# Redistribution and Insurance with Simple Tax Instruments*

**Sebastian Findeisen**

University of Mannheim and CEPR

**Dominik Sachs**

University of Cologne

January 9, 2015

## Abstract

We study optimal nonlinear taxation of labor income and linear taxation of capital income in a life-cycle framework with private information and idiosyncratic risk. We focus on simple history-independent tax instruments. We first analyze the welfare losses from this simplification as compared to optimal history-dependent policies. We find very small losses from restricting the complexity of savings wedges. Eliminating history dependence of labor wedges leads to larger welfare losses: moving from history dependence to age dependence yields approximately the same welfare losses as moving from age dependence to age independence and from nonlinear to linear income taxation. For optimal history-independent taxes, we provide a novel decomposition into a *redistribution* and an *insurance* component and a generalization of the top tax formula to dynamic environments. Capital taxation is desirable and yields sizable welfare gains, especially if labor income taxes are set below their optimal level.

*JEL*-**classification:** H21, H23

**Keywords:** Optimal Dynamic Taxation, Capital Taxation, First-Order Approach

# 1   Introduction

This paper characterizes Pareto optimal labor and capital income taxation with heterogenous individuals in a life cycle framework. Consistent with a large empirical literature, individuals face idiosyncratic labor income risk.[1] Consistent with real world tax policies, we characterize relatively simple tax policies that only condition on *current* earnings. The labor income tax is allowed to be fully nonlinear in the tradition of the seminal approach to optimal taxation by Mirrlees (1971). The tax on wealth (or equivalently capital income) is constrained to be linear in the tradition of dynamic Ramsey models.[2] To make the problem theoretically and computationally tractable, we employ a novel first-order approach for these simple policies in a dynamic, stochastic environment.

Since the pioneering theoretical work by Mirrlees (1971), a comprehensive literature has emerged which specializes on characterizing optimal labor income taxation. Recent papers have shown that labor supply elasticities and the distribution of income/abilities are the key forces determining optimal nonlinear income tax schedules (Piketty 1997, Diamond 1998, Saez 2001). A relatively recent literature, often called the *New Dynamic Public Finance* (*NDPF*), has expanded the classical approach and explicitly taken into account dynamics and risk. This has enabled the literature to make statements about savings distortions (Golosov, Kocherlakota, and Tsyvinski 2003), as well as to study the implications of idiosyncratic risk over the life cycle for optimal labor wedges (Golosov, Troshkin, and Tsyvinski 2013, Farhi and Werning 2013). The *NDPF* literature has focused on *history-dependent* labor income taxes, so that taxes paid on labor income can potentially depend on the whole history of past earnings.

In contrast, we restrict labor income taxation to be *history-independent* in this paper. This can be seen as the next logical step after the recent advances in the literature in exploring optimal taxation in dynamic economies and complements the *NDPF* literature and the static optimal taxation approach. One advantage of such an endeavor is that the instruments we characterize are within the realm of current tax practices.[3]

Formally, let $y_t$ be the income of an individual in period $t$ (or, equivalently, at age $t$) and $\theta_t$ be the productivity in that period. As emphasized by the *NDPF* literature, in the constrained optimal allocation, gross income is a function of the whole history of shocks $\theta^t = (\theta_1, \theta_2, ..., \theta_t)$ and these allocations are derived with dynamic mechanism design techniques. Decentral-

---

[1]Meghir and Pistaferri (2011) and Jappelli and Pistaferri (2010) provide recent surveys of the empirical literature.

[2]See, for example, Conesa, Kitao, and Krueger (2009) and Aiyagari (1995). Whereas labor income is taxed nonlinearly in most countries of the world, capital is often taxed at a linear rate. This practice is often linked to arbitrage opportunities in the case of capital. If individual A face a higher marginal tax rate on savings than individual B, there would be a deal where both could be better of by making individuals B save individual A's money. The assumption of linear capital taxes can therefore be grounded on the idea that the government cannot observe consumption on the individual level and is often made in the public finance literature. See Hammond (1987) for a more general theoretical discussion of that issue.

[3]Social insurance and tax practices in most countries typically combine *history-dependent* and *history-independent* elements. For personal income taxes, tax policies are arguably better described as being mostly history-independent, however, various forms of tax deductions also generate diverse forms of history dependence in the labor and capital income tax codes.

izing such an allocation typically requires taxes that condition on the history of incomes $y^t = (y_1, y_2, ..., y_t)$.[4] *History-independent* labor income taxes, i.e. taxes that condition only on $y_t$, can in general not implement the desired allocations. Thus, *history independence* places additional restrictions on allocations. To the best of our knowledge, no previous paper has so far investigated and characterized such arguably realistic optimal labor income tax systems in a dynamic and stochastic Mirrlees environment with a continuous type space.

We show that assuming preferences without income effects on labor supply makes this problem tractable. If labor income taxes are only a function of current income $y_t$, the income that individuals optimally choose in a decentralized economy only depends on their current productivity $\theta_t$ and not on accumulated wealth. For the allocation, this implies that income is solely a function of $\theta_t$ and not of $\theta^t$. A second advantage of this specification is that the Hessian matrix of the individual problem has a zero minor diagonal. This makes a first-order approach valid under a mild monotonicity condition on $y_t(\theta_t)$ as in the static Mirrlees model. As we show in the main body of the paper, these considerations make it possible to solve for optimal nonlinear labor and linear capital income taxes.

Having spelled out our formal approach, our analysis then proceeds in two broader steps. We first compare optimal history-independent taxes *theoretically and quantitatively* to more complex history-dependent policies but also to simpler linear policies. We provide welfare comparisons and also study how losses and gains from more complex (or even simpler) policies are distributed. In the second main part of the paper, we look in detail at the main properties and characteristics of optimal history-independent taxes.

**From History Dependence to Simpler Policies.** We start by comparing optimal history independent taxes *theoretically and quantitatively* to alternative tax structures. We begin with the solution to the dynamic mechanism design problem, in which allocations are restricted only by informational constraints (Golosov, Troshkin, and Tsyvinski 2013, Farhi and Werning 2013). We refer to this solution as the full optimum. We then proceed step by step, imposing additional constraints on capital and labor wedges in each step and thereby making the resulting tax system simpler. We theoretically characterize labor and capital wedges for the different scenarios, fleshing out in detail the implications of each additional restriction on complexity. Quantitatively, we explore how much welfare is lost at each stage, where our welfare criterion is Utilitarianism. We also provide a very detailed view on the distributional consequences of each scheme by looking at the welfare losses and gains across all income groups.

Starting from the full optimum, the first modification is to restrict the savings wedge to be linear and history-independent, while labor wedges remain history-dependent and nonlinear. This scenario has not been considered before in the literature and, importantly, it allows us investigate whether the power to raise welfare in the full optimum is stemming from the so-

---

[4]See Kocherlakota (2004) or Werning (2011). For a counterexample, see Albanesi and Sleet (2006) who show that for the special case of i.i.d. shocks simpler tax instruments suffice. See also our discussion in the related literature paragraph.

phistication of labor or capital wedges. The aggregate welfare losses from this restriction on simple capital wedges are very small. Next, we additionally restrict the ability of the planner to condition labor wedges on past skill shocks or earnings. We use our first-order approach to derive new formulas for optimal age-dependent income tax schedules and optimal linear savings taxes. Aggregate welfare losses measured in standard consumption equivalence relative to the full optimum are now significant, varying between 0.12% for log utility and 0.43% for a CRRA parameter of three.[5] This reveals that most of the benefits of the full optimum are explained by the history dependence and sophistication of the labor wedges and not by nonlinearities and history dependence in capital wedges.

Next, consistent with the previous steps, we further restrict labor taxes, and constrain them to depend on current income only. Arguably, many countries have labor tax systems, which share these features, namely that labor tax burdens are mostly determined by current income and labor income is taxed at nonlinear rates. Aggregate welfare losses relative to the full optimum vary between 0.23% for log utility and 0.70% with a CRRA parameter of three for optimal history-independent labor and linear capital taxes. Finally, our last simplification restricts the labor income tax to be linear and helps to understand the importance of nonlinearities of the labor income tax schedule. For this case, the aggregate welfare losses compared to the full optimum range from 0.29% to 1.09%.

Strikingly, one of the main conclusions of these welfare comparisons is that they imply approximately equal aggregate welfare differences: moving from the full optimum to age-dependent taxes, from age-dependent to age-independent, and from age-independent to linear tax rates entails welfare losses in each step that are very similar in size.

In our analysis, we also thoroughly inspect the distributional consequences of sophistication and simplicity of the different tax schemes by looking at the welfare changes for different initial income/skill groups. The most salient finding for this distributional perspective is that moving from an optimal nonlinear tax system to the linear one, leads to very unequally distributed losses. In the lower tail of the income distribution, the welfare losses from linearity restrictions can become very large and are around 4%.

**Optimal Simple Tax Instruments.** In the second main part of the paper, we look in detail at the main properties and characteristics of optimal history-independent taxes. First, our framework allows to investigate the interaction between labor and capital taxation, which is not possible in the static Mirrlees model. This interaction is, in general, also not studied in a dynamic Mirrlees approach, since both labor and capital wedges are part of the dynamic mechanism design solution. We examine optimal labor income taxes for a given capital tax and vice versa. Strikingly, we find that for given labor income taxes, optimal capital tax rates differ substantially, depending on how labor income taxes are set. This depends on two mechanisms. First, lower labor income taxes lead to a more concentrated distribution of wealth

---

[5]We compare our results on welfare gains and losses to other relevant papers (Weinzierl 2011, Farhi and Werning 2013) in the main text.

which increases the redistributive power of capital income taxes. Second, the lower degree of social insurance through labor income taxation, the stronger the desire to self insure in the form of precautionary savings and the lower the elasticity of savings w.r.t. capital income taxes. Conducting the opposing exercise, we find that optimal marginal labor income tax rates are decreasing in the level of capital taxation, the quantitative effect is very small, however.

Since the pathbreaking paper of Mirrlees (1971) an enormous literature has studied the static optimal tax problem. A natural question is to what extent the main prescriptions from the static model carry over to a dynamic model with income risk? First we show that an important difference to the static (or dynamic deterministic) perspective is that the so called mechanical effect of taxation – which measures the welfare effect from redistribution – does not only measure welfare gains from redistribution between ex-ante heterogeneous individuals but also from social insurance against idiosyncratic wage risk in dynamic economies with incomplete markets. We derive a decomposition of the mechanical effect into an insurance component and a redistribution component. The former is *independent* of redistributive preferences and is increasing in income risk and risk aversion.

