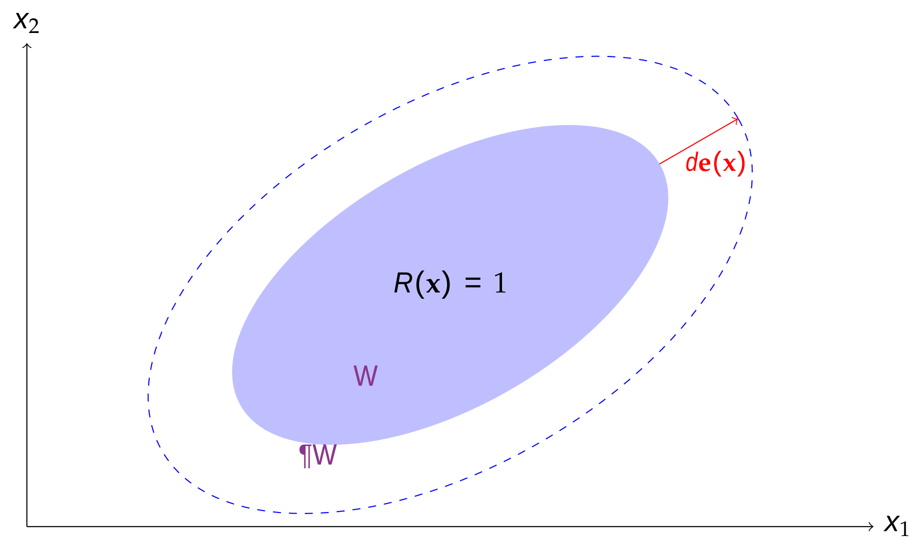


Figure 2: Intuition for Proposition 3

To clarify the economic intuition of Corollary 1, we now provide an heuristic derivation of Equation (20c) 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

[10] The symbol $\oint$ denotes a (hyper)-surface integral and $dS(\mathbf{x})$ is the corresponding measure.

the multidimensional case.[11] We consider a tax reform, illustrated in Figure 2, that consists of two parts:

i. **Inside the subset of incomes** Ω (shaded area in Figure 2): A lump sum perturbation (10a) that uniformly decreases tax liability by τ for all households with incomes $\mathbf{x} \in \Omega$ before the reform. 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 (14) and (15), 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 $\dot{\mathcal{E}}(\tau)$ is therefore given by:
$$-\left[1 - \overline{g}(\mathbf{x}) - \sum_{i=1}^{n} T_{x_i} \frac{\partial \overline{X}_i(\mathbf{x})}{\partial \rho}\right] h(\mathbf{x}).$$
Integrating these effects over all incomes $\mathbf{x}$ inside the shaded area Ω leads to minus the right-hand side of (20c).

ii. **Inside a ring of width** δ **around** Ω (area between the shaded area and the dashed curve in Figure 2): The tax gradient $(T_{x_1}, \ldots, T_{x_n})$ must change to ensure tax liabilities uniformly decrease by τ 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 $(T_{x_1}, \ldots, T_{x_n})$ has to be perturbed in a direction such that $R_{x_j}(\mathbf{X}(\mathbf{w})) = -e_j(\mathbf{x})/\delta$ for all $j \in \{1, \ldots, n\}$, where $(e_1(\mathbf{x}), \ldots, e_n(\mathbf{x}))$ 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 (14) and $R(\mathbf{X}(\mathbf{w})) \approx 0$ to approximate the contribution by taxpayers with initial income $\mathbf{x}$ inside the ring to the change in the government's objective $\dot{\mathcal{E}}(\tau)$ 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 (20c).

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 (20c).

Following Bourguignon and Spadaro (2012), Bargain et al. (2014a,b) and Jacobs et al. (2017), one can use Proposition 3 to reveal the social preferences that are consistent with the existing tax schedule. According to Equation (20a), the revealed marginal welfare weights are given by:
$$\hat{g}(\mathbf{x}) \stackrel{\text{def}}{\equiv} \left[1 - \sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \frac{\partial \overline{X}_i(\mathbf{x})}{\partial \rho}\right] + \frac{1}{h(\mathbf{x})} \sum_{j=1}^{n} \frac{\partial \left(\sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \frac{\partial \overline{X}_i(\mathbf{X}(\mathbf{w}))}{\partial \tau_j} h(\mathbf{x})\right)}{\partial x_j}. \tag{22}$$

[11] The derivation follows the graphical proof of Kleven et al. (2006) and Golosov et al. (2014). We extend their proof as we do not restrict our derivations to specifically chosen rectangular income subsets Ω. Note that the area Ω does not need to be convex, unlike what might be suggested by Figure 2.

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.7).

Proposition 4. *Under Assumptions 1 and 2:*

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

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

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) and Bierbrauer et al. (2020) 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 U and the tax schedule $T(\mathbf{x})$ satisfy the following conditions.*

i) The number n of different incomes is equal to the number p of unobserved characteristic, i.e. $n = p$.

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

iii) The mapping $\mathbf{w} \mapsto \left(S^1(c,\mathbf{x};\mathbf{w}),\dots,S^n(c,\mathbf{x};\mathbf{w})\right)$ defined on 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). Note that if Part *i*) and Part *iii*) of Assumption 2′ hold, then Part *ii*) also holds. Part *iii*) is thus more demanding than Part *ii*). We state Part *ii*) because it will be useful in the next section, when we compare Assumption 2′ with the assumptions required in the mechanism design approach. Part *iv*) of Assumption 2′ is more demanding than Assumption 1. It is necessary to ensure that behavioral responses $\partial X_i(\mathbf{w})/\partial t_j$, $\partial X_i(\mathbf{w})/\partial r$ 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.8). For $n = p = 1$, Parts *i*) *ii*) and *iii*) of Assumption 2′ are equivalent to the standard single crossing condition. For $n = p \geq 2$, when the utility function is additively separable:

$$U(c,\mathbf{x};\mathbf{w}) = g(c) - \sum_{i=1}^{n} u^i(x_i,w_i) \qquad \text{where :} \qquad g', u^i_{x_i}, u^i_{x_i,x_i} > 0 \neq u^i_{x_i,w_i} \tag{23}$$

then both Part *ii*) and Part *iii*) become equivalent to $u^i_{x_i,w_i} \neq 0$.[12] Combining Assumptions 1 and 2′, we obtain the following Lemma, which we prove in Appendix A.8.

Lemma 1. *Under Assumptions 1 and 2′, the mapping $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ is a continuously differentiable bijection from W into 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. To ensure that Assumption 2 holds, one can follow Jacquet and Lehmann (2021b) by assuming that Assumption 2′ holds with respect to an n-dimensional subset of types. Assumption 2 then holds by "pooling" the $p - n$ types of taxpayers who get the same combinations of incomes.

IV The Mechanism Design approach

While the tax perturbation approach used in Section III rigorously leads to the optimal tax formula in Proposition 3, one might doubt that this approach is consistent with the mechanism design approach pioneered by Mirrlees (1971, 1976). In the one-dimensional case, Saez (2001) demonstrates the validity of his approach by re-deriving his formula using the mechanism design approach. In this section, we do the same for the multidimensional case. We first explain how the traditional mechanism design approach leads to implementation problems in a multidimensional context. For this reason, we take a different route to derive the optimality conditions that were originally derived by Mirrlees (1976), in Proposition 5, thereby avoiding the implementation problems. Next, we derive a formulation of the optimal tax conditions in terms of welfare weights, behavioral elasticities and type densities, in the type space rather than the income space. This hybrid formulation allows us to derive conditions under which the tax perturbation and the mechanism design approaches are equivalent, and it will allow us to provide simulations for a fully optimized tax schedule in a multidimensional context. Finally, we provide conditions under which the optimal tax formula is not only necessary but also sufficient to characterize the solution of the government's problem and to ensure this solution is unique, in Proposition 6.

The mechanism design approach relies on the Taxation Principle (Hammond, 1979, Guesnerie, 1995) according to which it is equivalent for the government to select a 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:

$$\forall \mathbf{w}, \hat{\mathbf{w}} \in W: \qquad \mathcal{U}(\mathbf{w}) \stackrel{\text{def}}{\equiv} U\left(C(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right) \geq U\left(C(\hat{\mathbf{w}}), \mathbf{X}(\hat{\mathbf{w}}); \mathbf{w}\right). \tag{24}$$

To satisfy (24), the government must select an allocation that assigns a bundle $(c, \mathbf{x})$ to each type such that each taxpayer is (weakly) better off with the bundle assigned to her type than

[12] When the utility function takes the form (23), we get $S^i(c, \mathbf{x}; \mathbf{w}) = u^i_{x_i}(x_i, w_i)/g'(c)$. Assumption 2′ then amounts to demanding that the n one-dimensional mappings $w_i \mapsto u^i_{x_i}(x_i, w_i)/g'(c)$ are injective, which is guaranteed by $u^i_{x_i,w_i}$ being either everywhere positive or everywhere negative.

with any bundle assigned to any other type. Hammond (1979) shows that for each allocation $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ that verifies (24), there exists a tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ such that for each type $\mathbf{w}$, the solution of the taxpayer's program (1) corresponds to the bundle $(C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ designed for her.

Instead of dealing with the double continuum of inequalities in (24), Mirrlees (1971, 1976) adopts a First Order Mechanism Design approach (henceforth the FOMD). This approach consists of first considering only "smooth" allocations $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ and second to consider only the first-order incentive constraints:

$$\forall \mathbf{w} \in \Omega, \forall i \in \{1, ..., p\}: \qquad \frac{\partial U(\mathbf{w})}{\partial w_i} = U_{w_i}(C(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}). \tag{25}$$

The first-order incentive constraints (25) are obtained by applying the envelope theorem to the maximization of $U(C(\hat{\mathbf{w}}), \mathbf{X}(\hat{\mathbf{w}}); \mathbf{w})$ with respect to $\hat{\mathbf{w}}$ and demanding that the maximand equals $\mathbf{w}$.

In this section, we only consider "smooth" allocations that verify the following assumption.

Assumption 3. *The allocation $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ is continuously differentiable and verifies* (24).

Under Assumption 3, the FOMD consists of finding a continuously differentiable allocation that maximizes the government's Lagrangian:[13]

$$\iint_{\Omega} \left\{ \sum_{i=1}^{n} X_i(\mathbf{w}) - C(\mathbf{w}) + \frac{\Phi(U(\mathbf{w}); \mathbf{w})}{\lambda} \right\} f(\mathbf{w}) \mathrm{d}\mathbf{w} - E, \tag{26}$$

among continuously differentiable allocations that verify the first-order incentive constraints (25). We formulate the government's problem as choosing the mapping $\mathbf{w} \mapsto U(\mathbf{w})$ that maximizes the government's Lagrangian (26). For a given mapping $U(\mathbf{w})$, the allocation $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ follows from the gradient of $\mathbf{w} \mapsto U(\mathbf{w})$ using we can use the first-order incentive constraints (25).

Our approach differs from that of Mirrlees (1976) and Kleven et al. (2007), who choose the mapping $\mathbf{w} \mapsto (U(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ and treat $\mathbf{w} \mapsto C(\mathbf{w})$ as a residual. The reason we choose a different approach, is that the traditional approach hides a conceptual problem in the multidimensional context. To see this, let us consider an example. Assume that the utility function is additively separable, as in (23). We can then rewrite the first-order incentive constraints (25) as a system of Partial Differential Equations in $\mathbf{w} \mapsto U(\mathbf{w})$, for a given candidate allocation $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$:

$$\forall \mathbf{w} \in \Omega, \forall i \in \{1, ..., p\}: \qquad U_{w_i}(\mathbf{w}) = -u^i_{w_i}(X_i(\mathbf{w}), w_i). \tag{27}$$

If there is only one type, $p = 1$, integrating the Ordinary Differential Equation (27) for a given candidate mapping $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ provides the corresponding mapping $\mathbf{w} \mapsto U(\mathbf{w})$ (up to a constant). Conversely, when $p \geq 2$, solving the system of Partial Differential Equations (27) for

[13] This Lagrangian corresponds to Equation (6) in the perturbation approach, since by construction we have $T(\mathbf{x}) = \sum_{i=1}^{n} X_i(\mathbf{w}) - C(\mathbf{w})$.

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 $\mathbf{w} \mapsto Z_i(\mathbf{w}) \overset{\text{def}}{\equiv} u^i_{w_i}(X_i(\mathbf{w}), w_i)$ for all $i \in \{1, ..., p\}$. 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 $\mathbf{w} \mapsto U(\mathbf{w})$ has to 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.[14] To avoid the implementation problems that surface when optimizing $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ in multidimensional tax problems, we instead directly optimize the mapping $\mathbf{w} \mapsto U(\mathbf{w})$ and deduce the allocation $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ from the gradient of $\mathbf{w} \mapsto U(\mathbf{w})$ using the first-order incentive constraints (25). By directly choosing the mapping $\mathbf{w} \mapsto U(\mathbf{w})$, we automatically incorporate the implementability constraints.

