A Variational Approach to the Analysis of Tax Systems

Mikhail Golosov, Aleh Tsyvinski, and Nicolas Werquin*

December 12, 2014

Abstract

We develop a general method to study the effects of non-linear taxation in dynamic settings using variational arguments. We first derive general theoretical formulas that characterize the welfare effects of local tax reforms and, in particular, the optimal tax system, potentially restricted within certain classes (e.g., age-independent, linear, separable). These formulas are expressed in terms of intuitive parameters, such as the labor and capital income elasticities and the hazard rates of the income distributions. Second, we apply these formulas to various specific settings. In particular, we decompose the gains arising from each element of tax reform, starting from a simple baseline system, as the available tax instruments becomes more sophisticated. We further show that the design of tax systems obeys a common general principle, namely that more sophisticated tax instruments (e.g., age-dependent, non-linear, non-separable) allow the government to fine-tune the tax rates by targeting higher distortions to the segments of the population whose behavior responds relatively little to those taxes.

*Golosov: Princeton University; Tsyvinski and Werquin: Yale University. Golosov and Tsyvinski thank NSF for financial support. We thank audiences at Princeton, Yale, Society for Economics Dynamics annual meeting, NBER Summer Institute, NBER Public Economics Program meeting, Taxation Theory Conference (Cologne).

1 Introduction

Many counties use a complex system of taxes and transfers. Welfare and social insurance payments depend on individual earnings, which creates a complex nonlinear schedule of effective marginal labor and capital income tax rates. Figure 1 illustrates such patterns using the federal tax programs in the U.S.[1] Moreover, both the eligibility and the amount of payment often depends on the past history of labor earnings, assets, marital status, age and the number of children.

It is challenging to develop a theory of taxation that both allows for sufficiently rich tax functions and provides transparent, intuitive insights about the effect of taxes. The literature so far have mainly persued either of the following two approaches. The first approach imposes specific parametric functional form assumptions, and characterizes the optimal taxes in terms of intuitive measures of elasticities. This approach goes back to Ramsey (1927) and the modern application of this technique was introduced by Diamond and Mirrlees (1971), who restrict attention to linear taxes. The second approach imposes explicit informational restrictions on the government and characterizes the constrained optimum (e.g. Mirrlees 1971, Golosov, Kocherlakota, and Tsyvinski 2003). Both approaches have limitations. As far as the Ramsey approach is concerned, the vast majority of the literature restricts attention to linear taxes. A few papers that consider more general taxes typically choose specific functional forms that are easy to parameterize (e.g. power functions) and obtain analytical results only in special cases. The mechanism design approach is often sensitive to the assumptions on government's information set. The tax systems that emerge from it are often very complex, and the intuition for the economic forces that determine the size and the shape of the optimal taxes is not transparent.

In this paper we develop an alternative approach to the analysis of the effects of taxation that both preserves the transparency of the Ramsey approach and allows us to handle more complicated, nonlinear tax systems. Our approach is based on studying perturbations of a given non-linear tax system directly. We show that as long as the baseline tax system is sufficiently well behaved, the effect of perturbing the tax system can be expressed in terms of elasticities and hazard rates of income distributions that can be estimated in the data. Our method is sufficiently flexible to both allow us to restrict attention *a priori* to a given class of tax functions (e.g., non-linear taxes that do not depend on an individual's age and are separable between various incomes) and to study the sources and the magnitude of welfare gains that arise from using more sophisticated taxes (e.g., from introducing age- or history-dependent taxation).

We study a dynamic model, in which individuals' characteristics evolve over their lifetime.[2] The tax system consists of a sequence of tax functions which can be arbitrarily non-linear and joint in the entire history of labor and capital incomes. The generality of the tax functions

[1] This becomes even more complicated once state-level programs are taken into account, see Maag et al. (2012).

[2] In most of the paper we focus on a deterministic economy to make our approach transparent. In the last Section and in our working paper (Golosov, Tsyvinski, and Werquin 2014) we develop an extension to stochastic environments.

allows us to study the age-dependence and history dependence of taxes, non-linear taxation of capital income, and joint conditioning of taxes on labor and capital incomes. The first main contribution of the paper is to provide a general formula for the welfare effects of tax reforms in a compact and easily interpretable form. Our result is based on deriving the Gateaux differential of individual demands, government tax revenue, and social welfare. The complexity of the problem arises from the fact that an individual chooses his income as a function of the local characteristics of the tax system; but these tax rates themselves depend on the income that he chooses. Therefore, a local perturbation of the tax function faced by an individual leads him to adjust his income, which in turn induces a shift in the tax rates if the baseline tax system is non-linear, triggering further income adjustment. We provide a sufficient condition on the individual demand (namely, local Lipschitz continuity), which allows us to solve this circularity issue, and express the effects of general tax reforms only in terms of the local income and substitution effects at the individual level, and of the curvature of the baseline tax function. Importantly, these formulas are written only as a function of empirically observable and easily interpretable sufficient statistics.

We then show several applications of these results. First, we apply it to optimal taxation problems and show how it recovers the hallmark results on optimal linear commodity taxation of Diamond (1975) and non-linear labor taxation of static model of Mirrlees (1971), both of which are special cases of our general environment. Our formulas emphasize the insight that the same general principle underlies the two models, namely that more sophisticated (in this case, non-linear) tax instruments allow the government to better target the distortions associated with higher tax rates toward the segments of the populations that have either relatively small behavioral responses, or where relatively few individuals are affected. We then show that this fundamental principle can be generalized and applies to broader classes of environments. In particular, we derive several novel predictions such as the optimality conditions for the optimal non-linear capital income tax, or for the optimal labor tax on joint income of couples.

We next turn to the analysis of tax reforms, and refine our discussion of the close connection that exists between the effects of the various tax instruments (age-dependent, non-linear, joint taxes). We sequentially decompose the welfare gains of reforming existing, not necessarily optimal, tax systems as the tax instruments become more sophisticated. We show the effects of taking into account individuals' intertemporal optimization decisions, of allowing for age- and history-dependence, and of joint conditioning of labor and capital income. This sequential decomposition of increasingly sophisticated tax systems shows that the welfare effects of general tax reforms depend on aggregate measures of three key elements: the marginal social welfare weights, which summarize the government's redistributive objective; the labor and capital income elasticities and income effect parameters with respect to the marginal income tax rates, which capture the behavioral effects of taxes; and the properties of the labor and capital income distributions, namely the hazard rates of the marginal and joint distributions.

Finally, we show how one can use available empirical moments of income distributions and elasticities to quantify the welfare effects of small tax reforms. Unlike the traditional approach to measuring welfare gains, which requires solving often difficult maximization problems to find the optimum, our method is very transparent and can be done almost "by hand". It does not allow, however, to compute the gains from reforms that introduce large changes in the existing tax system.

Our approach is most closely related to and builds on the work of Piketty (1997) and Saez (2001). They used heuristic arguments to extend the techniques of Ramsey (1927) and Diamond and Mirrlees (1971) to non-linear taxation and obtain expressions for the optimal labor taxes in a unidimensional static model in terms of the elasticities of labor supply and income hazard rates. Our paper finds sufficient conditions for the more rigorous application of that approach and extends it to more general dynamic settings. This allows us to analyze such questions as non-linear capital taxation and joint taxation of incomes. We also show how this approach can be used beyond optimal taxation, as we apply it to analyze tax reforms and welfare gains from increased sophistication of tax systems. More broadly, our approach is also related to the sufficient statistics tax literature (e.g., Chetty 2009; Piketty, Saez, and Stancheva 2013). Similar to these papers, we express our tax formulas in terms of a small number of empirically observable parameters, which fully characterize the effects of taxes for a large set of underlying models, e.g., for very general utility functions, structures of heterogeneity, etc. Our application to capital taxation also builds on several insights of Piketty and Saez (2013) and Straub and Werning (2014).

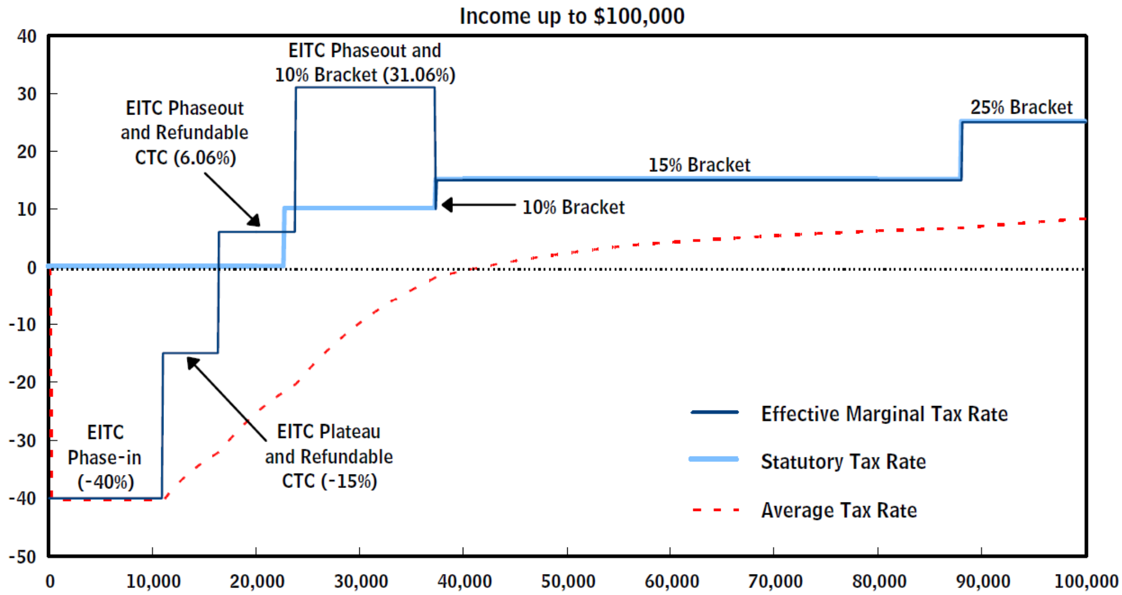
Our analysis of the evaluating increasingly sophisticated elements of the tax functions is most closely related to the growing literature on taxation within parametrically restricted sets of policies.[3] Conesa and Krueger (2009) and Conesa, Kitao, and Krueger (2009) quantitatively study optimal income tax within a Gouveia and Strauss (1994) class and a linear capital income tax. Kitao (2010) further incorporates the labor-dependent capital taxes in these models. Heathcote, Storesletten, and Violante (2014) characterize optimal taxes within a class where a government can control the degree of progressivity in dynamic models with idiosyncratic shocks and ex-ante heterogeneity in learning ability and disutility of work. A recent paper by Heathcote and Tsujima (2014) analyzes an environment in which groups of individuals can insure the shocks among themselves in addition to available private insurance, solve for the constrained optimum and explore whether parametric tax functions can come close to achieving those allocations. Huggett and Parra (2010) consider an optimal reform of the social security benefit function which is chosen in the class of piecewise-linear functions of average past earnings.

Our work is also related to the literature that studies tax reforms in dynamic settings. Weinziel (2012) is the closest in its approach to our analysis of the partial reform that introduces age-dependence.[4]Blundell and Shephard (2014) characterize numerically the optimal tax system

[3]For a more detailed review of the literature and of the relationship between our approach, the mechanism design approach to optimal taxation and the restricted tax functions approach see Golosov and Tsyvinski (2014).

[4]See also an early paper on this topic by Kremer (2002).

Figure 1: Effective Federal Tax Rates (source: CBO 2005)



in a complex dynamic environment. We complement their analysis of this issue by uncovering the theoretical forces which determine the effects of partial reforms.

A large recent "New Dynamic Public Finance" literature studies environments in which taxes are restricted only by explicit restrictions on government's information set. [5] Our approach is complementary. If we restrict attention to the classes of taxes considered in those paper, we obtain an alternative characterization of optimality conditions in terms of elasticities. More generally, it is easy to use our approach to analyze the tax systems using restricted tax instruments, e.g., non-linear but separable from labor income taxes, and quantify welfare gains from switching to more sophisticated taxes, e.g. gains from introducing joint taxation of capital and labor. More broadly, our paper relates to the literature on multidimensional screening problems, e.g., Kleven, Kreiner and Saez (2009) and Rothschild and Scheuer (2014). While they are able to solve this complex problem by collapsing their model to a one-dimensional problem, we use the variational method to analyze the environment with multiple types of shocks.

The rest of the paper is organized as follows. Section 2 describes our environment. Sections 3 and 4 derive the responses of individual income, tax revenue and social welfare to perturbations of the baseline tax system. Section 5 considers the applications of this approach to optimal taxation. Section 6 considers the applications to tax reforms and the decomposition of welfare gains from increasing sophistication of the tax system. Section 7 presents a brief overview of the extension of our analysis to the stochastic model.

[5] See, e.g. Albanesi and Sleet (2006), Farhi and Werning (2013), Golosov, Troshkin and Tsyvinski (2014), Albanesi (2011) and Shourideh (2012), Stantcheva (2014).

2 Environment

There is a measure one of agents in the economy. An agent lives for $S \leq \infty$ periods, and time is indexed by $s = 1, \ldots, S$. At the beginning of period $s = 1$, there is a draw of an exogenous vector of n characteristics $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^n$ for each individual. These idiosyncratic shocks can be, for instance, the individual's initial level of capital stock k_0, his sequence of productivities, tastes, interest rates (i.e., investment opportunities), etc. over his lifetime. The environment is deterministic: individuals know at the beginning of period $s = 1$ their entire vector of characteristics $\boldsymbol{\theta}$.[6]

Given the draw of vector $\boldsymbol{\theta}$, the individual chooses in each period $s \in \{1, \ldots, S\}$ a level of consumption c_s, labor income y_s, and savings or borrowings k_s which yield capital income z_{s+1} in period $s + 1$.[7] The utility function U can be a general, not necessarily time-separable, function of the vector of choices of consumption, labor income and capital income. We assume that the utility function is increasing and concave in each period's consumption (and capital income if it enters explicitly the utility function), decreasing and convex in each period's labor income, and twice differentiable in all of its variables. An example of the utility function which we use in several applications is $U = \sum_{s=1}^{S} \beta^{s-1} u\left(c_s, y_s/\theta_s\right)$. In this case, θ_s is a shock to the productivity of labor supply in period s.

In each period $s \in \{1, \ldots, S\}$, the government levies a tax T_s. The tax liability $T_s\left(\cdot\right)$ in period s is a non-linear function of the individual's entire history of labor incomes $\{y_{s'}\}_{s'=1}^{S}$ and capital incomes $\{z_{s'+1}\}_{s'=1}^{S}$.[8,9] The sequence of tax functions $\{T_s\left(\cdot\right)\}_{s=1}^{S}$ is known to an individual at the beginning of period $s = 1$, and the government can commit to it. The initial tax system $\mathscr{T}$ thus consists of a set of tax functions $T_s : \mathbb{R}_+^S \times \mathbb{R}^S \to \mathbb{R}$ for each period $s \in \{1, \ldots, S\}$, where each function $T_s\left(\cdot\right)$ maps a choice of labor and capital incomes $\mathbf{x} \in \mathbb{R}_+^S \times \mathbb{R}^S$ to a tax liability $T_s\left(\mathbf{x}\right) \in \mathbb{R}$. The tax function in period s, $T_s\left(\{y_{s'}\}_{s'=1}^{S}, \{z_{s'+1}\}_{s'=1}^{S}\right)$, is assumed twice continuously differentiable in all of its $2S$ variables, that is $T_s \in \mathcal{C}^2\left(\mathbb{R}_+^S \times \mathbb{R}^S, \mathbb{R}\right)$ for all $s \in \{1, \ldots, S\}$.[10]

[6] The deterministic environment allows us to show the main insights most transparently. We extend the analysis to the stochastic environment Golosov, Tsyvinski and Werquin (2014), and present an overview in Section 7.

[7] The capital income in period $s + 1$ can be written as $z_{s+1} = r_{s+1} k_s$, where the interest rate r_{s+1} in each period is exogenous. Our analysis allows the interest rate to be idiosyncratic, and thus the period-s savings k_s to yield any (deterministic) income z_{s+1} in the next period. The before-tax capital stock at the beginning of period $s + 1$ is then $k_s + z_{s+1}$.

[8] In a given period s, the planner can tax incomes earned in the future periods $s' > s$ because the model is deterministic. We assume here that initial capital k_0 is not taxed, because it is supplied inelastically and hence does not induce any behavioral effects. Our formulas can be trivially extended to the case where it can be taxed.

[9] Throughout the paper we consider only capital income taxes and not wealth taxes. The same approach can be used to analyze wealth taxation.

[10] In the deterministic model, we could without loss of generality write only one tax function, for instance $T_S\left(\cdot\right)$. Instead, we choose to define one tax function per period s, at the expense of slightly more cumbersome notation, to make it easier to discuss age-dependent taxes and capital taxes in Sections 5 and 6. This definition of taxes also has the advantage of extending naturally to stochastic settings; see Section 7.

The optimization problem of an individual with the vector of types $\boldsymbol{\theta}$ is:

$$\begin{aligned}\mathscr{U}_{\boldsymbol{\theta}}(\mathscr{T}) \equiv \max_{\{c_s, y_s, k_s, z_{s+1}\}_{1\le s\le S}} \quad & U\left(\{c_s\}_{1\le s\le S}, \{y_s\}_{1\le s\le S}, \{z_{s+1}\}_{1\le s\le S}, \boldsymbol{\theta}\right) \\ \text{s.t.} \quad & c_s + k_s = y_s + (k_{s-1} + z_s) - T_s\left(\{y_{s'}\}_{1\le s'\le S}, \{z_{s'+1}\}_{1\le s'\le S}\right), \;\; \forall s.\end{aligned} \tag{1}$$

We denote by $\mathbf{x}_{\boldsymbol{\theta}}(\mathscr{T})$ the argmax of this problem, i.e., the optimal choice of labor and capital incomes of the individual $\boldsymbol{\theta}$ as a function of the tax system $\mathscr{T}$. That is, we define the *individual income functional* as:

$$\mathbf{x}_{\boldsymbol{\theta}}(\mathscr{T}) = \left(y_{\boldsymbol{\theta},1}(\mathscr{T}), \dots, y_{\boldsymbol{\theta},S}(\mathscr{T}), z_{\boldsymbol{\theta},2}(\mathscr{T}), \dots, z_{\boldsymbol{\theta},S+1}(\mathscr{T})\right)'.$$

The optimal choices of consumption $\{c_{\boldsymbol{\theta},s}\}_{s=1}^{S}$ are then obtained from the budget constraints. The budget constraint in period s imposes that the sum of consumption c_s and savings k_s is no greater than the sum of labor income y_s and capital income $(k_{s-1} + z_s)$, net of the tax liability T_s.

We denote by $F_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ the c.d.f. of vectors $\boldsymbol{\theta} \in \Theta$, and $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ the corresponding density function. We also denote by $F_{\mathbf{x}}(\mathbf{x})$ and $f_{\mathbf{x}}(\mathbf{x})$ the c.d.f. and the p.d.f. of incomes $\mathbf{x} \in \mathrm{X} \subset \mathbb{R}_+^S \times \mathbb{R}^S$, given the tax system $\mathscr{T}$. For any choice of incomes $(x_1, \dots, x_s) \in \mathbb{R}^s$ with $s < 2S$, we also let $F_{x_1,\dots,x_s}(\cdot)$ and $f_{x_1,\dots,x_s}(\cdot)$ denote the marginal c.d.f. and the marginal p.d.f. of those variables. We assume that the sets Θ and X of vectors of types $\boldsymbol{\theta}$ and incomes $\mathbf{x}$ are compact in $\mathbb{R}^n$ and $\mathbb{R}_+^S \times \mathbb{R}^S$, respectively, and that the densities of types and incomes at the (piecewise smooth) boundaries $\partial\mathrm{X}$ and $\partial\Theta$ of the sets X and Θ are equal to zero.[11] We make the following assumption about the income vectors chosen by individuals with different types $\boldsymbol{\theta}$:

Assumption 1. *The map $\boldsymbol{\theta} \mapsto \mathbf{x}_{\boldsymbol{\theta}}(\mathscr{T})$ between the vector of types $\boldsymbol{\theta}$ and the vector of income choices $\mathbf{x}_{\boldsymbol{\theta}}$ (given the tax system $\mathscr{T}$) is injective. That is, if two individuals have a different vector of types $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$, they choose a different vector of incomes $\mathbf{x}_{\boldsymbol{\theta}}(\mathscr{T}) \neq \mathbf{x}_{\boldsymbol{\theta}'}(\mathscr{T})$.*

This assumption mainly simplifies the exposition but is not required for most of our results. We explain below how our main results are affected in the case where it does not hold, e.g., if the space of degrees of heterogeneity has a higher dimension than the space of income choices.

We define the present discounted value of tax revenue as a function of the tax system $\mathscr{T}$, or *tax revenue functional*, as

$$\mathscr{R}(\mathscr{T}) = \int_{\Theta} \left[\sum_{s=1}^{S} \beta^{s-1} T_s\left(\mathbf{x}_{\boldsymbol{\theta}}(\mathscr{T})\right)\right] dF_{\boldsymbol{\theta}}(\boldsymbol{\theta}), \tag{2}$$

where β is the marginal rate of transformation of resources across periods for the government, which we assume equal to the individual's discount factor. Tax revenue is thus the sum over

[11] In some applications of Sections 6 and 5, we let incomes evolve in the whole space $\mathbb{R}_+^S \times \mathbb{R}^S$. Our theory can be generalized to this case by using an increasing sequence of compact sets $X \subset \mathbb{R}_+^S \times \mathbb{R}^S$.

time $s \in \{1, \ldots, S\}$ and over individuals $\boldsymbol{\theta} \in \Theta$ of individual tax liabilities, taking into account the agents' optimizing behavior given the tax system $\mathscr{T}$.

We finally define the *social welfare functional* as a weighted average of the indirect utility functions of individual agents and the tax revenue, as a function of the tax system $\mathscr{T}$,

$$\mathscr{W}(\mathscr{T}) = \lambda^{-1}\left[(1-\alpha)\int_{\Theta} \mathcal{G}\left(\mathscr{U}_{\boldsymbol{\theta}}(\mathscr{T})\right) dF_{\boldsymbol{\theta}}(\boldsymbol{\theta}) + \alpha \mathcal{V}\left(\mathscr{R}(\mathscr{T})\right)\right], \tag{3}$$

for some $\alpha \in [0,1]$, where $\lambda \equiv \alpha \mathcal{V}'\left(\mathscr{R}(\mathscr{T})\right)$ denotes the shadow value of public funds. Here $\mathcal{V}\left(\mathscr{R}(\mathscr{T})\right)$ is a measure of the value of public goods that the government can provide with tax revenues $\mathscr{R}(\mathscr{T})$. The function $\mathcal{G} : \mathbb{R} \to \mathbb{R}$ is defined over lifetime utilities of the individuals, and is assumed continuously differentiable, increasing, and concave. The function $\mathcal{V} : \mathbb{R} \to \mathbb{R}$ is continuously differentiable and increasing. Note that normalizing equation (3) by the marginal value of public funds λ implies that social welfare is expressed in monetary units.

The tax system $\mathscr{T}$ considered so far is very general and allows for a rich set of non-linearities and non-separabilities in taxing different incomes at different dates. In practice, we are often interested in more restrictive classes of tax systems. For example, the classic Ramsey analysis restricts the tax functions to be separable and linear in each income (e.g., Ramsey 1927, Diamond and Mirrlees 1971, Diamond 1975). Another strand of the literature focuses on the analysis of separable but non-linear tax functions, (e.g., Mirrlees 1971, Diamond 1998, Saez 2001, Heathcote, Storesletten, and Violante 2014). More generally, the New Dynamic Public Finance literature (e.g., Kocherlakota 2005, Farhi and Werning 2012, Golosov, Troshkin, and Tsyvinski 2014) emphasizes the benefits of jointly taxing different incomes, namely labor and capital incomes within periods, or labor incomes across periods (history-dependent taxation), so that the tax rate on income i depends not only on its own level x_i, but also on the levels x_j of other incomes $j \neq i$. When we impose such constraints on the tax system $\mathscr{T}$, we say that $\mathscr{T}$ is "restricted within a class" (e.g., of linear separable, non-linear separable, etc., tax functions).

Our paper focuses on several conceptually distinct, but closely related questions. First, we analyze the revenue or welfare gains and losses of small perturbations of any *baseline tax system* $\mathscr{T}^0$. We refer to such changes as *tax reforms*. Suppose in particular that the tax system $\mathscr{T}^0$ is restricted within a certain class. By deriving the effects of reforms that keep the perturbed tax system within this class, we can shed light on the economic parameters that determine whether the existing tax system is (constrained) optimal, and derive the potential welfare gains obtained by reforming it. Moreover, we can analyze reforms that induce the tax system to leave its restricted class. For instance, we can introduce (a small amount of) non-linearity, age-dependence, history-dependence or joint taxation within a baseline linear, age-independent, or separable tax system. This allows us to sequentially decompose the gains arising from each additional element of reform as the tax code becomes more sophisticated. Second, we derive characterizations of the *optimal tax system*, or the optimum within a certain class. These two questions are closely related, because the characterization of the optimum is obtained by imposing that the net welfare effect of any tax reform is non-positive if the baseline tax system $\mathscr{T}^0$ is optimal.

3 Behavioral Effects of Tax Reforms

In this section, we formally define the admissible perturbations of the initial tax system, i.e., our tax reforms, and study their effect on individual behavior. We start with a *baseline tax system* $\mathscr{T}^0 = \left\{T_p^0\right\}_{1\leq p\leq S}$, and consider another tax system $\mathscr{H} = \{h_p\}_{1\leq p\leq S}$. The system $\mathscr{H}$ consists of a set of tax functions $h_p : \mathbb{R}_+^S \times \mathbb{R}^S \to \mathbb{R}$ for each period $p \in \{1, \ldots, S\}$, as defined in Section 2. Our goal is to analyze the revenue and welfare effects of reforming the baseline tax system $\mathscr{T}^0$ *"in the direction $\mathscr{H}$"*. Formally, for $\mu \in \mathbb{R}_+$, we then define the *perturbed tax system $\tilde{\mathscr{T}}$* as $\tilde{\mathscr{T}} = \mathscr{T}^0 + \mu\mathscr{H}$. That is, the perturbed tax function in any period p is given by $\tilde{T}_p = T_p^0 + \mu h_p$. We then derive the change in tax revenue or social welfare following this perturbation as $\mu \to 0$. Hence, we compute the *Gateaux differential* of the tax revenue and social welfare functionals, following the local perturbation of the baseline tax system $\mathscr{T}^0$ in the direction $\mathscr{H}$.

We can decompose this general perturbation $\mathscr{H}$ of the tax system into its period-p components $h_p : \mathrm{X} \to \mathbb{R}$, which only affect the period-p baseline tax function $T_p : \mathbb{R}_+^S \times \mathbb{R}^S \to \mathbb{R}$. The total effect of the perturbation $\mathscr{H}$ is then equal to the sum over periods p of the effects of the elementary perturbations h_p. Without loss of generality we can thus restrict the analysis to perturbations of a given period-p tax function $T_p(\cdot)$, and keep the rest of the baseline tax system $\mathscr{T}^0$ unchanged. We therefore define an *admissible perturbation* of the baseline tax function $T_p(\cdot)$ as a twice continuously differentiable function $h_p \in \mathcal{C}^2(\mathrm{X}, \mathbb{R})$. For any $\mu > 0$, we then define a perturbed function in period p as $\tilde{T}_p = T_p + \mu h_p$, and study the effects of this tax reform as $\mu \to 0$.

We say that the perturbation h_p is *restricted within a class* if it leaves the perturbed tax system $\tilde{\mathscr{T}}$ in the same class (e.g., of linear, separable, etc., tax systems) as the baseline tax system $\mathscr{T}^0$. As a first step towards deriving the effects on social welfare and tax revenue of the perturbation h_p, we characterize in this section its effects on the optimal individual behavior. That is, we compute the Gateaux differential of the individual income functional $\mathbf{x}_{\boldsymbol{\theta}}\left(\mathscr{T}^0\right)$ in the direction h_p.

We first characterize the solution to the problem (1) of individual $\boldsymbol{\theta}$, i.e., his choice of incomes $\mathbf{x}_{\boldsymbol{\theta}} \in \mathrm{X}$ given a tax system $\mathscr{T}$. Since we define perturbations of the tax function in each given period, we denote the vector of optimal choices of individual $\boldsymbol{\theta}$ under the baseline tax system either as $\mathbf{x}_{\boldsymbol{\theta}}(T_p)$ or as $\mathbf{x}_{\boldsymbol{\theta}}\left(\mathscr{T}^0\right)$, and under the perturbed tax system as $\mathbf{x}_{\boldsymbol{\theta}}(T_p + \mu h_p)$ or as $\mathbf{x}_{\boldsymbol{\theta}}\left(\tilde{\mathscr{T}}\right)$. We denote by $\{x_{\boldsymbol{\theta},j}\}_{1\leq j\leq 2S}$ the components of the income vector $\mathbf{x}_{\boldsymbol{\theta}}$, that is the labor incomes y_s and capital incomes z_s in each period $s \in \{1, \ldots, S\}$.

The main result of this section is that as long as $\mathbf{x}_{\boldsymbol{\theta}}$ is well-behaved (in a formal sense given in Assumption 2), the directional derivatives of $\mathbf{x}_{\boldsymbol{\theta}}$ are well-defined and can be expressed in terms of income and price elasticities. Before we state this result, we need to define these elasticities in settings where the consumers' budget constraints are non-linear.

3.1 Elasticities

We build our definitions of elasticities on the classical consumer demand theory that uses linear budget constraints. We start by defining the individual income responses to tax changes in an economy with linear budget constraints and then extend these definitions to non-linear constraints.