We then put our focus on two influential sets of results, which have proven to be extremely useful in understanding properties of the optimal static schedules. First, we investigate conditions determining the shape of optimal tax rates, as pioneered by Diamond (1998) for the static model. Second, we generalize the top tax formula by Saez (2001) to the multi-period environment, which is valid if the tail of the income distribution is Pareto distributed.

**Related Literature.** The present paper is related to the *NDPF* literature. Particularly, two recent articles by Golosov, Troshkin, and Tsyvinski (2013) and Farhi and Werning (2013) have characterized *optimal history-dependent labor wedges* and to this end made use of new dynamic first-order approaches to make the problem tractable.[6] In complementary work, Kapicka (2013) develops a first-order approach for a general Mirrleesian setting with persistent productivity shocks. These recent important advances are complementary to the present paper because we show how to make progress for *simpler, history independent tax systems* under the assumption of no-income effects. This parallels the contribution by Diamond (1998), who also assumes no-income effect to gain novel insights into the static Mirrlees model. His formulas are also the natural comparison for the optimal marginal tax rates formulas that we derive for the dynamic model. In an earlier contribution to the *NDPF*, Albanesi and Sleet (2006) show that a relatively simple labor tax system can implement the second-best in the *i.i.d. case*: wealth dependent labor income taxation. By contrast, this paper studies the case of persistent shocks.[7]

There is a small but growing literature on age-dependent income taxation. Most recent and related to our work, Weinzierl (2011) and Bastani, Blomquist, and Micheletto (2013) study

---

[6]Related to these papers and ours, Pavan, Segal, and Toikka (2013) characterize a first-order approach in very general, dynamic environments.

[7] Jacobs and Schindler (2012) show that in a two-period model with linear labor taxes, a similar role for the capital tax as in the *NDPF*-literature arises as capital taxes have the positive effect of boosting labor supply in the second period.

optimal age-dependent labor income taxation.[8] These papers focus on numerical results and work with a small discrete type space. Our innovation and contribution to this literature is that our first-order approach allows to study a continuous-type framework. We are, thus, able to optimize over a fully nonlinear labor income tax schedule that is well defined for each income level. We are able to characterize this tax schedule theoretically and numerically connecting our results precisely to the contributions by Diamond (1998) and Saez (2001) for a static framework and Golosov, Troshkin, and Tsyvinski (2013) for a *NDPF* framework.

Also studying age dependency as well as standard income taxes, Best and Kleven (2013) augment the canonical optimal tax framework by incorporating career effects into a two period model with certainty. By contrast, we place our focus on a risky and dynamic economy, a standard *NDPF* framework calibrated to empirical estimates of income risk, but leave out human capital.[9]

This paper is also related to Golosov, Tsyvinski, and Werquin (2014), who study general dynamic tax reforms and elaborate the welfare gains from the sophistication of the tax code such as age dependence, history dependence or joint taxation of labor and capital income. Similar as them, we study the design of taxes in dynamic environments by directly taking into account individual responses to taxes instead of using mechanism-design techniques.

**Structure.** This paper is organized as follows. In Section 2, we spell out the formal framework and briefly explain the calibration. As explained above, in Section 3, we step by step make the tax system simpler, allowing less degrees of sophistication, starting from the full optimum. At each step we provide a theoretical discussion and a detailed analysis of the implied welfare changes. Section 4 then explores optimal history-dependent taxes in depth. In Section 5 we conclude.

# 2    The Framework

## 2.1    Environment

We consider a life cycle framework with $T$ periods where individuals at any point in time $t$ are characterized by their productivity $\theta_t \in \Theta = [\underline{\theta}, \overline{\theta}]$.[10] Further, we denote the history of shocks by $\theta^t = (\theta_1, \theta_2, \ldots, \theta_t)$. Flow utility is given by

$$U(c_t, y_t, \theta_t) = U\left(c_t - \Psi\left(\frac{y_t}{\theta_t}\right)\right),$$

---

[8]Blomquist and Micheletto (2008) is an important earlier paper in this literature.

[9]Dynamic tax models with human capital are, for example, found in Kapicka and Neira (2013), Findeisen and Sachs (2013) and Stantcheva (2013). Kapicka (2006) looks at a dynamic deterministic environment with unobservable human capital and constrains the labor income tax to be history independent in a similar spirit as in our paper.

[10]For notational convenience, we assume that the support of $\theta_t$ does not change over time. Allowing for a moving support would be a straightforward extension.

where we assume $U' > 0, U'' < 0$, and the Inada condition $\lim_{x \to 0} U'(x) = \infty$, and $\Psi', \Psi'' > 0$. $c_t$ is consumption in period $t$ and $y_t$ is gross income in period $t$ and $\frac{y_t}{\theta_t}$ captures labor effort.

Abusing notation, we sometimes write the utility function or its derivatives as a function of the history of shocks only, i.e. $U(\theta^t)$, $U'(\theta^t)$ and $U''(\theta^t)$. When we present optimal tax formulas, we denote by $\varepsilon$ the labor supply elasticity and sometimes suppress the dependence of $\varepsilon$ on the current skill type, consistent with the utility functions we use for the simulations, where the elasticity is constant – see below. As our appendix reveals, the theoretical results are more general and apply also to utility functions where the elasticity changes with the skill type and taxes.

The yearly discount factor is denoted by $\beta$. Importantly, the functional form of $U$ eliminates income effects on labor supply, while allowing for risk-aversion.[11] This assumption is crucial for the tractability of the dynamic optimal tax problem with simple instruments as we describe in detail in Section 2.2.3. Eliminating income effects has also proven to be a key simplification in making progress on the theoretical and computational side in public finance models and especially in optimal tax problems, see e.g. Diamond (1998).

We assume that agents already differ in the first period. The conditional density function (*cdf*) of the initial distribution of productivities is denoted by $F_1(\theta_1)$ and captures the ex-ante heterogeneity of agents. The reader should think about this heterogeneity as the level of heterogeneity of individuals at age of roughly 25. In the following periods, productivities evolve stochastically over time according to a Markov process. The respective *cdf* is $F_t(\theta_t|\theta_{t-1})$. Denote by $h_t(\theta^t)$ the probability of history $\theta^t$, i.e. $h_t(\theta^t) = f_t(\theta_t|\theta_{t-1})f_{t-1}(\theta_{t-1}|\theta_{t-2}).....f_1(\theta_1)$. We assume that all conditional density functions are continuously differentiable. Denote by $\Theta^t$ the set of possible histories in $t$. Further, we denote by $f_t(\theta_t)$ the cross-sectional skill distribution at age $t$.

## 2.2 The Planner's Problem in the Different Scenarios

The preferences of the social planner are described by the set of Pareto weights $\{\tilde{f}_1(\theta_1)\}_{\theta_1 \in [\underline{\theta}, \overline{\theta}]}$. The cumulative Pareto weights are defined by $\tilde{F}_1(\theta_1) = \int_{\underline{\theta}}^{\theta_1} \tilde{f}_1(\tilde{\theta}_1)d\tilde{\theta}_1$. The set of weights are restricted such that $\tilde{F}_1(\overline{\theta}) = 1$. Different sets of Pareto weights refer to different points on the Pareto frontier. The set of weights where $\tilde{f}_1(\theta_1) = f_1(\theta_1) \; \forall \; \theta_1$, e.g., refers to the Utilitarian planner. Similar as $h_t(\theta^t)$, define $\tilde{h}_t(\theta^t) = f_t(\theta_t|\theta_{t-1})f_{t-1}(\theta_{t-1}|\theta_{t-2}).....\tilde{f}_1(\theta_1)$ to express the Pareto weight for individuals with history $\theta^t$. The planner can shift resources across periods at interest rate $r$. We assume $\beta(1 + r) = 1$, which does not come at the expense of generality but simplifies the exposition. We now describe the problems of the different policy scenarios considered in detail.

---

[11]The empirical literature using detailed micro data sets has typically not rejected a zero income elasticity on labor supply or found very small effects (see Gruber and Saez (2002) for the US or a recent paper by Kleven and Schultz (2013) using the universe of Danish tax records).

### 2.2.1 Full History Dependence

We start by stating the mechanism-design problem that is often referred to as the *Dynamic Mirrlees* or the *New Dynamic Public Finance Approach*. This case has recently been extensively studied by Farhi and Werning (2013) and Golosov, Troshkin, and Tsyvinski (2013). The purpose of our presentation of the problem is therefore not novelty but to set an upper bound for welfare to which we can compare the power of simpler tax instruments.

In this case, the planner assigns consumption-income bundles $c_t(\theta^t), y_t(\theta^t) \ \forall \ t = 1, ..., T$ and $\forall \theta^t \in \Theta^t$ that maximize welfare

$$\sum_{t=1}^T \beta^{t-1} \int_{\Theta^t} U\left(c_t(\theta^t) - \Psi\left(\frac{y_t(\theta^t)}{\theta_t}\right)\right) \tilde{h}_t(\theta^t) d\theta^t$$

subject to a resource constraint

$$\sum_{t=1}^T \frac{1}{R^{t-1}} \int_{\Theta^t} \left(y_t(\theta^t) - c_t(\theta^t)\right) h_t(\theta^t) d\theta^t$$

and subject to incentive compatibility constraints for each $\theta_1 \in \Theta$ and each reporting strategy $\sigma(\theta^t)$

$$U\left(c_1(\theta_1) - \Psi\left(\frac{y_1(\theta_1)}{\theta_1}\right)\right) + \sum_{t=2}^T \beta^{t-1} \int_{\Theta^t(\theta_1)} \left(U(c_t(\theta^t) - \Psi\left(\frac{y_t(\theta^t)}{\theta_t}\right)\right) h_t(\theta^t) d\theta^t \tag{1}$$

$$\geq U\left(c_1(\sigma(\theta_1)) - \Psi\left(\frac{y_1(\sigma(\theta_1))}{\theta_1}\right)\right) + \sum_{t=2}^T \beta^{t-1} \int_{\Theta^t(\theta_1)} \left(U(c_t(\sigma(\theta^t)) - \Psi\left(\frac{y_t(\sigma(\theta^t))}{\theta_t}\right)\right) h_t(\theta^t) d\theta^t$$

where $\Theta^t(\theta_1)$ is the set of all possible histories in period $t$ that contain $\theta_1$. Farhi and Werning (2013) and Golosov, Troshkin and Tsyvinski (2013) have shown how to make this problem tractable with a dynamic first-order approach that builds on Pavan, Segal, and Toikka (2013), Kapicka (2013) and others. See Appendix A.1 for the Lagrangian and the solution of this problem. In Section 3.1, we discuss theoretical and quantitative properties of this optimum.