In our mechanism design approach, we thus maximize the Lagrangian (6) with respect to the mapping $\mathbf{w} \mapsto U(\mathbf{w})$. The corresponding allocation $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ is characterized by the fact that it is incentive compatible for all taxpayers, and by the fact that it maximizes the resources $\sum_{i=1}^n x_i - c$ that are extracted from each taxpayer under the constraint that all types $\mathbf{w}$ receive utility $U(\mathbf{w})$. For each type $\mathbf{w}$, we thus take utility level $u = U(\mathbf{w})$ and the gradient $z_i = U_{w_i}(\mathbf{w})$ for all $i = 1, ..., p$ as given in the following resource-maximizing subprogram:

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

In what follows, we make the following assumption:[15]

Assumption 4. *Subprogram (28) admits a single solution denoted $\mathrm{X}_1(u, \mathbf{z}; \mathbf{w}), ..., \mathrm{X}_n(u, \mathbf{z}; \mathbf{w})$ which is twice continuously differentiable in $(u, \mathbf{z}; \mathbf{w})$.*

For individuals of type $\mathbf{w}$ enjoying utility u and facing a gradient of utility $\mathbf{z} = (z_1, ..., z_p)$, the government extracts resources equal to:

$$R(u, \mathbf{z}; \mathbf{w}) = \sum_{i=1}^n \mathrm{X}_i(u, \mathbf{z}; \mathbf{w}) - C(u, \mathrm{X}_1(u, \mathbf{z}; \mathbf{w}), ..., \mathrm{X}_n(u, \mathbf{z}; \mathbf{w}); \mathbf{w}).$$

One then has to choose the mapping $\mathbf{w} \mapsto U(\mathbf{w})$ to maximize:

$$\iint_{\mathbf{w}} \left[R\left(U(\mathbf{w}), \frac{\partial U(\mathbf{w})}{\partial w_1}, ..., \frac{\partial U(\mathbf{w})}{\partial w_p}; \mathbf{w} \right) + \frac{\Phi(U(\mathbf{w}); \mathbf{w})}{\lambda} \right] f(\mathbf{w}) d\mathbf{w} \equiv \mathcal{E}.$$

[14] The implementation problem that we here discuss is not about ignoring the second-order conditions associated to the right-hand side of (24). The problem is about the symmetry of second-order cross derivatives of $\mathbf{w} \mapsto U(\mathbf{w})$.

[15] Note that we make Assumption 4 from the outset. Mirrlees (1976, p. 342) notes only afterwards that his optimality condition does not characterize the optimum in a straightforward way, and reformulates it as a "second-order partial differential equation in" $\mathbf{w} \mapsto U(\mathbf{w})$. Kleven et al. (2007, p. 18) also warn of the implementability constraints, but do not provide a solution.

For given type $\mathbf{w}$ and utility gradient $\mathbf{z}$, the first-order incentive constraints (25) impose p limitations on the n-dimensional set of incomes. The set of incomes that verify the first-order incentive constraints (25), which constitute the choice set of Subprogram (28), is then typically of dimension $n - p$. If $n = p$, then this choice set is typically a singleton, which is then the solution of Subprogram (28). In this case, the utility level u, the utility gradient $\mathbf{z}$ and the type $\mathbf{w}$ together uniquely determine the corresponding incentive compatible income $\mathbf{x}$. On the other hand, u, $\mathbf{w}$ and $\mathbf{x}$ together yield U_{w_i} and thus via (25) a unique value for $\mathbf{z}$. In this case, Assumption 4 amounts to assuming that first-order incentive constraints (25) define a smooth one-to-one mapping:

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

$$(u, \mathbf{x}; \mathbf{w}) \mapsto \left(U_{w_1}\left(C(u, \mathbf{x}; \mathbf{w}), \mathbf{x}; \mathbf{w}\right), \ldots, U_{w_p}\left(C(u, \mathbf{x}; \mathbf{w}), \mathbf{x}; \mathbf{w}\right) \right)$$

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 Equation (23), Assumption 4′ is equivalent to $u^i_{x_i,w_i} \neq 0$. Hence, Assumption 4′ is a way to extend the single crossing condition in a multidimensional context.[16] If, on the other hand, there are more actions than types, i.e. if $n > p$, then the government has degrees of freedom beyond respecting the incentive constraints, which it then uses to maximize the resources extracted from each taxpayer.[17]

If Assumption 4 holds, the FOMD approach consists of selecting the twice differentiable utility profile $\mathbf{w} \mapsto U(\mathbf{w})$ that maximizes:

$$\iint_W \mathcal{L}\left(U(\mathbf{w}), U_{w_1}(\mathbf{w}), \ldots, U_{w_p}(\mathbf{w}); \mathbf{w}, \lambda \right) \mathrm{d}\mathbf{w} - E,$$

subject to incentive constraints (25), where $\mathcal{L}$ denotes the Lagrangian:

$$\mathcal{L}(u, \mathbf{z}; \mathbf{w}, \lambda) \stackrel{\text{def}}{\equiv} \left(R(u, \mathbf{z}; \mathbf{w}) + \frac{\Phi(u; \mathbf{w})}{\lambda} \right) f(\mathbf{w}). \tag{29}$$

Now consider the following perturbation of the utility function: $\mathbf{w} \mapsto U(\mathbf{w}) + t\, R(\mathbf{w})$, for any twice differentiable direction $R(\cdot)$. In the optimum with t close to zero, any such perturbation should not affect the value of the Lagrangian. In Appendix B.1, we show that this leads us to the Euler-Lagrange Equation:

$$\forall \mathbf{w} \in W: \qquad \mathcal{L}_u \langle \mathbf{w} \rangle = \sum_{j=1}^{n} \frac{\partial \mathcal{L}_{z_j} \langle \mathbf{w} \rangle}{\partial w_j}, \tag{30a}$$

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

[17] Several efficiency results in the literature rely on the existence of such degrees of freedom. The most famous example being the Atkinson and Stiglitz (1976) theorem. If, for instance, the type space is one-dimensional, $p = 1$, and U_w depends only on c and x_1, but not on $(x_2, \ldots, x_p)$, then marginal tax rates on $(x_2, \ldots, x_p)$ should be zero. More generally, as argued by Gauthier and Laroque (2009), if one includes externalities or public good provision in $(x_2, \ldots, x_p)$, 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.

and to the boundary conditions:

$$\forall \mathbf{w} \in \partial\Omega : \qquad 0 = \sum_{j=1}^{n} L_{z_j}\langle \mathbf{w}\rangle \, e_j(\mathbf{w}), \tag{30b}$$

where the notation $\langle \mathbf{w}\rangle$ is a shortcut to denote that functions $C(\cdot), \mathbf{X}(\cdot), U(\cdot)$ are evaluated at $\mathbf{w}$. We then obtain the following proposition which we prove in Appendix B.1:

Proposition 5. *Under Assumptions 3 and 4, the optimal utility profile* $\mathbf{w} \mapsto U(\mathbf{w})$ *has to verify:*

$$\forall \mathbf{w} \in \mathring{\Omega}, \quad \forall i \in \{1, \ldots, n\} : \quad \left(1 - S^i\langle \mathbf{w}\rangle\right) f(\mathbf{w}) = U_c\langle \mathbf{w}\rangle \sum_{j=1}^{n} q_j(\mathbf{w}) \, S^i_{w_j}\langle \mathbf{w}\rangle, \tag{31a}$$

$$\forall \mathbf{w} \in \mathring{\Omega} : \quad \sum_{j=1}^{n} \frac{\partial q_j}{\partial w_j}(\mathbf{w}) = \frac{1}{U_c\langle \mathbf{w}\rangle}\left[\frac{\Phi_u(U(\mathbf{w});\mathbf{w})}{\lambda} - f(\mathbf{w})\right] - \sum_{j=1}^{n} q_j(\mathbf{w}) \frac{U_{cw_j}\langle \mathbf{w}\rangle}{U_c\langle \mathbf{w}\rangle}, \tag{31b}$$

$$\forall \mathbf{w} \in \partial\Omega : \quad 0 = \sum_{j=1}^{n} q_j(\mathbf{w}) e_j(\mathbf{w}), \tag{31c}$$

where we denote:

$$\forall \mathbf{w} \in \Omega, \forall j \in \{1, \ldots, n\} : \qquad q_j(\mathbf{w}) \stackrel{def}{\equiv} -L_{z_j}\left(U(\mathbf{w}), U_{w_1}(\mathbf{w}), \ldots, U_{w_n}(\mathbf{w}); \mathbf{w}, \lambda\right). \tag{31d}$$

Equation (31a) corresponds to Equation (60) in Mirrlees (1976) and characterizes the optimal incomes $\mathbf{X}(\mathbf{w})$. Equation (31b) is the Euler-Lagrange equation characterizing the cost $q(w_j)$ of distorting the j^{th} component of the gradient of $\mathbf{w} \mapsto U(\mathbf{w})$ (see Equation (31d)). It corresponds to Equation (61) in Mirrlees (1976). Equation (31c) corresponds to the boundary conditions that have to hold on the boundary of the type space $\partial\Omega$. It corresponds to Equation (62) in Mirrlees (1976). Note that $q_j(\cdot)$ 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 from the first-order incentive constraint the implied perturbation of the allocation $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$, thus clarifies the economic meaning of $q_j(\cdot)$.

We now compare the necessary conditions obtained in Proposition 3 using the tax perturbation approach with the necessary conditions obtained in Proposition 5 using the FOMD. 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 (31b) can be rewritten as:

$$\forall \mathbf{w} \in \Omega : \left[1 - g(\mathbf{w}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial r}\right] f(\mathbf{w}) = \sum_{j=1}^{n} \frac{\partial \left\{\sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \, A_{j,i}(\mathbf{w}) \, f(\mathbf{w})\right\}}{\partial w_j}, \tag{32a}$$

while the Boundary conditions become:

$$\forall \mathbf{w} \in \partial\Omega : \quad \sum_{1 \le i,j \le n} T_{x_i}(\mathbf{X}(\mathbf{w})) \, A_{j,i}(\mathbf{w}) \, e_j(\mathbf{w}) = 0, \tag{32b}$$

where the matrix A is defined by:

$$A_{i,j} \quad _{i,j} \stackrel{\text{def}}{\equiv} \left[S^i_{w_j} \right]^{-1}_{i,j} = \left(\frac{\partial X_i(\mathbf{w})}{\partial w_j} \right)^{-1}_{i,j} \left(\frac{\partial X_i(\mathbf{w})}{\partial t_j} \right)_{i,j}. \tag{32c}$$

In Appendix A.9, we show how optimal tax formulas (32) 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 similar conditions for optimality as the FOMD approach that uses direct perturbations.

We now compare the strengths and weaknesses of the tax perturbation and the mechanism design approaches when the number n of incomes is equal to the number p of types. Assume that individual preferences verify Parts *i*), *ii*) and *iii*) of Assumptions 2′ and Assumption 4′.[18] 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 of Assumption 3. We show in Appendix A.2 that Assumptions 1 and 2′ together imply Assumption 3, but we cannot show the reverse. The tax perturbation approach thus appears to be slightly more demanding than the mechanism design approach. The tax perturbation can be easily adapted to the case where the number p of types is larger than the number n of incomes. Conversely, the mechanism design approach is typically tractable only in the case where $p \le n$.[19] Finally, our mechanism design approach makes it possible to derive a condition under which the necessary optimality conditions are also sufficient, as shown in the following Proposition (see the proof in Appendix B.3).

Proposition 6. *Under Assumptions 3 and 4, if for each type $\mathbf{w} \in W$ and each $l \in \mathbb{R}_+$ the mapping $(U, \mathbf{z}) \mapsto L(U, \mathbf{z}; \mathbf{w}, l)$ is concave and $\mathbf{w} \mapsto U^\star(\mathbf{w})$ verifies Equations (31), 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 $(U, \mathbf{z}) \mapsto L(U, \mathbf{z}; \mathbf{w}, l)$ is concave, and an allocation is found that verifies the necessary conditions, this allocation is the unique solution to the government's problem.

[18] Recall that when individual preferences are additively separable as in (23), Parts *i*), *ii*) and *iii*) of Assumption 2′ on the one hand and Assumption 4′ on the other hand are both equivalent to $u^i_{w_i,y_i} \neq 0$.

[19] See however Jacquet and Lehmann (2021c) when $n = 1 < p$.