Let $q_{j,s}$ be the after-tax price of good x_j (where $x_j = y_j$ for $1 \leq j \leq S$ and $x_j = z_{j+1}$ for $S+1 \leq j \leq 2S$) in period s and R_s be the exogenous lump-sum income. Define

$$\mathcal{H}_s(\mathbf{x}) \equiv R_s - \sum_{j=1}^{S} q_{j,s} y_j - \sum_{j=1}^{S} q_{S+j,s} z_{j+1}.$$

$\mathcal{H}_s(\mathbf{x})$ is a hyperplane defined by a vector that consists of $2S$ prices $q_{j,s}$ and the exogenous income R_s. We use $(\mathbf{q}, \mathbf{R})$ to denote the $2S \times S$-matrix of prices $\{q_{j,s}\}_{\substack{1 \leq j \leq 2S \\ 1 \leq s \leq S}}$ and the $1 \times S$-vector of exogenous incomes $\{R_s\}_{1 \leq s \leq S}$ in all S periods, and we let $\mathcal{H}^{(\mathbf{q},\mathbf{R})}$ denote the S hyperplanes defined by $(\mathbf{q}, \mathbf{R})$. The budget constraint of the consumer can be written compactly as $\mathbf{c} = \mathcal{H}^{(\mathbf{q},\mathbf{R})}(\mathbf{x})$, where $\mathbf{c} = \{c_s\}_{1 \leq s \leq S}$ is the vector of consumption choices in the S periods. Let $\hat{\mathbf{x}}_{\boldsymbol{\theta}}(\mathbf{q}, \mathbf{R})$ be the (Marshallian) demand of the individual $\boldsymbol{\theta}$ who faces a sequence of budget constraints defined by $\mathcal{H}^{(\mathbf{q},\mathbf{R})}$, and let $\hat{x}_{\boldsymbol{\theta},i}$ denote the i^{th} component of $\hat{\mathbf{x}}_{\boldsymbol{\theta}}$. We define the income and uncompensated price responses in the usual way. For example, the $2S \times 2S$-matrix $\boldsymbol{E}_s^{u,\boldsymbol{\theta}}$ of changes in uncompensated incomes x_i with respect to the period-s prices $q_{j,s}$, and the $2S \times 1$-vector $\boldsymbol{I}_s^{\boldsymbol{\theta}}$ of changes in incomes x_i with respect to the period-s virtual income R_s, are given by

$$\left[\boldsymbol{E}_s^{u,\boldsymbol{\theta}}\right]_{i,j} = \frac{\partial \hat{x}_{\boldsymbol{\theta},i}}{\partial q_{j,s}} \quad \text{and} \quad \left[\boldsymbol{I}_s^{\boldsymbol{\theta}}\right]_i = \frac{\partial \hat{x}_{\boldsymbol{\theta},i}}{\partial R_s}, \tag{4}$$

while the compensated (Hicksian) changes are defined through the Slutsky relationship

$$\left[\boldsymbol{E}_s^{c,\boldsymbol{\theta}}\right]_{i,j} = \left[\boldsymbol{E}_s^{u,\boldsymbol{\theta}}\right]_{i,j} + \left[\boldsymbol{I}_s^{\boldsymbol{\theta}}\right]_i \hat{x}_{\boldsymbol{\theta},j}. \tag{5}$$

At this stage this is a straightforward generalization of standard consumer theory (see Mas-Colell, Whinston and Green, 1995) when the purchase of good $x \in \{y, z\}$ in period j requires expenditures $q_{j,s} x_j$ in all periods s at prices that depend on s. This general formulation is helpful for our analysis of general tax systems and will allows us to capture the effects of, for example, history-dependence, when taxes on choices made in a given period depend also on other choices made in previous periods. Classical demand theory also imposes restrictions on how the responses of demand $\hat{x}_{\boldsymbol{\theta},i}$ to changes in income R_s and prices $q_{j,s}$ in different periods s are related to each other, but in principle we can allow those responses to be free parameters. For our purposes it is important to note that all the parameters in the matrices $\boldsymbol{E}^c$, $\boldsymbol{E}^u$ and $\boldsymbol{I}$

depend on the pair $(\mathbf{q},\mathbf{R})$ that defines the hyperplanes $\mathcal{H}^{(\mathbf{q},\mathbf{R})}$, and on the vector of types of the consumer, $\boldsymbol{\theta}$.

From standard consumer demand theory it is easy to infer the response of demand $\hat{\mathbf{x}}_{\boldsymbol{\theta}}$ to any perturbation of the hyperplane $\mathcal{H}^{(\mathbf{q},\mathbf{R})}$. To fix ideas, consider a pair $\left(\tilde{\mathbf{q}},\tilde{\mathbf{R}}\right)$ with the property that the column vectors $\tilde{\mathbf{q}}_s \equiv \{\tilde{q}_{j,s}\}_{j=1}^{2S}$ and the scalars $\tilde{R}_s$ are equal to zero for all but one s, and consider the response of $\hat{\mathbf{x}}_{\boldsymbol{\theta}}$ to the perturbation $\left(\mathbf{q}+\mu\tilde{\mathbf{q}},\mathbf{R}+\mu\tilde{\mathbf{R}}\right)$. This experiment corresponds to a perturbation of the linear budget constraint only in one period s. Standard arguments imply that response of demand $\hat{\mathbf{x}}_{\boldsymbol{\theta}}$ to such a perturbation around $(\mathbf{q},\mathbf{R})$ in the direction $\left(\tilde{\mathbf{q}},\tilde{\mathbf{R}}\right)$ for small μ, $\delta\hat{\mathbf{x}}_{\boldsymbol{\theta}}\left((\mathbf{q},\mathbf{R});\left(\tilde{\mathbf{q}},\tilde{\mathbf{R}}\right)\right)$, is given by

$$\delta\hat{\mathbf{x}}_{\boldsymbol{\theta}}\left((\mathbf{q},\mathbf{R});\left(\tilde{\mathbf{q}},\tilde{\mathbf{R}}\right)\right) \equiv \lim_{\mu\to 0}\frac{\hat{\mathbf{x}}_{\boldsymbol{\theta}}\left(\mathbf{q}+\mu\tilde{\mathbf{q}},\mathbf{R}+\mu\tilde{\mathbf{R}}\right)-\hat{\mathbf{x}}_{\boldsymbol{\theta}}\left(\mathbf{q},\mathbf{R}\right)}{\mu} = \boldsymbol{E}_s^{u,\boldsymbol{\theta}}\tilde{\mathbf{q}}_s + \boldsymbol{I}_s^{\boldsymbol{\theta}}\tilde{R}_s, \tag{6}$$

so that for all $i\in\{1,\ldots,2S\}$, the i^{th} component of the vector $\delta\hat{\mathbf{x}}_{\boldsymbol{\theta}}$ writes

$$\delta\hat{x}_{\boldsymbol{\theta},i} = \sum_{j=1}^{2S}\frac{\partial\hat{x}_{\boldsymbol{\theta},i}}{\partial q_{j,s}}\tilde{q}_{j,s} + \frac{\partial\hat{x}_{\boldsymbol{\theta},i}}{\partial R_s}\tilde{R}_s.$$

We can now extend these definitions to our environment with non-linear taxes. We define the hyperplane $\mathcal{H}_{\boldsymbol{\theta}}$ that is *tangent* to the individual $\boldsymbol{\theta}$'s budget constraint evaluated at the optimal income vector $\mathbf{x}_{\boldsymbol{\theta}}$. In particular let

$$\begin{aligned}
q_{j,s}\left(\mathbf{x}_{\boldsymbol{\theta}}\right) &\equiv \frac{\partial T_s\left(\mathbf{x}_{\boldsymbol{\theta}}\right)}{\partial x_j} && \text{if } x_j\notin\{y_s,z_s,z_{s+1}\},\\
q_{j,s}\left(\mathbf{x}_{\boldsymbol{\theta}}\right) &\equiv -\left(1-\frac{\partial T_s\left(\mathbf{x}_{\boldsymbol{\theta}}\right)}{\partial x_j}\right) && \text{if } x_j=y_s,\\
q_{j,s}\left(\mathbf{x}_{\boldsymbol{\theta}}\right) &\equiv -\left(\frac{1}{r_s}+1-\frac{\partial T_s\left(\mathbf{x}_{\boldsymbol{\theta}}\right)}{\partial x_j}\right) && \text{if } x_j=z_s,\\
q_{j,s}\left(\mathbf{x}_{\boldsymbol{\theta}}\right) &\equiv \frac{1}{r_{s+1}}+\frac{\partial T_s\left(\mathbf{x}_{\boldsymbol{\theta}}\right)}{\partial x_j} && \text{if } x_j=z_{s+1},\\
R_s\left(\mathbf{x}_{\boldsymbol{\theta}}\right) &\equiv \left\langle\nabla T_s\left(\mathbf{x}_{\boldsymbol{\theta}}\right),\mathbf{x}_{\boldsymbol{\theta}}\right\rangle - T_s\left(\mathbf{x}_{\boldsymbol{\theta}}\right).
\end{aligned}\tag{7}$$

where r_s is the (possibly idiosyncratic) interest rate, i.e. z_s/k_{s-1}.[12] The hyperplane $\mathcal{H}_{\boldsymbol{\theta}}$ defined by $\left(\mathbf{q}\left(\mathbf{x}_{\boldsymbol{\theta}}\right),\mathbf{R}\left(\mathbf{x}_{\boldsymbol{\theta}}\right)\right)$ is tangent to the budget constraint defined by taxes $\mathscr{T}^0$ at a point $\mathbf{x}_{\boldsymbol{\theta}}$. Since utility is strictly concave, the optimal income choice $\mathbf{x}_{\boldsymbol{\theta}}$ of individual $\boldsymbol{\theta}$ given the baseline non-linear tax function $\mathscr{T}^0$ coincides with the income vector $\hat{\mathbf{x}}_{\boldsymbol{\theta}}$ that he would choose if he faced the

[12] Note that some of these prices are negative, because they are those of (say) labor income rather than leisure. This allows us to express the individual's budget constraint as $\mathbf{c}=\mathcal{H}^{(\mathbf{q},\mathbf{R})}(\mathbf{x})$. The price on income x_j is equal to the marginal tax rate on good x_j plus a constant, so that the derivatives of income w.r.t. those prices are the same as the derivatives w.r.t. the marginal tax rates. Below we define the elasticities that we use in our applications, which are w.r.t. the net-of-tax rates rather than the prices when these are negative.

linear budget constraint defined by $\mathcal{H}_{\boldsymbol{\theta}}$. Assumption 1 ensures that for each $\mathbf{x}$ chosen under the tax $\mathscr{T}^0$ there is only one type $\boldsymbol{\theta}$ for whom it is the optimal choice. We use this observation to define for all s the behavioral responses at a vector $\mathbf{x}$ as

$$\boldsymbol{E}_s^c(\mathbf{x}) \equiv \boldsymbol{E}_s^{c,\boldsymbol{\theta}}, \boldsymbol{E}_s^u(\mathbf{x}) \equiv \boldsymbol{E}_s^{u,\boldsymbol{\theta}}, \boldsymbol{I}_s(\mathbf{x}) \equiv \boldsymbol{I}^{\boldsymbol{\theta}} \text{ for } \boldsymbol{\theta} \text{ such that } \mathbf{x} = \hat{\mathbf{x}}_\theta.$$

Responses are not defined for those $\mathbf{x}$ that are not chosen by any individual, but it turns out to be irrelevant for our analysis. Note that behavioral responses generally depend on the demand $\mathbf{x}$ at which they are evaluated.[13]

We can finally define in the usual way the compensated and uncompensated price *elasticities* $\zeta_{i,j,s}^{c,\boldsymbol{\theta}}$, $\zeta_{i,j,s}^{u,\boldsymbol{\theta}}$ and the income elasticities $\eta_{i,s}^{\boldsymbol{\theta}}$. In order to be consistent with the empirical literature that estimates such elasticities, we define in the case of linear budget constraints the elasticities of $\hat{x}_{\boldsymbol{\theta},i}$ with respect to the *modified prices* $\hat{q}_{j,s}$ that are defined as the (constant) marginal tax rates $\tau_{x_j,s} \equiv \frac{\partial T_s(\hat{\mathbf{x}}_{\boldsymbol{\theta}})}{\partial x_j}$ if $x_j \notin \{y_s, z_s\}$ and the net-of-tax rates $1 - \tau_{x_j,s}$ otherwise. Note that the behavioral responses to the modified prices are given by $\partial \hat{x}_{\boldsymbol{\theta},i}/\partial \hat{q}_{j,s} = \partial \hat{x}_{\boldsymbol{\theta},i}/\partial q_{j,s}$ if $x_j \notin \{y_s, z_s\}$, and $\partial \hat{x}_{\boldsymbol{\theta},i}/\partial \hat{q}_{j,s} = -\partial \hat{x}_{\boldsymbol{\theta},i}/\partial q_{j,s}$ otherwise. Similarly, we define the income elasticity $\eta_{i,s}^{\boldsymbol{\theta}}$ as the income response $\partial \hat{x}_{\boldsymbol{\theta},i}/\partial R_s$ weighted by the net-of-tax rate $1 - \tau_{x_i,s}$. Therefore, equation (6) can be re-written as

$$\delta \hat{x}_{\boldsymbol{\theta},i} = \sum_{j=1}^{2S} \frac{\partial \hat{x}_{\boldsymbol{\theta},i}}{\partial q_{j,s}} \tilde{q}_{j,s} + \frac{\partial \hat{x}_{\boldsymbol{\theta},i}}{\partial R_s} \tilde{R}_s = \sum_{j=1}^{2S} \frac{\hat{x}_{\boldsymbol{\theta},i}}{\hat{q}_{j,s}} \left[\pm \zeta_{i,j,s}^{u,\boldsymbol{\theta}}\right] \tilde{q}_{j,s} + \frac{1}{1 - \tau_{x_i,s}} \eta_{i,s}^{\boldsymbol{\theta}} \tilde{R}_s,$$

where $\pm$ stands for $+$ if $x_j \notin \{y_s, z_s\}$, and $-$ otherwise. The definitions of the income and price elasticities in the case of the non-linear budget constraints are extended as described above. We derive in the Appendix analytical expressions for all these elasticities and income effect parameters.[14]

3.2 The effect of non-linear perturbations

We now analyze the change in an individual's income vector $\mathbf{x}_{\boldsymbol{\theta}}$ in response to an admissible perturbation h_p of the period-p tax function T_p. The main difficulty in analyzing the effects of tax reforms is that the individual's demand $\mathbf{x}_{\boldsymbol{\theta}}$ depends on the characteristics (normal vector and intercept) of the tangent hyperplane $\mathcal{H}_{\boldsymbol{\theta}}$ that he faces, which in turn are determined by the vector of incomes $\mathbf{x}_{\boldsymbol{\theta}}$ that the individual optimally chooses. Therefore a perturbation of the individual's hyperplane has a direct effect on his demand, which in turn induces a shift in his hyperplane if the baseline tax function is non-linear. The key intermediate step of our analysis is to provide a sufficient condition on the individual demand which allows us to solve

[13] Our definitions are straightforward to generalize when Assumption 1 does not hold. In this case the definitions above are replaced by the average response by all types $\boldsymbol{\theta}$ who choose the vector $\mathbf{x}$ under the baseline tax schedule.

[14] In the general model, these expressions are of course complicated. We show in Section 6 that they significantly simplify if the utility function has no income effects on labor supply and the baseline tax system is separable.

this circularity issue. This condition, which states that the demand is "well behaved" in a formal sense, allows us to derive the change in income of each individual due to a perturbation using only the local income and substitution effects at the individual level, and the curvature of the baseline tax system. Specifically, we make the following assumption about the income functional $\mathbf{x}_{\boldsymbol{\theta}}$, which formalizes the idea that individuals' income choices do not change by a discrete amount in response to infinitesimal admissible perturbations of the initial tax system $\mathscr{T}^0$.[15]

Assumption 2. *The income functional $\mathbf{x}_{\boldsymbol{\theta}}(\cdot)$ is locally Lipschitz continuous in every direction h_p at the initial tax system T_p. That is, for any admissible perturbation $h_p \in \mathcal{C}^2(\mathrm{X}, \mathbb{R})$, there exist $\bar{\mu} > 0$ and $M > 0$ such that $\mu < \bar{\mu}$ implies $\|\mathbf{x}_{\boldsymbol{\theta}}(T_p + \mu h_p) - \mathbf{x}_{\boldsymbol{\theta}}(T_p)\| < M \times \mu$.*

Using Assumption 2, we are now able to derive formally the change in individual behavior in response to a perturbation h_p of the baseline tax system, that is, the Gateaux differential of the individual income function in the direction h_p. In response to this perturbation, all the labor and capital incomes chosen by the individual change simultaneously. We show that despite the apparent complexity of the problem, we can derive, using the matrix notations introduced above, a compact and transparent formula giving the change in individual's behavior following any such perturbation. We express these behavioral responses in terms of: (i) the elasticities and income effect parameters of the individual, and (ii) the local characteristics (gradient and Hessian) of the baseline tax function. We view the derivation of the formula in the next proposition and, most importantly, its compact and transparent representation, as one of the main contributions of this paper.

Proposition 1. *Suppose that Assumption 2 is satisfied. Then the income functional $\mathbf{x}_{\boldsymbol{\theta}}(\cdot)$ is Gateaux differentiable around the initial tax system. Its Gateaux differential at T_p in the direction h_p, $\delta\mathbf{x}_{\boldsymbol{\theta}}(T_p, h_p) \in \mathbb{R}^{2S}$, is given by:*

$$\delta\mathbf{x}_{\boldsymbol{\theta}}(T_p, h_p) = \left[\mathrm{i}_{2S} - \sum_{s=1}^{S} \boldsymbol{E}_s^c(\mathbf{x}_{\boldsymbol{\theta}}) D^2 T_s(\mathbf{x}_{\boldsymbol{\theta}})\right]^{-1} \left\{\boldsymbol{E}_p^c(\mathbf{x}_{\boldsymbol{\theta}}) \nabla h_p(\mathbf{x}_{\boldsymbol{\theta}}) - \boldsymbol{I}_p(\mathbf{x}_{\boldsymbol{\theta}}) h_p(\mathbf{x}_{\boldsymbol{\theta}})\right\}, \tag{8}$$

where i_{2S} the $2S \times 2S$ identity matrix, and $D^2 T_s(\mathbf{x})$ the Hessian of the tax function $T_s(\cdot)$.

Proof. See Appendix. □

We now sketch the main steps of the proof of Proposition 1. The individual's behavior under the baseline tax function T_p (resp., the perturbed tax function $\tilde{T}_p = T_p + \mu h_p$) is described by

[15] The same issue also arises in dual approach to optimal *linear* taxation developed by Diamond and Mirrlees (1971). In particular, to be able to express the effect of tax changes in terms of elasticities one needs to assume that consumer demand is differentiable in prices and lump-sum income at the optimal tax schedule. Differentiability of demand in classical consumer demand theory generally does not follow from differentiability and strict concavity of utility functions (Katzner 1968) and requires additional Lipschitzian assumptions on demand (Rader 1973, 1979). In our setting with non-linear taxes this assumption also rules out bunching (see, e.g., Rochet and Choné 1998).

the following system of first-order conditions

$$U_{x_j}\left(\mathcal{H}_{\boldsymbol{\theta}}\left(\mathbf{x}_{\boldsymbol{\theta}}\right), \mathbf{x}_{\boldsymbol{\theta}}, \boldsymbol{\theta}\right)=\sum_{s=1}^{S} q_{j, s}\left(\mathbf{x}_{\boldsymbol{\theta}}\right) U_{c_s}\left(\mathcal{H}_{\boldsymbol{\theta}}\left(\mathbf{x}_{\boldsymbol{\theta}}\right), \mathbf{x}_{\boldsymbol{\theta}}, \boldsymbol{\theta}\right), \tag{9}$$

where $q_{j,s}\left(\mathbf{x}_{\boldsymbol{\theta}}\right)$ are defined in (7). The first step is to write a Taylor expansion in the direction h_p of the perturbed set of equations, which yields the first-order (in μ) change in the income vector, $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}-\mathbf{x}_{\boldsymbol{\theta}}$. The local Lipschitz continuity of the income functional around the baseline tax system, Assumption 2, is the key to address the circularity issue discussed above, as it ensures that the change in income remains first-order in the size μ of the perturbation despite the feedback effect on demand generated by the endogenous shift of the tangent hyperplane along the baseline tax function. We obtain that the individual changes his income vector $\mathbf{x}_{\boldsymbol{\theta}}$ in response to the perturbation by an amount

$$d\mathbf{x}_{\boldsymbol{\theta}}=\left[\boldsymbol{E}_p^c\left(\mathbf{x}_{\boldsymbol{\theta}}\right) \nabla h_p\left(\mathbf{x}_{\boldsymbol{\theta}}\right)-\boldsymbol{I}_p\left(\mathbf{x}_{\boldsymbol{\theta}}\right) h_p\left(\mathbf{x}_{\boldsymbol{\theta}}\right)\right]+\left[\sum_{s=1}^{S} \boldsymbol{E}_s^c\left(\mathbf{x}_{\boldsymbol{\theta}}\right) D^2 T_s\left(\mathbf{x}_{\boldsymbol{\theta}}\right)\right] \times d\mathbf{x}_{\boldsymbol{\theta}}, \tag{10}$$

which leads to equation (8).

The first bracket of (10) is the direct effect of the perturbation. The perturbation increases marginal taxes at income level $\mathbf{x}_{\boldsymbol{\theta}}$ by $\nabla h_p\left(\mathbf{x}_{\boldsymbol{\theta}}\right)$ and the individual's response this induces is captured by the matrix of compensated elasticities $\boldsymbol{E}_p^c$. Moerover, the total tax liability at income $\mathbf{x}_{\boldsymbol{\theta}}$ is changed by $h_p\left(\mathbf{x}_{\boldsymbol{\theta}}\right)$, and the income effect of such a tax change is summarized by $\boldsymbol{I}_p$. The second bracket is the indirect effect due to the endogenous shift in the tax rates. Since the baseline tax system is non-linear, the consumer, who adjusts his demand in response to the original exogenous tax change, generally faces a different marginal tax rate after adjustment. The shift of the taxpayer demand $\mathbf{x}_{\boldsymbol{\theta}}$ along the non-linear tax function by $d\mathbf{x}_{\boldsymbol{\theta}}$ produces an additional change in the marginal rates in all periods $s \in \{1,\ldots,S\}$ equal to $d\left(\nabla T_s\right)\left(\mathbf{x}_{\boldsymbol{\theta}}\right)=\left(D^2 T_s\left(\mathbf{x}_{\boldsymbol{\theta}}\right)\right) d\mathbf{x}_{\boldsymbol{\theta}}$, and the behavioral responses from this change are captured by $\boldsymbol{E}_s^c$.

The important point to notice is that the consumer's response depends only on taxes and the perturbation evaluated at the consumer's optimal choice *in the baseline tax schedule*, before the perturbation. This allows us to express the response of any individual's choice of labor and capital incomes following any local perturbation of the baseline tax system, without having to solve for the optimization problem (1), and describe this response in terms of empirically observable and easily interpretable parameters. We can do so by substituting the values of the specific perturbation function h_p, i.e., its marginals $\nabla h_p\left(\mathbf{x}_{\boldsymbol{\theta}}\right)$ and its level $h_p\left(\mathbf{x}_{\boldsymbol{\theta}}\right)$ at point $\mathbf{x}_{\boldsymbol{\theta}}$, into equation (8). Importantly, this characterization is valid in very general settings and holds regardless of the specific dimensions of heterogeneity, utility functions, tax systems, etc., as long as the individual's first-order conditions and Assumption 2 are satisfied.

We conclude this section by providing two examples of application of formula (8). First, suppose that the baseline tax system is linear and separable, so that income x_p is taxed at the constant rate τ_p, for all $p \in \{1, \ldots, 2S\}$. Consider the perturbation $\mu h_p(\mathbf{x}) = \mu \tilde{\tau}_p x_p$ for all $\mathbf{x}$. This perturbation h_p corresponds to an increase in the marginal tax rate on good x_p by $\mu \tilde{\tau}_p$, with no change in lump-sum income. Let $h \equiv \sum_{p=1}^{2S} h_p$ denote the sum of these perturbations. Since the baseline tax system is linear, the second bracket in (10) is equal to zero and the behavioral response of income vector $\mathbf{x}$ as $\mu \to 0$ is equal, using the Slutsky equation (5), to

$$\delta \mathbf{x}(\mathscr{T}, h) = \sum_{p=1}^{2S} \boldsymbol{E}_p^u(\mathbf{x}) \cdot [0, \ldots, \tilde{\tau}_p, \ldots, 0]^{\mathrm{T}}.$$

The typical element of the vector $\delta \mathbf{x}(\mathscr{T}, h)$ can be written in elasticity form as

$$\delta x_i(\mathscr{T}, h) = -\sum_{p=1}^{2S} \frac{\tilde{\tau}_p}{1-\tau_p} x_i \zeta_{x_i, \hat{q}_p}^u(\mathbf{x}),$$

where $\zeta_{x_i, \hat{q}_p}^u(\mathbf{x}) \equiv \frac{1-\tau_p}{x_i} \frac{\partial x_i}{\partial(1-\tau_p)}$ is uncompensated elasticity of demand for good i with respect to a change in the net-of-tax rate (the "modified price" $\hat{q}_p$) on good x_p, evaluated at a vector $\mathbf{x}$.[16] This is just a standard expression that shows that individuals' response to an increase in linear tax rates is proportional to uncompensated behavioral responses.

Second, consider the static Mirrlees model ($S = 1$) with a non-linear labor income tax schedule $T(y)$, and let h_τ denote the perturbation of the baseline tax schedule defined by $h_\tau(y) = \max\{y - \hat{y}, 0\}$, for some fixed income level $\hat{y} \in \mathbb{R}$. For a given $\mu > 0$, the perturbed tax schedule is therefore given by $\tilde{T}(y) = T(y)$ if $y < \hat{y}$, and $\tilde{T}(y) = T(y) + \mu(y - \hat{y})$ if $y \geq \hat{y}$. Intuitively, this perturbation increases the marginal tax rate $T'(y)$ faced by individuals above the income threshold $\hat{y}$ by the same amount $\mu > 0$. Note that this perturbation introduces a kink in the tax system at $\hat{y}$, and hence strictly speaking it is not admissible. We smooth out the kink by defining instead the admissible perturbation $\tilde{h}_\tau$ as $\tilde{h}_\tau(y) = h_\tau(y)$ for all $y \notin [\hat{y} - u, \hat{y} + u]$ for some small $u > 0$, and letting $\tilde{h}_\tau$ be smooth and monotonic between $\hat{y} - u$ and $\hat{y} + u$. Applying formula (8), we obtain that an individual with income $y > \hat{y} + u$ adjusts his behavior in response to the perturbation $\tilde{h}_\tau$ by the amount:

$$\delta y(T, h_\tau) = -\frac{\frac{y}{1-T'(y)} \zeta_{y,w}^c(y)}{1 + \frac{y}{1-T'(y)} \zeta_{y,w}^c(y) T''(y)} - \frac{\frac{1}{1-T'(y)} \eta_y(y)}{1 + \frac{y}{1-T'(y)} \zeta_{y,w}^c(y) T''(y)} (y - \hat{y}),$$

[16] If the goods x represent the labor and capital incomes y, z, we have $x_p = y_p$ for $p \leq S$ and $x_p = z_p$ for $p \geq S+1$. Suppose that y_p and z_p are taxed in period p at rate $\tau_{y,p} \equiv \tau_p$ and $\tau_{z,p} \equiv \tau_{p+S-1}$, respectively. Using the definitions of the elasticities $\zeta_{i,j,s}^{u,\boldsymbol{\theta}}$ introduced in the previous section, we obtain that $\zeta_{x_i, \hat{q}_p}^u$ is equal to $\zeta_{i,p,p}^{u,\boldsymbol{\theta}}$ if $p \in \{1, \ldots, S\}$ (so that $x_p = y_p$, which is taxed in period p), and to $\zeta_{i,p,p-S+1}^{u,\boldsymbol{\theta}}$ if $p \in \{S+1, \ldots, 2S\}$ (so that $x_p = z_{p-S+1}$, which is taxed in period $p - S + 1$). The net-of-tax rates $\hat{q}_p$ are equal to $\hat{q}_{p,p}$ if $p \in \{1, \ldots, S\}$, and to $\hat{q}_{p-S+1,p}$ if $p \in \{S+1, \ldots, 2S\}$. For ease of notation, in the sequel, whenever labor and capital incomes earned in period p are taxed only in period p, in our expressions for the elasticities we denote by $w_p \equiv \hat{q}_p$ the modified price (net-of-tax rate) on period-p labor income y_p, and by $r_p \equiv \hat{q}_p$ the modified price (net-of-tax rate) on period-p capital income z_p.

where $\zeta^c_{y,w}$ is the compensated elasticity of y with respect to the net-of-tax rate on labor income (see footnote 16), and η_y is income effect of y with respect to the lump-sum income.[17] Next, consider the perturbation h_R defined by $h_R(y) = 1$, so that for a given $\mu > 0$, the perturbed tax schedule is therefore given by $\tilde{T}(y) = T(y) + \mu$. Intuitively, this perturbation increases the total tax liability faced by individual y by the amount $\mu > 0$. Applying formula (8), we obtain that an individual with income y adjusts his behavior in response to the perturbation h_R by the amount:

$$\delta y(T, h_R) = -\frac{\frac{1}{1-T'(y)}\eta_y(y)}{1 + \frac{y}{1-T'(y)}\zeta^c_{y,w}(y)\, T''(y)}.$$

4 Welfare Effects of Tax Reforms and Optimal Tax System

Having defined the perturbations and described the effects that they induce on individual behavior, we now derive the revenue and welfare effects of these tax reforms, and characterize the optimal tax system. Specifically, we start from a baseline tax system, which can be suboptimal or optimal. We locally perturb this tax system with tax reform, as defined above. Our first result (Proposition 2) describes the revenue and welfare effects of these *local tax reforms*. Formally, we compute these local effects as the Gateaux differentials of the revenue and social welfare functionals. These give the sign and the magnitude of the potential gains that arise from reforming the current, potentially suboptimal, tax code. If the perturbation yields a strictly positive (revenue or welfare) effect, then the corresponding tax reform is *(revenue or welfare)-improving* and should be implemented. The second result that our theory yields (Proposition 3) is a characterization of the *globally optimal tax function*. Specifically, the baseline tax system is optimal only if there is no local tax reform that yields a strict improvement. Characterizations of the revenue-maximizing or welfare-maximizing tax systems are therefore obtained by setting the Gateaux differentials of the corresponding functionals equal to zero for any admissible perturbation. Note finally that a similar reasoning yields a characterization of the *optimum tax system within a restricted class* (e.g., of linear, separable, etc., tax systems), by restricting the analysis to the perturbations within the corresponding class.