### 2.2.2 History-Dependent Labor Wedges and Linear Capital Wedges

The set of simple policy instruments that are at the heart of this paper differs from the full mechanism-design solution as follows: labor income taxes are constrained to be history-independent and capital taxes are constrained to be history-independent and linear. To better understand the welfare implications of each of these restrictions, we also consider an intermediate case were only the latter restriction applies.

In this intermediate case, the planner solves a problem that only differs from that in Section 2.2.1 by the following additional restriction:

$$\forall\ \theta^t:\ U'(\theta^t) = (1 - \tau_s)\int_\Theta U'(\theta^t, \theta_{t+1})dF_{t+1}(\theta_{t+1}|\theta_t). \tag{2}$$

Given the linearity constraint for the capital wedge, the planner has to respect an Euler equation that is adjusted by the linear wedge $\tau_s$ which itself is also a choice variable of the planner.[12] Golosov and Tsyvinski (2007) study a similar problem and investigate the desirability of a linear capital distortion in the presence of endogenous private insurance markets and general equilibrium effects. In Appendix A.2 the Lagrangian and the solution of this problem are stated. In Section 3.2, we discuss theoretical and quantitative properties of this optimum.

### 2.2.3 History Independence: Age-Dependent and Age-Independent Labor Income Taxation

We next relax the history dependence of the labor wedge. One consequence is that the labor wedge can now be directly interpreted as a marginal tax rate on labor income. Our notation in this section nests age-dependent and age-independent taxation. For the age-independent tax problem one has to drop the subscript $t$, i.e. write $\mathcal{T}$ and $y$ instead of $\mathcal{T}_t$ and $y_t$.

**Individual Problem Given Taxes.** Each period, individuals make a work and savings decision. Formally, the recursive problem of individuals given age-dependent labor income taxes $\{\mathcal{T}_j\}_{j=t,...,T}$ and a linear capital tax rate $\tau_s$ reads as:

$$V_t(\theta_t, a_t, \{\mathcal{T}_j\}_{j=t,...,T}, \tau_s) = \max_{a_{t+1},y_t} U\left(y_t - \mathcal{T}_t(y_t) + (1+r)(1-\tau_s)a_t - a_{t+1} - \Psi\left(\frac{y_t}{\theta_t}\right)\right)$$
$$+ \beta E_t\left[V_{t+1}(\theta_{t+1}, a_{t+1}, \{\mathcal{T}_j\}_{j=t+1,...,T}, \tau_s)\right], \tag{3}$$

where $a_1 = 0$ and $a_T \geq 0$.[13] Based on the assumption on preferences, the following lemma directly follows.

**Lemma 1.** *The optimal gross income $y_t$ that solves (3) is independent of assets, the capital tax rate $\tau_s$ and future labor income taxes $\{\mathcal{T}_j\}_{j=t+1,...,T}$. It is thus only a function of the current shock and of the current labor income tax: $y_t(\theta_t, \mathcal{T}_t)$.*

This greatly simplifies the optimal tax analysis. For the resulting allocation, this implies that $y_t$ is only a function of $\theta_t$ and not of $\theta^t$. The savings decision of individuals, in contrast, will depend on all state variables: $a_{t+1}(\theta_t, a_t, \{\mathcal{T}_j\}_{j=t,...,T}, \tau_s)$. Recursively inserting, one can also

---

[12]In our numerical simulations the problem always was concave in $\tau_s$.

[13]We allow agents to borrow up to natural debt limit (see Aiyagari (1994)). There are two differences to Aiyagari (1994): First, labor supply is endogenous; the minimal amount of future earnings in period $s$ is $\sum_{t=s}^T y_t(\underline{\theta})$. Second, individuals would actually not even borrow that much as repaying everything and consuming nothing would yield a negative argument in $U(\cdot)$ because of the disutility of labor. This can never be optimal as the marginal utility would be infinite whenever the argument in $U(\cdot)$ is zero under standard inada conditions. Therefore, in the absence of taxes, the maximal amount of debt is $\sum_{t=s}^T \left(y_t(\underline{\theta}) - \Psi\left(\frac{y_t(\underline{\theta})}{\underline{\theta}}\right)\right)$.

write $a_{t+1}(\theta^t, \{\mathcal{T}_j\}_{j=1,...,T}, \tau_s)$, which implies that assets are a function of the history of shocks: $a_{t+1}(\theta^t)$.

**The Social Planner's Problem.**   The social planner's problem reads as:

$$\max_{\{\mathcal{T}_j\}_{j=t,...,T}, \tau_s} \int_\Theta V_1(\theta_1, 0, \{\mathcal{T}_j\}_{j=1,...,T}, \tau_s) d\tilde{F}_1(\theta_1), \tag{4}$$

where $V_1(\theta_1, 0, \{\mathcal{T}_j\}_{j=1,...,T}, \tau_s)$ is the solution to (3) for each $\theta_1$ and subject to an intertemporal budget constraint:

$$\sum_{t=1}^T \frac{1}{(1+r)^{t-1}} \int_\Theta \mathcal{T}_t(y_t(\theta_t)) dF_t(\theta_t)$$

$$+ \sum_{t=2}^T \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \tau_s(1+r)a_t(\theta^{t-1}, \{\mathcal{T}_j\}_{j=1,...,T}, \tau_s) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \geq \mathcal{R}, \tag{5}$$

where $\mathcal{R}$ is some exogenous revenue requirement of the government and $a_t(\theta^{t-1}, \{\mathcal{T}_j\}_{j=1,...,T}, \tau_s)$ is the amount of savings of individuals with history $\theta^{t-1}$ that optimally follows from (3).

The presence of constraint (3) makes it difficult to tackle the problem with conventional Lagrangian or optimal control techniques. We argue next that (3) can be replaced by a set of first-order conditions for $a_{t+1}$ and $y_t$ and a monotonicity condition on $y_t$.

**First-Order Approach.**   In the remainder of this paper, we will suppress the dependence of assets and gross income on taxes. We will thus write $y_t(\theta_t)$ instead of $y_t(\theta_t, \mathcal{T}_t)$ and $a_t(\theta^{t-1})$ instead of $a_t(\theta^{t-1}, \{\mathcal{T}_j\}_{j=1,...,T}, \tau_s)$.

We now want to show how (3) can be replaced by two first-order conditions and a monotonicity constraint. The set of first-order conditions for the individual problem (3) are standard. For the labor supply decision, we have $\forall t$ and $\forall \theta_t \in \Theta$ :

$$1 - \mathcal{T}_t'(y_t(\theta_t)) = \Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right) \frac{1}{\theta_t}. \tag{6}$$

For the savings decision, we have $\forall t = 1, ..., T-1$ and $\forall \theta^t \in \Theta^t$ :

$$U'\left(y_t(\theta_t) - \mathcal{T}_t(y_t(\theta_t)) - a_{t+1}(\theta^t) + (1-\tau_s)a_t(\theta^{t-1}) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)$$

$$= (1-\tau_s) \int_\Theta U'\left(y_{t+1}(\theta_{t+1}) - \mathcal{T}_{t+1}(y_{t+1}(\theta_{t+1})) - a_{t+2}(\theta^t, \theta_{t+1})\right.$$

$$\left. + (1-\tau_s)a_{t+1}(\theta^t) - \Psi\left(\frac{y_{t+1}(\theta_{t+1})}{\theta_{t+1}}\right)\right) dF_{t+1}(\theta_{t+1}|\theta_t). \tag{7}$$

These conditions are only necessary and not sufficient for the agents' choices to be optimal. Due to the assumption about preferences, however, the second-order conditions are of particularly simple form. The derivative of the first-order condition of labor supply with respect to consumption, i.e. the cross derivative of the value function, is zero. By symmetry of the Hessian, the same holds for the derivative of the Euler equation with respect to labor supply. Thus, the minor diagonal of the Hessian matrix contains only zeros. For (6) and (7) to represent a maximum, only the second derivatives of the value function with respect to labor supply and consumption have to be $\leq 0$. For labor supply, a familiar argument from the standard Mirrlees model implies that this holds if and only if[14]

$$y_t'(\theta_t) \geq 0. \tag{8}$$

The second-order condition for savings is always fulfilled due to concavity of the utility function. Hence, (6) and (7) represent a maximum whenever $y_t'(\theta_t) \geq 0$. As $y_t'(\theta_t) \geq 0$ even implies global concavity, (6) and (7) represent a global maximum if $y_t'(\theta_t) \geq 0$ holds.

In a final step, we make use of a change of variables and define $M_t(\theta_t) = y_t(\theta_t) - \mathcal{T}_t(y_t(\theta_t))$. Applying this for (6) is still problematic as it contains $\mathcal{T}_t'$, however. To tackle this problem, we make use of the following derivative

$$\frac{\partial \left( y_t(\theta_t) - \mathcal{T}_t(y_t(\theta_t)) - \Psi \left( \frac{y_t(\theta_t)}{\theta_t} \right) \right)}{\partial \theta_t} = y_t'(\theta_t)(1 - \mathcal{T}_t'(y_t(\theta_t))) - \Psi' \left( \frac{y_t(\theta_t)}{\theta_t} \right) \left[ \frac{y_t'(\theta_t)}{\theta_t} - \frac{y_t(\theta_t)}{\theta_t^2} \right].$$

Inserting (6) into this derivative yields:

$$\frac{\partial \left( y_t(\theta_t) - \mathcal{T}_t(y_t(\theta_t)) - \Psi \left( \frac{y_t(\theta_t)}{\theta_t} \right) \right)}{\partial \theta_t} = \Psi' \left( \frac{y_t(\theta_t)}{\theta_t} \right) \frac{y_t(\theta_t)}{\theta_t^2}. \tag{9}$$

Thus, (9) is equivalent to (6). Now, we apply the change of variables $M_t(\theta_t) = y_t(\theta_t) - \mathcal{T}_t(y_t(\theta_t))$, which leads to a problem that can be solved with calculus of variation or optimal control:

**Proposition 1.** *Instead of choosing $\{\mathcal{T}_t\}_{t=1,\dots,T}$ and $\tau_s$ to maximize (4) subject to (5) and (3), the planner can also choose $\{\{M_t(\theta_t), y_t(\theta_t)\}_{\theta_t \in \Theta}, \{a_t(\theta^{t-1})\}_{\theta^{t-1} \in \Theta^{t-1}}\}_{t=1,\dots,T}$ and $\tau_s$ subject to (5), (9), (7) and (8), where $y_t(\theta_t) - \mathcal{T}_t(y_t(\theta_t)) = M_t(\theta_t)$.*

The approach can also be interpreted as a restricted direct mechanism that is augmented by a savings choice. Agents report their type $\theta_t$ and the planner assigns bundles $(M_t(\theta_t), y_t(\theta_t))$ – it is a restricted mechanism because the planner cannot make $M_t$ and $y_t$ conditional on the history of shocks.