V Numerical simulations

Because both the type space and the income space are multidimensional, the optimal tax formulas do not take the form of (systems 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, a change in the gradient of the tax function at one point of the income space, by changing tax liabilities near that point, triggers changes in the tax gradients elsewhere in a complicated way, as illustrated, for example, by Figure 2. 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. $n = p = 2$). Following Kleven et al. (2007), preferences over the couple's consumption c, female income x_f and male income x_m are assumed to be quasilinear in consumption, additively separable and isoelastic in each income:

$$U(c, x_f, x_m; w_f, w_m) = c - \frac{\varepsilon_f}{1+\varepsilon_f} x_f^{\frac{1+\varepsilon_f}{\varepsilon_f}} w_f^{-\frac{1}{\varepsilon_f}} - \frac{\varepsilon_m}{1+\varepsilon_m} x_m^{\frac{1+\varepsilon_m}{\varepsilon_m}} w_m^{-\frac{1}{\varepsilon_m}}. \tag{33}$$

Income effects are thus assumed away (i.e. $\partial X_f(\mathbf{w})/\partial r = \partial X_m(\mathbf{w})/\partial r = 0$). Moreover, if the tax schedule is additively separable, the cross responses are equal to zero (i.e. $\partial X_f(\mathbf{w})/\partial t_m = \partial X_m(\mathbf{w})/\partial t_f = 0$). 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 $\Phi(u, w_1, w_2) = 1 - \exp(-b\,u)/b$, where $b > 0$ stands for the degree of inequality aversion. For our baseline simulation, we select b such that the assumed 37% tax rate coincides with the optimal linear tax rate. This leads to $b = 0.0061$. 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 (29) becomes:

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

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 (20), (31) and (32). Let us explain why we choose (32). The optimal formula in (31) 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 (20a) is a *linear* second-order partial differential equation. However, it is defined in the income set X. Hence, if one solves the optimal tax formula (20a) using the same income set from one iteration to the next, which is required given the boundary conditions (20b), 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 (32) is linear, provided that the sufficient statistics $g(\mathbf{w})$ and $A(\mathbf{w})$ are taken as given. In addition, it is defined over the fixed type set W. Appendix C.1 describes the algorithm in more detail.

Here again, there is a difficulty. Equations (32a)-(32b) are defined in the type space, while $(T_{x_1}, ..., T_{x_n})$ stands for the gradient of tax liability with respect to incomes. However, one can rewrite (32a)–(32b) in terms of the gradient of tax liability in the skills-space by scaling matrix 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 (32a)–(32b) for given 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 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).

V.2 Results under the baseline calibration

(a) Full simulation set

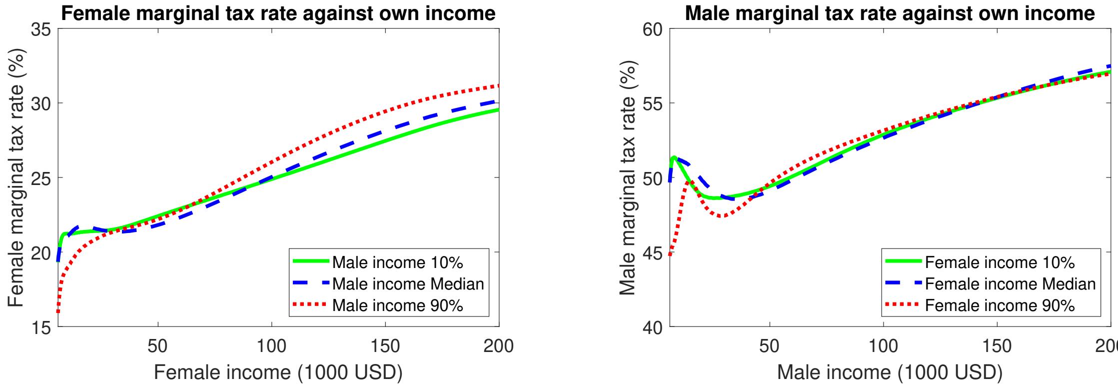
(b) Zoom on $x_f, x_m \leq \$200,000$

Figure 3: Isotax curves in the baseline case

Figure 3 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 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.

Strikingly, isotax curves are almost linear and parallel, except close to the boundaries. There, isotax curves are curved in order to satisfy boundary constraints (20b). This curvature pattern is most outspoken at high income levels where there are very few taxpayers. For lower incomes, the curvature only affects taxes 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 $(x_f, x_m) \mapsto T(x_f + a\, x_m)$ where a is optimized, we obtain a welfare gain from the current economy equal to 0.81%. The optimal value of a 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 *a*) is substantial, the welfare gain of relaxing the constraint that isotax curves have to 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 4: 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 4 that the optimal negative jointness result is not robust to using more realistic type densities with positive assortative matching. Figure 4a (resp. Figure 4b) 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 4a and 4b contradict this prediction, thereby rejecting the idea that negative jointness holds at the optimum.

V.3 Sensitivity analysis

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 *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.

(a) $\#_f = \#_m = 0.11$

(b) $\#_f = 0.43$, $\#_m = 0.11$

Figure 5: Isotax curves with different elasticities.

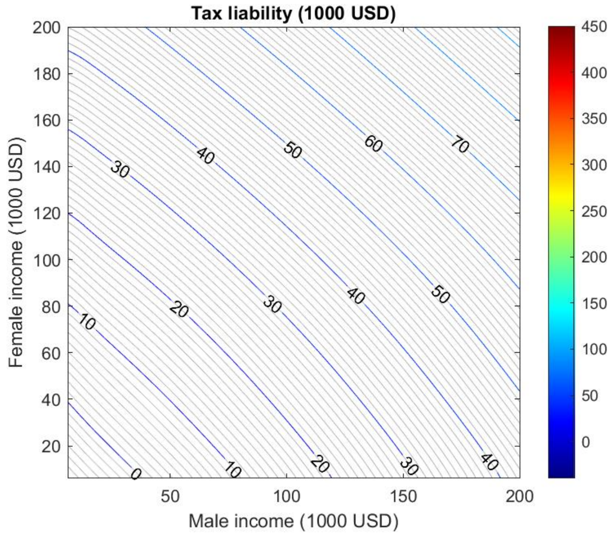
We investigate the validity of this conjecture in Figure 5. The left panel displays isotax curves when the two elasticities are equal $\#_f = \#_m = 0.11$, while the right panel shows the benchmark values $\#_f = 0.43$ for female and $\#_m = 0.11$ for male income.[20] As we conjectured, isotax curves are steeper when the two elasticities are different.

In Figure 6, we assume $\#_f = \#_m = 0.11$ in the left panel and $\#_f = \#_m = 0.43$ 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.

V.3.b Varying inequality aversion

We now investigate the sensitivity of the optimal tax schedule to the inequality aversion parameter b. In Figure 7, 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

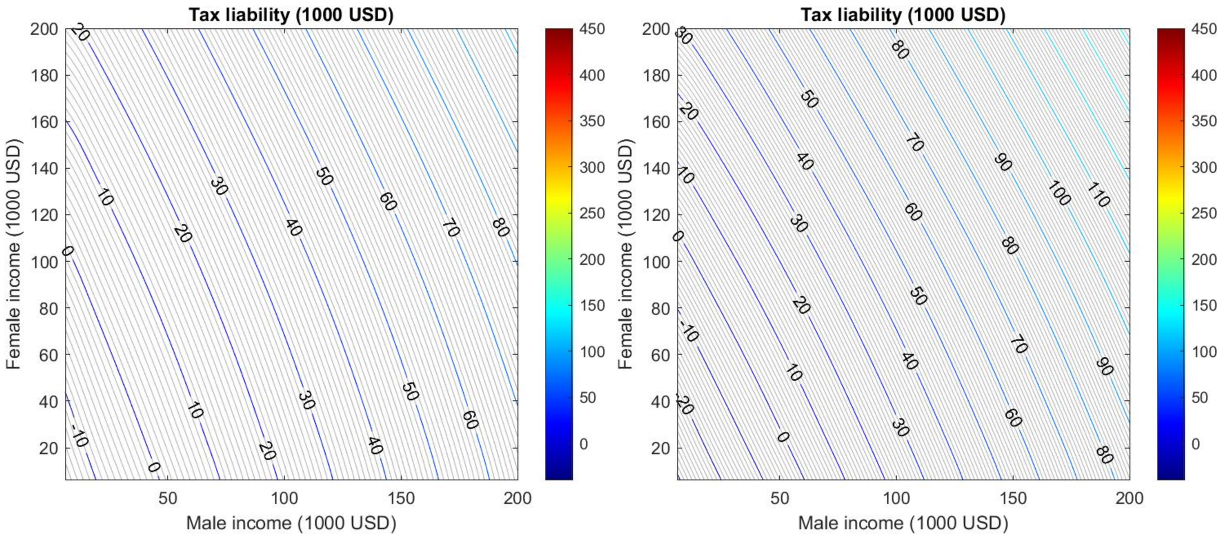
[20] When changing the elasticities, we keep the inequality aversion parameter at its baseline value of $b = 0.0061$.



(a) $\#_f = \#_m = 0.11$ (b) $\#_f = \#_m = 0.43$

Figure 6: Isotax curves with different elasticities.

its baseline value in the right panel. The shape 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) $b = 0.003$ (b) $b = 0.009$

Figure 7: 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 income space. In Figure 8 we compare the

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

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

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

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

Figure 8: Isotax curves with different domains

simulated isotax curves when both incomes are below $500,000 (Figures 8a and 8c), and when male income is below $500,000 while female income is below $800,000 (Figures 8b and 8d). As expected, changing the simulation domain has virtually no effect for incomes below $200,000 (Figures 8c and 8d). One difference between the simulations is that the larger domain adds some very rich taxpayers which triggers 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 8a and 8b). For instance, when female income is simulated up to $800,000, then the $130,000 isotax curve is concave everywhere (see Figure 8b). In this case, the shape of the isotax curve is affected by the boundary condition 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 8 thus con rms 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 verify that the tax perturbation and mechanism design approach 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 child care 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.

A Appendices on the 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 $c = C(u,x;w)$. Applying the implicit function theorem to the definition of $C(u,x;w)$, we find the gradient of the indifference sets:

$$C_{x_i}(u,x;w) = -\frac{U_{x_i}(C(u,x;w),x;w)}{U_c(C(u,x;w),x;w)}.$$

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

$$\begin{aligned} C_{x_i,x_j} &= -\frac{U_{x_i,x_j}U_c - U_{c,x_i}\frac{U_{x_j}}{U_c}U_c - U_{c,x_j}U_{x_i} + U_{c,c}\frac{U_{x_j}}{U_c}U_{x_i}}{U_c^2} \\ &= -\frac{U_{x_i,x_j}U_c^2 - U_{c,x_i}U_{x_j}U_c - U_{c,x_j}U_{x_i}U_c + U_{c,c}U_{x_j}U_{x_i}}{U_c^3}. \end{aligned}$$

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

$$\begin{aligned} S^i_{x_j} + S^j S^i_c &= \frac{U_{x_i,x_j}U_c - U_{c,x_j}U_{x_i}}{U_c^2} + \frac{U_{x_j}}{U_c}\frac{U_{c,x_i}U_c - U_{c,c}U_{x_i}}{U_c^2} \\ &= \frac{U_{x_i,x_j}U_c^2 - U_{c,x_i}U_{x_j}U_c - U_{c,x_i}U_{x_j}U_c + U_{c,c}U_{x_j}U_{x_i}}{U_c^3} = -C_{x_i,x_j}. \end{aligned} \tag{34}$$

The assumption that indifference sets are convex thus implies that the matrix $\left[S^i_{x_j} + S^j S^i_c\right]_{i,j}$ 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 Equations (9) with respect to t, x and w to get:

$$\left[C_{x_j x_i} + T_{x_j x_i}\right]_{j,i} [dx_i]_i = \left[R_{x_j}(X(w))\right]_j dt - \left[S^j_c\right]_j R(X(w))\, dt - \left[S^j_{w_k}\right]_{j,k} [dw_k]_k, \tag{35}$$

where the expressions are evaluated at $t = 0$, $x = X(w)$ and $c = C(w)$ and we use (3) and (34).

From (10b), a compensated reform of the j^{th} marginal tax rate is characterized by $R(X(w)) = 0$, $R_{x_j}(X(w)) = 1$ and $R_{x_k}(X(w)) = 0$ for $k \neq j$. Using (35), the matrix of compensated responses for type w is:

$$\left[\frac{\partial X_i(w)}{\partial t_j}\right]_{i,j} = \left[\frac{\partial X_j(w)}{\partial t_i}\right]_{i,j} = \left[C_{x_j x_i} + T_{x_j x_i}\right]^{-1}_{j,i}. \tag{36a}$$

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

From (10a), a lump-sum perturbation of the tax function is characterized by $R(X(w)) = 1$ and $R_{x_j}(X(w)) = 0$. Using (35), the vector of income responses of type w is therefore given by:

$$\left[\frac{\partial X_i(w)}{\partial r}\right]_i = \left[C_{x_j x_i} + T_{x_j x_i}\right]^{-1}_{j,i} \left[S^j_c\right]_j = \left[\frac{\partial X_i(w)}{\partial t_j}\right]_{i,j} \left[S^j_c\right]_j. \tag{36b}$$

Multiplying both sides of (35) by Matrix $\left[\mathcal{C}_{x_j x_i} + T_{x_j x_i}\right]^{-1}_{j,i}$ and using Equations (36) leads to (11).