We start by defining the social marginal welfare weights $g_s(\mathbf{x})$ that the planner assigns to agents with various income choices. These weights are defined such that the government is indifferent between having $g_s(\mathbf{x})$ more dollars of public funds in period s and giving one more dollar in period s to the taxpayers with choice vector $\mathbf{x}$. The smaller $g_s(\mathbf{x})$ is, the less the government values marginal consumption of individuals $\mathbf{x}$. We formally define the period-s social marginal welfare weight associated with an individual with the choice vector $\mathbf{x}$ (and type

[17] Using the definition of $\eta^{\boldsymbol{\theta}}_{i,s}$ above, we have $\eta_y \equiv \eta^{\boldsymbol{\theta}}_{1,1}$.

$\boldsymbol{\theta}$ such that $\mathbf{x}_{\boldsymbol{\theta}} = \mathbf{x}$) as[18]

$$g_s(\mathbf{x}) \equiv \frac{1-\alpha}{\lambda} \beta^{-(s-1)} \mathcal{G}'\left(\mathscr{U}_{\boldsymbol{\theta}}(\mathscr{T})\right) U_{c_s}(\boldsymbol{\theta}). \tag{11}$$

Intuitively, the envelope theorem implies that an additional dollar of revenue increases the individual's indirect utility by $d\mathscr{U} = U_{c_s}$, and social welfare increases by $d\left[\mathcal{G}(\mathscr{U})\right] = \mathcal{G}'(\mathscr{U})\, d\mathscr{U} = \mathcal{G}'(\mathscr{U})\, U_{c_t}$. We express this welfare gain in terms of the value of public funds (that is, in monetary units) by dividing this expression by the multiplier λ.

We now characterize the revenue and welfare effects of local tax reforms. Formally, we fix a period p and compute the Gateaux differential of the social welfare $\mathscr{W}(\cdot)$ and the tax revenue $\mathscr{R}(\cdot)$ following a perturbation of the baseline tax function T_p in the direction $h_p \in \mathcal{C}^2(\mathrm{X}, \mathbb{R})$. We show:

Proposition 2. *Suppose that Assumptions 1 and 2 are satisfied. The Gateaux differential of social welfare at the baseline tax system T_p in the direction h_p, is equal to*

$$\begin{aligned} \delta\mathscr{W}(T_p, h_p) = \int_{\mathrm{X}} \Big\{ & \left[\beta^{p-1}\left(1 - g_p(\mathbf{x})\right) - \boldsymbol{T}'(\mathbf{x})\, \boldsymbol{D}^{-1}(\mathbf{x})\, \boldsymbol{I}_p(\mathbf{x})\right] f_{\mathbf{x}}(\mathbf{x})\, h_p(\mathbf{x}) \\ & + \left[\boldsymbol{T}'(\mathbf{x})\, \boldsymbol{D}^{-1}(\mathbf{x})\, \boldsymbol{E}_p^c(\mathbf{x})\right] f_{\mathbf{x}}(\mathbf{x})\, \nabla h_p(\mathbf{x}) \Big\}\, d\mathbf{x}, \end{aligned} \tag{12}$$

where $\boldsymbol{D}(\mathbf{x}) \equiv \mathrm{i}_{2S} - \sum_{s=1}^{S} \boldsymbol{E}_s^c(\mathbf{x})\, D^2 T_s(\mathbf{x})$, and $\boldsymbol{T}'(\mathbf{x}) \equiv \sum_{s=1}^{S} \beta^{s-1} \left(\nabla T_s(\mathbf{x})\right)^{\mathrm{T}}$ is the discounted sum of the (transposed) gradients of the baseline tax functions $T_s(\cdot)$.[19] *This expression can be equivalently written as*

$$\begin{aligned} \delta\mathscr{W}(T_p, h_p) = \int_{\mathrm{X}} \Big\{ & \beta^{p-1}\left(1 - g_p(\mathbf{x})\right) f_{\mathbf{x}}(\mathbf{x}) - \boldsymbol{T}'(\mathbf{x})\, \boldsymbol{D}^{-1}(\mathbf{x})\, \boldsymbol{I}_p(\mathbf{x})\, f_{\mathbf{x}}(\mathbf{x}) \\ & - \nabla \cdot \left(\boldsymbol{T}'(\mathbf{x})\, \boldsymbol{D}^{-1}(\mathbf{x})\, \boldsymbol{E}_p^c(\mathbf{x})\, f_{\mathbf{x}}(\mathbf{x})\right) \Big\}\, h_p(\mathbf{x})\, d\mathbf{x}. \end{aligned} \tag{13}$$

The perturbation increases (resp., decreases), social welfare if $\delta\mathscr{W}(T_p, h_p) \geq 0$ (resp., ≤ 0). The Gateaux differential of the tax revenue functional, $\delta\mathscr{R}(T_p, h_p)$, is given by equations (12,13) in which $g_p(\mathbf{x})$ is replaced with 0. The perturbation increases (resp., decreases), tax revenue if $\delta\mathscr{R}(T_p, h_p) \geq 0$ (resp., ≤ 0).

Proof. See Appendix. □

Formulas (12) or (13) give the effects on social welfare of any local perturbation of the baseline tax system in the direction h_p. Equation (12) is obtained from the following formula,

[18] If Assumption 1 is not satisfied, then the social marginal welfare weight $g_s(\mathbf{x})$ should be defined as the average of the expression (11) over all individuals $\boldsymbol{\theta}$ who choose the same vector $\mathbf{x}$.

[19] For instance the first component of the $2S$-row vector $\boldsymbol{T}'(\mathbf{x})$ is the sum of the marginal tax rates on first-period labor income y_1 that the individual pays in every period of his life.

which follows from the definition (3) of social welfare and is formally derived in the Appendix:

$$\begin{aligned}\delta\mathscr{W}\left(T_p,h_p\right)=\lambda^{-1}\Bigg[\int_{\mathrm{X}}&\left\{\alpha\mathcal{V}'\left(\mathscr{R}\right)\beta^{p-1}-\left(1-\alpha\right)\mathcal{G}'\left(\mathscr{U}_{\mathbf{x}}\right)U_{c_p}\left(\mathbf{x}\right)\right\}h_p\left(\mathbf{x}\right)f_{\mathbf{x}}\left(\mathbf{x}\right)d\mathbf{x}\\&+\alpha\mathcal{V}'\left(\mathscr{R}\right)\int_{\mathrm{X}}\boldsymbol{T}'\left(\mathbf{x}\right)\delta\mathbf{x}\left(T_p,h_p\right)f_{\mathbf{x}}\left(\mathbf{x}\right)d\mathbf{x}\Bigg],\end{aligned}\tag{14}$$

where $\delta\mathbf{x}\left(T_p,h\right)$ denotes the Gateaux differential of the individual income functional derived in (8). The change in social welfare following the perturbation h_p comes from the effect on the individuals' utilities (the term that is multiplied by $\left(1-\alpha\right)\mathcal{G}'\left(\mathscr{U}_{\mathbf{x}}\right)$ in (14)) and the effect on the public goods through the change in tax revenue (the terms that are multiplied by $\alpha\mathcal{V}'\left(\mathscr{R}\right)$ in (14)). Equation (13) is then obtained by integrating (13) by parts, using our assumption that there is no mass of individuals at the boundary of the set X at the baseline tax system.

Intuitively, the first integral of equation (14) is the *mechanical effect*, net of the *welfare loss*, of the perturbation h_p, and the second integral is the *behavioral effect* of the tax reform. The mechanical effect captures the increase in government revenue due to the tax reform, assuming that individuals do not change their behavior in response to the perturbation. An individual with income $\mathbf{x}$ before the perturbation pays the additional tax liability $h_p\left(\mathbf{x}\right)$ in period p after the perturbation. By definition of the marginal social welfare weights (11), this induces a loss in social welfare, expressed in units of tax revenue, equal to $g_p\left(\mathbf{x}\right)h_p\left(\mathbf{x}\right)$. Summing over all individuals $\mathbf{x}\in\mathrm{X}$ using the density of incomes $f_{\mathbf{x}}\left(\mathbf{x}\right)$ yields the first integral in (14). Next, the behavioral effect of the perturbation captures the change in government revenue due to the behavioral response of individuals whose vector of labor and capital incomes $\mathbf{x}$ is affected by changes in the marginal tax rates or the virtual incomes. We derived in Proposition 1 the change $d\mathbf{x}=\delta\mathbf{x}\left(T_p,h_p\right)$ in each individual's income vector $\mathbf{x}$ induced by the perturbation h_p. This induces in turn a change in government's revenue in every period s, given by $d\left[T_s\left(\mathbf{x}\right)\right]=\left(\nabla T_s\left(\mathbf{x}\right)\right)^{\mathrm{T}}d\mathbf{x}$. The overall behavioral effect of the perturbation is thus equal to the second integral in (14). Finally, the effect on government revenue is identical to the effect on social welfare, except that we do not take into account the welfare loss of the perturbation described above. We call the perturbation h_p *budget-neutral* if $\delta\mathscr{R}\left(T_p,h_p\right)=0$.

Formula (13) (or equivalently (14)) allows to compute in a wide variety of settings the effects on social welfare of any local taxreform h_p of the baseline tax system, by simply substituting the values $h_p\left(\mathbf{x}\right)$ of the corresponding perturbation in the integral of (13). We analyze several examples of application of this result in Sections 5 and 6.

We can also use formula (13) to characterize the optimal tax system, or the optimum within a restricted class. Specifically, if the baseline tax system is optimal (possibly within a class) then there is no tax reform (within the corresponding class) that yields a positive effect on social welfare. Thus, by equating the Gateaux differential of social welfare for any such perturbation to zero, we obtain the optimum tax system. We obtain the following proposition:

Proposition 3. *Suppose that Assumptions 1 and 2 are satisfied. Then:*

- A necessary condition for the baseline tax function T_p to be optimal (resp., optimal within a class) is that, for any perturbation $h_p \in \mathcal{C}^2(\mathrm{X}, \mathbb{R})$ (resp., for any perturbation restricted within this class), we have

$$\delta \mathscr{W}(T_p, h_p) = 0. \tag{15}$$

- In particular, applying (15) to the class of separable linear perturbations,we obtain that the optimum separable linear tax system is characterized by, for all $p \in \{1, \ldots, 2S\}$:

$$\begin{aligned} 0 &= \int_{\mathrm{X}} \left\{ \beta^{p-1} (1 - g_p(\mathbf{x})) x_p + \boldsymbol{T}'(\mathbf{x}) \left[\boldsymbol{E}_p^c(\mathbf{x}) \right]_p - x_p \boldsymbol{T}'(\mathbf{x}) \boldsymbol{I}_p(\mathbf{x}) \right\} f_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x} \\ &= \int_{\mathrm{X}} \left\{ \beta^{p-1} (1 - g_p(\mathbf{x})) x_p + \boldsymbol{T}'(\mathbf{x}) \left[\boldsymbol{E}_p^u(\mathbf{x}) \right]_p \right\} f_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x}, \end{aligned} \tag{16}$$

where $\left[\boldsymbol{E}_p^c\right]_p$ is the p^{th} column of the matrix $\boldsymbol{E}_p^c$.

- In particular, applying (15) to the class of all admissible perturbations, we obtain that the baseline tax system is the full optimum if, for any compact volume $\mathrm{V} \subset \mathrm{X}$ with closed and piecewise smooth boundary $\mathrm{S} = \partial \mathrm{V}$, we have, for all $p \in \{1, \ldots, 2S\}$:

$$\begin{aligned} 0 = &\int_{\mathrm{V}} \beta^{p-1} (1 - g_p(\mathbf{x})) f_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x} - \int_{\mathrm{V}} \left(\boldsymbol{T}'(\mathbf{x}) \boldsymbol{D}^{-1}(\mathbf{x}) \boldsymbol{I}_p(\mathbf{x}) \right) f_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x} \\ &+ \int_{\partial \mathrm{V}} \left(\boldsymbol{T}'(\mathbf{x}) \boldsymbol{D}^{-1}(\mathbf{x}) \boldsymbol{E}_p^c(\mathbf{x}) \cdot \overrightarrow{\mathbf{n}}(\mathbf{x}) \right) f_{\mathbf{x}}(\mathbf{x})\, d\mathrm{S}(\mathbf{x}), \end{aligned} \tag{17}$$

where $\overrightarrow{\mathbf{n}}(\mathbf{x})$ is the inward-pointing unit normal vector of the closed surface S at point $\mathbf{x}$.

Proof. See Appendix. □

Proposition 3 has three parts. First, equation (15) formalizes the intuition that the baseline tax system is optimal (resp., optimal within a class) if no tax reform (resp., no tax reform that leaves the tax system within the corresponding class) induces a non-zero welfare gain. It is a standard first-order condition which should be satisfied by any perturbation (possibly within a restricted class), and thus provides a general characterization of the optimality of any tax system. The second and third parts of Proposition 3 show two examples of application. Equation (16) characterizes the optimal separable linear tax system, that is the set $\{\tau_{x_s}\}_{1 \leq s \leq 2S}$ of constant marginal tax rates on each income x_s. Note that the optimal linear tax system is such that the total mechanical (net of the welfare loss) and behavioral (elasticitiy and income) effects, averaged over the *whole* population of individuals $\mathbf{x} \in \mathrm{X}$, must sum to zero. This is because in a linear tax system, all the individuals face the same marginal tax rate, so that the feasible tax reforms increase the tax rates by the same amount for *every* individual in order to leave the perturbed tax system within this class. Moreover, the sum of the individual compensated elasticity effect and income effect in the first line of (16) yields the uncompensated elasticity term in the second line, from the Slutsky equation.

Equation (17) characterizes the fully optimal (in particular, non-linear and non-separable) tax system. We obtain this expression by imposing that every perturbation h_p yields a zero welfare effect, so that the integrand in equation (13) must be equal to zero pointwise. Integrating the resulting equation over any volume $\mathrm{V} \subset \mathrm{X}$ with closed boundary $\mathrm{S} = \partial \mathrm{V}$ must therefore have a zero effect. We then obtain formula (17) as a consequence of the divergence theorem, which separates the total behavioral effect of the tax reform into its components in the interior and on the surface of the volume V. To understand the intuition underlying this formula, suppose that the government wants to raise revenue by increasing uniformly and in a lump-sum way the tax liability of individuals with income in the region $\mathbf{x} \in \mathrm{V}$. This mechanically increases the government's revenue, since all the individuals in the region V now pay higher taxes; summing the individual mechanical effects (net of the welfare losses) over the region V yields the first integral in equation (17). Moreover, these individuals respond to the lump-sum increase in their tax liability by adjusting their incomes, as captured by the vector of income effect parameters $\boldsymbol{I}_p$; summing these behavioral effects over all the individuals in the region V yields the second integral in (17). Finally, the government can only raise the lump-sum tax liability in the region V by increasing the marginal tax rates of the individuals located on the boundary of V, that is those with income $\mathbf{x}$ on the surface $\mathrm{S} = \partial \mathrm{V}$. These individuals respond to this higher distortion by adjusting their incomes, as captured by the matrix of compensated elasticities $\boldsymbol{E}_p^c$; summing these behavioral effects over all the individuals on the boundary $\mathrm{S} = \partial \mathrm{V}$ yields the third integral in (17). Formula (17) thus shows that the optimal tax system is such that, for any region V of the space $\mathrm{X} \subset \mathbb{R}_+^S \times \mathbb{R}^S$, the elasticity effect induced by the additional distortion on the boundary $\partial \mathrm{V}$ exactly compensates the mechanical and the income effects due to the lump-sum tax increase inside the region V.

To show an example of application, consider the static Mirrlees model with a single income dimension $y \geq 0$, and apply formula (17) to the volume $\mathrm{V} = [\hat{y}, \infty)$, for some income level $\hat{y}$.[20] The boundary of V is the singleton $\partial \mathrm{V} = \{\hat{y}\}$; its inward pointing unit normal $\overrightarrow{\mathbf{n}}(\mathbf{x})$ is the real number 1. We obtain

$$\begin{aligned} 0 = & \int_{\hat{y}}^{\infty} (1 - g(y)) f_y(y)\, dy - \int_{\hat{y}}^{\infty} T'(y) \frac{1}{1 + \frac{y}{1-T'(y)} \zeta_{y,w}^c(y) T''(y)} \frac{\eta_y(y)}{1 - T'(y)} f_y(y)\, dy \\ & - \left[T'(\hat{y}) \frac{1}{1 + \frac{\hat{y}}{1-T'(\hat{y})} \zeta_{y,w}^c(\hat{y}) T''(\hat{y})} \frac{\hat{y} \zeta_{y,w}^c(\hat{y})}{1 - T'(\hat{y})} f_y(\hat{y}) \right]. \end{aligned} \tag{18}$$

This equation is the analogue of (17) for the static model, derived in Saez (2001). In particular, the third term of (18) (in the square brackets) is the analogue of the third term in (17), i.e., the integral over the boundary $\partial \mathrm{V}$. Intuitively, in order to raise the lump-sum tax liability of individuals with income $y \in [\hat{y}, \infty)$ (the region V, which generates a mechanical effect and an

[20] Rigorously, to apply formula (17), we need to work with a compact volume $[\hat{y}, \hat{y}']$, with an additional boundary $\{y = \hat{y}'\}$, on which the inward pointing normal is the real number -1, and let $\hat{y}' \to \infty$, using $\lim_{y \to \infty} f_y(y) = 0$. The matrix $\boldsymbol{D}(\mathbf{x})$ is equal to the real number $1 + \frac{y}{1-T'(y)} \zeta_{y,w}^c(\hat{y}) T''(y)$ in the static setting, the vector $\boldsymbol{T}'(\mathbf{x})$ is $T'(y)$, the matrix $\boldsymbol{E}_p^c$ is $-\frac{y}{1-T'(y)} \zeta_{y,w}^c(y)$ and the vector $\boldsymbol{I}_p$ is $\frac{1}{1-T'(y)} \eta_y(y)$.

income effect given by the first two terms of (18)), the government must increase the marginal tax rate at the income level $y = \hat{y}$ (the surface $\partial \mathrm{V}$, which generates an elasticity effect given by the third term of formula (18)). We discuss the economic intuition further in more detail in Section 5.3.

Equations (16) and (17) highlight the source of the gains that arise from using more sophisticated tax systems. In the case of the optimal separable linear tax system (16), the mechanical and behavioral effects of any feasible perturbation must cancel out *on average* over the whole population $\mathbf{x} \in \mathrm{X}$. On the other hand, in the case of the fully optimal tax system, these opposing forces must cancel out *pointwise*, that is over every region $\mathbf{x} \in \mathrm{V}$. Using more sophisticated tax instruments allows the government to "fine-tune" optimally the distribution of distortions within the population, whereas a linear tax system is constrained to imposing the same tax rate on every individual, and hence to balance the increase in tax revenue against a measure of the average distortion in the economy. The unrestricted government can thus choose appropriately the volume V so that the distortions induced by the higher marginal tax rates on the boundary $\partial \mathrm{V}$ are small relative to the benefits of higher lump-sum taxes in the interior of V, because either the fraction of individuals $f_{\mathbf{x}} d\mathrm{S}$ or the behavioral responses to distortions $\boldsymbol{D}^{-1} \boldsymbol{E}_p^c$ on the boundary $\partial \mathrm{V}$ are relatively small. Therefore, non-linear tax instruments allow the government to disentangle the compensated elasticity from the income effect, and target these two competing forces to different segments of the population. On the other hand, we saw in the case of the linear tax system that every individual must face both the elasticity and the income effect, leading to the uncompensated elasticity term in the second line of equation (16). We discuss this general principle in greater detail in Sections 5 and 6 below.

Equating the integrand of (13) to zero at each point $\mathbf{x}$ yields a partial differential equation system which, along with the individual's first-order conditions (9) characterizes the optimal tax system in terms of the endogenous distribution $f_{\mathbf{x}}$ of incomes $\mathbf{x} \in \mathrm{X}$. We can change variables to rewrite this PDE using the exogenous density $f_{\boldsymbol{\theta}}$ of types $\boldsymbol{\theta} \in \Theta$ instead.[21] Assume that individuals have $2S$ dimensions of characteristics, so that their vectors of types and incomes have the same dimension. We show in the Appendix that the optimal tax system is the solution to the partial differential equation:

$$\begin{aligned} 0 = & \left(1 - g_p\left(\mathbf{x}\left(\boldsymbol{\theta}\right)\right)\right) \frac{f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right)}{\det\left(J_{\mathbf{x}}\left(\boldsymbol{\theta}\right)\right)} + \boldsymbol{T}'\left(\mathbf{x}\left(\boldsymbol{\theta}\right)\right) \frac{J_{\mathbf{x}}\left(\boldsymbol{\theta}\right)}{\det\left(J_{\mathbf{x}}\left(\boldsymbol{\theta}\right)\right)} \left[J_{\mathbf{F}}^{-1}\left(\boldsymbol{\theta}\right) J_{\mathbf{F}}\left(T_p\right)\right] f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right) \\ & - \sum_{j=1}^{2S} \sum_{i=1}^{2S} \left[\left(J'_{\mathbf{x}}\left(\boldsymbol{\theta}\right)\right)^{-1}\right]_{i,j} \frac{\partial}{\partial \theta_i} \left\{\boldsymbol{T}'\left(\mathbf{x}\left(\boldsymbol{\theta}\right)\right) \frac{J_{\mathbf{x}}\left(\boldsymbol{\theta}\right)}{\det\left(J_{\mathbf{x}}\left(\boldsymbol{\theta}\right)\right)} \left[J_{\mathbf{F}}^{-1}\left(\boldsymbol{\theta}\right) J_{\mathbf{F}}\left(\boldsymbol{\tau}_p\right)\right] f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right)\right\}_j , \end{aligned} \tag{19}$$

where $J_{\mathbf{x}}\left(\boldsymbol{\theta}\right) = \left[\partial x_{\boldsymbol{\theta},i} / \partial \theta_j\right]_{1 \leq i,j \leq 2S}$ is the Jacobian matrix of the income function $\mathbf{x}\left(\boldsymbol{\theta}\right)$, $\det\left(J_{\mathbf{x}}\left(\boldsymbol{\theta}\right)\right)$

[21] Changing variables from $\mathbf{x}$ to $\boldsymbol{\theta}$ to characterize the fully optimal tax system is useful because the resulting partial differential equation does not feature the deformation matrix $\boldsymbol{D}\left(\mathbf{x}\right)$, so that we can solve directly for the marginal tax rates $\boldsymbol{T}'\left(\mathbf{x}\left(\boldsymbol{\theta}\right)\right)$. On the other hand, it is more useful to work with the distribution of incomes when deriving the welfare effects of local tax reforms of suboptimal tax systems, because it is observed given the current tax code.

is its determinant, and $J_{\mathbf{F}}(\boldsymbol{\tau}_p), J_{\mathbf{F}}(T_p), J_{\mathbf{F}}(\boldsymbol{\theta})$ are defined by $J_{\mathbf{F}}(\boldsymbol{\tau}_p) = \left[\partial F_i / \partial \tau_{x_j,p}\right]_{i,j}$, $J_{\mathbf{F}}(T_p) = \left[\partial F_i / \partial T_p\right]_i$, and $J_{\mathbf{F}}(\boldsymbol{\theta}) = \left[\partial F_i / \partial \theta_j\right]_{i,j}$, for the function F that represents the individual's first-order conditions as a non-linear system of $2S$ equations

$$F\left(\mathbf{x}_{\boldsymbol{\theta}}, \left\{\nabla T_s(\mathbf{x}_{\boldsymbol{\theta}})\right\}_{s=1}^{S}, \left\{T_s(\mathbf{x}_{\boldsymbol{\theta}})\right\}_{s=1}^{S}, \boldsymbol{\theta}\right) = 0. \tag{20}$$

In order to calculate these latter three matrices, we need to work with a specific model of heterogeneity, and write explicitly the system of first-order conditions in the form of (20). In the Appendix, we do so for a dynamic model in which the $2S$ sources of heterogeneity (i.e., the idiosyncratic vector $\boldsymbol{\theta}$) are the productivity of labor supply and the interest rate on the capital stock in each period. In particular, in the static Mirrlees model, we can easily compute the matrices $J_{\mathbf{F}}^{-1}(\boldsymbol{\theta})$, $J_{\mathbf{F}}(\boldsymbol{\tau}_p)$, $J_{\mathbf{F}}(T_p)$ by differentiating the individual's first order conditions, so that we obtain the following formula:

$$\begin{aligned} 0 = &\left(1 - g(\theta)\right) f_\theta(\theta) - \frac{T'(y_\theta)}{1 - T'(y_\theta)} \frac{\dot{y}_\theta}{y_\theta} \frac{\eta_y(y_\theta)}{1 + \zeta_{y,w}^{u}(y_\theta)} \theta f_\theta(\theta) \\ &+ \frac{d}{d\theta}\left\{\frac{T'(y_\theta)}{1 - T'(y_\theta)} \frac{\zeta_{y,w}^{c}(y_\theta)}{1 + \zeta_{y,w}^{u}(y_\theta)} \theta f_\theta(\theta)\right\}. \end{aligned} \tag{21}$$

Integrating this differential equation yields the characterization of the optimal marginal income tax rates $T'(y_\theta) / (1 - T'(y_\theta))$ derived by Mirrlees (1971). An advantage of writing the formula for the optimal tax schedule in the form (21) rather than in the original form of Mirrlees (1971) is that the explicit notation for the income and the substitution effects makes transparent the underlying economic effects that determine the optimal marginal taxes.

5 Applications to Optimal Taxation

In this section we discuss applications of our general analysis to optimal taxation. We first show how our results reproduce two canonical benchmarks in public finance: the optimal Ramsey tax formula of Diamond (1975), and the optimal non-linear income tax formula in a static economy due to Mirrlees (1971). We then apply our analysis to several environments to obtain novel insights about non-linear labor income taxation and capital income taxation in dynamic economies.

In Sections 5.1 to 5.4, we focus on *separable* tax systems. The tax function in period s depends on labor income y_s and capital income z_s as $T_{y,s}(y_s) + T_{z,s}(z_s)$. To simplify the notations, we let $\bar{x}_s \equiv \mathbb{E}[x_s]$ denote the average income $x_s \in \{y_s, z_s\}$ in period s in the economy. In some applications we take expectations conditional on vector $\mathbf{x}$ lying in a set V, in which case we denote $\bar{x}_s^{\mathrm{V}} \equiv \mathbb{E}[x_s | \mathbf{x} \in \mathrm{V}]$. With the exception of the last part of Section 6, in all our applications taxes on income earned in period s are assumed to be paid in period s. Hence we use $\zeta_{x_s,\hat{q}_p}^{c}, \zeta_{x_s,\hat{q}_p}^{u}$ to denote compensated and uncompensated elasticitities of income x_s to the net-of-tax rate on income x_p.[22] Similarly, $\eta_{x_s,p}$ denotes the effect of an increase in lump-sum income in period p on income x_s. Let $\bar{\zeta}^c_{x_s,\hat{q}_p}$, $\bar{\zeta}^u_{x_s,\hat{q}_p}$, $\bar{\eta}_{x_s,p}$ be the following weighted averages of the compensated price elasticity, uncompensated price elasticity and income effect parameter among all individuals, with additional superscript V notation if the elasticities are averages over $\mathbf{x} \in \mathrm{V}$:

$$\begin{aligned} \bar{\zeta}^{(\mathrm{V})}_{x_i,\hat{q}_p} &\equiv \int_{\mathbf{x}\in\mathrm{V}} \frac{x_i}{\bar{x}_p^{\mathrm{V}}} \zeta_{x_i,\hat{q}_p}(\mathbf{x})\, f_{\mathbf{x}}(\mathbf{x}\,|\mathbf{x}\in\mathrm{V})\, d\mathbf{x}, \\ \bar{\eta}^{(\mathrm{V})}_{x_s,p} &\equiv \int_{\mathbf{x}\in\mathrm{V}} \frac{\hat{q}_s}{1-\tau_{x_s,p}} \eta_{x_s,p}(\mathbf{x})\, f_{\mathbf{x}}(\mathbf{x}\,|\mathbf{x}\in\mathrm{V})\, d\mathbf{x}. \end{aligned} \tag{22}$$

As explained in footnote 16, in the notations for the elasticities we use the subscript w_p (resp, r_p) for the net-of-tax rate $\hat{q}_p$ on labor income y_p (resp., capital income z_{p-S+1}).

5.1 Optimal Commodity Taxation

As a first application of our theory, consider the analysis of Ramsey (1927) and Diamond (1975), who restrict the tax system to be separable and *linear* in each income, so that in each period s a consumer pays a proportional tax $T_s(\mathbf{x}) = \tau_s x_s + \tau_{S+s} x_{S+s}$ on the s-th and $S+s$-th argument of vector $\mathbf{x} = [y_1, .., y_S, z_2, ..., z_{S+1}]$. Applying formula (16), we obtain

$$\sum_{s=1}^{2S} \beta^{s-p} \frac{\tau_s}{1-\tau_p} \bar{\zeta}^u_{s,p} = 1 - \mathbb{E}\left[\frac{x_p}{\bar{x}_p} g_p\right] \text{ for } p = \{1, ..., 2S\}. \tag{23}$$

Define the net social marginal utility of income for individual $\mathbf{x}$ as

$$b_p(\mathbf{x}) \equiv g_p(\mathbf{x}) - \sum_{s=1}^{2S} \beta^{s-p} \frac{\tau_s}{1-\tau_s} \eta_{s,p}(\mathbf{x}),$$

and let $\bar{b}_p \equiv \mathbb{E}[b_p(\mathbf{x})]$. Using the Slutsky equations and rearranging the previous equation, we obtain, for all p,

$$\sum_{s=1}^{2S} \frac{\tau_s}{1-\tau_p} \bar{\zeta}^c_{s,p} = 1 - \bar{b}_p - \bar{b}_p \cdot \mathrm{cov}\left(\frac{b_p}{\bar{b}_p}, \frac{x_p}{\bar{x}_p}\right) \text{ for } p = \{1, ..., 2S\}.$$

This is Ramsey's formula with several consumers, first obtained by Diamond (1975).