We provide the Lagrangian and the first-order conditions for this problem in Appendix B.1. When solving for optimal policies, we do not incorporate the monotonicity constraint (8) into

---

[14]See, e.g., Salanié (2003, p.87 ff).

the Lagrangian, as is standard practice in the optimal tax literature. In the numerical simulations we ex-post check whether the monotonicity condition is fulfilled or not.

## 2.3   Calibration and Primitives

There is large literature on the estimation of earnings dynamics over the life cycle – see Meghir and Pistaferri (2011) and Jappelli and Pistaferri (2010) for recent surveys. For the parameterization of our model, we use the recent empirical approach taken by Karahan and Ozkan (2013). In their analysis, they estimate the persistence of permanent shocks as well as the variance of permanent and transitory income shocks for US workers. Innovatively, and in contrast to most previous work in this strand of the literature, they allow these parameters to be age-dependent and to change over the life cycle. They find two structural breaks in how the key parameters change over the life cycle, giving three age groups, in which income dynamics are governed by the same risk parameters.

We base our parameterization on their results. Given the estimates of Karahan and Ozkan (2013) for the evolution of income over the life cycle[15], we simulate millions of labor income histories. We describe this in more detail in Appendix C. After having simulated those earnings histories using a sufficient number of draws, we partition individuals into three age-groups, namely 24-36, 37-49 and 50-62, which represent periods one, two and three respectively. Last, we calibrate the cross sectional income distributions for each age group and the respective transition probabilities. Figure 8 in Appendix C shows the three cross-sectional income distributions for each age group. To complete the parametrization of the model, we calibrate all conditional skill distributions from their income counterparts, as pioneered by Saez (2001).[16]

We assume that the utility function is of the form

$$U(c, y, \theta) = \frac{\left( c - \frac{\left( \frac{y}{\theta} \right)^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}} \right)^{1-\rho}}{1-\rho}.$$

For the benchmark, we set $\varepsilon = 0.33$ (Chetty 2012). Our conclusions from the simulations are not sensitive to the choice of the labor supply elasticity. We let $\rho$ vary between 1 (log utility) and 3. The annual interest rate is 3% so in our simulations, we set $r = 1.03^{13} - 1$ and adjust the discount factor such that $\beta(1 + r) = 1$.

## 3   From The Full Optimum to Simpler Policies

We now compare optimal history independent taxes theoretically and quantitatively to more complex history-dependent policies but also to simpler linear policies. All derivations are del-

---

[15]We gratefully acknowledge that they shared some estimates with us that are not directly available from their paper.

[16]We back out the skill from the first-order condition of individual labor supply given an approximation of the current US-tax system, a linear tax rate of 30%.

egated to the appendix. We provide welfare comparisons using a Utilitarian social welfare function but also study how losses and gains from more complex (or even simpler) policies are distributed across different skill groups. We proceed step by step, starting out with the information-constrained optimum, making policies less sophisticated in each step.

## 3.1 Full History Dependence

The first case provides the planner with the most sophisticated instruments. The wedges we characterize are part of the solution to the mechanism design problem, where agents reveal their type in each period. The cross-sectional properties of optimal history-dependent distortions have been recently characterized by Golosov, Troshkin, and Tsyvinski (2013). Our formulas for utility functions without income effects can be seen as special cases of theirs. The purpose of this section is hence not novelty but to build the benchmark, from which we restrict the available instruments of the planner step by step. The labor wedge in the full optimum is given by:

$$
\frac{\tau_{Lt}^{F}(\theta^t)}{1 - \tau_{Lt}^{F}(\theta^t)} = \left(1 + \frac{1}{\varepsilon}\right) \frac{1}{\theta_t f_t(\theta_t|\theta_{t-1})} \int_{\theta_t}^{\bar{\theta}} \exp\left(\int_{\theta}^{x} \gamma_t(\theta^{t-1}, s) \frac{1}{s} ds\right)
$$
$$
U'(\theta^t) \left\{ f_t(x|\theta_{t-1}) \left( \frac{1}{U_t'(\theta^{t-1}, x)} - \frac{1}{U_{t-1}'(\theta^{t-1})} \right) \right.
$$
$$
+ \frac{\tau_{L,t-1}^{F}(\theta^{t-1})}{1 - \tau_{L,t-1}^{F}(\theta^{t-1})} \left(1 + \frac{1}{\varepsilon}\right)^{-1} \theta_{t-1} \frac{1}{U_{t-1}'(\theta^{t-1})} \frac{\partial f_t(x|\theta_{t-1})}{\partial \theta_{t-1}}
$$
$$
\left. - \frac{\tau_{L,t-1}^{F}(\theta^{t-1})}{1 - \tau_{L,t-1}^{F}(\theta^{t-1})} \left(1 + \frac{1}{\varepsilon}\right)^{-1} \theta_{t-1} \frac{1}{U_{t-1}'(\theta^{t-1})} f_t(x|\theta_{t-1}) \gamma_{t-1} \frac{1}{\theta_{t-1}} \right\} dx, \quad (10)
$$

where as in Golosov, Troshkin, and Tsyvinski (2013)[17]

$$
\gamma_t(\theta^t) = \Psi'\left(\frac{y_t(\theta^t)}{\theta_t}\right) \frac{y_t(\theta^t)}{\theta_t} \frac{U_t''(\theta^t)}{U_t'(\theta^t)} < 0.
$$

See Golosov, Troshkin, and Tsyvinski (2013) for a detailed theoreticaldiscussion of the labor wedge in the full optimum. Figure 1(a) shows the labor wedges in the second period for three different income histories (10th, 50th and 90th percentile in the earnings distribution when young). In line with Golosov, Troshkin, and Tsyvinski (2013), wedges are higher for higher earnings histories. The same is true for optimal savings wedges as illustrated in Figure 1(b).

---

[17]In Appendix A.1.1, we show how to precisely connect our formula to their formula.

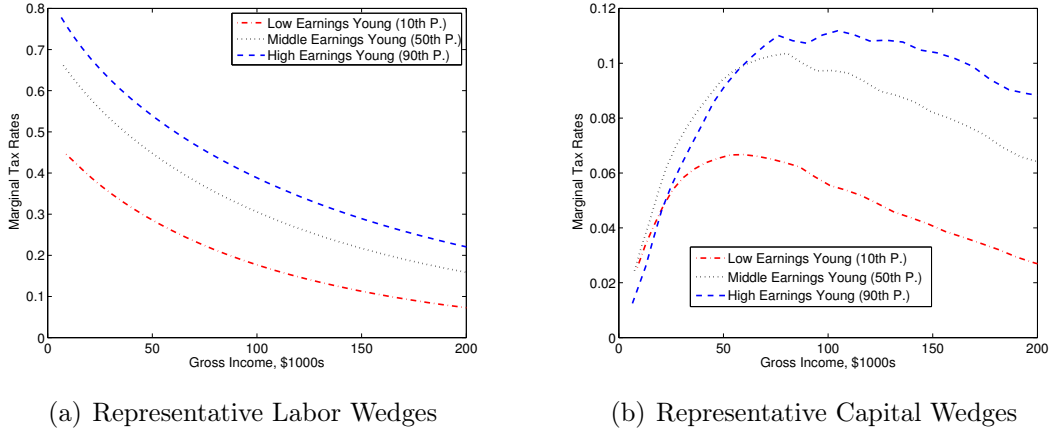| | |
|---|---|
| (a) Representative Labor Wedges | (b) Representative Capital Wedges |

Figure 1: Wedges in the Full Optimum (Middle-Aged, Conditional on Young Earnings)

## 3.2 History-Dependent Labor Wedges and Linear Capital Wedges

We now consider the policy scenario stated in Section 2.2.2 where only capital wedges are constrained to be linear and history-independent. The linearity restriction on the capital wedge has the following implication on optimal labor wedges:

$$
\frac{\tau_{Lt}^R(\theta^t)}{1-\tau_{Lt}^R(\theta^t)} = \frac{\tau_{Lt}^{F*}(\theta^t)}{1-\tau_{Lt}^{F*}(\theta^t)} + \left(1+\frac{1}{\varepsilon}\right)\frac{1}{\theta_t f_t(\theta_t|\theta_{t-1})}\int_{\theta_t}^{\overline{\theta}}\exp\left(\int_{\underline{\theta}}^x \gamma_t(\theta^{t-1},s)\frac{1}{s}ds\right)
$$

$$
U'(\theta^t)\Bigg\{\frac{U_t''(\theta^{t-1},x)}{U_t'(\theta^{t-1},x)}\left(\Delta\mu_t(\theta^{t-1},x)(1-\tau_s^R)f_t(x|\theta_{t-1}) + \mu_t(\theta^{t-1},x)(1-(1-\tau_s^R)f_t(x|\theta_{t-1}))\right)
$$

$$
-\beta f_t(x|\theta_{t-1})\frac{U_{t-1}''(\theta^{t-1})}{U_{t-1}'(\theta^{t-1})}\left(\Delta\mu_{t-1}(\theta^{t-1})(1-\tau_s^R)f_{t-1}(\theta_{t-1}|\theta_{t-2}) + \mu_{t-1}(\theta^{t-1})(1-(1-\tau_s^R)f_{t-1}(\theta_{t-1}|\theta_{t-2}))\right)\Bigg\}dx,
$$