Finally, the implicit function theorem ensures that the mapping $w \mapsto X(w)$ is differentiable for all $w \in \Omega$ with a Jacobian given by:

$$\left(\frac{\partial X_i(w)}{\partial w_k}\right)_{i,k} = \left[\mathcal{C}_{x_j x_i} + T_{x_j x_i}\right]^{-1}_{j,i} \left[S^j_{w_k}\right]_{j,k} = \left(\frac{\partial X_i(w)}{\partial t_j}\right)_{i,j} \left[S^j_{w_k}\right]_{j,k}. \tag{36c}$$

Equation (36c) shows that when the tax schedule verifies Assumption 1 and individual preferences verify Assumption 2', the ensuing allocation $w \mapsto X(w)$ verifies Assumption 3.

A.3 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 $\partial X_i^\star(w)/\partial r$ and $\partial X_i^\star(w)/\partial t_j$ denote the direct income and compensated responses of the incomes and let $\partial X_i^\star(w)/\partial w_j$ denote the direct responses to a change in types.

We now clarify the difference between direct and total responses. Let $\Delta_1 x$ 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 (35). We thus get direct responses ignoring the effects due to the non-linearity of the tax schedule:

$$\Delta_1 x = \left[\mathcal{C}_{x_i x_j}\right]^{-1}_{i,j} dB,$$

where dB is the column vector on the right-hand side of (35).

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

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

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

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

and so:

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

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

$$\begin{aligned}\Delta x &= \sum_{k=1}^{\infty} \Delta_k x = \sum_{k=1}^{\infty} \left(-\left(\frac{\partial X_i^\star}{\partial t_j}\right)_{i,j} [T_{x_i x_j}]_{i,j}\right)^{k-1} \Delta_1 x = \left(I_n + \left(\frac{\partial X_i^\star}{\partial t_j}\right)_{i,j} [T_{x_i x_j}]_{i,j}\right)^{-1} \Delta_1 x \\ &= \left(I_n + \left[\mathcal{C}_{x_i x_j}\right]^{-1}_{i,j} [T_{x_i x_j}]_{i,j}\right)^{-1} \left[\mathcal{C}_{x_i x_j}\right]^{-1}_{i,j} dB = \left[\mathcal{C}_{x_i x_j} + T_{x_i x_j}\right]^{-1}_{i,j} dB,\end{aligned}$$

where I_n denotes the identity matrix of rank n. We thus retrieve (35), 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:

$$\left(\frac{\partial X_i(w)}{\partial t_j}\right)_{i,j} = \left(I_n + \left(\frac{\partial X_i^\star}{\partial t_j}\right)_{i,j} [T_{x_ix_j}]_{i,j}\right)^{-1} \left(\frac{\partial X_i^\star(w)}{\partial t_j}\right)_{i,j}, \tag{37a}$$

$$\left(\frac{\partial X_i(w)}{\partial r}\right)_{i} = \left(I_n + \left(\frac{\partial X_i^\star}{\partial t_j}\right)_{i,j} [T_{x_ix_j}]_{i,j}\right)^{-1} \left(\frac{\partial X_i^\star(w)}{\partial r}\right)_{i}, \tag{37b}$$

$$\left(\frac{\partial X_i(w)}{\partial w_j}\right)_{i,j} = \left(I_n + \left(\frac{\partial X_i^\star}{\partial t_j}\right)_{i,j} [T_{x_ix_j}]_{i,j}\right)^{-1} \left(\frac{\partial X_i^\star(w)}{\partial w_j}\right)_{i,j}. \tag{37c}$$

Equations (36a) and (37a) imply that Part ii) of Assumption 1 is equivalent to assuming that the matrix $I_n + \left(\partial X_i^\star / \partial t_j\right)_{i,j} [T_{x_ix_j}]_{i,j}$ is positive definite despite the nonlinearity of the tax schedule.

A.4 Proof of Proposition 1

To find the derivative of (12) with respect to t, we add (14) to (15). We integrate the result over all types w to obtain (17). To obtain (16), we use the lump-sum perturbation (10a) in (17), i.e. we set $R(X(w)) = 1$ and $R_{x_j}(X(w)) = 0$ in (17).

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 $\kappa^R(t)$ denote the lump-sum transfer that ensures that the following tax perturbation keeps the government's budget balanced: $x \mapsto T(x) + t R(x) - \kappa^R(t)$. Let $(\partial \tilde{\mathcal{L}}^R(t)/\partial t)|_{t=0}$ denote the partial derivatives of the government's Lagrangian with respect to size t of the perturbation $w \mapsto T(x) + t R(x)$. Similarly, let $(\partial \hat{\Phi}^{R+\kappa}(t)/\partial t)|_{t=0}$, $(\partial \hat{B}^{R+\kappa}(t)/\partial t)|_{t=0}$ and $(\partial \hat{\mathcal{L}}^{R+\kappa}(t)/\partial t)|_{t=0}$ 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 $w \mapsto T(x) + t R(x) - \kappa^R(t)$. Let finally $(\partial \tilde{\mathcal{L}}^r(r)/\partial r)|_{r=0}$ denote the partial derivatives of the government's Lagrangian with respect to size r of the lump sum perturbation (10a). From (17), one get that:

$$\left.\frac{\partial \hat{\mathcal{L}}^{R+\kappa}(t)}{\partial t}\right|_{t=0} = \left.\frac{\partial \tilde{\mathcal{L}}^{R}(t)}{\partial t}\right|_{t=0} + \kappa'(t) \left.\frac{\partial \tilde{\mathcal{L}}^{r}(r)}{\partial r}\right|_{r=0}.$$

Since Equation (16) is equivalent to $(\partial \tilde{\mathcal{L}}^r(r)/\partial r)|_{r=0}$, we thus get:

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

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

$$\frac{1}{\lambda} \left.\frac{\partial \hat{\Phi}^{R+\kappa}(t)}{\partial t}\right|_{t=0} = \left.\frac{\partial \tilde{\mathcal{L}}^{R}}{\partial t}\right|_{t=0}.$$

The derivations in this proof are valid whenever Assumption 1 holds true, regardless of whether $n = p$, $n < p$ or $n > p$.

A.5 Optimal Tax for given isotax curves, Proof of Proposition 2

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

We first consider tax perturbations that preserve the isotax curves. Applying Equation (11) to the tax perturbation $x \mapsto \mathcal{T}(\mathcal{G}(x)) - tR(\mathcal{G}(x))$ leads to:

$$\left.\frac{\partial \tilde{X}_i^R(w,t)}{\partial t}\right|_{t=0} = \frac{\partial X_i(w)}{\partial \rho} R(\mathcal{G}(X(w))) + \sum_{i=1}^{n} \frac{\partial X_i(w)}{\partial \tau_j} \mathcal{G}_{x_j}(w)\, R'(\mathcal{G}(X(w)))\,. \tag{38}$$

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

$$\left.\frac{\partial \tilde{Y}^R(w,t)}{\partial t}\right|_{t=0} = \sum_{i=1}^{n} \mathcal{G}_{x_i}(X(w)) \left.\frac{\partial \tilde{X}_i^R(w,t)}{\partial t}\right|_{t=0}. \tag{39}$$

Applying Equation (38) to the tax liability perturbation $R(y) = 1$ and using (39) leads to (18a). Applying Equations (38) and (39) to the compensated perturbation $R(y) = y - Y(w)$ leads to (18b). Combining Equations (18a), (18b), (38) and (39), the response of taxable income to a generic tax perturbation $R(\cdot)$ is given by:

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

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

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

Using (15), the response of the perturbed Lagrangian (12) to a tax perturbation in the direction $R(\cdot)$ then is:

$$\begin{aligned}
\left.\frac{\partial \mathcal{L}^R(t)}{\partial t}\right|_{t=0} &= \iint_{w \in W} g(w) \left[-\left(1 + \mathcal{T}'(Y(w))\frac{\partial Y(w)}{\partial \rho}\right) R(Y(w)) \right. \\
&\quad \left. + \mathcal{T}'(Y(w)) \frac{\partial Y(w)}{\partial \tau} R'(Y(w))\right] f(w)dw \\
&= \int_{y \in \mathbb{R}_+} \left[\int_{Y(w)=y} g(w) \left[-\left(1 + \mathcal{T}'(y)\frac{\partial Y(w)}{\partial \rho}\right) R(y) \right.\right. \\
&\quad \left.\left. + \mathcal{T}'(y) \frac{\partial Y(w)}{\partial \tau} R'(y)\right] f(w|Y(w) = y)\, dw\right] m(y)dy,
\end{aligned} \tag{40}$$

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

$$\frac{\partial \overline{Y}(y)}{\partial t} \stackrel{\text{def}}{\equiv} \int_{Y(w)=y} \frac{\partial Y(w)}{\partial t} f(w|Y(w) = y)\, dw. \tag{41a}$$

Similarly, denote the mean of the income responses among taxpayers earning $Y(w) = y$ as:

$$\frac{\partial \overline{Y}(y)}{\partial r}(y) \stackrel{\text{def}}{\equiv} \int_{Y(w)=y} \frac{\partial Y(w)}{\partial r} f(w|Y(w) = y)\, dw. \tag{41b}$$

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

$$\bar{g}(y) \stackrel{\text{def}}{\equiv} \int_{Y(w)=y} g(w)\, f(w|Y(w) = y)\, dw. \tag{41c}$$

Equation (40) then simplifies to:

$$\left.\frac{\partial L^R(t)}{\partial t}\right|_{t=0} = \int_{y\in\mathbb{R}_+} \left\{\left[\bar{g}(y) - 1 + T'(y)\frac{\partial \overline{Y}(y)}{\partial r}\right] R(y) + T'(y)\frac{\partial \overline{Y}(y)}{\partial t} R'(y)\right\} m(y)dy. \tag{42}$$

Integrating by parts leads to:

$$\begin{aligned}\left.\frac{\partial L^R(t)}{\partial t}\right|_{t=0} = & \int_{y\in\mathbb{R}_+} \left\{\int_{z=y}^{\infty}\left[1 - \bar{g}(z) - T'(z)\frac{\partial \overline{Y}(z)}{\partial r}\right] m(z)dz + T'(y)\frac{\partial \overline{Y}(y)}{\partial t} m(y)\right\} R'(y)dy \\ & - R(0)\int_{y\in\mathbb{R}_+}\left[1 - \bar{g}(y) - T'(y)\frac{\partial \overline{Y}(y)}{\partial r}\right] m(y)dy. \end{aligned} \tag{43}$$

The effect of perturbation on the Lagrangian is nil for all directions R if and only if Equation (19b) and the following Equation:

$$\forall y: \quad -T'(y)\frac{\partial \overline{Y}(y)}{\partial t} m(y) = \int_{z=y}^{\infty}\left[1 - \bar{g}(z) - T'(z)\frac{\partial \overline{Y}(z)}{\partial r}\right] m(z)dz,$$

are valid. Rearranging terms using (19c) leads to (19a) if $T'(y) < 1$.

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

$$\hat{T}'(\hat{y}) = \frac{T' \circ a^{-1}(\hat{y})}{a'(a^{-1}(\hat{y}))} = \frac{T'(y)}{a'(y)}.$$

Differentiating both sides of $\hat{M}(a(y)) = M(y)$ leads to:

$$\hat{m}(\hat{y}) = \frac{m(y)}{a'(y)}.$$

Applying respectively (18) and (18b) to $\hat{Y}(w) = a(Y(w))$ leads to:

$$\frac{\partial \hat{Y}(w)}{\partial r} = a'(Y(w))\,\frac{\partial Y(w)}{\partial r} \quad \text{and} \quad \frac{\partial \hat{Y}(w)}{\partial t} = \left(a'(Y(w))\right)^2 \frac{\partial Y(w)}{\partial t}.$$

Hence

$$\hat{T}'(\hat{y})\,\frac{\partial \overline{\hat{Y}}(\hat{y})}{\partial r} = T'(y)\,\frac{\partial \overline{Y}(y)}{\partial r} \quad \text{and} \quad \hat{T}'(\hat{y})\,\frac{\partial \overline{\hat{Y}}(\hat{y})}{\partial t}\,\hat{m}(\hat{y}) = T'(y)\,\frac{\partial \overline{Y}(y)}{\partial t}\, m(y).$$

Therefore (A.5) and (19b) are equivalent in terms of y or in terms of $\hat{y}$.