5.2 Optimal Age-Independent Capital Income Tax Rates

In this section we study capital income taxes that are restricted to be linear and constant over many periods. Such taxes arise naturally in several cases. First, many applications impose an a

[22] As explained in footnote 16, using the notations of Section 3.1 $\hat{q}_p$ denotes the modified price $\hat{q}_{p,p} \equiv 1 - \tau_{y_p,p}$ if $p \leq S$, and $\hat{q}_{p,p-S+1} \equiv 1 - \tau_{z_{p-S+1},p-S+1}$ if $p \geq S+1$. In equation (22), note that $\tau_{x_s,p}$ is equal to 0 unless $x_s \in \{y_p, z_p\}$.

priori assumption that capital taxes do not depend on the time period, e.g., Conesa, Kitao and Krueger (2009). Second, the optimal asymptotic capital tax rate in infinite horizon economies, analyzed by Chamley (1986) and Judd (1985), is equivalent to a tax that is constant across time after the economy reaches the steady-state.

For our analysis we abstract from income effects on labor supply and assume that preferences are of the form

$$U = \sum_{s=1}^{S} \beta^{s-1} u \left(c_s - v \left(\frac{y_s}{\theta_s} \right) \right). \tag{24}$$

When labor supply has no income effects, the form of labor income taxes (linear, separable non-linear, or even history-dependent) is irrelevant for our main result.

First, note that when capital taxes can be chosen freely in each period, the optimal tax rate in period p satisfies Ramsey's formula (23), where the cross-partial elasticities $\bar{\zeta}^u_{y_s,r_p}$ between labor income and the capital income tax rates are equal to zero. If instead we exogenously restrict the tax rates to be constant across time, we can apply the general formulas (12) and (15) to obtain the following characterization of the optimal *age-independent* capital income tax rate τ_z:

Proposition 4. *Suppose that the utility function has no income effects on labor supply as defined in (24). The optimal age-independent capital income tax rate τ_z is then given by:*

$$\frac{\tau_z}{1-\tau_z} = \left(1 - \sum_{p=2}^{S} \gamma_p \mathbb{E} \left[\frac{z_p}{\overline{z}_p} g_p \right] \right) \frac{1}{\sum_{p=2}^{S} \gamma_p \mathring{\zeta}^u_p}, \tag{25}$$

where the weights γ_p and the compounded uncompensated elasticity $\mathring{\zeta}^u_p$ are equal to

$$\gamma_p = \frac{\beta^{p-1} \overline{z}_p}{\sum_{s=2}^{S} \beta^{s-1} \overline{z}_s}, \quad \textit{and} \quad \mathring{\zeta}^u_p = \sum_{s=2}^{S} \beta^{s-p} \bar{\zeta}^u_{z_s,r_p}.$$

Proof. See Appendix. □

The weight γ_p is the ratio between the mechanical effect of a linear perturbation of the capital income tax rate in period p only (the tax revenue generated in period p is proportional to the average capital income $\overline{z}_p$ in the economy in period p), and the total mechanical effect of the age-independent perturbation (which raises revenue in every period $s \geq 2$). The compounded elasticity $\mathring{\zeta}^u_p$ measures the behavioral effect of the period-p capital income tax change on capital income in all periods. This is a shorthand for all the cross-partial elasticities that appear on the left hand side of (23). Since tax rates do not depend on age in this application, the relevant behavioral elasticity is the sum of all $\mathring{\zeta}^u_p$ weighted by the fraction of the capital stock affected by the perturbation in period p, γ_p.

To illustrate the compounding effect arising from age-independent taxation, assume for simplicity that the capital income distribution is age-independent, so that $\gamma_p = \beta^{p-1}/\sum_{s=2}^{S}\beta^{s-1}$ is simply equal to a (normalized) discount factor. We now compare this compounded elasticity with the behavioral response of capital income to a *one-period* change in the tax rate, that is, a perturbation of the capital income tax rate in only one period. We show that compounding the elasticities over a longer horizon can either increase or lower the elasticity that is relevant for the optimal tax rate, depending on the relative strenghts of the income and substitution effects. First, observe that the compounded *uncompensated* elasticity can be written, using the Slutsky equation, as the sum of the compounded *compensated* elasticity and the compounded *income effect* parameter:

$$\sum_{p=2}^{S}\sum_{s=2}^{S}\gamma_p\beta^{s-p}\bar{\zeta}^u_{z_s,r_p} = \sum_{p=2}^{S}\sum_{s=2}^{S}\gamma_p\beta^{s-p}\left[\bar{\zeta}^c_{z_s,r_p} + \bar{\eta}_{z_s,p}\right].$$

The compensated elasticies of capital income are always positive, while the income effect parameter $\bar{\eta}_{z_s,p}$ is negative for $s \leq p$ and positive otherwise. The size of the compounded uncompensated elasticity thus depends on whether the substitution effect dominates the net income effect of a tax change. The analysis is the most stark if we follow Judd (1985) and assume that capital is being held only by the agents who have no labor income, and whose utility is then $u(c) = c^{1-\sigma}/(1-\sigma)$. To simplify calculations, assume further that $S = \infty$. Computing directly the compensated elasticities $\bar{\zeta}^c_{z_s,r_p}$ and the income effect parameters $\bar{\eta}_{z_s,p}$ when after-tax interest rates are equal to β^{-1}, we can compare the compounded elasticity to the elasticity of one-time tax change in period two. For concreteness, we state the following proposition comparing the effect of an age-independent tax change to a tax change in period two.

Proposition 5. *Assume that all the assumptions of this section are satisfied. Then the elasticity of capital income with respect to a change in the capital income tax rate in period two only, satisfies*

$$\begin{aligned}
\sum_{p=2}^{\infty}\sum_{s=2}^{\infty}\beta^{s-1}\bar{\zeta}^u_{z_s,r_p} &\geq \sum_{s=2}^{\infty}\beta^{s-1}\bar{\zeta}^u_{z_s,r_2}, \quad \textit{if } \sigma \leq 1, \\
\sum_{p=2}^{\infty}\sum_{s=2}^{\infty}\beta^{s-1}\bar{\zeta}^u_{z_s,r_p} &< \sum_{s=2}^{\infty}\beta^{s-1}\bar{\zeta}^u_{z_s,r_2}, \quad \textit{if } \sigma \textit{ is sufficiently large.}
\end{aligned} \tag{26}$$

Proof. See Appendix. □

Proposition 5 shows that compounding the elasticities may either increase or decrease the effective behavioral effect of capital income depending on the value of the intertemporal elasticity of substitution σ. Note that with our preferences, we have $\bar{\eta}_{z_s,p} = -\sigma\bar{\zeta}^c_{z_s,r_p}$ if $s \leq p$. This result builds on the insights developed by Straub and Werning (2014). One way to see the connection with their work is to consider a perturbation in capital taxes after some period P and take the limit as $P \to \infty$. In this case the positive income effects become negligible and we obtain the result of Proposition 6 in Straub and Werning (2014):

$$\lim_{P\to\infty} \beta^{-(P-1)} \sum_{p=P}^{\infty} \sum_{s=2}^{\infty} \beta^{s-1} \bar{\zeta}^{u}_{z_s,r_p} = \begin{cases} \infty, & \text{if } \sigma < 1, \\ -\infty, & \text{if } \sigma > 1. \end{cases} \tag{27}$$

Straub and Werning (2014) then use this insight to provide an intuition for their results that the optimal capital tax rate converges to zero in the long run steady state only if $\sigma < 1$ and that they remain positive and may even converge to infinity for $\sigma \geq 1$. Proposition 5 shows that the mechanisms they emphasize continue to operate for age-independent taxes even in the short run.

5.3 Optimal Non-Linear Labor Income Taxation

As a third application of our theory, consider the static model of optimal income taxation analyzed by Mirrlees (1971). Suppose that $S = 1$, there is no capital income and the individual utility is given by $u(c, y/\theta)$. We derived in equation (18) the optimal non-linear labor income tax schedule $T(y)$. We can rewrite this formula as:

$$\begin{aligned} 0 =& \mathbb{E}_{y\geq\hat{y}}\left[1-g\right] - \mathbb{E}_{y\geq\hat{y}}\left[\frac{T'(y)}{1-T'(y)+y\zeta^{c}_{y,w}(y)\,T''(y)}\eta_y(y)\right] \\ &- \frac{T'(\hat{y})}{1-T'(\hat{y})+\hat{y}\zeta^{c}_{y,w}(\hat{y})\,T''(\hat{y})}\zeta^{c}_{y,w}(\hat{y})\,\frac{\hat{y}f(\hat{y})}{1-F(\hat{y})}, \end{aligned} \tag{28}$$

Equation (28) is the formula obtained by Saez (2001). It formalizes his heuristic arguments that the optimal marginal tax rate on labor income $\hat{y}$ is driven by three forces: (i) the compensated elasticity of labor income $\zeta^{c}_{y,w}$ and the hazard rate $H_y(\hat{y}) \equiv \frac{\hat{y}f(\hat{y})}{1-F(\hat{y})}$ of the labor income distribution, which measure the distortions induced by the marginal tax rate at the income level $\hat{y}$; (ii) the average income effect parameter η_y for incomes above $\hat{y}$, which measure the behavioral effects of increased average taxes on those incomes; and (iii) the value of redistributing income away from individuals above $\hat{y}$, captured by $\mathbb{E}_{y\geq\hat{y}}(1-g)$.

We now discuss the connection between the formula obtained by Diamond (1975) in the Ramsey setting and that obtained by Saez (2001) in the Mirrlees setting. Formula (16) implies that the optimal linear tax schedule in the static model satisfies

$$\begin{aligned} 0 =& \mathbb{E}_{y\geq 0}\left[1-\frac{y}{\bar{y}}g\right] - \frac{\tau_y}{1-\tau_y}\bar{\zeta}^{u}_{y,w} \\ =& \mathbb{E}_{y\geq 0}\left[1-\frac{y}{\bar{y}}g\right] - \frac{\tau_y}{1-\tau_y}\bar{\eta}_y - \frac{\tau_y}{1-\tau_y}\bar{\zeta}^{c}_{y,w}, \end{aligned} \tag{29}$$

where $\tau_y \equiv T'(y)$, and where the second line follows from the Slutsky equations $\zeta^{u}_{y,w}(y) = \zeta^{c}_{y,w}(y) + \eta_y(y)$ for all $y \geq 0$. (Recall that $\zeta^{c}_{y,w}(y) > 0$ and $\eta_y(y) < 0$, so that the substitution effect of taxes tends to decrease tax revenue, while the income effects tend to increase it.) Formula (29) closely resembles the full optimum (28), with one key difference. The linear optimum cannot do better than balancing the mechanical effect of the perturbation with an *average* measure $\bar{\zeta}^{u}_{y,w}$ of uncompensated elasticities $\zeta^{u}_{y,w}(y)$ over the entire population $y \geq 0$, while the non-linear optimum (28) is able to disentangle the competing income and substitution components $\zeta^{c}_{y,w}(y), \eta_y(y)$ of the individual elasticity, and to allocate both effects to different segments of the population (cf. the discussion in Section 4 following Proposition 3). Specifically, a planner that can use non-linear tax instruments is able to better fine-tune the distortions in the population, by imposing a higher marginal tax rate to the incomes $y = \hat{y}$ where there is either a small fraction of individuals relative to those who pay the additional lump-sum tax (the hazard rate $H_y(\hat{y})$ is small), or where the behavioral response $\zeta^{c}_{y,w}(\hat{y})$ (resp., $\eta_y(\hat{y})$) to the increase in the marginal tax rate (resp., the total liability) is small (resp., large). We make this general principle of taxation more precise in the context of tax reforms in Section 6, where we show the tight connection between the linear and non-linear tax reforms when the baseline tax system is linear.

Equation (28) finally allows us to obtain asymptotic tax rates. Suppose that $\mathbb{E}_{y\geq\hat{y}}g$, $\zeta^{c}_{y,w}(\hat{y})$, $\eta_y(\hat{y})$, and $H_y(\hat{y})$ converge to the respective limits $g^{(\infty)}$, $\zeta^{c,(\infty)}_{y,w}$, $\eta^{(\infty)}_y$, and $H^{(\infty)}_y$ as $\hat{y} \to \infty$, and suppose moreover that $\hat{y}T''(\hat{y}) \to 0$ as $\hat{y} \to \infty$. We then obtain the top marginal tax rates as:

$$\lim_{\hat{y}\to\infty} \frac{T'(\hat{y})}{1 - T'(\hat{y})} = \frac{1 - g^{(\infty)}}{\zeta^{c,(\infty)}_{y,w} H^{(\infty)}_y + \eta^{(\infty)}_y}. \tag{30}$$

This expression is derived formally by Saez (2001b).

5.4 Optimal Non-linear Labor and Capital Income Taxation

The previous sections showed that the benefit of increasing the sophistication of the tax instruments in a static model come from the ability to spread the distortions within the population. In this section we extend this insight to dynamic settings. Moreover, we use this example to illustrate how our techniques can be applied to problems for which it is hard to obtain analytical results using standard techniques.

We consider a simplified version of the economy considered by Conesa, Kitao and Krueger (2009). These authors study a tax system that is separable and non-linear in capital and labor income. Analyzing such taxes is difficult with either the traditional Ramsey approach (due to the non-linearity in labor taxes) or with the mechanism design techniques (due to the lack of explicit informational microfoundations for this tax schedule). For this reason Conesa, Kitao and Krueger (2009) additionally impose parametric restrictions on tax functions and numerically optimize over those parameters in a sophisticated computational model. We show how this problem can be handled using our approach.

Suppose that $S = 2$ and the utility function has no income effects, as defined in (24). We are interested in deriving properties of the optimal taxes that are *separable*, *age-dependent*, and *non-linear*, so that the tax system consists of a non-linear labor income tax schedule $T_{y,1}(y_1)$ in period one, and of separable non-linear labor and capital tax schedules $T_{y,2}(y_2)+T_{z,2}(z_2)$ in period two.

We start by applying our general formulas (12) and (15) to the tax schedule on *labor income* in period $s \in \{1,2\}$, restricting the tax system to be separable between the various incomes. We obtain that the optimal labor income tax rate in period s at the income level $\hat{y}_s$ is given by

$$\begin{aligned} 0 =& \mathbb{E}_{y_s \geq \hat{y}_s}[1-g_s] - \frac{T'_{y,s}(\hat{y}_s)}{1-T'_{y,s}(\hat{y}_s)-\hat{y}_s\zeta^c_{y_s,w_s}(\hat{y}_s)T''_{y,s}(\hat{y}_s)}\bar{\zeta}^{c,(y_s=\hat{y}_s)}_{y_s,w_s}\frac{\hat{y}_s f(\hat{y}_s)}{1-F(\hat{y}_s)} \\ &- \beta^{2-s}\mathbb{E}_{y_s \geq \hat{y}_s}\left[\frac{T'_{z,2}(z)}{1-T'_{z,2}(z)-z\zeta^c_{z_2,r_2}T''_{z,2}(z)}\hat{\eta}_{z_2,s}\right], \end{aligned} \tag{31}$$

where $\hat{\eta}_{z_2,s} = \eta_{z_2,2}$ if $s=2$, and $\hat{\eta}_{z_2,s} = \left(1-T'_{z,2}(z)\right)\eta_{z_2,1}$ if $s=1$. Since we assume that there are no income effects on labor supply, the first line of this expression is simply a dynamic version of (28). Note that only the own-price elasticities of labor income play a role: the cross-price elasticities of labor income and the compensated elasticities of capital income with respect to the labor income tax rate are equal to zero because there are no income effects and the baseline tax system is separable. However, the second line of (31) shows that in the dynamic environment, additional considerations play a role in the determination of the optimal labor income tax rates in either period. Higher labor taxes in period two increase incentives to save, captured by $\eta_{z_2,2}$, and hence affect revenue from capital income taxes. If the optimal marginal tax rate on capital income is positive, this increases government revenue and creates a force to increase the labor income taxes in period two, relative to the static model. The opposite effect holds for labor income taxes in period one.

Next, we apply formulas (12) and (15) to the tax schedule on capital income in period two. We obtain that the optimal capital income tax rate at the income level $\hat{z}$ is characterized by

$$\begin{aligned} 0 =& \mathbb{E}_{z \geq \hat{z}}[1-g_2] - \mathbb{E}_{z=\hat{z}}\left[\frac{T'_{z,2}(\hat{z})}{1-T'_{z,2}(\hat{z})-\hat{z}\zeta^c_{z_2,r_2}T''_{z,2}(\hat{z})}\zeta^c_{z_2,r_2}(y_1,y_2,\hat{z})\frac{\hat{z}f_{\mathbf{x}}(y_1,y_2,\hat{z})}{1-F_{z_2}(\hat{z})}\right] \\ &- \mathbb{E}_{z \geq \hat{z}}\left[\frac{T'_{z,2}(z)}{1-T'_{z,2}(z)-z\zeta^c_{z_2,r_2}T''_{z,2}(z)}\eta_{z_2,2}\right]. \end{aligned} \tag{32}$$

The expectation operator in the second term of equation (32) appears because the elasticity $\zeta^c_{z_2,r_2}$ may be different for agents with a given value of capital income $z_2 = \hat{z}$, if they have different labor incomes y_1 and y_2. If the utility function is CARA, then the elasticities in the integrals do not depend on labor income and (32) is then conceptually identical to (28), since $\mathbb{E}_{z=\hat{z}}\left[\frac{\hat{z}f_{\mathbf{x}}(y_1,y_2,\hat{z})}{1-F_{z_2}(\hat{z})}\right]$ is equal to the hazard rate $H_{z_2}(\hat{z}_2) = \frac{\hat{z}f_{z_2}(\hat{z})}{1-F_{z_2}(\hat{z})}$; the only differences are that the relevant elasticity and income distribution are those of capital income (there are no effects on labor income because of the functional form of the utility function and the separability of the tax system). Therefore, formula (32) illustrates that the same general mechanisms that determine optimal labor income taxation also determine optimal capital income taxation. As in the case of labor income taxes, the size and the shape of the capital income tax schedule are determined by the hazard rates of the capital income distribution, and by the income and substitution effects of capital income in response to changes in the capital income tax rates.

The asymptotic marginal tax rate on capital income is given by the analogue of (30),

$$\lim_{\hat{z}\to\infty} \frac{T'_{2,z}(\hat{z})}{1-T'_{2,z}(\hat{z})} = \frac{1-g_2^{(\infty)}}{\zeta_{z_2,r_2}^{c,(\infty)} H_{z_2}^{(\infty)} - \eta_{z_2,2}^{(\infty)}}.$$

It can be futher shown (see Appendix) that if mobility at the top of the capital income distribution converges to zero, the same formula continues to apply for the top marginal tax rates in arbitrary S period economies. If, in addition, the capital income tax schedule T_z is restricted to be age-independent, all the parameters are replaced with their compounded analogues along the lines of the analysis in Section 5.2.

5.5 Optimal Joint Taxation

We now apply our theory to the analysis of the optimal *non-separable*, *non-linear* tax system. We illustrate this approach in a simple static framework of optimal taxation of couples. We assume that the household maximizes the total surplus, i.e., the total consumption minus the sum of disutilities of labor. Both individuals choose their labor supply on the intensive margin.[23] The couple's preferences over consumption and labor income are given by

$$\max_{c_1,c_2,y_1,y_2} \; u\left(c_1+c_2-\frac{1}{1+1/\zeta}\left(\frac{y_1}{\theta_1}\right)^{1+1/\zeta}-\frac{1}{1+1/\zeta}\left(\frac{y_2}{\theta_2}\right)^{1+1/\zeta}\right),$$

and its budget constraint is

$$c_1+c_2=y_1+y_2-T\left(y_1,y_2\right).$$

In the Appendix, we show by applying formula (19) to this environment that the optimal tax system is characterized by the following partial differential equation: for all $\boldsymbol{\theta}=(\theta_1,\theta_2)\in\mathbb{R}_+^2$,

$$0=\left(1-g\left(\boldsymbol{\theta}\right)\right)f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right)+\frac{\zeta}{1+\zeta}\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\partial y_{-i}}{\partial\theta_{-j}}\frac{\partial}{\partial\theta_j}\left\{\frac{\frac{\tau_1}{1-\tau_i}\frac{\partial y_{\boldsymbol{\theta},1}}{\partial\theta_i}+\frac{\tau_2}{1-\tau_i}\frac{\partial y_{\boldsymbol{\theta},2}}{\partial\theta_i}}{\frac{\partial y_{\boldsymbol{\theta},1}}{\partial\theta_1}\frac{\partial y_{\boldsymbol{\theta},2}}{\partial\theta_2}-\frac{\partial y_{\boldsymbol{\theta},1}}{\partial\theta_2}\frac{\partial y_{\boldsymbol{\theta},2}}{\partial\theta_1}}\theta_i f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right)\right\}, \qquad (33)$$

where the components of the Jacobian matrix $\frac{\partial y_{\boldsymbol{\theta},i}}{\partial\theta_j}$ evaluated at the type $\boldsymbol{\theta}$ can be easily expressed explicitly as a function of the tax rates (see Appendix for details), so that (33) gives a complete characterization of the optimal tax rates given the exogenous distribution of types

[23] Saez, Kleven, and Kreiner (2009) characterize the optimal joint tax system in the case where the secondary earner chooses labor supply on the participation margin only.

(θ_1, θ_2). Formula (33) is useful because it allows us to reduce the problem of finding the optimal joint tax system in the economy as the solution to a PDE which can be computed numerically, without the need to solve for a complicated individual optimization problem. Note finally that this PDE generalizes to the two-dimensional environment the differential equation obtained in the static model of individual-based taxation of Mirrlees (1971), that is

$$0 = (1 - g(\theta)) f_\theta(\theta) + \frac{d}{d\theta}\left\{\frac{\tau}{1-\tau}\frac{\zeta}{1+\zeta}\theta f_\theta(\theta)\right\}$$

which can be easily solved analytically to obtain the optimal tax rates of Diamond (1998).

6 Applications to Tax Reforms

In this section we use the tools developed in Section 4 to analyze the welfare gains of (small) reforms of the existing tax system. Computing such welfare gains is considerably simpler than solving for the optimal taxes as in Section 5. Finding the optimum requires either solving a partial differential equation (19) using explicit assumptions on the form of the utility function or applying equations derived in Proposition 3 with some assumptions on the values of the relevant elasticities at the optimal system.[24] On the other hand, the welfare effects of reforming the existing system depend on the labor and capital income elasticties that can be readily estimated empirically under the current tax system. Once these elasticities are known, the welfare gains can be computed directly using Proposition 2 without solving differential equations.

In this section we illustrate an application of this approach by considering a simple version of a lifecycle model. We assume that individuals live for S periods and have Greenwood, Hercowitz and Huffman (1988) preferences

$$\frac{1}{1-\sigma}\sum_{s=1}^{S}\beta^{s-1}\left(c_s - \frac{1}{1+1/\zeta_s}l_s^{1+1/\zeta_s}\right)^{1-\sigma}.$$

Note that we allow the elasticity of labor supply ζ_s to depend on age. We assume that the baseline tax system does not depend on the age of the individual and is separable between labor and capital incomes within and across time periods. We assume that capital income is taxed at a constant rate τ_z, while labor income y is taxed non-lineary with a tax function $T(\cdot)$:

$$T_s(\mathbf{x}) = T(y_s) + \tau_z z_s, \text{ for all } s = 1, ..., S. \tag{34}$$

Individuals are heterogeneous in their initial capital stock k_0 and face the same after-tax interest rate equal to β^{-1}. The individuals are also heterogeneous in labor productivity θ_s (so

[24] For example, the "sufficient statistics" literature (see, e.g. Chetty 2009) takes the latter approach and often assumes that elasticities evaluated at the current tax system provide a good approximation to the elasticities at the optimal tax schedule.

that labor income is $y_s = \theta_s l_s$) and this productivity may change over time. We choose this specification both because it allows us to illustrate the main effects transparently, and because these assumptions on preferences and taxes are often used in applied work on optimal taxation. It is straightforward to extend our methods to other specifications of taxes and preferences.

6.1 Separable Income Taxation

Our baseline system is separable between capital and labor taxes and in this section we consider only tax reforms that preserve this separability. We start with capital taxation. Recall from our discussion in Section 5.1 that the Gateaux differential of social welfare $\delta\mathscr{W}(\mathscr{T}, h_L)$ with respect an age-independent linear perturbation of the capital tax rate, $h_L(z) \equiv z$, can be written as

$$\delta\mathscr{W}(\mathscr{T}, h_L) = \sum_{p=1}^{S} \gamma_p \left\{ \mathbb{E}\frac{z_p}{\bar{z}_p}\left[1 - g_p\right] - \frac{\tau_z}{1-\tau_z}\mathring{\zeta}_p^c - \frac{\tau_z}{1-\tau_z}\mathring{\eta}_p \right\}.$$

This derivative represents the monetary value of the welfare gain (or loss, if it is negative) of a small increase in the tax rate on capital. This formula can be used directly with empirical estimates of the population-average elasticities $\mathring{\zeta}_p^c$ and $\mathring{\eta}_p$ to compute the gains from changes in the tax rates.

We can extend this analysis to quantify the gains of introducing non-linear capital taxes.[25] Specifically we design perturbations that isolate the effect of increasing the capital income tax rate only at the income level $\hat{z}$. We choose the numbers $\hat{z} > 0$, $\hat{z}' > \hat{z}$ and define the period-p perturbation h_p as $h_p(z) = (z - \hat{z})$ on $[\hat{z}, \hat{z}']$, and $h_p(z) = (\hat{z}' - \hat{z})$ on $[\hat{z}', \infty)$ for all $p \in \{1, \ldots, S\}$. As described in Section 2 (details in the Appendix), we appropriately smooth out the kinks that this perturbation generates at the points $\hat{z}$ and $\hat{z}'$. We finally define a sequence $\left\{h_p^n\right\}_{n\in\mathbb{N}}$ of such perturbations, with $(\hat{z}' - \hat{z}) \to 0$. At each point in the sequence, we compute the Gateaux differential of social welfare in that direction and focus on the limit as $n \to \infty$ and hence $\left\|h_p^n\right\| \to 0$.

This pertubation increases the marginal taxes on capital income in a small neighborhood of $\hat{z}$ and the average taxes on all capital incomes above $\hat{z}$. We evaluate the welfare gains normalized by the fraction of agents affected by perturbation, which we define as:

$$\Gamma(\hat{z}) = \frac{\lim_{n\to\infty} \sum_{p=1}^{S} \left\|h_p^n\right\|^{-1} \delta\mathscr{W}\left(\mathscr{T}, h_p^n\right)}{\sum_{p=1}^{S} \beta^{p-1}\left(1 - F_{z,p}(\hat{z})\right)}.$$

Applying our general formula (12) yields the following welfare effect:

$$\Gamma(\hat{z}) = \sum_{p=1}^{S} \gamma_{p,\hat{z}} \left\{ \mathbb{E}_{z_p \geq \hat{z}}\left[1 - g_p\right] - \frac{\tau_z}{1-\tau_z}\mathring{\zeta}_p^c \frac{\hat{z} f_{z_p}(\hat{z})}{1 - F_{z_p}(\hat{z})} - \frac{\tau_z}{1-\tau_z}\mathring{\eta}_p \right\}, \tag{35}$$

[25] A number of authors proposed non-linear, often progressive taxation of capital, e.g., Farhi and Werning (2010).

where the weights, compounded compensated elasticity and compounded income effect are respectively defined by:

$$\gamma_{p,\hat{z}} = \frac{\beta^{p-1}\left(1-F_{p,z}\left(\hat{z}\right)\right)}{\sum_{s=1}^{S}\beta^{s-1}\left(1-F_{z,s}\left(\hat{z}\right)\right)}, \quad \mathring{\zeta}_p^c = \sum_{s=2}^{S}\beta^{s-p}\bar{\zeta}_{z_s,r_p}^{c,(z_p=\hat{z})}, \text{ and } \quad \mathring{\eta}_p = \sum_{s=2}^{S}\beta^{s-p}\bar{\eta}_{z_s,p}^{(z_p\geq\hat{z})}.$$

We make several observations about this expression. First, we show in the Appendix that

$$\delta\mathscr{W}\left(\mathscr{T},h_L\right) = \int_0^\infty \Gamma\left(\hat{z}\right)d\hat{z}. \tag{36}$$

This equation states that the sum of the welfare gains from increases in the marginal tax rates at every given level of capital income $\hat{z}$ is equal to the welfare gain from a linear increase in marginal taxes on all incomes. As long as the function Γ is not constant with income, i.e., increasing tax rates at $\hat{z}$ yields different gains than doing so at $\hat{z}'$, non-linear capital taxes lead to higher welfare. Welfare gains are achieved by lowering tax rates at those levels of capital income for which $\Gamma\left(\hat{z}\right)$ is negative and increasing them at levels of capital incomme for which $\Gamma\left(\hat{z}\right)$ is positive. Bigger changes in tax rates are typically desirable for those $\hat{z}$ that have the largest values of $\left|\Gamma\left(\hat{z}\right)\right|$.