(11)

where $\mu_t(\theta^t)$ is the multiplier on the linear intertemporal wedge constraint (2), $\Delta\mu_t(\theta^{t-1},\theta) = \mu_t(\theta^{t-1},\theta) - \mu_{t-1}(\theta^{t-1})$ is the change in the multiplier for a given history and $\tau_{Lt}^{F*}(\theta^t)$ is labor wedge formula for the full optimum (10), where the autoregressive component from $\tau_{Lt}^{F*}(\theta^t)$ is evaluated at the past labor wedge from the restricted problem now.[18]

---

[18]Formally:

$$
\frac{\tau_{Lt}^{F*}(\theta^t)}{1-\tau_{Lt}^{F*}(\theta^t)} = \left(1+\frac{1}{\varepsilon}\right)\frac{1}{\theta_t f_t(\theta_t|\theta_{t-1})}\int_{\theta_t}^{\overline{\theta}}\exp\left(\int_{\underline{\theta}}^x \gamma_t(\theta^{t-1},s)\frac{1}{s}ds\right)
$$

$$
U'(\theta^t)\Bigg\{f_t(x|\theta_{t-1})\left(\frac{1}{U_t'(\theta^{t-1},x)} - \frac{1}{U_{t-1}'(\theta^{t-1})}\right)
$$

$$
+ \frac{\tau_{L,t-1}^R(\theta^{t-1})}{1-\tau_{L,t-1}^R(\theta^{t-1})}\left(1+\frac{1}{\varepsilon}\right)^{-1}\theta_{t-1}\frac{1}{U_{t-1}'(\theta^{t-1})}\frac{\partial f_t(x|\theta_{t-1})}{\partial\theta_{t-1}}
$$

$$
- \frac{\tau_{L,t-1}^R(\theta^{t-1})}{1-\tau_{L,t-1}^R(\theta^{t-1})}\left(1+\frac{1}{\varepsilon}\right)^{-1}\theta_{t-1}\frac{1}{U_{t-1}'(\theta^{t-1})}f_t(x|\theta_{t-1})\gamma_{t-1}\frac{1}{\theta_{t-1}}\Bigg\}dx.
$$

The labor wedge is equal to the formula from the full optimum plus additional terms (line 2 and 3), capturing the effect of the labor distortion on the linear capital wedge constraint (2). We can get a better understanding by looking at the first-order condition for $\tau_s^R$:

$$-\sum_{t=1}^{T-1}\int_{\Theta^t}\mu_t(\theta^t)\int_{\theta_{t+1}}U'_{t+1}(\theta^t,x)dF_{t+1}(x|\theta_t)d\theta^t = 0.$$

Note that $\mu_t(\theta^t)$ gives the welfare gain of slightly relaxing the Euler equation in the sense of letting an individual with history $\theta^t$ save slightly less. Thus, if $\mu_t(\theta^t)$ is positive, this implies that the planner would like to impose a higher savings wedge on individuals with history $\theta^t$. The opposite is true for a negative value of $\mu_t(\theta^t)$. Intuitively, imposing the same intertemporal wedge on all individuals forces the planner to restrict some types less than she would like and some types more than she would like. In the labor wedge formula (11), the planner takes into account how the labor wedge relaxes or tightens the Euler constraints. As it turns out, however, the quantitative implications of these additional terms, however, are relatively small. Optimal labor wedges are very similar to those in the full optimum, which is why we refrain from illustrating them in a graph. What differs more strongly are the capital wedges. The optimal linear capital wedge is 6.4%. In the full optimum, the average capital distortion was 4.6% in period one and 7.6% in period two. Intuitively, the planner now has to compromise by choosing some average value.

What are the welfare implications of restricting savings wedges to be linear and history-independent? We find that the implied aggregate welfare losses from this modification are very moderate; between 0.02% and 0.13% (constant relative risk aversion ranging from 1 to 3) in consumption equivalents. Also the distributional impact is relatively small, initially high-ability individuals tend to gain a bit as their savings wedges are in general reduced relative to the full optimum and the opposite holds for initially low-ability individuals. Like the aggregate losses, the losses across the skill distribution are very modest. In sum, the welfare consequences of restricting the complexity of optimal intertemporal wedges are small. In the next step we relax the sophistication of the labor wedge.

## 3.3   Age-Dependent Labor Taxation

We now remove history dependence in the labor wedge. The labor income tax can condition on age $t$. The capital tax is linear and age-independent, as the savings wedge was in the previous case. We now present a novel formula for the optimal age-dependent marginal tax rate in continuous models:

**Proposition 2.** *The optimal age-dependent marginal tax rate on labor income satisfies*

$$\frac{\mathcal{T}'_t(y_t(\theta_t))}{1-\mathcal{T}'_t(y_t(\theta_t))} = \left(1+\frac{1}{\varepsilon}\right)\frac{1}{\frac{1}{(1+r)^{t-1}}\theta_t f_t(\theta_t)} \times [\mathcal{M}_t(\theta_t)+\mathcal{S}_t(\theta_t)] \tag{12}$$

*where the mechanical effect $\mathcal{M}_t(\theta_t)$ is defined as:*

$$\mathcal{M}_t(\theta_t) = \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\theta_t}^{\overline{\theta}} dF_t(\tilde{\theta}_t|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}$$

$$- \frac{1}{\lambda(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\theta_t}^{\overline{\theta}} U'(\theta^{t-1}, \tilde{\theta}_t) dF_t(\tilde{\theta}_t|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1}. \qquad (13)$$

*and*

$$\mathcal{S}_t(\theta_t) = - \int_{\Theta^{t-1}} \int_{\theta_t}^{\overline{\theta}} \mu_t(\theta^{t-1}, \theta_s) U''(\theta^{t-1}, \theta_s) d\tilde{\theta}_t d\theta^{t-1}$$

$$+ (1 - \tau_s^A) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta_t}^{\overline{\theta}} U''(\theta^{t-1}, \theta_t) dF_t(\tilde{\theta}_t|\theta_{t-1}) d\theta^{t-1} \qquad (14)$$

*where $\mu_t(\theta^t)$ is the Lagrangian multiplier on the Euler equation of individuals with history $\theta^t$ and $\lambda$ is the Lagrangian multiplier on the government budget constraint, i.e. the marginal value of public funds. In Appendix B.1.3, we derive expressions for these multipliers. Further, we have $\mathcal{T}_t'(y_t(\underline{\theta})) = 0$, and $\mathcal{T}_t'(y_t(\overline{\theta})) = 0$ if $\overline{\theta} < \infty$.*

The mechanical effect $M_t(\theta_t)$ is similar as in the static model. An important difference is that it now also captures the insurance motive of taxation. We disentangle the welfare gains from redistribution and insurance in Section 4.2. $\mathcal{S}_t(\theta_t)$ captures the impact of labor income taxes on the Euler equation constraints (7). For $\tau_s = 0$ this additional term is equal to zero because the value of the Lagrangian multipliers on the Euler equations is equal to zero. Relaxing or tightening the Euler equations has no first-order effect on welfare because it does not affect incentives to supply labor. Only for $\tau_s \neq 0$, relaxing or tightening the Euler equations has a first-order impact on welfare through the implied change in capital tax revenue.[19]

Alternatively, one can derive (12) with tax perturbation methods (Golosov, Tsyvinski, and Werquin 2014).[20] In that case, term $\mathcal{S}_t(\theta_t)$ reads as:

$$\mathcal{S}_t(\theta_t) = \tau_s^A \sum_{j=2}^{T} (1+r)^{j-2} \int_{\Theta^{j-1}} \int_{\theta_t}^{\overline{\theta}} \frac{\partial a_j(\theta^{j-1})}{\partial \mathcal{T}_t(y_t(\tilde{\theta}_t))} d\tilde{\theta}_t h_{j-1}(\theta^{j-1}) d\theta^{j-1}. \qquad (15)$$

We briefly sketch this perturbation argument. Consider – starting from the optimal tax system– an infinitesimal increase of the marginal tax rate $\Delta\mathcal{T}'$ at an infinitesimal income interval with length $\Delta y_t(\theta_t)$ around income level $y_t(\theta_t)$. Given that the tax system was optimal, this should have no first-order effect on welfare. The impact on welfare can be decomposed into three effects. First, there is a mechanical welfare effect from taking money from all young individuals with income $> y_t(\theta_t)$ is given by $\mathcal{M}_t(\theta_t) \times \Delta\mathcal{T}'\Delta y_t(\theta_t)$. It depends on the redistributive

---

[19]By contrast, the value of the Lagrangian multipliers on the Euler equations are non-zero also for $\tau_s = 0$ in Section 3.2. In the case of history-dependent labor distortions, relaxing the Euler equations always has a first-order impact on incentive constraints.

[20]See also Piketty (1997) and Saez (2001) who study the tax perturbation method in a static environment.
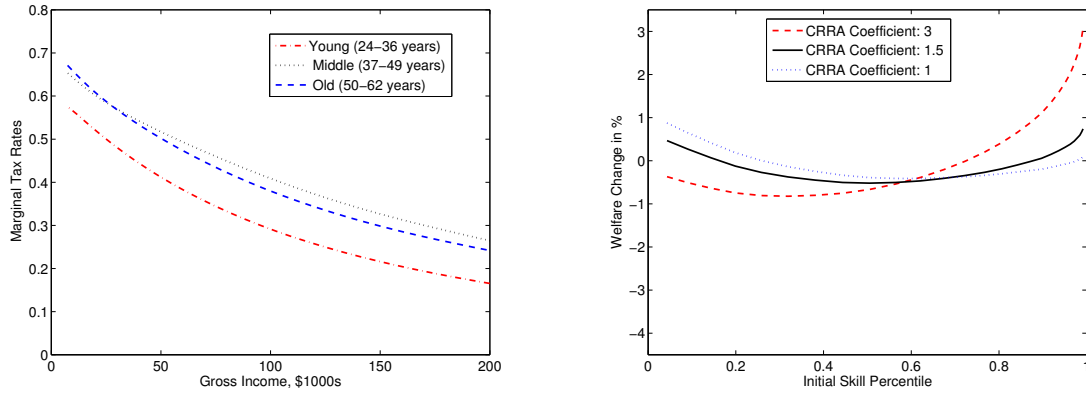
(a) Optimal Age-Dependent Tax Schedules

(b) Welfare Changes From History-Dependent Labor Wedges To Age-Dependent Taxation

Figure 2: Welfare Changes From Restricting Labor Wedges

preferences, i.e. the Pareto weights, of the planner, the degree of risk aversion as well as on the share of individuals with income $> y_t(\theta_t)$. The planner trades off this *mechanical effect* against a loss in tax revenue which is induced by lower labor supply of individuals of type $\theta_t$. It can be shown that this is captured by $LS(\theta_t) = -\lambda \frac{\mathcal{T}_t'(y_t(\theta_1))}{1-\mathcal{T}_t'(y_t(\theta_t))}\theta_t f_t(\theta_t)\frac{\varepsilon}{\varepsilon+1} \times \Delta\mathcal{T}'\Delta y_t(\theta_t)$. Further, a tax change also changes revenue from the savings tax. This is captured by $\mathcal{S}_t(\theta_t) \times \Delta\mathcal{T}'\Delta y_t(\theta_t)$. Note that the summation in the savings term is over the index $j$, since a tax change in $t$ might affect savings in every period of the life cycle. We show in Appendix B.2.3 how $\mathcal{S}_t(\theta_t)$ can also be understood as an income effect on savings triggered by changes in labor taxes. The induced effects on labor supply and savings behavior have no first-order effect on welfare because of the envelope theorem. Setting $LS(\theta_t) + [\mathcal{M}_t(\theta_t) + \mathcal{S}_t(\theta_t)] \times \Delta\mathcal{T}'\Delta y_t(\theta_t) = 0$ yields the optimal tax formula.