A.6 Optimal tax formula in the income space, Proof of Proposition 3

In Appendix A.4, we show that equation (17) holds under Assumption 1. We can rewrite Equation (17) in terms of the income density $h(\cdot)$ (which is well defined under Assumption 2), rather than the type density $f(\cdot)$ to obtain:

$$\frac{\partial \widetilde{\mathcal{L}}^R(t)}{\partial t}\bigg|_{t=0} = \iint_{\mathcal{X}} \Bigg\{ \Bigg[\bar{g}(X(w)) \quad 1+\sum_{i=1}^{n} T_{x_i}(x)\frac{\partial X_i(x)}{\partial \rho} \Bigg] R(x) \tag{44}$$

$$+ \sum_{1\le i,j\le n} T_{x_i}(x)\frac{\partial X_i(x)}{\partial t_j} R_{x_j}(x) \Bigg\} h(x)\,dx.$$

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

$$\frac{\partial \widetilde{\mathcal{L}}^R(t)}{\partial t}\bigg|_{t=0} = \oint_{\partial\mathcal{X}} \sum_{1\le i,j\le n} T_{x_i}(x)\frac{\partial X_i(X(w))}{\partial t_j} h(x)\, e_j(x)\, R(x)\, dS(x) \tag{45}$$

$$\iint_{\mathcal{X}} \left\{ \left[1-\bar{g}(X(w)) \quad \sum_{i=1}^{n} T_{x_i}(x)\frac{\partial X_i(x)}{\partial \rho} \right] h(x) + \sum_{j=1}^{n} \frac{\partial \left(\sum_{i=1}^{n} T_{x_i}(x)\frac{\partial X_i(X(w))}{\partial t_j} h(x) \right)}{\partial x_j} \right\} R(x)\,dx.$$

If the tax schedule $T(\cdot)$ is optimal, Equation (45) has to equal 0 for all possible directions $R(\cdot)$. This is only possible if the Euler-Lagrange Partial Differential Equation (20a) and the boundary conditions (20b) are both satisfied.

A.7 Proof of Proposition 4

Equations (22) and (45) are valid when Assumption 1 and Assumption 2 hold true. According to Equations (5), (6) and (15), removing the term $\bar{g}(X(w))$ from Equation (45) provides the effects of a tax perturbation on government's revenue:

$$\frac{\partial \widetilde{\mathcal{B}}^R(t)}{\partial t}\bigg|_{t=0} = \oint_{\partial\mathcal{X}} \sum_{1\le i,j\le n} T_{x_i}(x)\frac{\partial X_i(X(w))}{\partial t_j} h(x)\, e_j(x)\, R(x)\, dS(x)$$

$$\iint_{\mathcal{X}} \left\{ \left[1 \quad \sum_{i=1}^{n} T_{x_i}(x)\frac{\partial X_i(x)}{\partial \rho} \right] h(x) + \sum_{j=1}^{n} \frac{\partial \left(\sum_{i=1}^{n} T_{x_i}(x)\frac{\partial X_i(X(w))}{\partial t_j} h(x) \right)}{\partial x_j} \right\} R(x)\,dx,$$

which, given (22), can be simplified to:

$$\frac{\partial \widetilde{\mathcal{B}}(t)}{\partial t}\bigg|_{t=0} = \oint_{\partial\mathcal{X}} \sum_{1\le i,j\le n} T_{x_i}(x)\frac{\partial X_i(X(w))}{\partial t_j} h(x)\, e_j(x)\, R(x)\, dS(x) \quad \iint_{\mathcal{X}} \beta(x)\, R(x)\, h(x)\,dx, \tag{46}$$

Note that the right-hand side of (22), and thus also $\beta(\cdot)$, is continuous with respect to x. Let $x^\star$ be an income bundle in the interior of $\mathcal{X}$ such that $\beta(x^\star) < 0$. By continuity of $\beta(\cdot)$, there exists $r > 0$ such that the ball of radius r around $x^\star$ remains in the interior of $\mathcal{X}$ and $\beta(\cdot)$

remains negative everywhere in this ball. Consider then a tax perturbation $x \mapsto T(x) - \tau R(x)$ where $R(\cdot)$ is twice continuously differentiable, positive inside the ball of radius r around $x^\star$ and nil otherwise.[21] Hence, $\bar{g}(x)\, R(x)$ is negative inside the ball of radius r around $x^\star$ and nil outside.

The first term in the right-hand side of (46) 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 $\tau > 0$ therefore generates tax revenue. Moreover, for incomes x inside the ball of radius r around $x^\star$, utility increases since there $R(x)$ is positive so perturbed tax liability $T(x) - \tau R(x) < T(x)$ 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 $\bar{g}(x) \geq 0$ for all $x \in \mathcal{X}$.

A.8 Proof of Lemma 1

Given that $\mathcal{X}$ is defined as the range of the typeset $\mathcal{W}$ under the allocation $w \mapsto X(w)$, it is sufficient to show that the mapping $w \mapsto X(w)$ is injective to establish that it is a bijection. Assume there exists $x \in \mathcal{X}$ and $w, \hat{w} \in \mathcal{W}$ such that $X(w) = X(\hat{w}) = x$. From Assumption 1, the first-order conditions (3) have to be verified both at $(c, x; w)$ and at $(c, x; \hat{w})$, so we get $S^i(c, x, w) = S^i(c, x, \hat{w})$ for all $i \in \{1, \dots, n\}$. According to Part $iii)$ of Assumption 2', these n equalities imply that $w = \hat{w}$. Differentiability of $w \mapsto X(w)$ is ensured under Assumption 1 by the implicit function theorem applied to (3). Part $ii)$ of Assumption 2' then ensures the Jacobian of $w \mapsto X(w)$ is invertible (see Equation (36c) in Appendix A.2).

Because the mapping $w \mapsto X(w)$ is injective, we get that $\bar{g}(X(w)) = g(w)$, $\partial\overline{X_i}(X(w))/\partial t_j = \partial X_i(w)/\partial t_j$ and $\partial\overline{X_i}(X(w))/\partial \rho = \partial X_i(w)/\partial \rho$. According to Equations (7), (36a) and (36b), $g(w)$, $\partial X_i(w)/\partial t_j$ and $\partial X_i(w)/\partial \rho$ are continuously differentiable functions of c, x, 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 $w \mapsto X(w)$ is continuously differentiable and invertible, and because of Part $iv)$ of Assumption 2', $\partial\overline{X_i}(x)/\partial t_j$, $\partial\overline{X_i}(x)/\partial \rho$ and $\bar{g}(x)$ are continuously differentiable in x. Finally, the income density is given by:

$$h(X(w)) = \frac{f(w)}{\det\left(\frac{\partial X_i(w)}{\partial w_j}\right)_{i,j}}, \tag{47}$$

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

[21] For instance, one can take $R(x) = \int_{\|x - x^\star\|}^{r} u^2 (u - r)^2 du$ inside the ball and zero outside.

A.9 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, (17), in the type space rather than in the income space. To reparametrize the direction of a tax perturbation as a function of types, define:

$$\hat{R}(w) \overset{\text{def}}{=} R(X(w)).$$

Differentiating both sides with respect to w_j yields:

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

In matrix notation, the latter equality becomes:

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

where we use Parts i) and ii) of Assumption 2' and Equation (36c) to ensure that matrix $\left(\partial X_i(w)/\partial w_j\right)_{i,j}$ is invertible. Using the symmetry of the matrix of compensated effects $\left(\partial X_i(w)/\partial t_j\right)_{i,j}$, we can rewrite the last term of Equation (17):

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

where the last Equality follows from (36c). Using the definition of matrix $A_{i,j}(w)$ in (32c), Equation (17) can be rewritten as:

$$\frac{\partial \mathcal{L}^R(t)}{\partial t}\Big|_{t=0} = \iint_W \left\{ \left[g(w) - 1 + \sum_{i=1}^{n} T_{x_i}(X(w)) \frac{\partial X_i(w)}{\partial r} \right] \hat{R}(w) - \sum_{1 \le i,j \le n} T_{x_i}(X(w))\, A_{j,i}(w)\, \hat{R}_{w_j}(w) \right\} f(w)\,dw.$$

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

$$\begin{aligned}
\frac{\partial \mathcal{L}(t)}{\partial t}\Big|_{t=0} = & -\oint_{\partial W} \sum_{1 \le i,j \le n} T_{x_i}(X(w))\, A_{j,i}(w)\, e_j(w)\, f(w)\, \hat{R}(w)\, dS(w) \\
& - \iint_W \left\{ \left[1 - g(w) - \sum_{i=1}^{n} T_{x_i}(X(w)) \frac{\partial X_i(w)}{\partial r} \right] f(w) - \sum_{j=1}^{n} \frac{\partial \sum_{i=1}^{n} T_{x_i}(X(w))\, A_{j,i}(w)\, f(w)}{\partial w_j} \right\} \hat{R}(w)\, dw.
\end{aligned}$$

This partial derivative is equal to zero for any direction of tax perturbation $\hat{R}(\cdot)$ if and only if the Euler-Lagrange Equation (32a) and Boundary conditions (32b) are verified.

B Appendices on the First-Order Mechanism Design approach (FOMD)

B.1 Proof of Proposition 5

Let R be a twice differentiable function defined over Ω into $\mathbb{R}$. We consider the effects of perturbing the utility profile $\omega \mapsto U(\omega)$ in the direction R. Consider the perturbed Lagrangian, with the unperturbed Lagrangian defined by (29):

$$\mathcal{E}^R(t) \stackrel{\text{def}}{\equiv} \iint_\Omega \mathcal{L}\left(U(\omega)+tR(\omega), U_{\omega_1}(\omega)+tR_{\omega_1}(\omega), \ldots, U_{\omega_p}(\omega)+tR_{\omega_p}(\omega); \omega, \lambda\right) d\omega. \quad (48)$$

Applying the chain rule and denoting $\langle\omega\rangle$ as a shortcut to denote that a function is evaluated at $\left(U(\omega), U_{\omega_1}(\omega), \ldots, U_{\omega_p}(\omega); \omega, \lambda\right)$, we obtain:

$$\left.\frac{\partial \mathcal{E}^R(t)}{\partial t}\right|_{t=0} = \iint_\Omega \left\{ \mathcal{L}_u\langle\omega\rangle R(\omega) + \sum_{j=1}^p \mathcal{L}_{z_j}\langle\omega\rangle R_{\omega_j}(\omega) \right\} d\omega.$$

Applying integration by parts using the divergence theorem leads to:

$$\left.\frac{\partial \mathcal{E}^R(t)}{\partial t}\right|_{t=0} = \iint_\Omega \left\{ \mathcal{L}_u\langle\omega\rangle - \sum_{j=1}^p \frac{\partial \mathcal{L}_{z_j}\langle\omega\rangle}{\partial \omega_j} \right\} R(\omega)d\omega + \oint_{\partial\Omega} \sum_{j=1}^p \mathcal{L}_{z_j}\langle\omega\rangle \, e_j(\omega)\, R(\omega)\, dS(\omega).$$

At the optimal allocation, the latter expression is nil for any perturbation R. This implies that the boundary conditions (30b) must hold together with the Euler-Lagrange Equation (30a). Using (31d) leads to (31c), and Euler-Lagrange equation (30a) becomes:

$$\forall \omega \in \Omega: \quad \sum_{j=1}^p \frac{\partial q_j(\omega)}{\partial \omega_j} = \mathcal{L}_u\langle\omega\rangle. \quad (49)$$

Using incentive compatibility constraint (25), we can rewrite Lagrangian (29):

$$\left[\sum_{i=1}^p X_i(\omega) - C\left(U(\omega), X(\omega); \omega\right) + \frac{\Phi(U(\omega);\omega)}{\lambda}\right] f(\omega) = \quad (50)$$
$$\mathcal{L}\left(U(\omega), U_{\omega_1}\left(C(U(\omega),X(\omega);\omega), X(\omega);\omega\right), \ldots, U_{\omega_n}\left(C(U(\omega),X(\omega);\omega), X(\omega);\omega\right); \omega, \lambda\right).$$

Differentiating both sides of (50) with respect to $X_i(\omega)$ and using (2) and (31d) leads to:

$$\left(1 - S^i\langle\omega\rangle\right) f(\omega) = \sum_{j=1}^p q_j(\omega)\left[U_{x_i\omega_j}\langle\omega\rangle + S^i\langle\omega\rangle U_{c\omega_j}\langle\omega\rangle\right],$$

which leads to (31a) given that $S^i_{\omega_j} = (U_{c\omega_j}U_{x_i} - U_{x_i\omega_j}U_c)/U_c^2 = -\left[U_{x_i\omega_j} + S^i U_{c\omega_j}\right]/U_c$. Differentiating (50) with respect to $U(\omega)$ and using $C_u = 1/U_c$ and (31d) leads to:

$$\left(-\frac{1}{U_c\langle\omega\rangle} + \frac{\Phi_u(U(\omega);\omega)}{\lambda}\right) f(\omega) = \mathcal{L}_u\langle\omega\rangle - \sum_{j=1}^p q_j(\omega)\frac{U_{c,\omega_j}\langle\omega\rangle}{U_c\langle\omega\rangle}. \quad (51)$$

Substituting (49) into (51) yields (31b).