We now explore which factors affect the variablility of $\Gamma\left(\cdot\right)$ in $\hat{z}$. There are three terms in the expression for Γ that capture (i) the welfare gains and losses from redistribution, $\mathbb{E}_{z_p\geq\hat{z}}\left[1-g_p\right]$; (ii) the behavioral effects due to higher marginal tax rates, $\mathring{\zeta}_p^c\frac{\hat{z}f_{z_p}(\hat{z})}{1-F_{z_p}(\hat{z})}$; and (iii) the behavioral effects due to higher average tax rates, $\mathring{\eta}_p$. The behavioral effects can be measured empirically or deduced theoretically from the assumed functional form of the utility function. In general, if the government is redistributive, $\mathbb{E}_{z_p\geq\hat{z}}\left[1-g_p\right]$ is increasing in $\hat{z}$. If at the top of the wealth distribution individuals do not earn any labor income, then the elasticity $\mathring{\zeta}_p^c$ is constant and equal to σ^{-1},[26] and the functional form for the utility function implies that $\mathring{\eta}_p$ is independent of $\hat{z}$. The last component in equation (35) that depends on $\hat{z}$ is the hazard rate of the capital income (or wealth) distribution $H_{z,p}\left(\hat{z}\right) \equiv \frac{\hat{z}f_{z_p}(\hat{z})}{1-F_{z_p}(\hat{z})}$. If this hazard rate has a weakly decreasing tail for all high enough $\hat{z}$, then we obtain that Γ is increasing in $\hat{z}$. In this case the same increase in the marginal tax rate generates larger welfare benefits if applied to higher levels of income. This suggests that progressive capital taxation is desirable for high levels of capital income.

The conclusion about the desirability of progressive capital tax reform is sensitive to assumptions about the hazard rate of the distribution of capital income F_z and the behavior of

[26]More generally, with labor income we can show that the elasticity of capital depends on $\hat{z}$ through:

$$\bar{\zeta}_{z_s,r_p}^{c,(z_p=\hat{z})} \propto \frac{1}{\hat{z}}\sigma^{-1}\mathbb{E}\left[c_p - \frac{l_p^{1+1/\zeta}}{1+1/\zeta}\left|z_p=\hat{z}\right.\right].$$

If labor income is equal to zero, then assuming that the average capital income in period $(p+1)$ among individuals with period-p capital income $\hat{z}$ is exactly $\hat{z}$, i.e., $\mathbb{E}\left[z_{p+1}\left|z_p=\hat{z}\right.\right]=\hat{z}$, it is easy to show (writing $c_p = Rk_{p-1}-k_p$) that $\bar{\zeta}_{z_s,r_p}^{c,\left(z_p=\hat{z}\right)}$ is a constant independent of $\hat{z}$ and $\mathring{\zeta}_p^c = \sigma^{-1}$.

the elasticity of capital income $\mathring{\zeta}_p^c$ as a function of $\hat{z}$. For example, if we assume that F_z is Pareto-log-normal rather than Pareto after a given finite threshold, then the hazard rate $H_{p,z}$ is increasing at the tail and for a wide range of parameters it is possible to show that $\mathring{\zeta}_p^c H_{z,p}(\hat{z})$ is increasing. In this case whether Γ increases or decreases in the right tail depends on the relative strenths of the redistributive versus behavioral effects. The behavioral effect, which favors regressive reforms in this case, dominates if the government is very redistributive and assigns low weights to individuals with a lot of wealth. For example, if the government is Rawlsian or, more generally, assigns Pareto weights 0 to all agents above a certain wealth threshold, then $\mathbb{E}_{z_p \geq \hat{z}}[1 - g_p] = 1$ for all $\hat{z}$ sufficiently high. In this case Γ is a decreasing function in the right tail.[27] To illustrate the quantitative gains from higher marginal taxes for high income, assume that initial capital distribution has a Pareto tail with coefficient 1.5 and that capital income tax rates in the baseline system are 50%.[28] Then in the limit as $\hat{z} \to \infty$ the welfare gains from increasing marginal rates are equal to $1 - 1.5\sigma^{-1}$. Therefore they are positive if the elasticity of the intertemporal substitution is less than 2/3.

Analogous arguments apply to reforms of labor income taxation. Welfare gains from increasing marginal taxes at income level $\hat{y}$ are given by

$$\Gamma(\hat{y}) = \sum_{p=1}^{S} \Gamma_p(\hat{y}) \equiv \sum_{p=1}^{S} \gamma_{p,\hat{y}} \left\{ \mathbb{E}_{y_p \geq \hat{y}}[1 - g_p] - \frac{T'(\hat{y})\zeta_p}{1 - T'(\hat{y}) + \hat{y}\zeta_p T''(\hat{y})} \frac{\hat{y} f_{y_p}(\hat{y})}{1 - F_{y_p}(\hat{y})} - \frac{\tau_z}{1 - \tau_z} \mathring{\eta}_p \right\}, \tag{37}$$

which shows the same three forces that we discussed in capital income taxation. As in the case of capital taxation, variability of Γ in $\hat{y}$ determines the welfare gains from non-linear tax reform. A useful benchark to consider is the one in which the baseline labor income taxes are linear, individual productivity is constant, and the government is highly redistributive, e.g. Ralwsian. In this case, the terms $\frac{T'(\hat{y})\zeta_p}{1-T'(\hat{y})+\hat{y}\zeta_p T''(\hat{y})}$ and $\mathring{\eta}_p$ are constant and $\mathbb{E}_{y_p \geq \hat{y}}[1 - g_p]$ is equal to 1 for almost all $\hat{y}$. In this case the shape of the gains Γ is determined by the shape of $H_y(\hat{y}) \equiv \frac{\hat{y} f_y(\hat{y})}{1-F_y(\hat{y})}$. If the hazard rate of labor income is inversely U-shaped, as documented for example by Saez (2001) for the U.S., then the benefits from increasing marginal tax rates are U-shaped.

We can use equation (37) to illustrate the sources of gains from switching from age-independent to age-dependent taxation. In particular, $\Gamma_p(\hat{y})$ measures the welfare gains from changing tax rates only for individuals of age p who earn income $\hat{y}$. The total gain $\Gamma(\hat{y})$ is equal to the sum of the age-dependent gains, $\Gamma_p(\hat{y})$. The more Γ_p varies in p, the larger the gain are from age-dependent labor taxation. We can thus use expressioon (37) to illustrate some recent arguments in favor of age-dependent taxation. Kremer (1999) argued for age-dependent labor taxation on

[27] We are not aware of any empirical work that systematically documents empirical properties of the hazard rate $H_{z,p}(\hat{z})$ of capital income. Saez (2001) used the IRS tax return data to study the hazard rate for wage income and showed that that it exhibits an inverse U-shaped pattern.

[28] The numbers are chosen to capture some stylized facts about distribution of income and tax rates in the U.S. Nirei and Souma (2007) estimate the Pareto tail of wealth distribution in the U.S. to be 1.5, while Saez estimates the Pareto tail of labor income (wages) to be 2 around 2. Prante and John (2013) argue that top effective marginal tax rates in the U.S. for both labor and interest income are about 50%.

the ground that the labor supply elasticity ζ_p varies systematically with age p, while Weinzierl (2011) explored variation in the age-dependent distribution of income, that can be summarized by $H_{y,p}(\hat{y}) \equiv \frac{\hat{y} f_{y_p}(\hat{y})}{1-F_{y_p}(\hat{y})}$. The quantitative magnitude of this variation can be calculated directly using (37). This formula also shows that, as long as the capital income tax rate is not zero, the age-dependent income effect of savings $\mathring{\eta}_p$ introduces an additional source of variability in welfare gains, as discussed in Section 5.4. In our economy it is easy to show that

$$\frac{1}{1-\tau_z}\mathring{\eta}_p = \beta r \left(\frac{1}{1-\beta} - \frac{\beta^S}{1-\beta^S} S - p \right).$$

To give a sense of the magnitude of these numbers, we suppose that $S = 40$, $\beta = 0.97$, and $\tau_z = 50\%$ and calculate that the behavioral effects on savings from higher labor income taxes at age p increase revenue by:

$$-\frac{\tau_z}{1-\tau_z}\mathring{\eta}_p = \begin{cases} -0.47 & \text{if } p = 1 \text{ (age 21)}, \\ 0.10 & \text{if } p = 20 \text{ (age 40)}, \\ 0.70 & \text{if } p = 40 \text{ (age 60)}. \end{cases}$$

This indicates that, all else being equal, positive taxation of savings favors higher labor taxes later in life. Note that these numbers are substantial, reaching an additional benefit of ¢73 per dollar increase in the statutory tax liability.

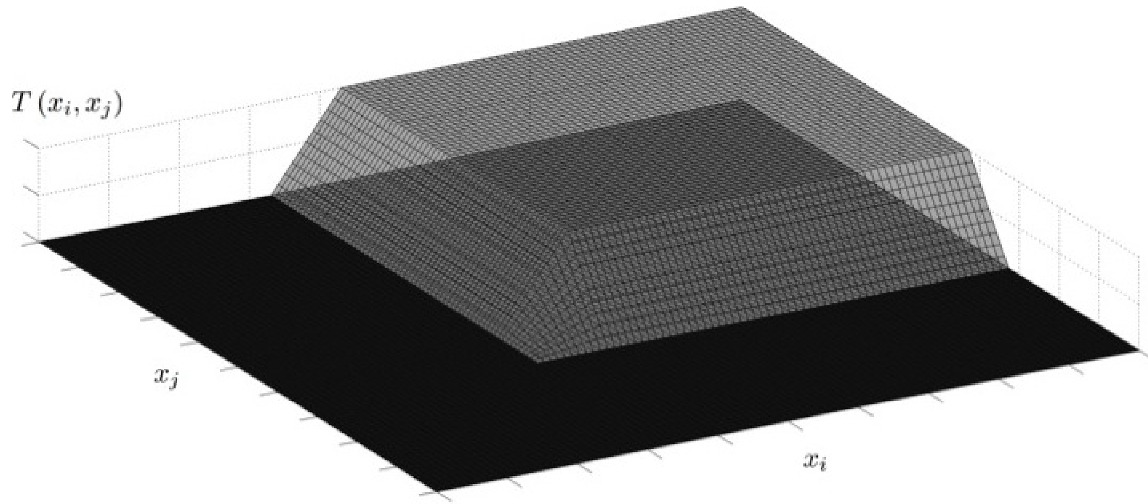
6.2 Joint Income Taxation

The baseline tax system (34) is initially separable between incomes, both across periods (there is no history-dependence) and within periods (labor and capital incomes are not jointly taxed). In this section, we characterize the welfare gains from introducing joint taxation. That is, we allow taxes on income x_i to depend not only on the level of x_i, but also on the levels of other incomes x_j. Such taxes arise in several different contexts. In the U.S., many social insurance programs and the Social Security system condition their payments both on current labor earnings and on the history of past earnings. Some programs are also often asset-tested, i.e., individuals are eligible to participate if their labor earnings are low and their assets are below a certain treshold. Finally, the individual tax bill depends jointly on income from labor and capital.

The non-separable tax reforms that we consider consist of increasing the marginal tax rate on income x_i at the level $\hat{x}_i$ (hence the average tax rates increase on $x_i \geq \hat{x}_i$) *conditional* on earning more than the threshold $\hat{x}_j$ of income x_j, i.e., $x_j \geq \hat{x}_j$. Since we consider perturbations that leave the tax function continuous, this reform also raises marginal tax rates on income x_j at level $\hat{x}_j$, conditional on $x_i \geq \hat{x}_i$. This joint pertubation is shown in Figure 2, where the dark (resp., light) surface represents the baseline (resp., perturbed) tax function.

We showed in the previous sections that the welfare effects of *separable* perturbations are

Figure 2: Joint Perturbations



determined by the fraction of individuals above the base of the perturbation relative to the fraction at the base, summarized by the hazard rate H_{x_i} of the distribution of income x_i. The generalization of the hazard rate to two dimensions of income (x_i, x_j) is captured by

$$H_{x_i,x_j}\left(\hat{x}_i \,|\hat{x}_j\right) \equiv \frac{\hat{x}_i \int_{\hat{x}_j}^{\infty} f_{x_i,x_j}\left(\hat{x}_i, x_j\right) dx_j}{\int_{\hat{x}_i}^{\infty} \int_{\hat{x}_j}^{\infty} f_{x_i,x_j}\left(x_i, x_j\right) dx_i dx_j}.$$

The denominator is the fraction of agents who face an increase in their average tax liability. The numerator is the fraction of agents who face an increase in their marginal tax rate on income x_i, scaled by the income threshold $\hat{x}_i$.

We first consider introducing joint taxation of labor income across periods, i.e., history-dependence. That is, we increase the marginal tax rate on labor income $\hat{y}_p$ in period p conditional on $y_{p-1} \geq \hat{y}_{p-1}$. The welfare gain of this tax reform is given by

$$\begin{aligned}
\Gamma_{\hat{y}_p,\hat{y}_{p-1}} =& \mathbb{E}_{\substack{y_{p-1}\geq \hat{y}_{p-1} \\ y_p \geq \hat{y}_p}}\left[1-g_p\right] - \frac{\tau_z}{1-\tau_z}\mathring{\eta}_p \\
&- \frac{T'\left(\hat{y}_p\right)\zeta_p}{1-T'\left(\hat{y}_p\right)+\hat{y}_p\zeta T''\left(\hat{y}_p\right)} H_{y_{p-1},y_p}\left(\hat{y}_p \,|\hat{y}_{p-1}\right) \\
&- \frac{T'\left(\hat{y}_{p-1}\right)\zeta_{p-1}}{1-T'\left(\hat{y}_{p-1}\right)+\hat{y}_{p-1}\beta\zeta_{p-1}T''\left(\hat{y}_{p-1}\right)} H_{y_{p-1},y_p}\left(\hat{y}_{p-1} \,|\hat{y}_p\right).
\end{aligned} \tag{38}$$

The first three terms of expression (38) are the analogue of the period-p age-dependent perturbation of the labor income tax rates discussed in Section 6.1, the main difference being that now the region over which the individual effects of the perturbation are summed is further restricted to the households earning more than $\hat{y}_{p-1}$ in period $(p-1)$. The last term is a novel term that appears because this perturbation distorts the labor supply decisions around the $\hat{y}_{p-1}$ treshold in period $(p-1)$.

The benefits of the joint pertubation come from two sources. First, by conditioning redistribution on past income, the government can better target its redistributive effort, as summarized in the term $\mathbb{E}_{y_{p-1}\geq\hat{y}_{p-1},y_p\geq\hat{y}_p}\left[1-g_p\right]$. Conditional on a given level of earnings y_p in period p, society generally values differently the welfare of households who have a different history of labor earnings in the previous periods. History-dependence in taxation allows the government to tailor taxes to those social preferences. Second, conditioning taxes on past earnings allows the government to raise more tax revenue with less distortions.

To illustrate the latter effect, suppose that for all p, the marginal distribution of income y_p has a Pareto tail with coefficient a_p, so that for high $\hat{y}_p$ we have

$$\mathbb{P}\left(y_p\geq\hat{y}_p\right)=c_p\cdot\left(\hat{y}_p\right)^{-a_p}.$$

Furthermore, assume that joint distribution of y_{p-1} and y_p at the tails can be summarized by the (survival) Clayton copula[29]

$$\mathbb{P}\left(y_{p-1}\geq\hat{y}_{p-1},y_p\geq\hat{y}_p\right)=\left(\left[\mathbb{P}\left(y_{p-1}\geq\hat{y}_{p-1}\right)\right]^{-\rho}+\left[\mathbb{P}\left(y_p\geq\hat{y}_p\right)\right]^{-\rho}-1\right)^{-\rho} \tag{39}$$

for $\rho>0$. The limit as $\rho\to0$ represents the case where y_p and y_{p-1} are comonotone (in particular, perfectly correlated), that is, all the agents with a given income in period $p-1$ also earn the same income in period p. The limit as $\rho\to\infty$ represents the case where labor earnings in the two periods are drawn independently from each other. In this case the conditional hazard rates are given by

$$H_{y_{p-1},y_p}\left(\hat{y}_p|\hat{y}_{p-1}\right)=\frac{a_p\left[\mathbb{P}\left(y_p\geq\hat{y}_p\right)\right]^{-1/\rho}}{\left[\mathbb{P}\left(y_p\geq\hat{y}_p\right)\right]^{-1/\rho}+\left[\mathbb{P}\left(y_{p-1}\geq\hat{y}_{p-1}\right)\right]^{-1/\rho}-1}.$$

Suppose that the Pareto coefficient a_p and the elasticity of labor supply ζ_p are independent of age p, that there are no savings (or that $\tau=0$), and that the baseline labor income taxes are chosen to maximize tax revenue collected from the agents with sufficiently high earnings. Under these assumptions, using the analysis of Section 5, the marginal tax rates on high incomes are constant and satisfy $T'\left(y\right)/\left(1-T'\left(y\right)\right)=\left(a\zeta\right)^{-1}$. In this case the joint perturbation for sufficiently high labor incomes in both periods yields a revenue effect equal to

$$\Gamma_{\hat{y}_p,\hat{y}_{p-1}}=1-\left(a\zeta\right)^{-1}\frac{\left[\mathbb{P}\left(y_{p-1}\geq\hat{y}_{p-1}\right)\right]^{-1/\rho}+\left[\mathbb{P}\left(y_p\geq\hat{y}_p\right)\right]^{-1/\rho}}{\left[\mathbb{P}\left(y_{p-1}\geq\hat{y}_{p-1}\right)\right]^{-1/\rho}+\left[\mathbb{P}\left(y_p\geq\hat{y}_p\right)\right]^{-1/\rho}-1}.$$

This expression implies that $\Gamma_{\hat{y}_p,\hat{y}_{p-1}}<0$ for all ρ, which implies that the separable tax system is not optimal. Specifically, a joint perturbation that decreases the marginal tax rates on incomes $\hat{y}_{p-1}$ and $\hat{y}_p$, and hence reduces the average tax rates for individuals with incomes $y_{p-1}\geq\hat{y}_{p-1}$

[29] This joint distribution is a generalization of the bivariate Pareto distribution, obtained for $a_p=a_{p-1}=\rho$.

in period $p-1$ *and* $y_p \geq \hat{y}_p$ in period p, jointly allows to raise additional revenue, starting from the optimal separable tax schedule. Note also that $\Gamma_{\hat{y}_p,\hat{y}_{p-1}} \to 0$ and as $\rho \to 0$, so that the gains from history-dependence disappear if each agent's income is the same in both periods.

The arguments above can be generalized to other forms of joint taxation. For example, the welfare effects of jointly taxing labor and capital incomes within period p at the joint income threshold $(\hat{y}_p, \hat{z}_p)$ are given by

$$\begin{aligned}\Gamma_{\hat{y}_p,\hat{z}_p} =& \mathbb{E}_{\substack{y_p \geq \hat{y}_p \\ z_p \geq \hat{z}_p}} [1-g_p] - \frac{\tau}{1-\tau}\mathring{\eta}_p \\ &- \frac{T'(\hat{y}_p)\zeta}{1-T'(\hat{y}_p)+\hat{y}_p \zeta T''(\hat{y}_p)} H_{y_{p-1},y_p}(\hat{y}_p \,|\hat{z}_p) \\ &- \frac{\tau_z}{1-\tau_z}\left\{\sum_{s=2}^{S} \beta^{s-p} \bar{\zeta}^{c,(z_p=\hat{z}_p, y_p \geq \hat{y}_p)}_{z_s,r_p}\right\} H_{y_{p-1},y_p}(\hat{z}_p \,|\hat{y}_p),\end{aligned} \tag{40}$$

where $\zeta^c_{z_s,r_p}$ is the elasticity of period-s capital income with respect to the period-p net-of-tax rateon capital income. Formula (40) is formally similar to equation (38), the relevant conditional hazard rates in this case being those of the joint distribution of labor and capital incomes, and the relevant elasticities being those of capital income with the usual compounding effect discussed in Sections 5.2 and 6.1. Note that these elasticities are in general different for individuals with different labor and capital incomes, and are therefore averaged over the region where the capital income tax rate is perturbed. More generally, for different preferences or a non-separable baseline income tax system, the elasticity parameters $\bar{\zeta}^c_{y_s,w_p}, \bar{\zeta}^c_{z_s,r_p}, \eta_{z_s,p}$ would all depend on the individuals' earnings histories, implying an additional source of benefits from using a non-separable tax system: the government can impose higher distortions in the regions where these elasticities are smaller, with an additional degree of "fine-tuning" relative to the separable case.

7 Overview of the Stochastic Model

In this section we briefly discuss the derivation of some of the results in the stochastic model. We only give an outline of the derivation here, the details are collected in our companion paper (Golosov, Tsyvinski, and Werquin 2014). For the clarity of the exposition, we consider the case where the horizon is $T=2$ periods, but our results generalize to the case $T \leq \infty$. In period one, an individual knows his first-period type, or productivity, $\theta_1 \in [0,\infty)$, and his initial capital stock $k_0 \in \mathbb{R}$. He then chooses his first-period consumption $c_1 \geq 0$, labor income $y_1 \geq 0$, and savings or borrowings $k_1 \in \mathbb{R}$ to carry over to period two (yielding capital income $z_2 \in \mathbb{R}$ in period two). For simplicity assume that the interest rate is the same for all individuals, so that capital income z_2 is known with certainty in period one given savings k_1. In period two, he draws his second-period productivity $\theta_2 \in [0,\infty)$. For all $\theta_1 \in \mathbb{R}_+$, the second-period type θ_2 is drawn from an exogenous distribution $F_{\theta_2|\theta_1}(\cdot)$ whose density $f_{\theta_2|\theta_1}(\cdot)$ is strictly positive on $\mathbb{R}_+$. The individual then chooses his second-period consumption $c_2 \geq 0$ and labor income $y_2 \geq 0$. Given his initial draw (k_0, θ_1), he thus chooses his first-period labor income and savings $y_1(k_0, \theta_1)$, $k_1(k_0, \theta_1)$, and a set of second-period incomes contingent on the second-period productivity $\{y_2(k_0, \theta_1, \theta_2) : \theta_2 \in \mathbb{R}_+\}$ in order to maximize the expected discounted value of his utility. The income vector $\mathbf{x}$ of an individual with initial capital and productivity (k_0, θ_1) thus has a two plus a continuum of rows, corresponding to the continuum of possible draws of θ_2 in period two.

In each period $s = 1, 2$, the government levies a tax T_s. The first-period tax function T_1 is a function of the individual's first-period labor income y_1 and capital k_1 only. (The government cannot tax second-period labor income y_2 in period one, as y_2 depends on the value of θ_2 that the individual will draw in period two, and hence is not known in period one.) The second-period tax function T_2 is a function of the individual's entire history of labor incomes $\{y_1, y_2\}$ and capital income z_2. The assumptions about the tax functions are identical to those we made in the deterministic model. Social welfare is then a weighted sum of individuals' expected indirect utilities $\mathscr{U}(k_0, \theta_1)$.

It is important to note that there are many more marginal tax rates and virtual incomes that are relevant for the individual than in the deterministic model. Since θ_2, and hence y_2 and $T_2(\cdot,\cdot,\cdot)$, are unknown when y_1 and k_1 are chosen, the two decision variables (y_1, k_1) depend on the set of all possible marginal tax rates and virtual incomes that the individual may end up facing in period two, depending on his type θ_2. Thus, y_1 and k_1 depend on the whole set $\left\{\left(\tau_2\left(y_1, \mathbf{x}_2^2, z_2\right), R_2\left(y_1, \mathbf{x}_2^2, z_2\right)\right) : \mathbf{x}_2^2 \in \mathbb{R}_+\right\}$, parametrized by the possible values $\mathbf{x}_2^2$ of second-period incomes that the individual may end up choosing in period two. Moreover, even though y_2 is chosen after a value of θ_2 has been drawn (say θ_2^*), $y_2(\theta_2^*)$ does not depend only on the marginal tax rate and virtual income that he ends up actually facing (i.e., $\tau_2(y_1, y_2(\theta_2^*), z_2)$), unless the utility function has no income effects. This is because y_1 and k_1, which have been chosen before the draw (taking into account the probabilities of all possible draws of θ_2), are not in general the optimal values given this particular draw θ_2^*, and this in turn affects the choice of $y_2(\theta_2^*)$. We thus obtain that for all $\theta_2^* \in \mathbb{R}_+$, $y_2(\theta_2^*)$ depends on the entire set of marginal tax rates and virtual incomes $\left\{\left(\tau_2\left(y_1, \mathbf{x}_2^2, z_2\right), R_2\left(y_1, \mathbf{x}_2^2, z_2\right)\right) : \mathbf{x}_2^2 \in \mathbb{R}_+\right\}$. In particular, when we perturb the tax function in the second period, $T_2(\cdot,\cdot,\cdot)$, at a given point $\mathbf{x}^2 = \left(y_1, \mathbf{x}_2^2, z_2\right)$, *all* the choice variables, $(y_1, \{y_2(\theta_2) : \theta_2 \in \mathbb{R}_+\}, z_2)$, adjust, even if the individual turns out not to be affected at all by the perturbation (i.e., even if $y_2(\theta_2^*) \neq \mathbf{x}_2^2$). This is the main conceptual difficulty that needs to be addressed in the stochastic model.

We first define the elasticities of labor incomes y_1, $\{y_2(\theta_2) : \theta_2 \in \mathbb{R}_+\}$ and savings k_1 with respect to the marginal tax rates on y_1 and k_1 that the individual actually faces in period one: $\tau_{1,y_1}, \tau_{1,k_1}$. We then define the elasticities of y_1, $\{y_2(\theta_2) : \theta_2 \in \mathbb{R}_+\}$ and k_1 with respect to all the marginal tax rates $\left\{\tau_{2,y_1}\left(\mathbf{x}^2\right), \tau_{2,y_2}\left(\mathbf{x}^2\right), \tau_{2,z_2}\left(\mathbf{x}^2\right) : \mathbf{x}^2 = \left(y_1, \mathbf{x}_2^2, z_2\right) \in \mathbb{R}_+^2 \times \mathbb{R}\right\}$ that the individual can possibly face in period two, depending on the possible values $\mathbf{x}_2^2$ of second-period incomes that the individual may end up choosing in period two. Similarly we first define the income effect parameters of y_1, $\{y_2(\theta_2) : \theta_2 \in \mathbb{R}_+\}$ and k_1 with respect to the individual's virtual income in period one, R_1. We then define the income effect parameters of y_1, $\{y_2(\theta_2) : \theta_2 \in \mathbb{R}_+\}$ and k_1 with respect to all the virtual incomes that the individual can possibly face in period two, $\left\{R_2\left(\mathbf{x}^2\right) : \mathbf{x}^2 = \left(y_1, \mathbf{x}_2^2, z_2\right) \in \mathbb{R}_+^2 \times \mathbb{R}\right\}$. We thus need to consider many more elasticities and income effect parameters than in the deterministic setting. These elasticities (e.g., of labor income y_2 with respect to the marginal tax rate at level $y_2' \neq y_2$) are new to the literature on taxation. We derive explicit analytical expressions for all these elasticities, as we did in the deterministic setting. They resemble those in the deterministic setting, except that they are weighted by the probabilities of earning the second-period income where the tax rate is perturbed.

We then go on to derive the behavioral responses to perturbations. The results are proved in the same way as in the deterministic setting, but the added degree of complexity we just described makes the derivations more involved both theoretically and conceptually. The formulas we obtain are accordingly more complex. Remarkably, however, we show that we can define the elasticity matrices, as well as the gradients and Hessians of the tax functions, in a way that allows to write the formula in a similar compact way as (8) in the deterministic model (details are in the Appendix). The proof and intuition of this formula follows the same steps as those of (8). We show that the change $d\mathbf{x}$ in the income vector $\mathbf{x}$ following a general perturbation $\left(d\tau_1, dR_1, d\tau_2\left(\mathbf{x}^2\right) dR_2\left(\mathbf{x}^2\right)\right)$ of the baseline tax system is given by:

$$\begin{aligned} d\mathbf{x} = & \left\{\mathfrak{i} - \boldsymbol{E}_{\tau_1}^c(\mathbf{x})\left(D^2 T_1(\mathbf{x}_1)\right) - \int_0^\infty \boldsymbol{E}_{\tau_2(\mathbf{x}^{2\prime})}^c(\mathbf{x})\left(D^2 T_2\left(\mathbf{x}^{2\prime}\right)\right) d\mathbf{x}_2^{2\prime}\right\}^{-1} \\ & \times\left[\left(\boldsymbol{E}_{\tau_1}^c(\mathbf{x})\, d\tau_1 + \boldsymbol{I}_{R_1}(\mathbf{x})\, dR_1\right) + \left(\boldsymbol{E}_{\tau_2(\mathbf{x}^2)}^c(\mathbf{x})\, d\tau_2\left(\mathbf{x}^2\right) + \boldsymbol{I}_{R_2(\mathbf{x}^2)}(\mathbf{x})\, dR_2\left(\mathbf{x}^2\right)\right)\right]. \end{aligned} \tag{41}$$

As an illustration of these results, we show how the revenue effects of reforming the baseline tax system of Section 6 write in the stochastic model, when the utility function has no income effects and is CRRA. A non-linear separable perturbation of the first-period labor income tax schedule at point $\hat{y}_1$ yields the following change in government revenue:

$$\Gamma_{1,y}(\hat{y}_1) = 1 - \frac{T_1'(\hat{y}_1)}{1 - T_1'(\hat{y}_1) + \hat{y}_1 \zeta T_1''(\hat{y}_1)} \zeta \frac{\hat{y}_1 f_{y_1}(\hat{y}_1)}{1 - F_{y_1}(\hat{y}_1)} - \beta \frac{\tau_z}{1 - \tau_z} \bar{\eta}_{z_2, R_1}^{(y_1 \geq \hat{y}_1)}. \tag{42}$$

Formula (42) shows that the revenue effect of perturbing the first-period labor income tax rate in the stochastic model is formally similar to the effect in the deterministic model. However, we show that uncertainty about second-period productivity implies that the income effect parameter on savings in the stochastic model is equal to $\left.\frac{\partial k_1}{\partial R_1}\right|_S = \left(u_1'' + \beta R^2 \mathbb{E}\left[u_2'' \mid \theta_1\right]\right)^{-1} u_1''$, and hence is *smaller* than in the deterministic model, $\left.\frac{\partial k_1}{\partial R_1}\right|_D = \left(1 + \beta^{-1/\sigma} R^{1-1/\sigma}\right)^{-1}$. This implies that the gain from decreasing the labor income tax rate in period one is smaller in the stochastic model than in the deterministic model; the latter thus provides an upper bound for the gains of age-dependence.