Obtaining expressions for $\frac{\partial a_j(\theta^{j-1})}{\partial \mathcal{T}_t(y_t(\tilde{\theta}_t))}$ in (15) is quite complex in a stochastic environment. Using (14) provides a way to analytically capture these effects that we also exploit in our numerical simulations. In Appendix B.2.2, we show the equivalence between (15) and (14) for a three-period economy. Deriving savings responses for the three-period case is already quite involved, which shows that using (14) for computational analysis is significantly simpler. For intuition and interpretation on the other hand, (15) is more suitable. We therefore stick to this expression in the main text.

In Figure 2, we plot the optimal marginal labor income tax rates. Labor income taxes are lowest for the *young*. The other tax schedules for the middle and the old lie closer to each other. Taxes on the middle-aged are higher than on the old for most of the income support, but the quantitative difference is smaller. What are the underlying economic intuitions for these results? The first driving force for the pattern of age-dependent taxes is the way how hazard-rates of the skill distribution differ among age groups. A well known result in optimal nonlinear taxation is that the hazard rates of the income and skill distributions are both very

16

informative statistics for the optimal pattern of marginal tax rates. The higher the ratio $\frac{1-F}{f\theta}$, the higher are optimal tax rates, ceteris paribus. In the calibration based on age-dependent income risk processes, these ratios are highest for the middle, lowest for the old, and lie in-between for the young for most parts of the income support. In addition, in a dynamic and risky economy, age-dependent taxation has the additional power to provide insurance against income shocks; we provide a decomposition of the mechanical effect into an insurance and a redistribution effect in Section 4.2. This insurance motive counteracts the hazard-rate force for low taxes on the old and explains that taxes on the old are higher than taxes on the young. Our result that the young should face the lowest taxes is consistent with Weinzierl (2011). He finds that taxes are highest for the old, whereas in our case they are similar to taxes on the middle-aged due to offsetting tagging and insurance effects.

The optimal linear tax on the capital stock is now 5.5%, i.e. slightly lower than the optimal linear capital wedge in restricted optimum in Section 3.2. In Section 4.1, we explain how the tax rate on the capital stock in each period can be transformed into an annual tax on capital income.

The next questions we ask is how much aggregate welfare is lost and who bears what share of that loss resulting from the restriction of labor wedges to only depend on age and current income and not on the history of incomes? Depending on risk-aversion, the losses range between 0.12% and 0.43% of consumption. Figure 2(b) illustrates the distributional effects. Interestingly, the welfare changes are U-shaped in initial skill type. What is the intuition for this heterogeneity in welfare changes? On the one hand, history dependence allows for more redistribution and on the other hand it allows for more targeted redistribution. Initially high-ability types gain from simplicity because of the lower overall level of redistribution. Eliminating history dependence drastically reduces the possibilities of the planner to engage in tagging and tailor labor wedges to conditional skill distributions. Initially high skilled people therefore face much higher labor wedges over the life cycle on average with history dependence than without.

Initially low-ability types also profit from simplicity because redistribution is less targeted; at least for moderate levels of risk-aversion. A lot of redistribution in the full optimum is being carried out by differential lump sum payments across different histories. Redistribution in the full optimum is, hence, very targeted. With only age-dependency, redistribution through differentiated transfers is by definition not possible and higher lump sum transfers always strongly benefit the initially-low skilled.

## 3.4 Age-Independent Labor Taxation

Next we restrict the ability of the planner to condition labor income taxes on age. Optimal marginal tax rates follow similar forces as in the age-dependent case. For ease of exposition, we only state the formula in terms of a tax perturbation:[21]

$$\frac{\mathcal{T}'(y(\theta))}{1 - \mathcal{T}'(y(\theta))} = \left(1 + \frac{1}{\varepsilon}\right) \frac{1}{\theta f^*(\theta)} \times \left[\sum_{t=1}^{T} \mathcal{M}_t(\theta) + \mathcal{S}^I(\theta)\right], \tag{16}$$

where

$$f^*(\theta) = \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} f_t(\theta|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}, \tag{17}$$

$$\mathcal{S}^I(\theta) = \tau_s^I \sum_{j=2}^{T} (1+r)^{j-2} \int_{\Theta^{j-1}} \int_{\underline{\theta}}^{\overline{\theta}} \frac{\partial a_j(\theta^{j-1})}{\partial \mathcal{T}(y(\tilde{\theta}))} d\tilde{\theta} h_{j-1}(\theta^{j-1}) d\theta^{j-1}. \tag{18}$$

In Section 4, we focus in detail on the general properties and shape of age-independent non-linear labor taxes. There, we also derive a result for the optimal top tax rate, if the right tail of the income distribution is Pareto distributed. As before in the case with age-dependent taxes, it easy to show that at the bottom:

$$\mathcal{T}'(y(\underline{\theta})) = 0.$$

Conceptually, the major difference to age dependence is simply that labor supply incentives, the mechanical effects and the savings effect are averaged across age groups – as a consequence, optimal tax rates in Figure 3(a) lie in between the age-dependent rates. The optimal linear tax rate on the capital stock is 4.7%, hence slightly lower than for optimal age-dependent labor income taxes.

The aggregate welfare losses from getting rid of age-dependency range from 0.11% to 0.27%. Remarkably, these losses are very similar as the in the previous simplification. Who wins and who loses from age independence (as compared to age dependence)? Interestingly, all individuals except for the lower quintile, lose from the simplification. Moreover, the losses are increasing in the initial skill level and at the top of the distribution become an order of magnitude larger than the aggregate Utilitarian losses. What explains this? Age-dependent taxes are higher for the middle and old aged relative to the young for insurance reasons. In the age-independent case, taxes are therefore higher for the young than they would be with age-dependent taxation. This increase in taxes for the young also redistributes from the initially high skilled to to the initially low skilled, which is why the losses are larger for initially skilled individuals. In his study on age-dependent taxes, Weinzierl (2011) finds that it is mostly the low skilled who gain from age-dependent and sophisticated policies. Interestingly, our results here

---

[21]For a version of the formula in terms of the Lagrangian multipliers on the Euler equation, see Appendix B.2.1.

(a) Optimal nonlinear labor income taxes

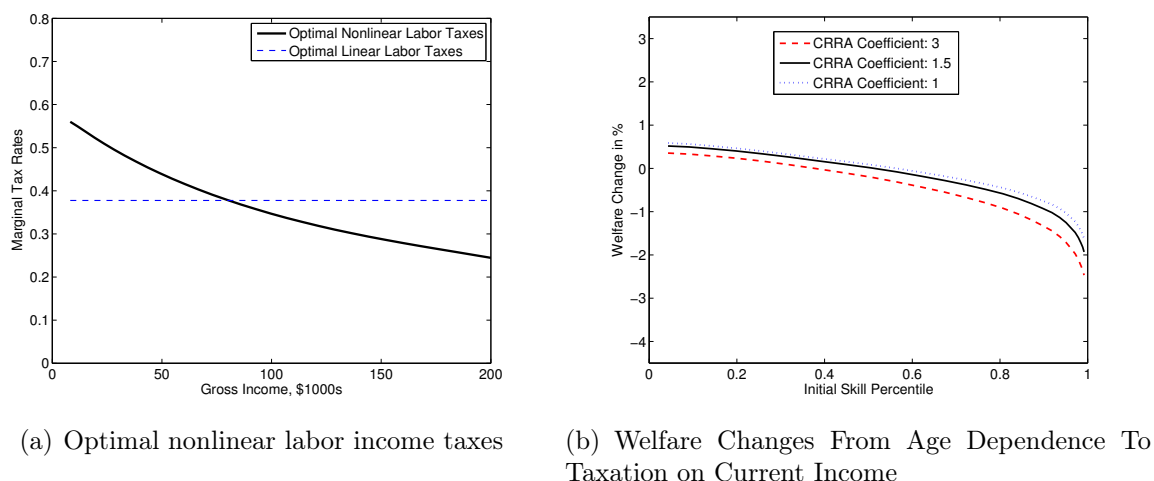(b) Welfare Changes From Age Dependence To Taxation on Current Income

Figure 3: Welfare Changes From Restricting Labor Wedges
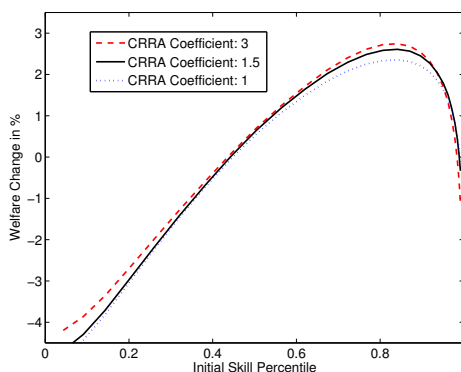


Figure 4: Loss from Linear Taxes Relative To Optimal Nonlinear

differ slightly in this regard. In line with Weinzierl (2011), we find that age dependence delivers roughly half of the welfare gains from history dependence over age-independent policies.[22]

**Comparison To Linear Taxes.** The last question we want to answer is how nonlinear history-independent labor taxes compare to a simple flat tax system in the dynamic life cycle model? The aggregate welfare losses are significant and vary between 0.06% and 0.39%. In Figure 4 we illustrate the distributional consequences of the linearity restriction, compared to the optimal nolinear policy. Low ability individuals suffer large welfare losses of 4% and more. This is because the nonlinear scheme entails much more redistribution through the lump sum element. There is a break point above the 40th percentile, where individuals actually gain from linearity, as their tax burdens are brought down by the flat tax. Although these individuals lose some insurance from the linearity, they still benefit on net terms. Finally, there is another breakpoint and for very high ability types around the 90th percentile, the change in expected

---

[22]The most suitable comparison here is log-utility because Weinzierl (2011) quantitatively studies log-utility.

| | $\rho = 1$ | $\rho = 1.5$ | $\rho = 3$ |
|---|---|---|---|
| History-Dependent Labor Wedges and Linear Capital Wedges | - 0.02% | - 0.03% | - 0.13% |
| Age-Dependent Labor Taxes | - 0.12% | - 0.20% | - 0.43% |
| Age-Independent Labor Taxes | - 0.23% | - 0.34% | - 0.70% |
| Linear Labor Taxes | - 0.29% | - 0.48% | - 1.09% |

Table 1: Utilitarian Welfare Relative to the Full Optimum with Simpler Instruments

lifetime utility decreases with skill level. Very high skilled individuals are better off under the optimal nonlinear scheme This is because for sufficiently high income levels, the life cycle tax burden is lower under the linear scheme since optimal nonlinear marginal tax rates are decreasing, see Figure 3(a).