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

Using (3), Equation (31a) leads to:

$$T'_{x_i}(X(w))\, f(w) = \sum_{j=1}^{p} \mu_j(w)\, S^i_{w_j}\langle w\rangle, \tag{52}$$

where we denote $\mu_j(w) \overset{\text{def}}{=} q_j(w)\, U_c(C(w), X(w); w)$. This can be rewritten $[T'_{x_i}(X(w))]_i\, f(w) = \left[S^i_{w_j}\right]_{i,j} \left[\mu_j(w)\right]_j$ in matrix notation, which leads to: $\left[\mu_j(w)\right]_j = \left[S^i_{w_j}\right]^{-1}_{i,j} [T'_{x_i}(X(w))]_i\, f(w)$. Using (32c), we therefore get:

$$\forall w \in \Omega, \forall i \in \{1, ..., p\} \qquad \mu_i(w) = \sum_{j=1}^{n} A_{i,j}(w)\, T'_{x_j}(X(w))\, f(w). \tag{53}$$

Combining Equation (31c) with (53) thus leads to (32b). Using Equation (7), Equation (31b) implies that:

$$\begin{aligned}
\sum_{j=1}^{p} \frac{\partial \mu_j(w)}{\partial w_j} &= (1 - g(w))\, f(w) - \sum_{j=1}^{p} q_j(w)\, U_{cw_j}\langle w\rangle \\
&\quad + \sum_{j=1}^{p} q_j(w) \left[U_{cc}\langle w\rangle \frac{\partial C(w)}{\partial w_j} + \sum_{i=1}^{n} U_{cx_i}\langle w\rangle \frac{\partial X_i(w)}{\partial w_j} + U_{cw_j}\langle w\rangle \right] \\
&= (1 - g(w))\, f(w) + \sum_{j=1}^{p} q_j(w) \left[U_{cc}\langle w\rangle \frac{\partial C(w)}{\partial w_j} + \sum_{i=1}^{n} U_{cx_i}\langle w\rangle \frac{\partial X_i(w)}{\partial w_j} \right].
\end{aligned} \tag{54}$$

Differentiating $C(w) = \mathcal{C}(U(w), X(w); w)$ with respect to w_j and using $\mathcal{C}_u = 1/U_c$, $\mathcal{C}_{x_i} = -U_{x_i}/U_c$, $\mathcal{C}_{w_j} = -U_{w_j}/U_c$ and (25) leads to:

$$\frac{\partial C(w)}{\partial w_j} = -\frac{U_{w_j}\langle w\rangle}{U_c\langle w\rangle} - \sum_{i=1}^{n} \frac{U_{x_i}\langle w\rangle}{U_c\langle w\rangle} \frac{\partial X_i(w)}{\partial w_j} \qquad \frac{U_{w_j}\langle w\rangle}{U_c\langle w\rangle} = \sum_{i=1}^{n} \frac{U_{x_i}\langle w\rangle}{U_c\langle w\rangle} \frac{\partial X_i(w)}{\partial w_j}.$$

Plugging this equality into (54) leads to

$$\begin{aligned}
\sum_{j=1}^{p} \frac{\partial \mu_j(w)}{\partial w_j} &= (1 - g(w))\, f(w) + \sum_{j=1}^{p} \sum_{i=1}^{n} q_j(w) \left(U_{cx_i}\langle w\rangle - \frac{U_{x_i}\langle w\rangle}{U_c\langle w\rangle} U_{cc}\langle w\rangle \right) \frac{\partial X_i(w)}{\partial w_j} \\
&= (1 - g(w))\, f(w) - \sum_{j=1}^{p} \sum_{i=1}^{n} \mu_j(w)\, S^i_c\langle w\rangle \frac{\partial X_i(w)}{\partial w_j}.
\end{aligned} \tag{55}$$

Substituting Equation (36c) into (55) yields:

$$\sum_{j=1}^{p} \frac{\partial \mu_j(w)}{\partial w_j} = (1 - g(w))\, f(w) + \sum_{j=1}^{p} \sum_{i=1}^{n} \sum_{k=1}^{n} \mu_j(w)\, S^i_c\langle w\rangle \frac{\partial X_i(w)}{\partial t_k} S^k_{w_j}(w). \tag{56}$$

Substituting Equation (36b) into (56) and using $\frac{\partial X_i(w)}{\partial t_k} = \frac{\partial X_k(w)}{\partial t_i}$ yields:

$$\sum_{j=1}^{p} \frac{\partial \mu_j(w)}{\partial w_j} = (1 - g(w))\, f(w) - \sum_{j=1}^{p} \sum_{k=1}^{n} \mu_j(w)\, S^k_{w_j}(w) \frac{\partial X_k(w)}{\partial \rho}. \tag{57}$$

Plugging (52) into (57) leads to:

$$\sum_{j=1}^{p} \frac{\partial \mu_j(\omega)}{\partial \omega_j} = \left(1 - g(\omega) - \sum_{k=1}^{n} T_{x_k}(X(\omega)) \frac{\partial X_k(\omega)}{\partial \rho}\right) f(\omega). \tag{58}$$

Plugging (53) into (58) leads to (32a).

The last equality in (32c) follows from Equation (36c) of Appendix A.2.

B.3 Proof of Proposition 6

If $(u,z) \mapsto \mathcal{L}(u,z;\omega,\lambda)$ is concave then for any perturbation p, the function $t \mapsto \Lambda^R(t)$ defined in (48) is concave. Let $\omega \mapsto U(\omega)$ be another utility profile that verifies Equations (31a) and take the perturbation $R(\omega) = U(\omega) - U^{\star}(\omega)$. As the utility profile $\omega \mapsto U(\omega)$ verifies Equations (31), we get that function $t \mapsto \Lambda^R(t)$ admits a zero derivative at $t = 0$ and is concave. So $\Lambda^R(0) > \Lambda^R(1)$ and $U^{\star}(\cdot)$ provides a strictly higher welfare than $U(\cdot)$.

If two distinct allocations $\omega \mapsto U^{\star}(\omega)$ and $\omega \mapsto U(\omega)$ verify Equations (31) 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 Equations (31).

C Appendix on the Numerical Simulations

C.1 Simulations for the full optimum

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

$$J(\omega) \stackrel{\text{def}}{\equiv} \begin{pmatrix} \frac{\partial X_1(\omega)}{\partial \omega_1} & \frac{\partial X_1(\omega)}{\partial \omega_2} \\ \frac{\partial X_2(\omega)}{\partial \omega_1} & \frac{\partial X_2(\omega)}{\partial \omega_2} \end{pmatrix}^{-1}.$$

Given the mapping $\omega \mapsto T(\omega)$ and the allocation $\omega \mapsto X(\omega)$, we find the marginal tax rates for a type-ω taxpayer:

$$T_{x_i}(X(\omega)) = \sum_{k=1}^{n} T_{\omega_k}(\omega) J_{k,i}(\omega). \tag{59}$$

Taking into account that individual preferences (33) do not feature income effects, we rewrite optimal tax condition (32a) in the type space:

$$(1 - g(\omega)) f(\omega) = \sum_{j=1}^{p} \frac{\partial \left(\sum_{1 \le i,k \le n} T_{\omega_k}(\omega) J_{k,i}(\omega) A_{j,i}(\omega) f(\omega) \right)}{\partial \omega_j},$$

with boundary conditions:

$$\forall \omega \in \partial\Omega \quad : \quad \sum_{1 \le i,j,k \le n} T_{\omega_k}(\omega) \frac{\partial X_k^{-1}(X(\omega))}{\partial x_i} A_{j,i}(\omega) e_j(\omega) = 0.$$

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 $x \mapsto T^{(\ell)}(x)$. Starting from the tax function $x \mapsto T^{(\ell)}(x)$, we use the individual first-order conditions to calculate the corresponding allocation $w \mapsto X^{(\ell)}(w)$, and the corresponding inverse Jacobian $w \mapsto J^{(\ell)}(w) \stackrel{\text{def}}{=} \partial {X_k^{(\ell)}}^{-1}(X(w)) / \partial x_i$.

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

$$(1 - g(w))\, f(w) = \sum_{j=1}^{p} \frac{\partial \left(\sum_{1 \leq i,k \leq n} T_{w_k}^{(\ell+1)}(w)\, J_{k,i}^{(\ell)}(w)\, A_{j,i}(w)\, f(w) \right)}{\partial w_j}, \tag{60a}$$

with boundary conditions:

$$\forall w \in \partial\Omega \quad : \quad \sum_{1 \leq i,j,k \leq n} T_{w_k}^{(\ell+1)}(w)\, J_{k,i}^{(\ell)}(w)\, A_{j,i}(w)\, e_j(w) = 0. \tag{60b}$$

In Equations (60), welfare weights $g(w)$ are computed endogenously through (7), and matrices $A(w)$ through (32c), both as functions of the allocation $w \mapsto X(w)$. The allocation is computed from marginal tax rates $x \mapsto T_{x_i}(x)$. Marginal tax rates are deduced from $w \mapsto T_{w_k}^{(\ell+1)}(w)$ and from $w \mapsto J^{(\ell)}(w)$ using (59). By keeping the Jacobian $w \mapsto J^{(\ell)}(w)$ fixed, the Partial Differential Equation remains solvable by MATLAB.

3. We repeat these steps until the process converges to a fixed point $x \mapsto T(x)$. As convergence criterion, we require that for more than 0.1% 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.[22]

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

While solving the partial differential equation (60a) for $T^{(\ell+1)}$, MATLAB's solver will inspect different candidate solutions $T^{(\ell+1)}$ with corresponding partial derivatives $T_{w_j}^{(\ell+1)}$. Unavoidably, some candidates will correspond through (59) 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 $T^{(\ell+1)}$ 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 (59) 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 $g(w)$ and

[22] 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.

A(w):

$$\forall w: \quad T_{x_j}^{(\ell+1)}(X(w)) \quad \begin{cases} \dfrac{T_{x_j}^{(\ell+1)}(X(w))}{T_{x_j}^{(\ell+1)}(X(w)) + 1 - T_{x_j}^{(\ell)}(X(w))} & \text{if } T_{x_j}^{(\ell+1)}(X(w)) \geq 0, \\ T_{x_j}^{(\ell+1)}(X(w)) & \text{if } T_{x_j}^{(\ell+1)}(X(w)) < 0. \end{cases} \tag{61}$$

Equation (61) ensures that $T_{x_j}^{(\ell+1)}(X(w)) < 1$, given that $T_{x_l}^{(\ell)} < 1$. Moreover, $T_{x_j}^{(\ell+1)}(X(w))$ is continuous and increasing in $T_{x_j}^{?(\ell+1)}(X(w))$. Finally, if the algorithm converges, it converges to the correct schedule $w \mapsto T(w)$, i.e. if $T_{x_j}^{?(\ell+1)}(X(w)) = T_{x_j}^{(\ell)}(X(w))$, then one has $T_{x_j}^{(\ell+1)}(X(w)) = T_{x_j}^{?(\ell+1)}(X(w)) = T_{x_j}^{(\ell)}(X(w))$.

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. In order to have sufficient detail where necessary, e.g. near the boundaries and where most households are, we use a transformation of the types. We use the following utility function:

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

where $W_i(w_i)$ are transformations of the individual abilities w_i. For given observations of the incomes and for given marginal tax rates, we find for an optimizing individual:

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

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:

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

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:

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

With an evenly spaced grid for w, the grid for $W(w)$ will be more detailed where $dW_i(w_i)/dw_i$ 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:

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

The inverse Jacobian matrices $J(w)$ of the allocation $w \mapsto X(w)$ are approximated for distances $dw_k = 10^{-7}$ 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.

C.2 Individual taxation

In this subsection we describe our algorithm for finding the optimal individual tax system. With some abuse of notation we will write the individual-tax system as:

$$\tilde{T}(x) = T(x_f) + T(x_m).$$

With our utility function (33), the labor supply decisions of the couple can be decomposed as the sum of two programs, one for each spouse, according to:

$$\max_{x_f} \; x_f - T(x_f) - \frac{\varepsilon_f}{1+\varepsilon_f} x_f^{\frac{1+\varepsilon_f}{\varepsilon_f}} + \max_{x_m} \; x_m - T(x_m) - \frac{\varepsilon_m}{1+\varepsilon_m} x_m^{\frac{1+\varepsilon_m}{\varepsilon_m}}.$$

Hence, female (male) labor supply x_f (x_m) depends only on female (male) type w_f (w_m). Denote then $W_i(x)$ the reciprocal $w_i \mapsto X_i(w_i)$. From Jacquet and Lehmann (2021b), the optimal-tax formula is given by:

$$\frac{T'(x)}{1-T'(x)} = \frac{\sum_i \int_{w=W_i(x)}^{\overline{w}_i} (1-\tilde{g}_i(w))\, f_i(w)dw}{\varepsilon_f\, W_f(x)\, f_f(W_f(x)) + \varepsilon_m\, W_m(x)\, f_m(W_m(x))}, \tag{62}$$

where $\tilde{g}_i(w_i)$ denotes the average welfare weight of a couple in which spouse i has ability w_i and $f_i(w_i)$ is the marginal distribution of w_i.

Our strategy is to apply a fixed-point algorithm to (62). That is, we calculate the right-hand side of (62) under a given tax schedule. This allows us to update the tax schedule, which in turn updates the right-hand side of (62), and so on.