A non-linear separable perturbation of the second-period labor income tax schedule at point

$\hat{y}_2$ yields the following change in government revenue:

$$\Gamma_{2,y}\left(\hat{y}_2\right)=1-\frac{T'_{y,2}\left(\hat{y}_2\right)}{1-T'_{y,2}\left(\hat{y}_2\right)+\hat{y}_2\zeta T''_{y,2}\left(\hat{y}_2\right)}\zeta\frac{\hat{y}_2 f_{y_2}\left(\hat{y}_2\right)}{1-F_{y_2}\left(\hat{y}_2\right)}-\frac{\tau_z}{1-\tau_z}\bar{\eta}_{z_2,R_2(y_2\geq\hat{y}_2)}, \tag{43}$$

where $\bar{\eta}_{z_2,R_2(y_2\geq\hat{y}_2)}$ is the aggregate change in capital income in the economy when an additional dollar is distributed lump-sum in period two, uniformly among all the individuals whose labor income in period two is above $\hat{y}_2$, that is

$$\bar{\eta}_{z_2,R_2(y_2\geq\hat{y}_2)}\equiv\int_{\mathbb{R}_+}\int_{\hat{y}_2}^{\infty}\int_{\mathbb{R}}\eta^{(\mathbf{x}_1)}_{z_2,R_2\left(\mathbf{x}_2^2\right)}\frac{f_{\mathbf{x}_1}\left(y_1,k_1\right)}{1-F_{y_2}\left(\hat{y}_2\right)}dy_1 d\mathbf{x}_2^2 dk_1.$$

Note that every individual (with choice vector $\mathbf{x}_1=(y_1,k_1)$ in period one) reacts to this change by adjusting their savings, i.e. $\eta^{(\mathbf{x}_1)}_{z_2,R_2\left(\mathbf{x}_1,\mathbf{x}_2^2\right)}\neq 0$, because they have positive probability of earning more that $\hat{y}_2$ in the second period. However, only those with second-period income $\hat{y}_2$ under the baseline tax system change their second-period income. Formula (43) shows that the revenue effect of perturbing the second-period labor income tax rate in the stochastic model is formally similar to the effect in the deterministic model. However, we show that the savings effect in the stochastic setting, $\bar{\eta}_{z_2,R_2(y_2\geq\hat{y}_2)}$, is strictly larger than in the deterministic setting, $\bar{\eta}^{(y_2\geq\hat{y}_2)}_{z_2,R_2}$. Hence the revenue gains from increasing the labor income tax rates in period two are smaller in the stochastic model than in the deterministic model.

A non-linear separable perturbation of the capital income tax schedule at point $\hat{z}_2$ yields the following change in government revenue:

$$\Gamma_{2,z}\left(\hat{z}_2\right)=1-\frac{\tau_z}{1-\tau_z}\bar{\zeta}^{c,(\hat{z}_2)}_{z_2,r_2}\frac{\hat{z}_2 f_{z_2}\left(\hat{z}_2\right)}{1-F_{z_2}\left(\hat{z}_2\right)}-\frac{\tau_z}{1-\tau_z}\bar{\eta}^{(z_2\geq\hat{z}_2)}_{z_2,R_2}, \tag{44}$$

where $\bar{\eta}^{(z_2\geq\hat{z}_2)}_{z_2,R_2}$ is the average income effect parameter of capital income with respect to a *certain* increase in period-two virtual income, among individuals with capital income $z_2\geq\hat{z}_2$, that is,

$$\bar{\eta}^{(z_2\geq\hat{z}_2)}_{z_2,R_2}\equiv\int_{\mathbb{R}_+}\int_{\hat{z}_2}^{\infty}\eta^{(\mathbf{x}_1)}_{z_2,R_2}\frac{f_{\mathbf{x}_1}\left(y_1,k_1\right)}{1-F_{z_2}\left(\hat{z}_2\right)}dy_1 dk_1.$$

Formula (43) shows that the revenue effect of perturbing the second-period labor income tax rate in the stochastic model is formally similar to the effect in the deterministic model. However, we show that the savings effect in the stochastic setting, $\bar{\eta}_{z_2,R_2(y_2\geq\hat{y}_2)}$, is strictly larger than in the deterministic setting, $\bar{\eta}^{(y_2\geq\hat{y}_2)}_{z_2,R_2}$. Hence the revenue gains from increasing the labor income tax rates in period two are smaller in the stochastic model than in the deterministic model. However, we show that the average compensated capital income elasticity in the stochastic model, $\bar{\zeta}^{c,(\hat{z}_2)}_{z_2,r_2}$, is positive but smaller than its counterpart in the deterministic model. Similarly, the average income effect parameters in the stochastic model, $\bar{\eta}^{(z_2)}_{z_2,R_2}$, are negative and smaller than their counterparts in the deterministic model. Thus, on the one hand, the increase in the tax rate induces a smaller decrease in capital income (in the stochastic model) for individuals with $z_2 = \hat{z}_2$; on the other hand, the increase in the lump-sum tax liability induces a larger increase in capital income (in the stochastic model) for individuals with $z_2 \geq \hat{z}_2$. Therefore the revenue gains from increasing the capital income tax rates in period two in the stochastic model are larger than in the deterministic model.

8 Conclusion

We identify a condition on individual demand under which the effects of taxation on individual behavior, tax revenue, and social welfare of can be expressed in terms of empirically observable and easily interpretable parameters, namely the labor and capital income elasticities, the multivariate hazard rates of the income distributions, and the marginal social welfare weights. Applying these formulas to various settings, we show that optimal taxes and the effects of tax reforms obey common general principles, and that the benefits of using sophisticated tax instruments come from the ability to fine-tune the distortions to the segments of the population who respond relatively little to taxes.

We leave two important extensions for future research. First, our numerical applications were meant to provide rough orders of magnitude of the forces at play in a few examples. It would be valuable to do more extensive numerical welfare calculations, estimating the fundamental parameters that enter our tax formulas using micro data. Second, we believe our approach is useful to analyze problems which may be difficult to tackle directly, e.g., multidimensional mechanism design models. However, an open question is to find a condition on the primitives of the model such that our assumption on individual demand is satisfied.

References

[1] **Albanesi, Stefania**, "Optimal Taxation of Entrepreneurial Capital with Private Information", *Working Paper*, 2011

[2] **Blundell, Richard and Andrew Shephard**, "Optimal Taxation in an Empirical Life-Cycle Model of Labor Supply", *Working Paper*, 2014

[3] **Bonhomme, Stephane and Jean-Marc Robin**, "Modeling Individual Earnings Trajectories using Copulas with an Application to the Study of Earnings Inequality: France, 1990-2002", *Working Paper*, 2003

[4] **Chamley, Christophe**, "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives", *Econometrica*, 1986

[5] **Chetty, Raj**, "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods", *Annual Review of Economics*, 2009

[6] **Chetty, Raj**, "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply", *Econometrica*, 2012

[7] **Chetty, Raj**, "Time Trends in the Use of Administrative Data for Empirical Research", *Unpublished*, 2012

[8] **Conesa, Juan Carlos, Sagiri Kitao, and Dirk Krueger**, "Taxing Capital? Not a Bad Idea After All!", *American Economic Review*, 2009

[9] **Congressional Budget Office**, "Effective Marginal Tax Rates on Labor Income", *CBO paper*, 2005

[10] **Dearden, Lorraine, Emla Fitzsimons, Alissa Goodman and Greg Kaplan**, "Estimating Lifetime Earnings Distributions Using Copulas", *Working Paper*, 2006

[11] **Diamond, Peter**, "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates", *American Economic Review*, 1998

[12] **Diamond, Peter**, "A Many-Person Ramsey Tax Rule", *Journal of Public Economics*, 1975

[13] **Diamond, Peter, and James Mirrlees**, "Optimal Taxation and Public Production II: Tax Rules", *American Economic Review*, 1971

[14] **Farhi, Emmanuel and Ivan Werning**, "Insurance and Taxation over the Life-Cycle", *Review of Economic Studies*, 2012

[15] **Farhi, Emmanuel and Ivan Werning**, "Progressive Estate Taxation", *Quarterly Journal of Economics*, 2010

[16] **Golosov, Mikhail, Narayana Kocherlakota and Aleh Tsyvinski**, "Optimal Indirect and Capital Taxation", *Review of Economic Studies*, 2003

[17] **Golosov, Mikhail, Maxim Troshkin and Aleh Tsyvinski**, "Redistribution and Social Insurance", *Working Paper*, 2013

[18] **Golosov, Mikhail, Aleh Tsyvinski and Nicolas Werquin**, "A Variational Approach to the Analysis of Tax Systems: The Stochastic Model", *Working Paper*, 2014

[19] **Greenwood, Jeremy, Zvi Hercowitz, and Gregory Huffman**, "Investment, Capacity Utilization, and the Real Business Cycle", *American Economic Review*,1988

[20] **Gruber, Jonathan and Emmanuel Saez**, "The Elasticity of Taxable Income: Evidence and Implications", *Journal of Public Economics*, 2002

[21] **Guvenen, Fatih, Serdar Ozkan, Fatih Karahan, and Jae Song**, "What Do Data on Millions of US Workers Say About Lifecycle Earnings Risk?", *Working Paper*, 2014a

[22] **Guvenen Fatih, Serdar Ozkan, and Jae Song,** "The Nature of Countercyclical Income Risk", *Journal of Political Economy*, 2014b

[23] **Heathcote, Jonathan, Kjetil Storesletten, and Gianluca Violante**, "Optimal Tax Progressivity: An Analytical Framework", *Working Paper*, 2014

[24] **Hendren, Nathaniel**, "The Policy Elasticity", *Working* Paper*,* 2013

[25] **Judd, Ken**, "Redistributive Taxation in a Simple Perfect Foresight Model", *Journal of Public Economics*, 1985

[26] **Keane, Michael, and Richard Rogerson**, "Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom." *Journal of Economic Literature*, 2012

[27] **Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez**, "The Optimal Income Taxation of Couples", *Econometrica*, 2009

[28] **Kocherlakota, Narayana**, "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 2005

[29] **Kremer, Michael**, "Should Taxes Be Independent of Age?", *Unpublished Paper*, 2002

[30] **Maag, Elaine, C. Eugene Steuerle, Ritadhi Chakravarti, and Caleb Quakenbush**, "How Marginal Tax Tates Affect Families at Different Levels of Poverty", *National Tax Journal*, 2012

[31] **Mirrlees, James**, "An Exploration in the Theory of Optimal Income Taxation", *Review of Economic Studies*, 1971

[32] **Nirei, Makoto, and Wataru Souma**, "A Two Factor Model of Income Distribution Dynamics", *Review of Income and Wealth*, 2007

[33] **Piketty, Thomas**, "La Redistribution Fiscale Contre le Chômage", *Revue Française d'Economie*, 1997

[34] **Piketty, Thomas and Emmanuel Saez**, "A Theory of Optimal Inheritance Taxation", *Econometrica*, 2013

[35] **Piketty, Thomas, Emmanuel Saez and Stefanie Stancheva**, "Optimal taxation of Top Incomes: A Tale of Three Elasticities", *American Economic Journal: Economic Policy*, 2014

[36] **Prante, Gerald and Austin John**, "Top Marginal Effective Tax Rates by State and by Source of Income, 2012 Tax Law vs. 2013 Tax Law (as enacted in ATRA)", *Working Paper*, 2013

[37] **Ramsey, Frank**, "A Contribution to the Theory of Taxation", *The Economic Journal*, 1927

[38] **Rothschild, Casey and Florian Scheuer**, "A Theory of Income Taxation under Multidimensional Skill Heterogeneity", *Working Paper*, 2014

[39] **Saez, Emmanuel**, "Using Elasticities to Derive Optimal Income Tax Rates", *Review of Economic Studies*, 2001

[40] **Saez, Emmanuel**, "Essays on the Economics of Income Taxation", *M.I.T. Manuscript*, 2001b

[41] **Saez, Emmanuel, Joel Slemrod and Seth Giertz**, "The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review", *Journal of Economic Literature*, 2012

[42] **Shourideh, Ali**, "Optimal Taxation of Wealthy Individuals", *Working Paper*, 2012

[43] **Straub, Ludwig, and Ivan Werning**, "A Reappraisal of Chamley-Judd Zero Capital Taxation Results", *Working Paper*, 2014

[44] **Weinzierl, Matthew**, The Surprising Power of Age-Dependent Taxes, *Review of Economic Studies*, 2011

A Proofs of Sections 3 and 4

In this section, we provide the proofs of the results of Sections 3 and 4 from the main text. We first derive analytical expressions for all the elasticities and income effect parameters in the general model, as well as under the assumptions of Section 6. We then prove the results of Propositions 1, 2 and 3, and provide details for the derivation of various results in the text.

A.1 Elasticities and Income Effect Parameters

We start by providing analytical expressions for the elasticities and income effect parameters in the general model of Sections 4 to 2. To derive these expressions, we differentiate the system of first-order conditions (9) of the individual's problem with respect to the marginal or net-of tax rates $\left\{\hat{q}_{x_j,s}\right\}_{1\leq s,j\leq S}$ and the virtual incomes $\{R_s\}_{1\leq s\leq S}$. For a given individual, define r_{s+1} as the exogenous interest rate that he faces on his capital k_s, so that his capital income in period $s+1$ is equal to $z_{s+1}=r_{s+1}k_s$. We can write the individual's first-order conditions as

$$\begin{aligned}
&U_{x_j}\left(\left\{-\sum_{i=1}^{2S}q_{x_i,t}x_{\boldsymbol{\theta},i}+R_t\right\}_{1\leq t\leq S},\{y_{\boldsymbol{\theta},t}\}_{1\leq t\leq S},\{z_{\boldsymbol{\theta},t+1}\}_{1\leq t\leq S},\boldsymbol{\theta}\right)\\
&=\sum_{s=1}^{S}q_{x_j,s}U_{c_s}\left(\left\{-\sum_{i=1}^{2S}q_{x_i,t}x_{\boldsymbol{\theta},i}+R_t\right\}_{1\leq t\leq S},\{y_{\boldsymbol{\theta},t}\}_{1\leq t\leq S},\{z_{\boldsymbol{\theta},t+1}\}_{1\leq t\leq S},\boldsymbol{\theta}\right),
\end{aligned}\tag{45}$$

We then use the Slutsky equations to obtain the compensated elasticities from the uncompensated elasticities and the income effect parameters. Define the $2S\times 2S$ matrix A by:

$$[A]_{s,j}\equiv -U_{x_s,x_j}-\sum_{t=1}^{S}\sum_{q=1}^{S}q_{x_s,t}q_{x_j,q}U_{c_t,c_q}+\sum_{t=1}^{S}q_{x_s,t}U_{c_t,x_j}+\sum_{q=1}^{S}q_{x_j,q}U_{x_s,c_q}.$$

Define also the $2S$-vectors $B_{\tau_{p,x_i}}$, B_{R_p}, and $B^c_{\tau_{p,x_i}}$, for any $1\leq p\leq S$ and $1\leq i\leq 2S$, by

$$\begin{aligned}
\left[B^u_{\tau_{p,x_i}}\right]_s &\equiv \sum_{t=1}^{S}q_{x_s,t}U_{c_t,c_p}x_i-U_{x_s,c_p}x_i-U_{c_p}\mathbf{1}_{\{i=s\}},\quad \forall s\in\{1,\ldots,2S\},\\
\left[B_{R_p}\right]_s &= -\sum_{t=1}^{S}q_{x_s,t}U_{c_t,c_p}+U_{x_s,c_p},\quad \forall s\in\{1,\ldots,2S\},\\
\left[B^c_{\tau_{p,x_i}}\right]_s &\equiv -U_{c_p}\mathbf{1}_{\{i=s\}},\quad \forall s\in\{1,\ldots,2S\}.
\end{aligned}$$

We can then write the the uncompensated and compensated income elasticities and the income effect parameters as

$$\zeta^{u,c}_{x_j,\hat{q}_{x_t,s}}=\pm\frac{\hat{q}_{x_t,s}}{x_j}\left[A^{-1}B^{u,c}_{\tau_{s,x_t}}\right]_j,\quad\text{and}\quad \eta_{x_j,R_s}=\hat{q}_{x_j,s}\left[A^{-1}\times B_{R_s}\right]_j,\tag{46}$$

where $\pm = +$ if $x_t \in \{z_s, r_s k_{s-1}\}$, and $\pm = -$ otherwise. Note that the components of the elasticity matrices and the income effect vectors (4) are the *partial derivatives* of the (compensated or uncompensated) demands, and not directly the *elasticities* and income effect parameters we just derived.

For concreteness we show how to apply these formulas to the static model ($S = 1$). Differentiating the first-order conditions

$$U_y\left(\left(1-\tau\right)y + R, y, \theta\right) = -\left(1-\tau\right)U_c\left(\left(1-\tau\right)y + R, y, \theta\right)$$

implies the following expressions for the elasticities (46):

$$\zeta_{y,1-\tau}^{u} = \frac{U_y/y - \left(U_y/U_c\right)^2 U_{cc} + \left(U_y/U_c\right)U_{cy}}{U_{yy} + \left(U_y/U_c\right)^2 U_{cc} - 2\left(U_y/U_c\right)U_{cy}}, \quad \eta_{y,R} = \frac{-\left(U_y/U_c\right)^2 U_{cc} + \left(U_y/U_c\right)U_{cy}}{U_{yy} + \left(U_y/U_c\right)^2 U_{cc} - 2\left(U_y/U_c\right)U_{cy}}.$$

In this case, the matrix A defined above is minus the denominator of these two expressions, and the vectors B_τ^u, B_R are respectively $\left(1-\tau\right)y$ and $-\left(1-\tau\right)^{-1}$ times the numerators of these two expressions.

We now show how these expressions simplify under the assumptions of Section 6. That is, we assume that the utility function is time-separable, has no income effects on labor supply, and that the baseline tax system is separable and linear in capital income. In this case, the $S \times S$ upper-left submatrix of $\boldsymbol{E}_s^{c,(\mathbf{x}_{\boldsymbol{\theta}})}$ is diagonal, and its upper-right and lower-left submatrices are zero. Moreover, the first S components of the income effect vector $\boldsymbol{I}_s^{(\mathbf{x}_{\boldsymbol{\theta}})}$ are equal to zero. Thus, for every period $p \in \{1, \ldots, S\}$ where the tax system is perturbed, the only non-zero compensated elasticities and income effect parameters are: (i) the compensated elasticities of labor incomes y_s with respect to the labor income tax rates in the current period $\tau_{y_s,s}$, i.e., the S parameters $\zeta_{y_s,1-\tau_{y_s,s}}^c$; (ii) the compensated elasticities of capital incomes z_s with respect to all of the capital income tax rates $\tau_{z_t,p}$, i.e., the S^2 parameters $\zeta_{z_s,1-\tau_{z_t,p}}^c$; (iii) the income effect parameters on capital incomes z_s, i.e., the S parameters η_{z_s,R_p}. The formulas above show that the labor income elasticities are given by:

$$\zeta_{y_s,1-\tau_{y_s,s}}^c = \frac{v'\left(y_s/\theta_s\right)}{\left(y_s/\theta_s\right)v''\left(y_s/\theta_s\right)}.$$

Suppose either that the utility function is CRRA, i.e. $u\left(x\right) = x^{1-\sigma}/\left(1-\sigma\right)$, and let $\alpha \equiv \beta^{-1/\sigma}R^{1-1/\sigma}$, or that the utility function is CARA, i.e. $u\left(x\right) = -\gamma^{-1}\exp\left(-\gamma x\right)$, and let $\alpha = R$. We obtain that the only non-zero compensated capital income elasticities are given by

$$\frac{\partial z_s^c}{\partial\left(1-\tau_{z_p,p}\right)} = \left(\frac{-u_p'}{u_p''}\right)\frac{r^2 R^{s-p-2}}{\sum_{i=0}^{S-1}\alpha^i}\begin{cases}\left(\sum_{i=0}^{S-p}\alpha^i\right)\left(\sum_{i=p-s+1}^{p-1}\alpha^i\right), & \text{if } s \leq p,\\ \left(\sum_{i=0}^{S-s}\alpha^i\right)\left(\sum_{i=1}^{p-1}\alpha^i\right), & \text{if } s \geq p+1,\end{cases} \tag{47}$$

and the only non-zero income effect parameters are given by

$$\frac{\partial z_s}{\partial R_p} = \frac{rR^{s-p-1}}{\sum_{i=0}^{S-1}\alpha^i}\begin{cases} -\left(\sum_{i=S-s+1}^{S-1}\alpha^i\right), & \text{if } s \le p, \\ \left(\sum_{i=0}^{S-s}\alpha^i\right), & \text{if } s \ge p+1. \end{cases} \tag{48}$$

Note that in the CARA case, $\frac{-u_p'}{u_p''}$ is simply equal to γ^{-1}.

A.2 Proofs of Propositions 1 to 3

We first prove the existence of the Gateaux differential of the income functional and show Proposition 1.

Proof of Proposition 1. We first show that the income functional $\mathbf{x}_{\boldsymbol{\theta}}(\cdot)$ is Gateaux differentiable around the initial tax system T_p. Denote by $\mathbf{x}_{\boldsymbol{\theta}} \equiv \mathbf{x}_{\boldsymbol{\theta}}(T_p)$, resp. $\tilde{\mathbf{x}}_{\boldsymbol{\theta}} \equiv \mathbf{x}_{\boldsymbol{\theta}}(T_p + \mu h)$, the income vector chosen by an individual $\boldsymbol{\theta}$ given the baseline tax system T_p, resp. the perturbed tax system in the direction h, $T_p + \mu h$. The vectors $\mathbf{x}_{\boldsymbol{\theta}}$ and $\tilde{\mathbf{x}}_{\boldsymbol{\theta}}$ are the solution to the respective systems of the first-order conditions (9), where the map $F : \mathbb{R}^{2S} \times \mathbb{R}^{2S} \times \mathbb{R} \to \mathbb{R}^{2S}$ is continuously differentiable. For any $j \in \{1, \ldots, 2S\}$, let F_j denote the j^{th} component of F. Writing the first-oder conditions both at the baseline and the perturbed tax system yields, for all j,

$$\begin{aligned} 0 = F_j &\left(\tilde{\mathbf{x}}_{\boldsymbol{\theta}}, \left\{\tau_{x_t,p}(\tilde{\mathbf{x}}_{\boldsymbol{\theta}}) + \mu\frac{\partial h}{\partial x_t}(\tilde{\mathbf{x}}_{\boldsymbol{\theta}})\right\}_{1\le t\le 2S}, T_p(\tilde{\mathbf{x}}_{\boldsymbol{\theta}}) + \mu h(\tilde{\mathbf{x}}_{\boldsymbol{\theta}})\right) \\ &- F_j\left(\mathbf{x}_{\boldsymbol{\theta}}, \{\tau_{x_t,p}(\mathbf{x}_{\boldsymbol{\theta}})\}_{1\le t\le 2S}, T_p(\mathbf{x}_{\boldsymbol{\theta}})\right), \quad \forall j = 1, \ldots, 2S. \end{aligned} \tag{49}$$

Define the matrix $M = (m_{j,s})_{1\le j,s\le 2S}$ as

$$m_{j,s} = \frac{\partial F_j}{\partial x_{\boldsymbol{\theta},s}} + \sum_{t=1}^{2S}\frac{\partial F_j}{\partial \tau_{x_t,p}}\frac{\partial \tau_{x_t,p}(\mathbf{x}_{\boldsymbol{\theta}})}{\partial x_{\boldsymbol{\theta},s}} + \frac{\partial F_j}{\partial T_p}\frac{\partial T_p(\mathbf{x}_{\boldsymbol{\theta}})}{\partial x_{\boldsymbol{\theta},s}},$$

the vectors $N_{x_t} = (n_{j,x_t})_{1\le j\le 2S}$ for all $t \in \{1, \ldots, 2S\}$ as $n_{j,x_t} = \frac{\partial F_j}{\partial \tau_{x_t,p}}$, and the vector $N_T = (n_{j,T})_{1\le j\le 2S}$ as $n_{j,T} = \frac{\partial F_j}{\partial T_p}$. Assumption 2 implies that $\|\tilde{\mathbf{x}}_{\boldsymbol{\theta}} - \mathbf{x}_{\boldsymbol{\theta}}\| = O(\mu)$ as $\mu \to 0$. Moreover, we have $\|\mu h(\tilde{\mathbf{x}}_{\boldsymbol{\theta}}) - \mu h(\mathbf{x}_{\boldsymbol{\theta}})\| = o(\mu)$ as $\mu \to 0$. A first-order Taylor expansion of (49) as $\mu \to 0$, i.e., of the perturbed system of first-order conditions around the initial system, thus writes:

$$\frac{1}{\mu}(\tilde{\mathbf{x}}_{\boldsymbol{\theta}} - \mathbf{x}_{\boldsymbol{\theta}}) = -\sum_{t=1}^{2S}\left\{M^{-1}N_{x_t}\right\}\frac{\partial h(\mathbf{x}_{\boldsymbol{\theta}})}{\partial x_t} - \left\{M^{-1}N_T\right\}h(\mathbf{x}_{\boldsymbol{\theta}}) + o_{\mu\to 0}(1).$$

This shows the existence of the Gateaux differential $\delta\mathbf{x}_{\boldsymbol{\theta}}(T_p, h) \in \mathbb{R}^{2S}$ of the income functional $\mathbf{x}_{\boldsymbol{\theta}}(\cdot)$ at T_p with increment h. We then express the Gateaux differential of the income functional as a function of the elasticity matrices and vectors of income effect parameters. To do so, we derive the change $\tilde{\mathbf{x}}_{\boldsymbol{\theta}} - \mathbf{x}_{\boldsymbol{\theta}}$ in the individual's choice vector by writing the first-order Taylor approximation of the post-perturbation system of first order conditions (45) around the solution $\mathbf{x}_{\boldsymbol{\theta}}$ to the initial system. Using the explicit expressions for the elasticities and income effect parameters derived in (46), we obtain

$$(\tilde{\mathbf{x}}_{\boldsymbol{\theta}} - \mathbf{x}_{\boldsymbol{\theta}}) = \left[\mathrm{i}_{2S} - \sum_{s=1}^{S} \boldsymbol{E}_{s}^{c,(\mathbf{x}_{\boldsymbol{\theta}})} \left(D^2 T_s\left(\mathbf{x}_{\boldsymbol{\theta}}\right)\right)\right]^{-1} \left\{\boldsymbol{E}_{p}^{c,(\mathbf{x}_{\boldsymbol{\theta}})} \nabla h\left(\mathbf{x}_{\boldsymbol{\theta}}\right) + \boldsymbol{I}_{\mathbf{x},R_p}^{(\mathbf{x}_{\boldsymbol{\theta}})} h\left(\mathbf{x}_{\boldsymbol{\theta}}\right)\right\}.$$

This concludes the proof of Proposition 1. □

We then show Proposition 2, which gives expressions for the Gateaux differentials of the tax revenue and social welfare functionals.

Proof of Proposition 2. Consider an admissible perturbation h_p of the baseline tax function T_p, so that the perturbed tax function is $T_p + \mu h_p$. For any $\boldsymbol{\theta}$, letting $\mathbf{x}_{\boldsymbol{\theta}} \equiv \mathbf{x}_{\boldsymbol{\theta}}\left(T_p\right)$ and $\tilde{\mathbf{x}}_{\boldsymbol{\theta}} \equiv \mathbf{x}_{\boldsymbol{\theta}}\left(T_p + \mu h_p\right)$, a Taylor approximation yields

$$\begin{aligned} & \left[T_p + \mu h_p\right]\left(\tilde{\mathbf{x}}_{\boldsymbol{\theta}}\right) - T_p\left(\mathbf{x}_{\boldsymbol{\theta}}\right) \\ =& \mu \left\langle \nabla T_p\left(\mathbf{x}_{\boldsymbol{\theta}}\right), \delta \mathbf{x}_{\boldsymbol{\theta}}\left(T_p, h_p\right)\right\rangle + \mu h_p\left(\mathbf{x}_{\boldsymbol{\theta}}\left(T_p\right)\right) + o\left(\mu\right). \end{aligned}$$

Similarly, using the envelope theorem and the local Lipschitz continuity of the income function (Assumption 2), we get

$$\mathcal{G}\left(\mathscr{U}_{\boldsymbol{\theta}}\left(T_p + \mu h_p\right)\right) - \mathcal{G}\left(\mathscr{U}_{\boldsymbol{\theta}}\left(T_p\right)\right) = \left(-\frac{\lambda}{1-\alpha}\beta^{p-1} g_p\left(\mathbf{x}_{\boldsymbol{\theta}}\right) h_p\left(\mathbf{x}_{\boldsymbol{\theta}}\right)\right)\mu + o\left(\mu\right).$$

Using the compactness of the set X and assuming that the integrand is twice continuously differentiable, we thus obtain that the change in social welfare is equal to

$$\begin{aligned} & \mathscr{W}\left(T_p + \mu h_p\right) - \mathscr{W}\left(T_p\right) \\ =& \mu\lambda \int_X \left\{\beta^{p-1}\left(1 - g_p\left(\mathbf{x}\right)\right) h_p\left(\mathbf{x}\right) + \left\langle \sum_{s=1}^{S} \beta^{s-1} \nabla T_s\left(\mathbf{x}_{\boldsymbol{\theta}}\right), \delta \mathbf{x}\left(T_p, h_p\right)\right\rangle\right\} f_{\mathbf{x}}\left(\mathbf{x}\right) d\mathbf{x} + o\left(\mu\right). \end{aligned}$$

This proves formula (12). Letting $\tilde{\boldsymbol{T}}'\left(\mathbf{x}\right) \equiv \boldsymbol{T}'\left(\mathbf{x}\right) \boldsymbol{D}^{-1}\left(\mathbf{x}\right)$ and using the fact that the density of incomes is equal to zero on the boundary ∂X of the set X, we can integrate by parts the integral involving $\nabla h_p\left(\mathbf{x}\right)$ in this expression to get

$$\int_{\mathrm{X}} \left[\tilde{\boldsymbol{T}}'\left(\mathbf{x}\right) \boldsymbol{E}_{p}^{c,(\mathbf{x})} f_{\mathbf{x}}\left(\mathbf{x}\right)\right] \nabla h_p\left(\mathbf{x}\right) d\mathbf{x} = -\int_{\mathrm{X}} \nabla \cdot \left[\tilde{\boldsymbol{T}}'\left(\mathbf{x}\right) \boldsymbol{E}_{p}^{c,(\mathbf{x})} f_{\mathbf{x}}\left(\mathbf{x}\right)\right] h_p\left(\mathbf{x}\right) d\mathbf{x}.$$

This proves formula (13). □

Next we prove Proposition 3.