## 3.5 Summary of Welfare Losses from Simplicity

Summing up, the main contribution and goal of Section 3 is to understand the role of simplicity in the form of history independence relative to more complex tax systems. In particular, we contribute to and complement the *NDPF* literature. Table 1 summarizes the results. History dependence in the capital wedge has small welfare consequences, esp. if risk aversion is moderate – on average (i.e. simply averaged over the three CRRA scenarios) the welfare loss is only 0.06%, as can be seem from the first line in the table. For labor wedges, the interesting conclusion from the aggregate welfare analysis is that each "degree of sophistication" for labor taxes approximately adds the same welfare benefits: Starting from a linear tax system, nonlinear taxation adds approximately the same welfare improvements as moving from age-independent to age-dependent taxation and from age dependence to history dependence. The average welfare losses relative to the full optimum across the three levels of risk aversion we report are 0.25% for age dependence, 0.42% for age independence and 0.62% for linear taxes. This implies relative welfare losses of 0.19% from removing history dependence (0.25%-0.06%), 0.17% from removing age dependence (0.42%-0.25%), and 0.20% from removing nonlinearities in the labor tax code (0.62%-0.42%). Including higher levels for $\rho$ than 3 in the analysis gives similar patterns of approximately equal welfare differences. The relative losses from linearity, however, become more important the larger the level of risk aversion.

Our results are consistent with Farhi and Werning (2013), who also relate linear age-independent policies to the full optimum. Comparing our results with the caveat in mind that they use different preferences and a different calibration, they find a welfare loss of 0.30% using log utility for optimal age-independent linear policies, close to our number of 0.29%. As our results show, however, the welfare losses are strongly increasing in risk aversion, especially for linear policies. Restricting their analysis to linear labor wedges, Farhi and Werning (2013) also find that the welfare losses from age independence are about twice as large as under age dependence. We find a similar result for nonlinear policies when risk aversion is low ($\rho$ around

1), however, when risk aversion is larger age-independent taxes gain ground relative to age-dependent policies. In the broader picture, our analysis contributes to the previous literature by showing that under age independence, nonlinear income taxes can substantially improve upon the simple linear schemes. This is potentially important as age dependence may never be fully realized in real world tax policies.

We also investigate the distributional consequences of simplicity and find that they vary significantly by scenario. Removing history dependence in labor wedges leads to U-shaped welfare changes in initial skill type. Initial skill types around the median lose because of the simplification. Initially high or low skill types may actually gain from this simplification. Eliminating age dependence leads to welfare losses which are increasing in initial skill type. The lowest initial skill types gain from this simplification. The reason is that in the age-independent case, taxes are higher for the young than they would be with age-dependent taxation. This increase in taxes for the young also redistributes from the initially high skilled to to the initially low skilled, which is why the losses are larger for initially skilled individuals. Finally, by far the largest distributional impact comes from restricting labor taxes to be linear. As the linear schemes offer relatively little redistribution, the low skilled suffer losses which are an order of magnitude larger than the aggregate losses.

# 4    A Characterization of Optimal Simple Tax Instruments

In the previous sections we have provided the formal foundations for studying history-independent tax functions in dynamic stochastic environments and have related it to more complex allocations in terms of welfare.

For policy implications it is of first-order importance to understand the properties of optimal history-independent taxes. How should governments design history-independent tax schedules? What is the role of capital taxation if labor income taxes are constrained to condition on current income only? In this section, we elaborate on these questions in greater detail. In Section 4.1, we use our approach to study the optimal capital tax rate for some given income tax function and vice versa. In Section 4.2, we provide a comprehensive decomposition of the *mechanical effect* into a redistribution and an insurance component, where the latter provides a meaningful lower bound on optimal taxes. In Section 4.3, we show how optimal properties of history independent labor income taxes relate to the properties of the optimal static income tax function. One important difference between the static and the dynamic perspective is the mentioned insurance motive for income taxes in dynamic stochastic environments. We show that this can lead to either large or small differences in optimal static and dynamic tax schedules, depending on the social welfare function.

## 4.1 The Interaction Between Capital and Labor Income Taxes

An advantage of our history-independent tax approach is that it also allows to study optimal capital taxation given some labor income tax schedule and vice versa. Studying this interaction between capital and labor income taxation is not possible in the static Mirrlees model. This interaction is, in general, also not studied in a dynamic Mirrlees approach, since both labor and capital wedges are part of the dynamic mechanism design solution.

### 4.1.1 Optimal Capital Taxes for Given Labor Taxes

The formula for the optimal linear capital tax rate is given by:

$$
\frac{\tau_s}{1-\tau_s} = \frac{\sum_{t=2}^{T} \frac{1}{(1+r)^{t-2}} \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \left[ h_{t-1}(\theta^{t-1}) - \int_{\Theta} \frac{U'(\theta^{t-1},\tilde{\theta}_t) dF_t(\tilde{\theta}_t|\theta_{t-1})}{\lambda} \tilde{h}_{t-1}(\theta^{t-1}) \right] d\theta^{t-1}}{\sum_{t=2}^{T} \frac{1}{(1+r)^{t-2}} \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \zeta_{a_t,1-\tau_s}(\theta^{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}},
\tag{19}
$$

where $\zeta_{a_t,1-\tau_s}(\theta^{t-1})$ is the elasticity of savings elasticity with respect to (one minus) the tax rate on capital for individuals of history $\theta^{t-1}$. Formula (19) is very general because it holds irrespective of the choice of the labor income tax schedule. We provide a brief intuitive derivation. Assume that starting from the optimal capital tax rate, the government slightly increases it. This small change will mechanically increase government's revenue in present value terms by

$$
\sum_{t=2}^{T} \frac{1}{(1+r)^{t-2}} \int_{\Theta^{t-1}} a_t(\theta^{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}.
\tag{20}
$$

The tax increase decreases utility of individuals. This impact on the planner's objective (in terms of public funds) is given by

$$
\sum_{t=2}^{T} \beta^{t-1}(1+r) \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \int_{\underline{\theta}}^{\overline{\theta}} \frac{U'(\theta^{t-1},\tilde{\theta})}{\lambda} dF_t(\tilde{\theta}|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1}.
\tag{21}
$$

It also influences the savings decision of individuals, which has no first-order impact on individual utility but on public funds, which is given by:

$$
\begin{aligned}
& \sum_{t=2}^{T} \frac{\tau_s}{(1+r)^{t-2}} \int_{\Theta^{t-1}} \frac{\partial a_t(\theta^{t-1})}{\partial \tau_s} h(\theta^{t-1}) d\theta^{t-1} \\
& = \sum_{t=2}^{T} -\frac{1}{(1+r)^{t-2}} \frac{\tau_s}{1-\tau_s} \int_{\Theta^{t-1}} \zeta_{a_t,1-\tau_s}(\theta^{t-1}) a_t(\theta^{t-1}) h(\theta^{t-1}) d\theta^{t-1}.
\end{aligned}
\tag{22}
$$

Note that for $\tau_s = 0$ this effect is of second order indicating that increasing or decreasing $\tau$ from zero has no first-order incentive costs and a non-zero capital tax is desirable whenever $(20)|_{\tau=0} + (21)|_{\tau=0} \neq 0$. For $\tau \neq 0$, however, it holds that $(22) \neq 0$. Optimality of $\tau_s$ then requires $(20) + (21) + (22) = 0$, which yields (19) (assuming $\beta(1+r) = 1$). Developing a novel dynamic tax reform approach, Golosov, Tsyvinski, and Werquin (2014) look at the welfare effects of an increase of a linear capital tax rate starting from any given tax system and obtain a formula similar to (19). In Appendix B.1.2 we also provide a first-order condition for the optimal linear capital tax rate in terms of the Lagrangian multiplier functions on the Euler equations.

Whereas formula (19) applies for optimal and suboptimal labor income taxes, its quantitative implications are sensitive with respect to the labor income tax schedule. On the one hand, the more progressive the labor income tax function, the less concentrated is wealth, which lowers the power of capital taxes for redistribution and insurance as captured by the numerator of (19). On the other hand, the lower the level of labor income taxation, the less insured individuals are against labor income risk and the stronger the need for self-insurance through savings. A strong need for self-insurance implies a lower responsiveness of savings with respect to capital taxes. Both effects call for higher capital taxation if labor income taxes are lower.

We quantitatively explore this issue and consider the optimal capital tax given (i) the optimal labor income tax (as in Section 3.4), (ii) an approximation of the current US labor income tax schedule[23], (iii) zero labor income taxes and (iv) a very high progressive labor income tax code which is as in (ii) but marginal tax rates are increased by 50 percentage points at each point. In terms of the amount of redistribution carried through the labor tax system, scenario (iv) features the most progressive schedule, followed by the optimal schedule (i), and then the current schedule (iii) and finally the zero tax case.