Equation (33) allows us to calculate maximized utility as a function of type, $U(w_f,w_m)$. This in turn allows us to find the value of λ by noting that with quasi-linear utility, in absence of income effects, Equation (16) can be written as:

$$\lambda = \int_{\underline{w}_f}^{\overline{w}_f}\int_{\underline{w}_m}^{\overline{w}_m} F'(U(w_f,w_m))\, f(w_f,w_m)dw_m dw_f. \tag{63}$$

The solution for λ from (63) allows us to find welfare weights as:

$$g(w_f,w_m) = \frac{F'(U(w_f,w_m))}{\lambda}.$$

Finally, the average welfare weight of a couple in which spouse i has ability w_i can be found through integration:

$$\tilde{g}_i(w_i) = \frac{\int_{\underline{w}_j}^{\overline{w}_j} g(w_f,w_m)\, f(w_f,w_m)dw_j}{f_i(w_i)} \quad \text{for} \quad i \neq j. \tag{64}$$

This gives us all the necessary elements to calculate the right-hand side of (62) for a given tax schedule on the basis of the primitives of our model.

We solve our fixed-point algorithm over a fixed income range $X = [\underline{w},\overline{w}]$ where $\underline{w}$ is the minimum ability level in the sample, and $\overline{w}$ the maximum ability level. Note that for our utility function $x_i = w_i$ when marginal tax rates equal 0. Hence, the range of income X is consistent with our type-space for any optimal-tax function that satisfies the boundary constraint $T'(\underline{w}) = T'(\overline{w}) = 0$. Moreover, note that in our simulations $\underline{w}_f = \underline{w}_m$ and $\overline{w}_f = \overline{w}_m$, such that the minimum and maximum ability level are the same for males and females.

We now have all the ingredients in place to provide a step-by-step description of our fixed-point algorithm:

1. Create a grid with S linearly spaced incomes $\{x^s\}_{s=1}^S$ between $\underline{w}$ and $\overline{w}$.

2. Start a loop. Denote the tax function in iteration k by $T^{(k)}(x)$. The first tax function is given by $T^{(0)}(x) = T^{(0)}(0) = E$. That is, a lump-sum tax equal to exogenous expenditure. Note that the initial tax function satisfies the boundary conditions $T^{(0)\prime}(\underline{w}) = T^{(0)\prime}(\overline{w}) = 0$. Moreover, equations (62)–(63) jointly imply that an updated tax function also satisfies the boundary conditions. Therefore, we can be sure that in each iteration of the loop $\underline{w}_i = \underline{x}_i$ and $\overline{w}_i = \overline{x}_i$.

3. For each income level x^s, use the first-order conditions (3) to find the ability level of spouse i that would choose x^s under tax function $T^{(k)}(x)$. We again use the skill transformation described in Subsection C.1 , to increase the detail of our simulations near the boundaries of the domain and where most of the population resides. This provides a grid of abilities for each spouse $W_i = \{w_i^s\}_{s=1}^S$.

4. Expand W_i to create an $S \times S$ mesh of ability levels W by matching each $w_i^s \in W_i$ to the full vector $\{w_j^s\}_{s=1}^S$ for $i \neq j$. Note that because i.) preferences are quasi-linear in consumption, ii.) preferences are separable, iii.) the tax function is separable, and iv.) the skill transformations are separable, each spouse with ability w_i^s will earn x_i^s independent of the ability of the partner. This allows us to map ability to income for every type in the mesh W. We will refer to this mapping as $X_i(w_i)$.

5. Find the marginal densities $f_i(w_i)$ and use Equation (64) to find the welfare weight $\tilde{g}_i(w_i)$ for each agent in W.[23]

6. Use Equation (62) to find new marginal tax rates $(T'(x))^{new}$.

7. Update marginal tax rates by taking a weighted average between old and new tax rates:

$$(T'(x))^{(k+1)} = p(T'(x))^{(k)} + (1-p)(T'(x))^{new}. \tag{65}$$

8. Now we determine the lump-sum tax. Gross tax liabilities as a function of income for the points on our grid are given by:

$$T_{gross}^{(k+1)}(x^s) = \int_{\min_i(\underline{w}_i)}^{x^s} (T'(x))^{(k+1)} dx. \tag{66}$$

[23] Throughout this subsection, and the next subsection all numerical integration occurs over a discretized grid using Simpson's rule as an approximation of the integral.

Total tax receipts are thus:

$$B = \int_{\underline{w}_f}^{\overline{w}_f}\int_{\underline{w}_m}^{\overline{w}_m} \left(T^{(k+1)}_{gross}(X_f(w_f,w_m)) + T^{(k+1)}_{gross}(X_m(w_f,w_m)) \right) f(w_f,w_m)\,dw_m dw_f. \tag{67}$$

Hence, the intercept of the tax function consistent with the marginal tax rates equals $T^{(k+1)}(0) = E - B$. Together, the intercept and the marginal tax rates provide sufficient information to create the new tax function $T^{(k+1)}(x) = T^{(k+1)}(0) + T^{(k+1)}_{gross}(x)$.

The algorithm converges if the update in the marginal tax rate is small. Otherwise, reset the counter $k = k+1$ and loop back to step 3.

C.3 Linear isotax curves and joint taxation

In this section we describe our algorithm for finding the optimal-tax function with linear isotax curves where taxable income is defined as:

$$y \equiv a\, x_f + x_m,$$

and a is a parameter that determines the slope of the isotax curve. We want to find the optimal tax function $T(y)$ as well as the optimal slope a. Our strategy is to apply a fixed-point algorithm to equation (19a) for a given slope a. We then use a grid search to find the optimal value for a. For the optimal joint taxation, we fix $a = 1$.

To apply the fixed-point algorithm we first need to express each term in (19a) as a function of the primitives of our model, and for a given tax function. More specifically, we are looking for an expression of i.) the distribution of joint income $m(y)$, ii.) the elasticity of joint income $\epsilon(y)$, and iii.) welfare weights $g(y)$.

We first consider the joint density function $m(y)$ which we find by initially defining the cumulative density $M(y)$, and then differentiating with respect to y.

We consider a rectangular income space. In a rectangular income space the set of couples with joint income less than or equal to y can take four forms. In the first case the set forms a triangle with vertices $(\underline{x}_f,\underline{x}_m), (\underline{x}_f, y - a\underline{x}_f), ((y - \underline{x}_m)/a, \underline{x}_m)$. The second case is where the space forms a trapezoid with vertices $(\underline{x}_f,\underline{x}_m), (\underline{x}_f,\overline{x}_m), ((y-\overline{x}_m)/a,\overline{x}_m), ((y-\underline{x}_m)/a,\underline{x}_m)$. The third case is also a trapezoid with vertices: $(\underline{x}_f,\underline{x}_m), (\overline{x}_f,\underline{x}_m), (\overline{x}_f, y - a\overline{x}_f), (\underline{x}_f, y - a\underline{x}_f)$. The final case is a pentagon with vertices $(\underline{x}_f,\underline{x}_m), (\overline{x}_f,\underline{x}_m), (\overline{x}_f, y - a\overline{x}_f), ((y-\overline{x}_m)/a,\overline{x}_m), (\underline{x}_f,\overline{x}_m)$. We can write the cumulative density on the set with joint incomes smaller than or equal to y more succinctly by using a double integral:

$$M(y) \equiv \int_{\underline{x}_m}^{\min(y - a\underline{x}_f,\overline{x}_m)} \left(\int_{\underline{x}_f}^{\min\left(\frac{y-x_m}{a},\overline{x}_f\right)} h(x_f,x_m)\,dx_f \right) dx_m. \tag{68}$$

In the integral, the four cases are subsumed in the upper bound of the integrals. If for a particular value of y, $y - a\underline{x}_f \leq \overline{x}_m$ and $(y - \underline{x}_m)/a \leq \overline{x}_m$ we are effectively integrating over a triangle. If $y - a\underline{x}_f > \overline{x}_m$ and $(y - \underline{x}_m)/a \leq \overline{x}_m$, we are in the first trapezoidal case, and so on.

The density function $m(y)$ is defined as:

$$m(y) \equiv \frac{dM(y)}{dy}.$$

Given the double integral in (68) we need to apply Leibniz' rule for integration twice. To simplify this step, first define:

$$f(y, x_m) \equiv \int_{\underline{x}_f}^{\min\left(\frac{y-x_m}{a}, \bar{x}_f\right)} h(x_f, x_m)\, dx_f.$$

Applying Leibniz' Rule a first time, we arrive at:

$$\begin{aligned} m(y) &= \frac{d \int_{\underline{x}_m}^{\min(y-a\underline{x}_f, \bar{x}_m)} f(y, x_m)\, dx_m}{dy} \\ &= \frac{\partial \min(y - a\underline{x}_f, \bar{x}_m)}{\partial y} f(y, \min(y - a\underline{x}_f, \bar{x}_m)) + \int_{\underline{x}_m}^{\min(y-a\underline{x}_f, \bar{x}_m)} \frac{\partial f(y, x_m)}{\partial y} dx_m. \end{aligned}$$

Now apply Leibniz' rule a second time to arrive at:

$$\begin{aligned} m(y) &= 1(y - a\underline{x}_f < \bar{x}_m) \int_{\underline{x}_f}^{\min\left(\frac{y - \min(y - a\underline{x}_f, \bar{x}_m)}{a}, \bar{x}_f\right)} h(x_f, \min(y - a\underline{x}_f, \bar{x}_m))\, dx_f \\ &\quad + \int_{\underline{x}_m}^{\min(y-a\underline{x}_f, \bar{x}_m)} \frac{\partial \int_{\underline{x}_f}^{\min\left(\frac{y-x_m}{a}, \bar{x}_f\right)} h(x_f, x_m)\, dx_f}{\partial y} dx_m \qquad (69) \\ &= \int_{\underline{x}_m}^{\min(y-a\underline{x}_f, \bar{x}_m)} \frac{\partial \int_{\underline{x}_f}^{\min\left(\frac{y-x_m}{a}, \bar{x}_f\right)} h(x_f, x_m)\, dx_f}{\partial y} dx_m, \qquad (70) \end{aligned}$$

where $1(\cdot)$ is an indicator function that equals one if its argument is true, and zero otherwise. In (69), the first term drops for the following reason. First, if $\bar{x}_m \leq y - a\underline{x}_f$ the indicator function equals zero. Second, if $\bar{x}_m < y - a\underline{x}_f$ the upper bound to the integral simplifies to: $\min\left((y - (y - a\underline{x}_f))/a, \bar{x}_f\right) = \underline{x}_f$, which equals the lower bound of the integral.

Equation (70) describes $m(y)$ in terms of the primitives of our model and the tax function, since $h(\cdot)$ is given by equation (47).

Next we consider the elasticity of joint income $\varepsilon_y(y)$. The maximization problem of couples that face tax schedule $T(y)$, and have quasi-linear and separable preferences is given by:

$$\max_{x_f, x_m} x_f + x_m - T(ax_f + x_m) - v(x_f, w_f) - v(x_m, w_m).$$

First-order conditions are:

$$\begin{aligned} 1 - aT'(ax_f + x_m) &= v_{x_f}(x_f, w_f), \\ 1 - T'(ax_f + x_m) &= v_{x_m}(x_m, w_m). \end{aligned}$$

Perturbing the marginal tax rate by t and totally differentiating with respect to x_f, x_m and t, we arrive at:

$$\begin{aligned} a\,dt &= (v_{x_f,x_f} + a^2 T'')dx_f + aT''dx_m, \qquad (71) \\ dt &= aT''dx_f + (v_{x_m,x_m} + T'')dx_m. \qquad (72) \end{aligned}$$

Using Cramer's rule we have:

$$
\begin{aligned}
\frac{dx_f}{dt} &= \frac{-a(v_{x_m,x_m} + T'') + aT''}{(v_{x_f,x_f} + a^2T'')(v_{x_m,x_m} + T'') - a^2(T'')^2} \\
&= \frac{-av_{x_m,x_m}}{v_{x_f,x_f}v_{x_m,x_m} + v_{x_f,x_f}T'' + a^2T''v_{x_m,x_m}}, \\
\frac{dx_m}{dt} &= \frac{-(v_{x_f,x_f} + a^2T'') + a^2T''}{v_{x_f,x_f}v_{x_m,x_m} + v_{x_f,x_f}T'' + a^2T''v_{x_m,x_m}} \\
&= \frac{-v_{x_f,x_f}}{v_{x_f,x_f}v_{x_m,x_m} + v_{x_f,x_f}T'' + a^2T''v_{x_m,x_m}}.
\end{aligned}
$$

Therefore, the elasticity of joint income as a function of the individual incomes is given by:

$$
\begin{aligned}
\varepsilon^y(x_f, x_m) &= \frac{dy}{dt}\frac{1-T'}{y} = \frac{d(ax_f + x_m)}{dt}\frac{1-T'}{ax_f + x_m} \\
&= -\frac{a^2v_{x_m,x_m} + v_{x_f,x_f}}{v_{x_f,x_f}v_{x_m,x_m} + v_{x_f,x_f}T'' + a^2T''v_{x_m,x_m}}\frac{v_{x_m}}{ax_f + x_m}. \qquad (73)
\end{aligned}
$$

Hence, the average elasticity of a couple with joint income y is given by:

$$
\varepsilon^y(y) = \frac{1/a \int_{\max(\underline{x}_m, y-a\bar{x}_f)}^{\min(y-a\underline{x}_f, \bar{x}_m)} \varepsilon^y(\frac{y-x_m}{a}, x_m)h(\frac{y-x_m}{a}, x_m)dx_m}{m(y)}. \qquad (74)
$$

This in turn can be expressed in terms of the primitives of our model, and the tax function by using equation (73) for $\varepsilon^y(x_f, x_m)$ and the definition of the utility function (33) to determine its derivatives.