Proof of Proposition 3. A necessary condition for the social welfare functional $\mathscr{W}(\cdot)$ to have an extremum at T_p is $\delta\mathscr{W}(T_p, h) = 0$, for all h (see, e.g., Luenberger 1969). From equation (13), this implies that the integrand must be equal to zero pointwise, that is for all $\mathbf{x} \in \mathrm{X}$,

$$\left(\beta^{p-1}\left(1-g_p(\mathbf{x})\right)-\boldsymbol{T}'(\mathbf{x})\,\boldsymbol{D}^{-1}(\mathbf{x})\,\boldsymbol{I}_p^{(\mathbf{x})}\right) f_{\mathbf{x}}(\mathbf{x})-\nabla\cdot\left(\boldsymbol{T}'(\mathbf{x})\,\boldsymbol{D}^{-1}(\mathbf{x})\,\boldsymbol{E}_p^{c,(\mathbf{x})} f_{\mathbf{x}}(\mathbf{x})\right)=0.$$

Integrating this equation on the volume V with closed boundary $\mathrm{S} = \partial\mathrm{V}$ and using the divergence theorem, we obtain formula (17). Finally, we obtain formula (16) by using the separable linear perturbations $h_p(\mathbf{x}) = x_p$ and equation (12). □

We finally prove the formulas which express the optimal tax system as a function of the distribution of types $\boldsymbol{\theta}$.

Proof of formula (19). Differentiating the i^{th} first-order condition (9) with respect to θ_j for $j \in \{1, \ldots, 2S\}$ yields

$$\sum_{s=1}^{2S} m_{i,s}\frac{\partial x_{\boldsymbol{\theta},s}}{\partial\theta_j} = -\frac{\partial F_i}{\partial\theta_j} \Rightarrow J_{\mathbf{x}}(\boldsymbol{\theta}) = -M^{-1}J_{\mathbf{F}}(\boldsymbol{\theta}),$$

where the matrix M is the same as in the proof of Proposition 1, and $J_{\mathbf{x}}(\boldsymbol{\theta}), J_{\mathbf{F}}(\boldsymbol{\theta})$ are the matrices $[\partial x_{\boldsymbol{\theta},i}/\partial\theta_j]_{1\le i,j\le 2S}$ and $[\partial F_i/\partial\theta_j]_{1\le i,j\le 2S}$ respectively. Similarly, differentiating the first-order conditions (9) with respect to the variables $\left\{\tau_{x_j,p}\right\}_{1\le j\le 2S}$ and T_p yields:

$$J_{\mathbf{x}}(\boldsymbol{\tau}_p) = -M^{-1}J_{\mathbf{F}}(\boldsymbol{\tau}_p), \text{ and } J_{\mathbf{x}}(T_p) = -M^{-1}J_{\mathbf{F}}(T_p),$$

where $J_{\mathbf{x}}(\boldsymbol{\tau}_p), J_{\mathbf{F}}(\boldsymbol{\tau}_p)$ are the matrices $\left[\partial x_{\boldsymbol{\theta},i}/\partial\tau_{x_j,p}\right]_{1\le i,j\le 2S}$ and $\left[\partial F_i/\partial\tau_{x_j,p}\right]_{1\le i,j\le 2S}$ respectively, and $J_{\mathbf{F}}(T_p)$ is the vector $[\partial F_i/\partial T_p]_{1\le i\le 2S}$. But we have $J_{\mathbf{x}}(\boldsymbol{\tau}_p) = \boldsymbol{D}^{-1}(\mathbf{x})\,\boldsymbol{E}_p^{c,(\mathbf{x})}$ and $J_{\mathbf{x}}(T_p) = \boldsymbol{D}^{-1}(\mathbf{x})\,\boldsymbol{I}_p^{(\mathbf{x})}$. We use these expressions to write the deformation matrix $\boldsymbol{D}(\mathbf{x})$ as a function of the Jacobian matrix $J_{\mathbf{x}}(\boldsymbol{\theta})$, and $J_{\mathbf{F}}(\boldsymbol{\theta}), J_{\mathbf{F}}(\boldsymbol{\tau}_p), J_{\mathbf{F}}(T_p)$. Using the change of variables formula $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \det(J_{\mathbf{x}}(\boldsymbol{\theta}))\,f_{\mathbf{x}}(\mathbf{x}(\boldsymbol{\theta}))$ in the equation

$$\begin{aligned}0 =& \beta^{p-1}\left(1-g_p(\mathbf{x})\right) f_{\mathbf{x}}(\mathbf{x}) - \boldsymbol{T}'(\mathbf{x})\,\boldsymbol{D}^{-1}(\mathbf{x})\,\boldsymbol{I}_p^{(\mathbf{x})} f_{\mathbf{x}}(\mathbf{x})\\ &- \nabla_{\mathbf{x}}\cdot\left(\boldsymbol{T}'(\mathbf{x})\,\boldsymbol{D}^{-1}(\mathbf{x})\,\boldsymbol{E}_p^{c,(\mathbf{x})} f_{\mathbf{x}}(\mathbf{x})\right)\end{aligned} \tag{50}$$

and the chain rule, we obtain (19).

Now, consider the model with idiosyncratic productivities $\{\theta_1, \ldots, \theta_S\}$ and interest rates $\{\theta_{S+1}, \ldots, \theta_{2S-1}\}$. Let $\boldsymbol{l}_{\boldsymbol{\theta}}$ denote the vector of labor supplies $y_{\boldsymbol{\theta},s}/\theta_s$ and capital stocks $k_{\boldsymbol{\theta},s}$. We can write the first-order conditions of the individual problem as

$$\mathbf{x}_{\boldsymbol{\theta}} = \boldsymbol{\theta}\circ\boldsymbol{l}_{\boldsymbol{\theta}}\left(\left\{\theta_j\hat{\tau}_{y_j,s}\right\}_{\substack{1\le j\le 2S\\1\le s\le S}}, \left\{\theta_{S+j-1}\hat{\tau}_{z_j,s}\right\}_{\substack{2\le j\le S\\1\le s\le S}}, \left\{R_s\right\}_{1\le s\le S}\right),$$

where $\hat{\tau}_{x_j,s}$ is the marginal tax rate on income x_j (if $x_j \in \{y_s, z_s\}$) or the next-of-tax rate otherwise, and $\circ$ is the element-wise multiplication. Differentiating this system of equations with respect to θ_j for $1 \le j \le 2S$ yields the Jacobian matrix

$$J_{\mathbf{x}}(\boldsymbol{\theta}) = \boldsymbol{D}^{-1}(\mathbf{x})\left[\frac{\mathbf{x}}{\boldsymbol{\theta}} \circ \left(\mathrm{i}_{2S} + \sum_{s=1}^{S} \zeta_{\mathbf{x},\hat{q}_s}^{u,(\mathbf{x})}\right)\right],$$

where $(\mathbf{x}/\boldsymbol{\theta})$ denotes the matrix $[x_{\boldsymbol{\theta},i}/\theta_j]_{1\le i,j\le 2S}$, $\zeta_{\mathbf{x},\hat{q}_s}^{u,(\mathbf{x})}$ is the matrix of uncompensated elasticities with respect to the marginal and net-of-tax rates, and $\circ$ is the element-wise multiplication of matrices. Changing variables as before yields

$$J_{\mathbf{F}}^{-1}(\boldsymbol{\theta})\, J_{\mathbf{F}}(T_p) = -\left[\frac{\mathbf{x}}{\boldsymbol{\theta}} \circ \left(\mathrm{i}_{2S} + \sum_{s=1}^{S} \zeta_{\mathbf{x},\hat{q}_s}^{u,(\mathbf{x})}\right)\right]^{-1} \boldsymbol{I}_p^{(\boldsymbol{\theta})},$$

$$J_{\mathbf{F}}^{-1}(\boldsymbol{\theta})\, J_{\mathbf{F}}(\boldsymbol{\tau}_p) = \left[\frac{\mathbf{x}}{\boldsymbol{\theta}} \circ \left(\mathrm{i}_{2S} + \sum_{s=1}^{S} \zeta_{\mathbf{x},\hat{q}_s}^{u,(\mathbf{x})}\right)\right]^{-1} \boldsymbol{E}_p^{c,(\boldsymbol{\theta})}.$$

In particular, in the static Mirrlees model, the first-order condition (9) writes $F\left[\frac{x_\theta}{\theta}, \theta\left(1 - T'(x_\theta)\right), R(x_\theta)\right] \equiv F[l, \tau, R] = 0$ with $F[l,\tau,R] = \tau u_c(\tau l + R, l) + u_l(\tau l + R, l)$. It is then straightforward to compute $\frac{\partial F}{\partial l}$, $\frac{\partial F}{\partial \tau}$, and $\frac{\partial F}{\partial R}$. Note moreover that $J_{\mathbf{x}}(\boldsymbol{\theta}) = \det\left(J_{\mathbf{x}}(\boldsymbol{\theta})\right) = \dot{x}(\theta)$. Differentiating the first-order-condition with respect to θ then yields

$$\frac{\dot{x}_\theta}{x_\theta} = \frac{\frac{1}{\theta^2}\frac{\partial F}{\partial l} - \frac{1}{x_\theta}\left(1 - T'(x_\theta)\right)\frac{\partial F}{\partial \tau}}{\frac{1}{\theta}\frac{\partial F}{\partial l} + \left(-\theta\frac{\partial F}{\partial \tau} + \frac{\partial F}{\partial R}x_\theta\right)T''(x_\theta)} \Rightarrow \dot{x}_\theta^{-1}\frac{1}{1 + \frac{x_\theta \zeta_{x,1-\tau}^c}{1-T'(x_\theta)}T''(x_\theta)} = \frac{1}{\frac{x_\theta}{\theta}\left(1 + \zeta_{x,1-\tau}^u\right)}.$$

This expression is identical to that in Lemma 1 in Saez (2001). □

B Proofs of Sections 5 and 6

B.1 Proofs of Section 5

We start by deriving the formulas for the known results in the literature: optimal commodity taxes and non-linear labor income taxes.

Proofs of Sections 5.3 and 5.1. Formula (23) follows from using the Slutsky equation and rearranging the terms in equation (23). Formula (28) follows from (17) applied to the region $[\hat{y}, \infty)$, or directly from rearranging equation (18). Formula (30) follows from (28) under the assumptions made in the text. □

We now characterize the optimum linear capital income tax schedule.

Proof of Propositions 4 and 5. Consider a separable linear perturbation $h_p(\mathbf{x}) = z_p$ of the capital income tax rate *in every period* $p \geq 2$. The welfare effect of these perturbations, $\delta\mathscr{W}\left(\tau_z, \{h_p\}_{p\geq 2}\right)$, is given by the sum (for $p = 2, \ldots, S$) of the effects of each of the period-p perturbations h_p, $\delta\mathscr{W}(\tau_z, h_p)$. Applying Proposition 2, we obtain that the welfare effect of this perturbation is given by

$$\begin{aligned}
&\delta\mathscr{W}\left(\tau_z, \{h_p\}_{p\geq 2}\right)\\
&=\sum_{p=2}^{S}\left\{\int_{\mathbb{R}_+^S\times\mathbb{R}^S}\beta^{p-1}\left(1-g_p(\mathbf{x})\right)z_p f_{\mathbf{x}}(\mathbf{x})\,d\mathbf{x}+\int_{\mathbb{R}_+^S\times\mathbb{R}^S}\sum_{s=2}^{S}\beta^{s-1}\tau_z\left[\delta\mathbf{x}(\tau_z,h_p)\right]_{S+s-1}f_{\mathbf{x}}(\mathbf{x})\,d\mathbf{x}\right\}\\
&=\sum_{p=2}^{S}\left\{\beta^{p-1}\bar{z}_p\left(1-\mathbb{E}\left[g_p(\mathbf{x})\frac{z_p}{\bar{z}_p}\right]\right)-\int_{\mathbb{R}_+^S\times\mathbb{R}^S}\sum_{s=2}^{S}\beta^{s-1}\frac{\tau_z}{1-\tau_z}z_s\zeta^{u,(\mathbf{x})}_{z_s,r_p}f_{\mathbf{x}}(\mathbf{x})\,d\mathbf{x}\right\}\\
&=\left(\sum_{s=2}^{S}\beta^{s-1}\bar{z}_s\right)\times\sum_{p=2}^{S}\frac{\beta^{p-1}\bar{z}_p}{\sum_{s=2}^{S}\beta^{s-1}\bar{z}_s}\left\{1-\mathbb{E}\left[g_p\frac{z_p}{\bar{z}_p}\right]-\frac{\tau_z}{1-\tau_z}\sum_{s=2}^{S}\beta^{s-p}\bar{\zeta}^u_{z_s,r_p}\right\}.
\end{aligned}$$

Equating this expression to zero leads the optimal capital income tax rate (25). (Note that it would be straightforward to characterize the optimal *affine* tax schedule, by considering *revenue-neutral* perturbations of the capital income tax rate τ_z and the virtual income R (uniform lump-sum rebate of the tax revenue generated by the increase in the tax rate), and equating their effect to zero.)

Now consider the case where the perturbation is implemented in every period $p = p_1, \ldots, p_2$. Under the assumptions of Proposition **5**, the expressions (47) and (48) imply:

$$\begin{aligned}
\bar{\zeta}^c_{z_s,r_p} &= \sigma^{-1}(R-1)\begin{cases}R^{s-p-1}-R^{-p}, & \text{if } s\leq p,\\ R^{-1}-R^{-p}, & \text{if } s\geq p+1,\end{cases}\\
\bar{\eta}_{z_s,R_p} &= \qquad (R-1)\begin{cases}R^{-p}-R^{s-p-1}, & \text{if } s\leq p,\\ R^{-p}, & \text{if } s\geq p+1.\end{cases}
\end{aligned}$$

Hence, the compounded uncompensated elasticities are equal to:

$$\begin{aligned}
\sum_{s=2}^{\infty}\beta^{s-1}\bar{\zeta}^u_{z_s,r_2} &= \sigma^{-1}(1-\beta)\beta+(2\beta-1)\beta,\\
\sum_{p=2}^{\infty}\sum_{s=2}^{\infty}\frac{\beta^{s-1}}{\sum_{p=2}^{\infty}\beta^{p-1}}\bar{\zeta}^u_{z_s,r_p} &= \sigma^{-1}+\beta-1.
\end{aligned}$$

Result (26) follows. Moreover, we obtain

$$\beta^{-(P-1)}\sum_{p=P}^{\infty}\sum_{s=2}^{\infty}\beta^{s-1}\bar{\zeta}^u_{z_s,r_p} = \frac{\beta}{1-\beta}\sigma^{-1}+\left(\sigma^{-1}-1\right)(P-1),$$

from which (27) follows. Finally, for $S<\infty$ (still assuming $\beta R=1$), we have

$$\frac{\partial z_s}{\partial R_p}=\frac{r}{1-R^S}\begin{cases}R^{S+s-p-1}-R^{S-p}, & \text{if } s\leq p,\\ R^{s-p-1}-R^{S-p}, & \text{if } s\geq p+1,\end{cases}$$

which implies $\sum_{p=1}^{S}\sum_{s=1}^{S}\beta^{s-1}\frac{\partial z_s}{\partial R_p}=0$. □

We now prove the results of Section 5.4, i.e., the optimal non-linear, age-dependent, separable tax system in a two-period economy.

Proofs of formulas (31) and (32). Under the assumptions of this section, we have

$$\boldsymbol{T}'(\mathbf{x})\,\boldsymbol{D}^{-1}(\mathbf{x})\,\boldsymbol{E}_p^{c,(\mathbf{x})}=\begin{pmatrix} T_1'(y_1)\dfrac{-\frac{y_1}{1-\tau_{y_1,1}}\zeta_{y_1,w_1}^{c,(\mathbf{x}_{\boldsymbol{\theta}})}\mathbb{I}_{\{p=1\}}}{1+\frac{y_1}{1-\tau_{y_1,1}}\zeta_{y_1,w_1}^{c,(\mathbf{x}_{\boldsymbol{\theta}})}T_1''(y_1)}\\ \beta T_{y,2}'(y_2)\dfrac{-\frac{y_2}{1-\tau_{y_2,2}}\zeta_{y_2,w_2}^{c,(\mathbf{x}_{\boldsymbol{\theta}})}\mathbb{I}_{\{p=2\}}}{1+\frac{y_2}{1-\tau_{y_2,2}}\zeta_{y_2,w_2}^{c,(\mathbf{x}_{\boldsymbol{\theta}})}T_{y,2}''(y_2)}\\ \beta T_{z,2}'(z_2)\dfrac{-\frac{z_2}{1-\tau_{z_2,2}}\zeta_{z_2,r_2}^{c,(\mathbf{x}_{\boldsymbol{\theta}})}\mathbb{I}_{\{p=2\}}}{1+\frac{z_2}{1-\tau_{z_2,2}}\zeta_{z_2,r_2}^{c,(\mathbf{x}_{\boldsymbol{\theta}})}T_{z,2}''(z_2)}\end{pmatrix},$$

$$\boldsymbol{T}'(\mathbf{x})\,\boldsymbol{D}^{-1}(\mathbf{x})\,\boldsymbol{I}_p^{(\mathbf{x})}=\beta T_{z,2}'(z_2)\frac{1}{1+\frac{z_2}{1-\tau_{z_2,2}}\zeta_{z_2,r_2}^{c,(\mathbf{x}_{\boldsymbol{\theta}})}T_{z,2}''(z_2)}\frac{\eta_{z_2,R_p}^{(\mathbf{x}_{\boldsymbol{\theta}})}}{1-\tau_{z_2,p}},$$

and $\overrightarrow{\mathbf{n}}(\mathbf{x})$ is the 3-vector whose only non-zero component is equal to 1 and is in the first (resp., second, third) row if $\hat{x}=\hat{y}_1$ (resp., $\hat{y}_2,\hat{z}$). Application of formula (17) to the region $\mathrm{V}=[\hat{x},\infty)\times\mathbb{R}^2$, for $\hat{x}\in\{\hat{y}_1,\hat{y}_2,\hat{z}_2\}$ and $p=1,2,2$ respectively, and dividing by $1-F_{x_p}(\hat{x})$, yields formulas (31) and (32). □

Next, we derive the optimal asymptotic capital income tax rate in a non-linear tax system.

Optimal Asymptotic Capital Income Tax Rate. Assume that the baseline tax system is separable and age-independent, but non-linear in capital income. For simplicity, we also assume that the distribution of capital income is stationary, and that it is Pareto distributed at the tail with coefficient a_z. Here we let the utility function have income effects on labor supply, and assume that the labor income tax rate τ_y is constant and age-independent. Next, we assume the convergence toward constants of the (average) marginal social welfare weights, $\mathbb{E}_{z_p\geq\hat{z}}\left[g_p\frac{z_p}{\hat{z}}\right]\xrightarrow[\hat{z}\to\infty]{}\bar{g}_p^{(\infty)}$, of the elasticities, $\bar{\zeta}_{z_s,r_p}^{u,(z_p\geq\hat{z})}\xrightarrow[\hat{z}\to\infty]{}\bar{\zeta}_{z_s,r_p}^{u,(\infty)}$ and $\bar{\eta}_{z_s,R_p}^{(z_p\geq\hat{z})}\xrightarrow[\hat{z}\to\infty]{}\bar{\eta}_{z_s,R_p}^{(\infty)}$, and of the marginal tax rates at the top of the capital income distribution, $T_z'(z)\xrightarrow[z\to\infty]{}\tau_z^{\infty}$. Moreover, we assume that $T_z''(\cdot)$ converges to zero fast enough, i.e., for all $p\geq 2$,

$$\sup_{\{\mathbf{x}:z_p\geq\hat{z}\}}\left|z_s\zeta_{z_s,r_p}^{c,(\mathbf{x})}T_z''(z_p)\right|\xrightarrow[\hat{z}\to\infty]{}0.$$

Finally, we assume that there is "no mobility at the top": as the threshold capital income level $\hat{z} \to \infty$, individuals with capital income $z_s \geq \hat{z}$ in a given period s have capital income $z_p \geq \hat{z}$ in all periods $p \geq 2$. Intuitively, individuals at the top of the capital income distribution in a given period stay there forever. This ensures that as $z_p \to \infty$ for any p, all the components of the matrix $\boldsymbol{F}_z(\mathbf{x})$ (defined below) converge to zero, and that all the marginal tax rates $T'_z(z_s)$ converge to τ_z^∞.

Consider a sequence, indexed by $\hat{z} > 0$, of separable perturbations of the capital income tax rate in every period $p \geq 2$, that are linear above the threshold $\hat{z}$. That is, for all p we define $h_p(\mathbf{x}) = \max\{z_p - \hat{z}, 0\}$. The Gateaux differential of social welfare writes:

$$
\begin{aligned}
\delta\mathscr{W}\left(\mathscr{T}, \{h_p\}_{p\geq 2}\right) &= \sum_{p=2}^{S} \beta^{p-1} \left\{ \int_{\hat{z}}^{\infty} (1 - g_p(\mathbf{x}))(z_p - \hat{z}) f_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x} \right\} \\
&+ \sum_{p=2}^{S} \left\{ \int_{\hat{z}}^{\infty} \int_{\mathbb{R}_+^S \times \mathbb{R}^{S-2}} \left[\boldsymbol{T}'_z(\mathbf{x})\right] \left[\mathfrak{i}_{S-1} + \boldsymbol{F}_z(\mathbf{x})\right]^{-1} \left[\boldsymbol{E}_{p,z_p}^{c,(\mathbf{x})} - (z_p - \hat{z}) \boldsymbol{I}_p^{(\mathbf{x})}\right] f_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x} \right\} \\
&+ \sum_{p=2}^{S} \left\{ \int_{\hat{z}}^{\infty} \int_{\mathbb{R}_+^S \times \mathbb{R}^{S-2}} \left[\tau_y \sum_{s=2}^{S} \beta^{s-1} \left(-\frac{y_s \zeta_{y_s,r_p}^{c,(\mathbf{x})}}{1 - \tau_{z_p,p}} - \frac{\eta_{y_s,R_p}^{(\mathbf{x})}}{1 - \tau_{y_s,p}} (z_p - \hat{z}) \right) \right] f_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x} \right\},
\end{aligned}
$$

where $[\boldsymbol{T}'_z(\mathbf{x})]$ is the $(S-1)$-row vector with components $\beta^{s-1} T'_z(z_s)$ and $[\boldsymbol{F}_z(\mathbf{x})]$ is the $(S-1) \times (S-1)$-matrix with components $T''_z(z_j) \frac{\partial z_i^c}{\partial \tau_{j,z_j}}$ for $i, j \geq 2$. Thus, letting $\hat{z} \to \infty$ and imposing $\lim \frac{\delta\mathscr{W}(T, \{h_p\}_{p\geq 2})}{(1-F(\hat{z}))\hat{z}} = 0$, we obtain the following characterization of the optimal asymptotic capital income tax rate:

$$
\begin{aligned}
\frac{\tau_z^\infty}{1 - \tau_z^\infty} =& \frac{\left(\frac{a_z}{a_z - 1} - 1\right)\left(1 - \sum_{p=2}^{S} \gamma_{p,z} \bar{g}_{p,\mathtt{NL}}^{(\infty)}\right)}{\frac{a_{rk}}{a_{rk}-1} \sum_{p,s=2}^{S} \gamma_{p,z} \bar{\zeta}_{z_s,r_p}^{u,(\infty)} - \sum_{p,s=2}^{S} \gamma_{p,z} \bar{\eta}_{z_s,R_p}^{(\infty)}} \\
& - \frac{\tau_y}{1 - \tau_z^\infty} \frac{\frac{a_z}{a_z-1} \sum_{p,s=2}^{S} \gamma_{p,z} \bar{\zeta}_{y_s,r_p}^{u,(\infty)} - \sum_{p,s=2}^{S} \gamma_{p,z} \bar{\eta}_{y_s,R_p}^{(\infty)}}{\frac{a_z}{a_z-1} \sum_{p,s=2}^{S} \gamma_{p,z} \bar{\zeta}_{z_s,r_p}^{u,(\infty)} - \sum_{p,s=2}^{S} \gamma_{p,z} \bar{\eta}_{z_s,R_p}^{(\infty)}},
\end{aligned}
$$

where $\gamma_{p,z} = \beta^{p-1} / \sum_{s=2}^{S} \beta^{s-1}$. Note that in the case where the utility function has no income effects on labor supply, the second line of this expression is equal to zero. □

We now prove the result of Section 5.5, i.e., the joint taxation of couples.

Proof of formula (33). We follow the same steps as in the derivation of formula (19). Letting $\tau_i \equiv \frac{\partial T}{\partial y_i}$ and $\tau_{ij} \equiv \frac{\partial^2 T}{\partial y_i \partial y_j}$, the Gateaux differential of individual income in a direction h,

$\delta \mathbf{y}_{\boldsymbol{\theta}}(T,h)$, writes

$$\begin{aligned}\delta \mathbf{y}_{\boldsymbol{\theta}}(T,h) =& \frac{1}{(1-\tau_1+y_1\zeta\tau_{11})(1-\tau_2+y_2\zeta\tau_{22})-y_1y_2\zeta^2\tau_{12}^2} \\ & \times \begin{pmatrix} -(1-\tau_2+y_2\zeta\tau_{22})\,y_1\zeta & (y_1\zeta\tau_{12})\,y_2\zeta \\ (y_2\zeta\tau_{12})\,y_1\zeta & -(1-\tau_1+y_1\zeta\tau_{11})\,y_2\zeta \end{pmatrix} \begin{pmatrix} \frac{\partial h(\mathbf{y}_{\boldsymbol{\theta}})}{\partial y_1} \\ \frac{\partial h(\mathbf{y}_{\boldsymbol{\theta}})}{\partial y_2} \end{pmatrix}.\end{aligned}$$

Applying formula (50), we obtain that the revenue-maximizing tax function satisfies the following PDE:

$$\begin{aligned} f_{\mathbf{y}}(y_1,y_2) =& \frac{\partial}{\partial y_1}\left\{\frac{-(\tau_1y_1\zeta)(1-\tau_2+y_2\zeta\tau_{22})+(\tau_{12}y_1\zeta)(\tau_2y_2\zeta)}{(1-\tau_1+y_1\zeta\tau_{11})(1-\tau_2+y_2\zeta\tau_{22})-(\tau_{12}y_1\zeta)(\tau_{12}y_2\zeta)}f_{\mathbf{y}}(y_1,y_2)\right\} \\ &+\frac{\partial}{\partial y_2}\left\{\frac{-(1-\tau_1+y_1\zeta\tau_{11})(\tau_2y_2\zeta)+(\tau_1y_1\zeta)(\tau_{12}y_2\zeta)}{(1-\tau_1+y_1\zeta\tau_{11})(1-\tau_2+y_2\zeta\tau_{22})-(\tau_{12}y_1\zeta)(\tau_{12}y_2\zeta)}f_{\mathbf{y}}(y_1,y_2)\right\}.\end{aligned}$$

To rewrite the PDE in terms of the distribution of types, first notice that in this model, the incomes as functions of types are given by:

$$\begin{aligned} y_1 =& \theta_1^{1+\zeta}(1-\tau_1)^{\zeta}, \\ y_2 =& \theta_2^{1+\zeta}(1-\tau_2)^{\zeta}. \end{aligned} \tag{51}$$

and the Jacobian matrix $J_{\mathbf{y}}(\boldsymbol{\theta})$ writes:

$$\begin{aligned}\begin{pmatrix} \frac{\partial y_1}{\partial \theta_1} & \frac{\partial y_1}{\partial \theta_2} \\ \frac{\partial y_2}{\partial \theta_1} & \frac{\partial y_2}{\partial \theta_2}\end{pmatrix} =& \frac{1}{(1-\tau_1+\zeta y_1\tau_{11})(1-\tau_2+\zeta y_2\tau_{22})-(\zeta y_1\tau_{12})(\zeta y_2\tau_{12})} \\ \times \begin{pmatrix} (1-\tau_1)\left(1-\tau_2+\left(\frac{1+\zeta}{\theta_1}y_1\right)(\zeta y_2\tau_{22})\right) & -(1-\tau_2)(\zeta y_1\tau_{12})\left(\frac{1+\zeta}{\theta_2}y_2\right) \\ -(1-\tau_1)(\zeta y_2\tau_{12})\left(\frac{1+\zeta}{\theta_1}y_1\right) & (1-\tau_2)\left(1-\tau_1+(\zeta y_1\tau_{11})\left(\frac{1+\zeta}{\theta_2}y_2\right)\right)\end{pmatrix}&, \end{aligned} \tag{52}$$

Therefore, we have

$$\delta \mathbf{y}_{\boldsymbol{\theta}}(T,h) = \begin{pmatrix} -\frac{\theta_1}{1-\tau_1}\frac{\zeta}{1+\zeta}\frac{\partial y_1}{\partial \theta_1} & -\frac{\theta_2}{1-\tau_2}\frac{\zeta}{1+\zeta}\frac{\partial y_1}{\partial \theta_2} \\ -\frac{\partial y_2}{\partial \theta_1}\frac{\theta_1}{1-\tau_1}\frac{\zeta}{1+\zeta} & -\frac{\theta_2}{1-\tau_2}\frac{\zeta}{1+\zeta}\frac{\partial y_2}{\partial \theta_2}\end{pmatrix}\begin{pmatrix} \frac{\partial h(\mathbf{y}_{\boldsymbol{\theta}})}{\partial y_1} \\ \frac{\partial h(\mathbf{y}_{\boldsymbol{\theta}})}{\partial y_2}\end{pmatrix}.$$

The optimal tax system is thus characterized by:

$$\begin{aligned} 0 =& (1-g(\mathbf{y}))f_{\mathbf{y}}(\mathbf{y}) - \nabla_{\mathbf{y}}\cdot\begin{pmatrix}\left\{-\frac{\tau_1}{1-\tau_1}\frac{\zeta}{1+\zeta}\theta_1\frac{\partial y_1}{\partial \theta_1}-\frac{\tau_2}{1-\tau_1}\frac{\zeta}{1+\zeta}\theta_1\frac{\partial y_2}{\partial \theta_1}\right\}f_{\mathbf{y}}(\mathbf{y}) \\ \left\{-\frac{\tau_1}{1-\tau_2}\frac{\zeta}{1+\zeta}\theta_2\frac{\partial y_1}{\partial \theta_2}-\frac{\tau_2}{1-\tau_2}\frac{\zeta}{1+\zeta}\theta_2\frac{\partial y_2}{\partial \theta_2}\right\}f_{\mathbf{y}}(\mathbf{y})\end{pmatrix}' \\ =& (1-g(\boldsymbol{\theta}))f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) + \frac{\zeta}{1+\zeta}\sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\partial y_{-i}}{\partial \theta_{-j}}\frac{\partial}{\partial \theta_j}\left\{\frac{\frac{\tau_1}{1-\tau_i}\frac{\partial y_1}{\partial \theta_i}+\frac{\tau_2}{1-\tau_i}\frac{\partial y_2}{\partial \theta_i}}{\frac{\partial y_1}{\partial \theta_1}\frac{\partial y_2}{\partial \theta_2}-\frac{\partial y_1}{\partial \theta_2}\frac{\partial y_2}{\partial \theta_1}}\theta_i f_{\boldsymbol{\theta}}(\boldsymbol{\theta})\right\}, \end{aligned} \tag{53}$$

where the second equality follows from the change of variables from $\mathbf{y}$ to $\boldsymbol{\theta}$. Equations (51), (52), and (53) form a PDE system whose solution is the optimal tax system. □

B.2 Proofs of Section 6

We now provide the proofs of the results of Sections 6.1 and 6.2. We first characterize the welfare effects of reforming the labor income tax system.

Proofs of formulas (37) and (36). Consider first the effects of reforming the marginal labor income tax rate at point $\hat{y}$ in period p. The perturbation we consider is as follows. We choose the numbers $\hat{y} > 0$, , and define for every p a perturbation $\tilde{h}_p \in \mathcal{C}^2(\mathbb{R}_+)$, with $\tilde{h}_p(y) = 0$ on $[0, \hat{y}]$, $\tilde{h}_p(y) = (y - \hat{y})$ on $[\hat{y}, \hat{y}']$, and $\tilde{h}_p(y) = (\hat{y}' - \hat{y})$ on $[\hat{y}', \infty)$. We obtain a smooth perturbation h_p from $\tilde{h}_p$ by letting $h_p = \tilde{h}_p$ except on the intervals $\left[\hat{y} - \frac{u}{2}, \hat{y} + \frac{u}{2}\right]$ and $\left[\hat{y}' - \frac{u}{2}, \hat{y}' + \frac{u}{2}\right]$, for some small $u > 0$, where we take h_p monotonic. We then consider a sequence $\left\{h_p^n\right\}_{n\in\mathbb{N}}$ of such perturbations, with $(\hat{y}' - \hat{y}) \to 0$, $u \to 0$, and $u = o(\hat{y}' - \hat{y})$. Applying formula (8), we obtain that the effect of this perturbation h_p on the individual income choices is given by

$$\delta y_{\boldsymbol{\theta},p}^{(\mathbf{x})}\left(T_p, h_p^n\right) = -\frac{y_p \zeta_{y_p,w_p}^{c,(\mathbf{x})}}{1 - T_y'(y_p) + y_p \zeta_{y_p,w_p}^{c,(\mathbf{x})} T_y''(y_p)}, \text{ for all } y_p \in \left[\hat{y} + \frac{u}{2}, \hat{y}' - \frac{u}{2}\right],$$

$$\delta z_{\boldsymbol{\theta},s}^{(\mathbf{x})}\left(T_p, h_p^n\right) = -\left(\hat{y}' - \hat{y}\right) \frac{\eta_{z_s,R_p}^{(\mathbf{x})}}{1 - \tau_{z_s,p}}, \text{ for all } s, \text{ for all } y_p \geq \hat{y}' + \frac{u}{2}.$$

Applying formula (12) and taking the limit of the Gateaux differentials of social welfare as $(\hat{y}' - \hat{y}), \Delta\tau \to 0$, we get

$$\begin{aligned}
&\frac{1}{1 - F_{y_p}(\hat{y})} \frac{\delta \mathscr{W}\left(T_p, h_p^n\right)}{(\hat{y}' - \hat{y})} \\
\xrightarrow[n\to\infty]{} \quad & \beta^{p-1} \int_{\hat{y}}^{\infty} (1 - g_p(y)) \frac{f_{y,p}(y)}{1 - F_{y_p}(\hat{y})} dy \\
& - \beta^{p-1} \int_{\mathbb{R}_+^{S-1} \times \mathbb{R}^{S-1}} T_y'(\hat{y}) \frac{\hat{y} \zeta_{y_p,w_p}^{c,(\hat{y},\mathbf{x}_{-p})}}{1 - T_y'(\hat{y}) + \hat{y} \zeta_{y_p,w_p}^{c,(\hat{y},\mathbf{x}_{-p})} T_y''(\hat{y})} \frac{f_{\mathbf{x}}(\hat{y}, \mathbf{x}_{-p})}{1 - F_{y_p}(\hat{y})} d\mathbf{x}_{-p} \\
& - \sum_{s=2}^{S} \beta^{s-1} \int_{\hat{y}}^{\infty} \int_{\mathbb{R}_+^{S-1} \times \mathbb{R}^{S-1}} \tau_z \frac{\eta_{z_s,R_p}^{(\mathbf{x})}}{1 - \tau_{z_s,p}} \frac{f_{\mathbf{x}}(\mathbf{x})}{1 - F_{y_p}(\hat{y})} d\mathbf{x}.
\end{aligned}$$

Noting that $\zeta_{y_p,w_p}^{c,(\hat{y},\mathbf{x}_{-p})} = \zeta$, that $\eta_{z_s,R_p}^{(\mathbf{x})}$ is independent of $\mathbf{x}$ since the utility function is CRRA and the baseline tax system is separable, and using the definition of $\bar{\eta}_{z_s,R_p}$, we obtain

$$\begin{aligned}
\frac{1}{1 - F_{y_p}(\hat{y})} \frac{\delta \mathscr{W}\left(T_p, h_p^n\right)}{(\hat{y}' - \hat{y})} \xrightarrow[n\to\infty]{} \quad & \beta^{p-1}\left(1 - \mathbb{E}_{y\geq\hat{y}}[g_p]\right) - \frac{\tau_z}{1 - \tau_z} \sum_{s=2}^{S} \beta^{s-1} \bar{\eta}_{z_s,R_p} \\
& - \beta^{p-1} \frac{T_y'(\hat{y})\zeta}{1 - T_y'(\hat{y}) + \hat{y}\zeta T_y''(\hat{y})} \frac{\hat{y} f_{y_p}(\hat{y})}{1 - F_{y_p}(\hat{y})}.
\end{aligned}$$

Summing over periods p and normalizing by $\sum_{p\geq 1} \beta^{p-1}\left(1 - F_{y_p}(\hat{y})\right)$ yields (37). Finally, to obtain the gains of age-dependence assuming that the income distribution is stationary, define the "savings effect" as $\mathscr{S}_p \equiv -\frac{\tau_z}{1-\tau_z}\sum_{s=2}^{S}\beta^{s-1}\bar{\eta}_{z_s,R_p}$. Since $\partial z_s/\partial R_p < 0$ for all $s \leq p$ and $\partial z_s/\partial R_p > 0$ for all $s \geq p+1$, we obtain that the sequence $\left\{\beta^{-(p-1)}\mathscr{S}_p : p = 1, \ldots, S\right\}$ is increasing, with $\mathscr{S}_1 < 0$ and $\mathscr{S}_S > 0$. Hence there exists p^* such that the revenue gains of the period-p separable perturbation are strictly smaller (resp., larger) in the dynamic model than in the static model for $p \leq p^*$ (resp., $p > p^*$). Moreover, the revenue gains of the separable perturbation that increases lump-sum the tax liability above $\hat{y}$ by \$1 in period p are smaller than the gains from the perturbation that increases the tax liability above β by $\$\beta^{-(p'-p)}$ in period $p' > p$, yielding gains from age-dependent taxes.)

Suppose that the baseline marginal labor income tax rate is constant, i.e., $T_y'(\hat{y}) = \tau_y$ for all $\hat{y}$. Then we obtain the following relationship between the linear and the non-linear tax reforms:

$$\begin{aligned}
&\int_0^\infty \delta\mathscr{W}(\mathscr{T}, h_{\hat{y}})\, d\hat{y} \\
=&\int_0^\infty \left\{\sum_{p=1}^{S}\beta^{p-1}\left(\left(1 - F_{y_p}(\hat{y})\right) - \int_{\hat{y}}^\infty g_p(y)\, f_{y_p}(y)\, dy\right.\right. \\
&\qquad \left.\left. - \frac{\tau_y}{1-\tau_y}\zeta\hat{y} f_{y_p}(\hat{y}) - \frac{\tau_z}{1-\tau_{p,z_s}}\sum_{s=2}^{S}\beta^{s-1}\eta_{z_s,R_p}\left(1 - F_{y_p}(\hat{y})\right)\right)\right\} d\hat{y} \\
=&\sum_{p=1}^{S}\beta^{p-1}\left(\int_0^\infty \hat{y} f_{y_p}(\hat{y})\, d\hat{y}\right)\left(1 - \frac{\int_0^\infty \hat{y} g_p(\hat{y})\, f_{y_p}(\hat{y})\, d\hat{y}}{\int_0^\infty \hat{y} f_{y_p}(\hat{y})\, d\hat{y}} - \frac{\tau_y}{1-\tau_y}\zeta - \frac{\tau_z}{1-\tau_{z_s,p}}\sum_{s=2}^{S}\beta^{s-1}\eta_{z_s,R_p}\right) \\
=&\delta\mathscr{W}(\mathscr{T}, h_L),
\end{aligned}$$

which proves (36). □

We next characterize the welfare effects of reforming the capital income tax schedule.

Proof of formula (35). A reasoning identical to that leading to formula (37), noting that the elasticities $\zeta_{y_s,r_p}^{c,(\mathbf{x})}$ are all equal to zero, shows that the welfare effect of a non-linear perturbation $h_{p,\hat{z}}$ implemented at point $\hat{z}$ in period p is equal to

$$\begin{aligned}
\frac{1}{1-F_{z_p}(\hat{z})}\delta\mathscr{W}(\mathscr{T}, h_{p,\hat{z}}) = \quad & \beta^{p-1}\int_{\hat{z}}^\infty \left(1 - g_p(z)\right)\frac{f_{z_p}(z)}{1-F_{z_p}(\hat{z})}dz \\
& - \sum_{s=2}^{S}\beta^{s-1}\int_{\mathbb{R}_+^S\times\mathbb{R}^{S-2}} \frac{\tau_z}{1-\tau_z} z_s \zeta_{z_s,r_p}^{c,\left(\hat{z},\mathbf{x}_{-(S+p-1)}\right)} \frac{f_{\mathbf{x}}\left(\hat{z},\mathbf{x}_{-(S+p-1)}\right)}{1-F_{z_p}(\hat{z})} d\mathbf{x}_{-(S+p-1)} \\
& - \sum_{s=2}^{S}\beta^{s-1}\int_{\hat{z}}^\infty\int_{\mathbb{R}_+^S\times\mathbb{R}^{S-2}} \frac{\tau_z}{1-\tau_{z_s,p}}\eta_{z_s,R_p}^{(\mathbf{x})}\frac{f_{\mathbf{x}}(\mathbf{x})}{1-F_{z_p}(\hat{z})}d\mathbf{x}.
\end{aligned}$$

Formula (35) follows. □

We now characterize the welfare effects of joint tax reforms.

Proof of formulas (40) and (38). Fix $d \geq 2$ directions $(x_1, \ldots, x_d)$ of the space $\mathbb{R}_+^S \times \mathbb{R}^S$ and the income threshold $\bar{\mathbf{x}}_d = (\hat{x}_1, \ldots, \hat{x}_d)$. We define the d-multilinear perturbation h_p of the baseline tax function T_p as $h_p(\mathbf{x}) = 0$ if $x_j \leq \hat{x}_j$ for all $j \in \{1, \ldots, d\}$, $h_p(\mathbf{x}) = (x_i - \hat{x}_i)\, d\tau$ if $x_i \in [\hat{x}_i, \hat{x}_i + d\hat{x}]$ for some $i \in \{1, \ldots, d\}$ *and* $x_j \geq \hat{x}_j + d\hat{x}$ for all $j \in \{1, \ldots, d\} \setminus \{i\}$, and $h_p(\mathbf{x}) = d\tau d\hat{x}$ if $x_j \geq \hat{x}_j + d\hat{x}$ for all $j \in \{1, \ldots, d\}$. We complete this definition on the remaining regions of the space (hypercubes of size $d\hat{x}$) by making h_p continuous and multilinear on each of these regions, e.g., for $d = 2$, $h_p(x_1, x_2)$ is of the form $c_{12}(x_1 - \hat{x}_1)(x_2 - \hat{x}_2)$. (More precisely, we consider a smooth approximation of these perturbations, as in Section 6.1.) For simplicity, we let $d = 2$ and consider a joint perturbation in period two in the directions (y_1, y_2), at point $(\hat{y}_1, \hat{y}_2)$. Note that

$$\zeta_{y_1, \tau_{21}}\big|_{\tau_{21}=0} = -\frac{\zeta}{1 + (1 - \tau_z)\, r} = -\beta\zeta.$$

Let $\bar{F}_{y_1,y_2}(\hat{y}_1, \hat{y}_2)$ denote the measure of individuals above $(\hat{y}_1, \hat{y}_2)$. Applying our general formula yields:

$$\begin{aligned}
\frac{\delta\mathscr{W}\left(\mathscr{T}, h_{2,(\hat{y}_1,\hat{y}_2)}\right)}{\beta\bar{F}_{y_1,y_2}(\hat{y}_1, \hat{y}_2)} =& 1 - \int_{\hat{y}_1}^{\infty}\int_{\hat{y}_2}^{\infty} g_2(y_1, y_2)\, \frac{f_{y_1,y_2}(y_1, y_2)}{\bar{F}_{y_1,y_2}(\hat{y}_1, \hat{y}_2)} dy_1 dy_2 \\
&- \sum_{s=2}^{S} \beta^{s-2} \int_{\hat{y}_1}^{\infty}\int_{\hat{y}_2}^{\infty} \tau_z \frac{\eta_{z_s,R_2}}{1 - \tau_{2,z_s}} \frac{f_{y_1,y_2}(y_1, y_2)}{\bar{F}_{y_1,y_2}(\hat{y}_1, \hat{y}_2)} dy_1 dy_2 \\
&- \int_{\hat{y}_2}^{\infty} \beta^{-1} T_1'(\hat{y}_1)\, \frac{\hat{y}_1\beta\zeta}{1 - T_1'(\hat{y}_1) + \hat{y}_1\beta\zeta T_1''(\hat{y}_1)} \frac{f_{y_1,y_2}(\hat{y}_1, y_2)}{\bar{F}_{y_1,y_2}(\hat{y}_1, \hat{y}_2)} dy_2 \\
&- \int_{\hat{y}_1}^{\infty} T_2'(\hat{y}_2)\, \frac{\hat{y}_2\zeta}{1 - T_2'(\hat{y}_2) + \hat{y}_2\zeta T_2''(\hat{y}_2)} \frac{f_{y_1,y_2}(y_1, \hat{y}_2)}{\bar{F}_{y_1,y_2}(\hat{y}_1, \hat{y}_2)} dy_1.
\end{aligned}$$

Using the definitions of the conditional hazard rates, we obtain formula (38). We similarly obtain the expression for $\frac{\delta\mathscr{W}\left(\mathscr{T}, h_{2,(\hat{y}_2,\hat{z}_2)}\right)}{\beta\bar{F}_{y_2,z_2}(\hat{y}_2,\hat{z}_2)}$, i.e., formula (38), the only difference being the compounding of the capital income elasticities. □

We finally show some results about the Clayton copula, used in equation (39).

Generalized Clayton copula. The generalized Clayton copula with correlation parameters (d, ρ), with $d \geq 1$ and $\rho \in (0, \infty)$, is defined as

$$C(u, v) = \left\{ \left[\left(u^{-1/\rho} - 1\right)^d + \left(v^{-1/\rho} - 1\right)^d \right]^{1/d} + 1 \right\}^{-\rho}.$$

Kendall's tau[30] and the coefficients of lower and upper tail dependence are given by:

$$\rho_\tau = 1 - \frac{2}{\left(2+\frac{1}{\rho}\right)d}, \;\; \lambda_l = \lim_{q\to 0}\frac{C(q,q)}{q} = 2^{-\rho/d}, \;\; \lambda_u = 2 + \lim_{q\to 0}\frac{C(1-q,1-q)-1}{q} = 2-2^{1/d}.$$

If the marginal distributions are Pareto distributed, $\bar{F}_{x_j}(x_j) = \alpha_j\left(\frac{x_j}{c_j}\right)^{-a_j}$, the log-survival c.d.f. obtained from the generalized Clayton copula writes

$$\ln \bar{F}_{x_1,x_2}(x_1,x_2) = -\rho \ln\left\{1+\left[\left(\alpha_1^{-1/\rho}\left(\frac{x_1}{c_1}\right)^{a_1/\rho}-1\right)^d + \left(\alpha_2^{-1/\rho}\left(\frac{x_2}{c_2}\right)^{a_2/\rho}-1\right)^d\right]^{1/d}\right\}.$$

In the case where $d=1$, the i^{th} component of multivariate hazard ratio vector (for $i=1,2$) is equal to:

$$-\hat{y}_i\frac{\partial \ln \bar{F}_{y_1,y_2}(\hat{y}_1,\hat{y}_2)}{\partial y_i} = \frac{\hat{y}_i\int_{\hat{y}_{-i}}^{\infty} f_{y_1,y_2}(\hat{y}_i,y_{-i})\,dy_{-i}}{\bar{F}_{y_1,y_2}(\hat{y}_1,\hat{y}_2)} = \frac{a_i\left[\bar{F}_{y_i}(\hat{y}_i)\right]^{-1/\rho}}{\left[\bar{F}_{y_1}(\hat{y}_1)\right]^{-1/\rho}+\left[\bar{F}_{y_2}(\hat{y}_2)\right]^{-1/\rho}-1}.$$

□

C Notations for the Stochastic Model

In the stochastic model outlined in Section 7 (see Golosov, Tsyvinski, and Werquin 2014), we define the marginal tax rates and virtual incomes in period one as

$$\begin{aligned}\tau_{1,x_j} &\equiv \frac{\partial T_1(y_1,k_1)}{\partial x_j}, \quad \forall x_j \in \{y_1,k_1\},\\ R_1 &\equiv \tau_{1,y_1}y_1 + \tau_{1,k_1}k_1 + k_0 - T_1(y_1,k_1),\end{aligned}$$

and in period two as

$$\begin{aligned}\tau_{2,x_j}\left(y_1,\mathbf{x}_2^2,z_2\right) &\equiv \frac{\partial T_2\left(y_1,\mathbf{x}_2^2,z_2\right)}{\partial x_j}, \quad \forall x_j \in \{y_1,y_2,z_2\},\\ R_2\left(y_1,\mathbf{x}_2^2,z_2\right) &\equiv \tau_{2,y_1}y_1 + \tau_{2,y_2}\mathbf{x}_2^2 + \tau_{2,z_2}z_2 - T_2\left(y_1,\mathbf{x}_2^2,z_2\right).\end{aligned}$$

[30] Kendall's tau is defined as follows. Consider two random variables $\tilde{x}_1,\tilde{x}_2$, independent of x_1,x_2, but with the same joint distribution. Then $\rho_\tau(x_1,x_2) \equiv \mathbb{E}\left[\text{sign}\left((x_1-\tilde{x}_1)\cdot(x_2-\tilde{x}_2)\right)\right]$.

We define the choice vector $\mathbf{x}$ of an individual with type (k_0, θ_1) in period one as:

$$\mathbf{x}_{(k_0,\theta_1)} = \begin{pmatrix} y_1(k_0,\theta_1) \\ y_2(k_0,\theta_1,\underline{\theta}_2) \\ \vdots \\ y_2(k_0,\theta_1,\theta_2) \\ \vdots \\ y_2\left(k_0,\theta_1,\bar{\theta}_2\right) \\ k_1(k_0,\theta_1) \end{pmatrix}.$$

Note that this vector has a continuum of interior rows, corresponding to all the possible values for $\theta_2 \in \left[\underline{\theta}_2, \bar{\theta}_2\right]$. It will also be the case for the matrices that we define below. However, we show that all the usual operations on vectors and matrices generalize naturally to this case.

We define the vector of income effect parameters as

$$\boldsymbol{I}_{R_1}^{(\mathbf{x})} = \begin{pmatrix} \partial y_1/\partial R_1 \\ \partial y_2(\underline{\theta}_2)/\partial R_1 \\ \vdots \\ \partial y_2\left(\bar{\theta}_2\right)/\partial R_1 \\ \partial z_2/\partial R_1 \end{pmatrix}, \quad \boldsymbol{I}_{R_2(\mathbf{x}^2)}^{(\mathbf{x})} = \begin{pmatrix} \partial y_1/\partial R_2\left(\mathbf{x}^2\right) \\ \partial y_2(\underline{\theta}_2)/\partial R_2\left(\mathbf{x}^2\right) \\ \vdots \\ \partial y_2\left(\bar{\theta}_2\right)/\partial R_2\left(\mathbf{x}^2\right) \\ \partial z_2/\partial R_2\left(\mathbf{x}^2\right) \end{pmatrix}.$$

We define the matrix of compensated elasticities with respect to the first-period marginal tax rates $\tau_{1,y_1}, \tau_{1,k_1}$ as

$$\boldsymbol{E}_{\tau_1}^{c,(\mathbf{x})} = \begin{pmatrix} \partial y_1/\partial \tau_{1,y_1} & 0 & \cdots & 0 & \partial y_1/\partial \tau_{1,k_1} \\ \partial y_2(\underline{\theta}_2)/\partial \tau_{1,y_1} & 0 & \cdots & 0 & \partial y_2(\underline{\theta}_2)/\partial \tau_{1,k_1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \partial y_2\left(\bar{\theta}_2\right)/\partial \tau_{1,y_1} & 0 & \cdots & 0 & \partial y_2\left(\bar{\theta}_2\right)/\partial \tau_{1,k_1} \\ \partial z_2/\partial \tau_{1,y_1} & 0 & \cdots & 0 & \partial z_2/\partial \tau_{1,k_1} \end{pmatrix},$$

and the matrix of compensated elasticities with respect to the second-period marginal tax rates $\tau_{2,y_1}\left(\mathbf{x}^2\right), \tau_{2,y_2}\left(\mathbf{x}^2\right), \tau_{2,z_2}\left(\mathbf{x}^2\right)$, at point $\mathbf{x}^2 = \left(\mathbf{x}_1^2, \mathbf{x}_2^2, \mathbf{x}_3^2\right) = \left(y_1, \mathbf{x}_2^2, z_2\right)$, as

$$\boldsymbol{E}_{\tau_2(\mathbf{x}^2)}^{c,(\mathbf{x})} = \begin{pmatrix} \frac{\partial y_1}{\partial \tau_{2,y_1}(\mathbf{x}^2)} & 0 & \cdots & 0 & \frac{\partial y_1}{\partial \tau_{2,y_2}(\mathbf{x}^2)} & 0 & \cdots & 0 & \frac{\partial y_1}{\partial \tau_{2,z_2}(\mathbf{x}^2)} \\ \frac{\partial y_2(\underline{\theta}_2)}{\partial \tau_{2,y_1}(\mathbf{x}^2)} & 0 & \cdots & 0 & \frac{\partial y_2(\underline{\theta}_2)}{\partial \tau_{2,y_2}(\mathbf{x}^2)} & 0 & \cdots & 0 & \frac{\partial y_2(\underline{\theta}_2)}{\partial \tau_{2,z_2}(\mathbf{x}^2)} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial y_2(\bar{\theta}_2)}{\partial \tau_{2,y_1}(\mathbf{x}^2)} & 0 & \cdots & 0 & \frac{\partial y_2(\bar{\theta}_2)}{\partial \tau_{2,y_2}(\mathbf{x}^2)} & 0 & \cdots & 0 & \frac{\partial y_2(\bar{\theta}_2)}{\partial \tau_{2,z_2}(\mathbf{x}^2)} \\ \frac{\partial z_2}{\partial \tau_{2,y_1}(\mathbf{x}^2)} & 0 & \cdots & 0 & \frac{\partial z_2}{\partial \tau_{2,y_2}(\mathbf{x}^2)} & 0 & \cdots & 0 & \frac{\partial z_2}{\partial \tau_{2,z_2}(\mathbf{x}^2)} \end{pmatrix},$$

where the only non-zero interior column of $\boldsymbol{E}^{c,(\mathbf{x})}_{\mathbf{x},\tau_2(\mathbf{x}^2)}$ is the one indexed by θ_2^*, where θ_2^* is such that $y_2\left(k_0,\theta_1,\theta_2^*\right)=\mathbf{x}_2^2$.

Next, we define the gradient vectors of the tax functions as

$$DT_1\left(y_1,k_1\right)=\begin{pmatrix}\frac{\partial T_1}{\partial y_1}\left(y_1,k_1\right)\\0\\\vdots\\0\\\frac{\partial T_1}{\partial k_1}\left(y_1,k_1\right)\end{pmatrix},\quad DT_2\left(y_1,\mathbf{x}_2^2,z_2\right)=\begin{pmatrix}\frac{\partial T_2}{\partial y_1}\left(y_1,\mathbf{x}_2^2,z_2\right)\\0\\\vdots\\0\\\frac{\partial T_2}{\partial y_2}\left(y_1,\mathbf{x}_2^2,z_2\right)\\0\\\vdots\\0\\\frac{\partial T_2}{\partial z_2}\left(y_1,\mathbf{x}_2^2,z_2\right)\end{pmatrix},$$

where the only non-zero element in the (continuum of) interior rows of $DT_2\left(y_1,\mathbf{x}_2^2,z_2\right)$ is in the row indexed by θ_2^*, where θ_2^* is such that $y_2\left(k_0,\theta_1,\theta_2^*\right)=\mathbf{x}_2^2$.

We finally define the Hessian matrices as

$$D^2T_1\left(y_1,k_1\right)=\begin{pmatrix}\frac{\partial^2 T_1}{\partial y_1^2}\left(y_1,k_1\right)&0&\cdots&0&\frac{\partial^2 T_1}{\partial y_1\partial k_1}\left(y_1,k_1\right)\\0&0&\cdots&0&0\\\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&\cdots&0&0\\\frac{\partial^2 T_1}{\partial y_1\partial k_1}\left(y_1,k_1\right)&0&\cdots&0&\frac{\partial^2 T_1}{\partial k_1^2}\left(y_1,k_1\right)\end{pmatrix},$$

and

$$D^2T_2\left(\mathbf{x}^2\right)=\begin{pmatrix}\frac{\partial^2 T_2}{\partial y_1^2}\left(\mathbf{x}^2\right)&0&\cdots&0&\frac{\partial^2 T_2}{\partial y_1\partial y_2}\left(\mathbf{x}^2\right)&0&\cdots&0&\frac{\partial^2 T_2}{\partial y_1\partial z_2}\left(\mathbf{x}^2\right)\\0&0&\cdots&0&0&0&\cdots&0&0\\\vdots&\vdots&\ddots&\vdots&\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&\cdots&0&0&0&\cdots&0&0\\\frac{\partial^2 T_2}{\partial y_1\partial y_2}\left(\mathbf{x}^2\right)&0&\cdots&0&\frac{\partial^2 T_2}{\partial y_2^2}\left(\mathbf{x}^2\right)&0&\cdots&0&\frac{\partial^2 T_2}{\partial y_2\partial z_2}\left(\mathbf{x}^2\right)\\0&0&\cdots&0&0&0&\cdots&0&0\\\vdots&\vdots&\ddots&\vdots&\vdots&\vdots&\ddots&\vdots&\vdots\\0&0&\cdots&0&0&0&\cdots&0&0\\\frac{\partial^2 T_2}{\partial y_1\partial z_2}\left(\mathbf{x}^2\right)&0&\cdots&0&\frac{\partial^2 T_2}{\partial y_2\partial z_2}\left(\mathbf{x}^2\right)&0&\cdots&0&\frac{\partial^2 T_2}{\partial z_2^2}\left(\mathbf{x}^2\right)\end{pmatrix},$$

where the only non-zero elements in the (continuum of) interior rows (resp., columns) of $D^2T_2\left(\mathbf{x}^2\right)$ are in the row (resp., column) indexed by θ_2^*, where θ_2^* is such that $y_2\left(k_0,\theta_1,\theta_2^*\right)=\mathbf{x}_2^2$.

The perturbations $\left(d\tau_1,d\tau_2\left(\mathbf{x}^2\right)\right)$ of the marginal tax rates faced by an individual $\left(y_1,\mathbf{x}^2,k_1\right)$

that we condider in formula (41) are defined as the changes in the gradient vectors defined above, that is

$$d\tau_1 = \begin{pmatrix} d\tau_{1,y_1} & 0 & \dots & 0 & d\tau_{1,k_1} \end{pmatrix}',$$

$$d\tau_2\left(\mathbf{x}^2\right) = \begin{pmatrix} d\tau_{2,y_1} & 0 & \dots & 0 & d\tau_{2,y_2} & 0 & \dots & 0 & d\tau_{2,z_2} \end{pmatrix}',$$

where the only non-zero element of $d\tau_2\left(\mathbf{x}^2\right)$ is indexed by θ_2^*, that is the second period type such that $z_2\left(k_0, \theta_1, \theta_2^*\right) = \mathbf{x}_2^2$.