We can use the individual first-order condition (3) to find utility as a function of the incomes. In turn, (63) gives us the Lagrange multiplier. Welfare weights are then $g(x_f, x_m) = F'(U(x_f, x_m))/\lambda$. Finally, the welfare weight of a couple with joint income y is given by:

$$
g(y) = \frac{1/a \int_{\max(\underline{x}_m, y-a\bar{x}_f)}^{\min(y-a\underline{x}_f, \bar{x}_m)} g(\frac{y-x_m}{a}, x_m)h(\frac{y-x_m}{a}, x_m)dx_m}{m(y)}. \qquad (75)
$$

Next we describe the implementation of our algorithm. The general idea is to start with a given tax schedule, which allows us to calculate the right-hand side of (19a). This in turn updates the tax schedule, and so on until the algorithm converges to a specific tax schedule.

More specifically, our algorithm takes the following steps:

1. Initialize an outer loop by choosing the slope of the isotax curves along a grid of points α between 0 and 1 and setting $a = \alpha/(1-\alpha)$. Here $\alpha = .5$ represents the case of unweighted joint income, $\alpha < .5$ implies female income is taxed more than male income, and vice versa for $\alpha > .5$.

2. Create a mesh of incomes $\{x_f^s, x_m^s, y^s\} = \{x_f^s, x_m^s, ax_f^s + x_m^s\}_{s=1}^S$ between bottom income $\{\underline{w}_f, \underline{w}_m, a\underline{w} + \underline{w}_m\}$ and top incomes $\{\bar{w}_f, \bar{w}_m, a\bar{w} + \bar{w}_m\}$.

   Our grid has a rectangular structure. Specifically, we first select 100 linearly spaced values of x_f between $\underline{w}_f$ and $\bar{w}_f$. Then for each value of x_f we choose 100 values of $x_m \in \underline{w}_m$ and

$\overline{w}_m$ creating a total of $S = 100\ \ 100 = 10,0000$ representative couples. We create this grid such that there is more detail close to the boundaries of the domain, similar to what we did in the full simulations (see Appendix C.1). Furthermore, we slightly round the corners of the income domain at $(\underline{w}_f, \overline{w}_m)$ and $(\overline{w}_f, \underline{w}_m)$, to prevent the sharp corners from causing discontinuities in the Lagrangian for the optimal tax problem. [24]

3. Initialize an inner loop by setting a $T^0(y) = E$.

4. For each point in the grid calculate the corresponding type (w^i_f, w^i_m), the density $h(x^i_f, x^i_m)$, the elasticity $\#^y(x^i_f, x^i_m)$ and the utility $U(x^i_f, x^i_m)$.

5. Use (63) for l and use this to determine welfare weights $g(x_f, x_m)$.

6. Numerically integrate equations (70), (73) and (75) to get $m(y), g(y), e^y(y)$.

7. Evaluate the right-hand side of (19a) and nd updated marginal tax rates $T^0(y^i)^{new}$

8. Update marginal tax rates by taking a weighted average between new and previous marginal tax rates:

$$T^0(y^i)^{(k+1)} = z\ \ T^0(y^i)^{(k)} + (1\ \ \ z)\ \ T^0(y^i)^{new},$$

where z is a number between 0 and 1.

9. Calculate gross tax receipts $T_{gross}(y) = \int_{\underline{y}}^{y} (T^0(w))^{(k+1)} dw$ and gross tax revenue revenue $= \int_{\underline{y}}^{\overline{y}} T_{gross}(y)m(y)dy$. The intercept of the tax function is given by $T^{(k+1)}(0) = E\ \ \ Revenue$ Together, the intercept and the marginal tax rates provide suf cient information to create the new tax function $T^{(k+1)}(x) = T^{(k+1)}(0) + T^{(k+1)}_{gross}(x)$.

10. Evaluate whether marginal tax rates are suf ciently close between iterations. If they are, the inner loop has converged. If they are not close, return to step 4 using the new tax system $T^{(k+1)}(y)$ and update the counter $k = k + 1$. When determining the starting function for the next iteration, we smooth the tax function and we linearly extrapolate it for the bottom 3.98% of the population rather than letting the marginal tax rates go to zero. This extrapolation and smoothing prevents sharp swings in the second-order derivative of the tax function in intermediate iterations of the algorithm.

11. Once the inner loop has converged, we evaluate welfare under the optimal tax system and store it. Return to step 2 for a different value of a until all have been tried.

12. The algorithm returns the value of a that provides the highest level of welfare as well as the corresponding welfare level.

[24] Income is endogenous to the tax function. However, for our utility function, when the marginal tax rate equals zero, we nd $x_i = w_i$. Moreover, from the boundary conditions we know that optimal marginal tax rates at the top and bottom of the income distribution equal 0. Therefore, the top and the bottom of the income space are independent of the tax system if the tax system satis es the boundary conditions. The rest of boundaries of the income space however can change from one iteration to the other.

C.4 Comparisons of social welfare

In order to find a money metric to compare social welfare between different simulations, we invert the social welfare function to find the evenly distributed level of utility that gives rise to the measured level of social welfare:

$$u = -\frac{\log[b(1-F)]}{b}.$$

Since we use a quasi-linear social welfare function, differences in disposable income give rise to equal differences in utility levels.

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.

— and —, *Lectures on public economics: Updated edition*, Princeton University Press, 2015.

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, "Tax-benefit revealed social preferences in Europe and the US," *International Tax Public Finance*, 2014, *21*, 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.

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

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

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

Gauthier, Stéphane and Guy Laroque, "Separability and public finance," *Journal 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, "Dynamic Tax Reforms," NBER Working Papers 20 780 2014.

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

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.

_ and _ , "Optimal tax problems with multidimensional heterogeneity: A Mechanism Design Approach," *Social Choice and Welfare*, 2021, p. Forthcoming.

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.

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.

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

Scheuer, Florian and Iván Werning, "Mirrlees meets Diamond-Mirrlees," NBER Working Papers 22076 2016.

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

A Mathematical Appendix not for publication

Studying multidimensional screening problems requires mathematical tools that are unusual to most economists. For the sake of completeness, this appendix states some results from the mathematics literature that are useful for our paper. Hence, this appendix is not intended for publication. For additional results and for proofs, an excellent reference is Basov (2005).

A mapping from a set X of $\mathbf{R^n}$ onto $\mathbf{R}$ is a *scalar field*. A mapping from a set X of $\mathbf{R^n}$ onto $\mathbf{R}^n$ is a *vector field*. When a scalar field $\varphi : \mathbf{x} \mapsto \varphi(\mathbf{x})$ is differentiable, the vector of its partial derivatives $(\partial\varphi/\partial x_1, ..., \partial\varphi/\partial x_n)$ is called the *gradient*. The gradient of a scalar field is therefore a vector field. When a vector field is the gradient of a scalar field, this vector field is said to be *conservative*. When $n = 1$, any continuous (vector) field is conservative, simply because any continuous function over an interval of R is integrable. This result does not easily generalize when $n > 1$. This is because if a scalar field is twice continuously differentiable, the Hessian matrix of its second order derivative has to be symmetric. This imposes a necessary condition on the partial derivatives of a vector field to be conservative. The following theorem ensures that symmetry of second-order cross derivatives is also sufficient.

Theorem 1. *Let $\mathbf{f} : \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_n(\mathbf{x})) \in \mathrm{R}^n$ be a continuously differentiable vector field defined over a compact subset X of R^n. Then $\mathbf{f}$ is conservative if and only if:*

$$\forall \mathbf{x} \in X \quad \text{and} \quad \forall (i,j) \in \{1, ..., n\}^2 \qquad : \qquad \frac{\partial f_i}{\partial x_j}(\mathbf{x}) = \frac{\partial f_j}{\partial x_i}(\mathbf{x}).$$

We now state the divergence theorem, which generalizes to multiple dimensions a corollary of the fundamental theorem of calculus in one dimension according to which $f(b) - f(a) = \int_a^b f'(x)\mathrm{d}x$.

Theorem 2 (Divergence theorem)**.** *Let $\mathbf{f} : \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_n(\mathbf{x})) \in \mathrm{R}^n$ be a continuously differentiable vector field defined over a compact subset X of R^n. Let Ω be a connected compact subset of X which admits a continuously differentiable boundary denoted $\partial\Omega$. For any $\mathbf{x}$ on the boundary $\partial\Omega$ of Ω, let $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), ..., e_n(\mathbf{x}))$ denote the outward unit surface normal vector to the boundary of Ω. Then:*

$$\iint_{\mathbf{x}\in\Omega} \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(\mathbf{x})\, \mathrm{d}S(\mathbf{x}) = \oint_{\mathbf{x}\in\partial\Omega} \sum_{i=1}^{n} f_i(\mathbf{x}) e_i(\mathbf{x})\, \mathrm{d}\mathbf{x}.$$

The Divergence theorem can be used to perform integration by parts in multiple dimensions.

Theorem 3 (Multidimensional integration by parts)**.** *Let $\mathbf{f} : \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_n(\mathbf{x})) \in \mathrm{R}^n$ and $a : \mathbf{x} \mapsto a(\mathbf{x}) \in \mathrm{R}$ be respectively a vector and a scalar field that are continuously differentiable over a compact subset Ω of R^n, with a smooth boundary $\partial\Omega$ and let $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), ..., e_n(\mathbf{x}))$ be the outward unit surface normal vector defined on the boundary $\partial\Omega$. Then:*

$$\iint_{\Omega} \left(\sum_{i=1}^{n} f_i(\mathbf{x}) \frac{\partial a}{\partial x_i}(\mathbf{x}) \right) \mathrm{d}\mathbf{x} = \oint_{\mathbf{x}\in\partial\Omega} \sum_{i=1}^{n} a(\mathbf{x}) f_i(\mathbf{x}) e_i(\mathbf{x}) \mathrm{d}S(\mathbf{x}) - \iint_{\Omega} \left(\sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(\mathbf{w}) a(\mathbf{x}) \right) \mathrm{d}\mathbf{x}.$$

Proof: Let $\mathbf{Q}(\mathbf{x}) = (Q_1(\mathbf{x}), ..., Q_n(\mathbf{x})) \stackrel{\text{def}}{=} a(\mathbf{x})\mathbf{f}(\mathbf{x}) = (a(\mathbf{x}) f_1(\mathbf{x}), ..., a(\mathbf{x}) f_n(\mathbf{x}))$. Since $(\partial Q_i / \partial x_i)(\mathbf{x}) = (\partial a / \partial x_i)(\mathbf{x}) f_i(\mathbf{x}) + a(\mathbf{x})(\partial f_i / \partial x_i)(\mathbf{x})$, applying the divergence theorem to $\mathbf{Q}$ gives the result.

Theorem 4 (Integral form of a divergence PDE)**.** *Let $a : \mathbf{x} \mapsto a(\mathbf{x}) \in \mathbf{R}$ and $\mathbf{f} : \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_n(\mathbf{x})) \in \mathrm{R}^n$ be respectively a continuously differentiable scalar and vector fields defined over a compact subset X of R^n. Then f verifies the divergence partial differential equation:*

$$\sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(\mathbf{x}) = a(\mathbf{x}),$$

if and only if it verifies the integrated divergence equation:

$$\oint_{\mathbf{x} \in \partial \Omega} \sum_{i=1}^{n} f_i(\mathbf{x}) e_i(\mathbf{x}) \, \mathrm{d}S(\mathbf{x}) = \iint_{\mathbf{x} \in \Omega} a(\mathbf{x}) \, \mathrm{d}\mathbf{x},$$

for any compact subset Ω of X which admits a continuously differentiable boundary denoted $\partial\Omega$ with an outward unit surface normal vector to the boundary of Ω denoted $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), ..., e_n(\mathbf{x}))$.

Proof: This result is obtained by applying the Divergence theorem to left-hand side of the *divergence* partial differential equation. For the reciprocal, applying the Divergence theorem to left-hand side of the integrated *divergence* equation leads to:

$$\iint_{\mathbf{x} \in \Omega} \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(\mathbf{x}) \, \mathrm{d}\mathbf{x} = \iint_{\mathbf{x} \in \Omega} a(\mathbf{x}) \, \mathrm{d}\mathbf{x}.$$

As this equality should be verified for any smooth compact subset of X, and f is assumed continuously differentiable, then the two integrand must be equal everywhere inside Ω, which leads to the divergence PDE.