Table 2 summarizes the results. We recalculate $\tau_s$ such that it can be interpreted as an annual tax rate on capital income, following Farhi, Sleet, Werning, and Yeltekin (2013).[24] There are very large differences across the four scenarios. The optimal tax rate on annual capital income varies between 3% and 51%. In line with the argumentation from above, the less progressive the labor income tax schedule the higher the optimal capital tax rate and also the implied welfare gains from capital taxation.[25] Wealth inequality increases if the labor tax code is less

---

[23]Here, we use the Gouveia-Strauss specification of Guner, Kaygusuz, and Ventura (2013), who provide different parametric approximations of the US tax system.

[24]As Farhi, Sleet, Werning, and Yeltekin (2013), we calculate the tax on annual capital income as a relative geometric average $\tau_{CI,annual}$ according to

$$1 - \tau_{CI,annual} = \frac{((1-\tau_s)*(1+r))^{1/T} - 1}{(1+r)^{1/T} - 1},$$

where $T = 13$ is the number of years per model period and and $r = 1.03^{13} - 1$.

[25]For brevity, we do not provide comparative statics w.r.t. the CRRA coefficient. A higher CRRA coefficient increases both, the size and the welfare gain of capital taxation, see also an earlier version of this paper (Findeisen and Sachs 2014).

| | Form of Labor Taxation | | | |
|---|---|---|---|---|
| | Optimal Nonlinear (Mirrlees) | Current System | Zero Labor Taxes | High and Progressive |
| Optimal Tax Rate on Annual Capital Income | 14.82% | 31.28% | 51.18% | 2.97% |
| Welfare Gain From Taxing Capital | 0.05% | 0.23% | 0.83% | $\approx 0.00\%$ |

Table 2: Optimal Capital Tax Rates For Given Labor Tax Systems (Utilitarian Planner, CRRA Coefficient of 1.5)

redistributive: going from the case with zero labor taxes to the high tax scenario (iv), the wealth share held by the top 10% of the wealth distribution in the last period increases from 64% to over 80%.

As argued above, a more progressive labor income tax schedule calls for lower capital income taxation not only because of lower wealth concentration but also because individuals need to provide less self-insurance through precautionary savings and should therefore be more elastic w.r.t. capital taxes in their savings decision. To disentangle these effects, we decompose (19) to see what part of the variation in optimal capital tax rates across the four scenarios is driven by the the desire to redistribute wealth, captured by the numerator, and behavioral responses, captured by the denominator. Going from the zero labor tax case to the high tax scenario (iv), the numerator decreases by a factor around six and the denominator increases by a factor around three. Wealth inequality and the strength of behavioral responses, hence, both matter in explaining the variation in optimal capital tax rates and the decomposition results imply a larger role for the wealth inequality component.

### 4.1.2  Optimal Labor Income Taxes for Given Capital Taxes

The optimal labor income tax schedule is still described by (16) – the formula holds for any level of the capital tax rate $\tau_s$. The optimal labor income tax schedule is nevertheless influenced by the degree of capital taxation because the level of the capital tax rate impacts the terms $\mathcal{M}_t(\theta_t)$ and $\mathcal{S}^I(\theta)$. The higher the capital tax rate, the weaker the desire for redistribution through labor income taxes as captured by $\mathcal{M}_t(\theta_t)$. The impact on $\mathcal{S}^I(\theta)$ is ambiguous.

Our quantitative exploration reveals that the level of the capital tax rate is relatively unimportant for the design of optimal labor income taxes. For example, if we compare the two extreme cases (i) zero capital taxes and (ii) very high capital taxes of 18% (which was the highest level in Table 2), the difference in the optimal labor income tax schedule is relatively small. The schedules differ by at most two percentage points.

This result strongly contrasts with the finding that the optimal capital tax rate is strongly influenced by the labor income tax schedule. What is driving this asymmetry? The reason is that capital taxes hardly influence the redistributive power of labor income taxes. In particular,

the distribution of labor income which is of key importance for optimal marginal tax rates, is not affected by the level of capital income taxation. The level of the capital tax only influences the mechanical effects (13) and the savings effect (18). However, these forces are of second order numerically. By contrast, the redistributive power of capital taxes heavily depends on inequality in savings, which in turn is strongly affected by the labor income tax code.

### 4.1.3 A Naive Planner: Using a Static Labor Tax Formula

We now look at a planner who ignores the direct impact of labor taxes on savings tax revenue. In other words, the planner does not take into account the direct effect of capital tax rates on labor taxes. The planner in this scenario uses the optimal tax formula (16) with (18) set to zero. Thus, a government is considered that does not take into account the impact on public funds through the implied savings responses of individuals. Our simulation reveals that both, the effects on the optimal tax schedule and on welfare of using this static formula is minuscule. The welfare loss is only 0.01% in consumption equivalents. This in line with the findings from Section 4.1.2 that the level of the capital tax does hardly influence optimal labor tax rates.

## 4.2 Disentangling Redistribution and Insurance Motives

We now show how the role of income taxes in a dynamic environment can be cleanly decomposed into an insurance and a redistribution component.[26] Specifically, we decompose the mechanical effect as defined in Section 3.3. Recall its definition for an increase of the marginal tax rate at income level $y(\theta)$:

$$M_t(\theta) = \int_{\Theta^{t-1}} \int_\theta^{\bar\theta} \left( \frac{\lambda h_t(\theta^{t-1})}{(1+r)^{t-1}} - \frac{U'(\theta^{t-1}, \theta_t)\tilde h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} \right) dF_t(\theta_t|\theta_{t-1})d\theta^{t-1}. \tag{23}$$

$M_1(\theta)$ measures the redistributive gain from income taxation in period one. $M_t(\theta)$ with $t > 1$ captures both, welfare gains of taxation from redistribution between ex-ante heterogeneous individuals and insurance against idiosyncratic uncertainty in the $t$-th period. Whereas the gains from redistribution depend on the particular set of Pareto weights, gains from insurance are independent of the welfare criterion – this is formalized in the next proposition:

**Proposition 3.** *The mechanical effect can be decomposed into two parts:*

$$\mathcal{M}_t(\theta) = \mathcal{M}_t^I(\theta) + \mathcal{M}_t^R(\theta),$$

---

[26]Relatedly, Boadway and Sato (2012) derive a formula for the optimal marginal tax rate in a static setting with heterogeneity and uncertainty. They also show how their formula addresses the desire to redistribute *and* to provide insurance. Their timing is different, however. In their setup, individuals do not perfectly know the gross income they will earn when making their labor supply decision because gross income will be a function of labor supply and a stochastic term.

*where*

$$\mathcal{M}_t^I(\theta) = \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \left[ (1 - F_t(\theta|\theta_{t-1})) - \mathcal{CU}(\theta^{t-1};\theta) \right] h_{t-1}(\theta^{t-1}) d\theta^{t-1} \qquad (24)$$

*and*

$$\mathcal{M}_t^R(\theta) = \int_{\Theta^{t-1}} \left( \frac{\lambda h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} - \frac{\tilde{h}_{t-1}(\theta^{t-1}) \int_{\theta_t} U'(\theta^{t-1},\theta_t)) dF_t(\theta_t|\theta_{t-1})}{(1+r)^{t-1}} \right) \times \mathcal{CU}(\theta^{t-1};\theta) d\theta^{t-1}$$

$$(25)$$

*and*

$$\mathcal{CU}(\theta^{t-1};\theta) = \frac{\int_{\underline{\theta}}^{\overline{\theta}} U'(\theta^{t-1},\theta_t) dF_t(\theta_t|\theta_{t-1})}{\int_{\underline{\theta}}^{\overline{\theta}} U'(\theta^{t-1},\theta_t) dF_t(\theta_t|\theta_{t-1})}.$$

It can easily be shown that adding (24) and (25) yields (23). Whereas (25) depends on the set of Pareto weights, (24) is independent of the redistributive preferences. In the following we use an intuitive perturbation argument to show that $\mathcal{M}_t^I(\theta)$ indeed captures the mechanical insurance value of taxation and $\mathcal{M}_t^R(\theta)$ captures the mechanical impact of taxation on welfare through redistribution between ex-ante heterogeneous agents.

For this purpose we slightly reinterpret the classical tax perturbation method. For ease of exposition, we consider the case without capital taxation.[27] Assume that $\mathcal{T}'(y(\theta))$ is increased in a way that everyone with income above $y(\theta)$ now pays one more dollar of labor income taxes. Then, assume that the additional tax revenue generated by this increase is redistributed in a lump sum fashion; this lump sum increase has no first-order impact on welfare via the implied responses of savings behavior.[28] The uniform lump-sum tax decrease is given by

$$\Delta(\theta) = \frac{1 - F^*(\theta)}{\sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}}},$$

where

$$1 - F^*(\theta) = 1 - F_1(\theta) + \sum_{t=2}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} (1 - F_t(\theta|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}. \qquad (26)$$

Individuals with $\theta_1 < \theta$ will enjoy higher period one consumption of $\Delta(\theta)$, whereas period one consumption for individuals with $\theta_1 \geq \theta$ will decrease by $1 - \Delta(\theta)$. Whether period $t$ consumption will be increased or decreased depends on the realization of the shock. From a period $t - 1$ perspective, the reform of the tax function in period $t$ will provide an insurance

---

[27]The presence of a non-zero capital tax $\tau_s$ has not impact on the decomposition in Proposition 3, however, the perturbation argument is a bit more involved. In Appendix B.3 we show how the perturbation argument differs in that case.

[28]The reason is that the change in behavior has no first-order impact on individual utilities by the envelope theorem and that there is no effect on the government budget because wealth taxes are zero. An equivalent option would be to redistribute the lump sum in the first period. But also in the presence of a savings tax a slightly related tax reform can be constructed to obtain the decomposition of the mechanical effect. See Appendix B.3.

value. To obtain an expression for this insurance value, define for each type $\theta^{t-1}$ a 'constant utility term':

$$\mathcal{CU}(\theta^{t-1};\theta) = \frac{\int_{\underline{\theta}}^{\bar{\theta}} U'(\theta^{t-1},\theta_t)dF_t(\theta_t|\theta_{t-1})}{\int_{\underline{\theta}}^{\bar{\theta}} U'(\theta^{t-1},\theta_t)dF_t(\theta_t|\theta_{t-1})}.$$

The numerator captures the (expected) utility loss in period $t$ due to the tax increase (absent the lump-sum tax adjustment). Dividing it by the expected marginal utility in period $t$ says by how much consumption had to be increased in period $t$, for every possible realization of the $t$-period shock, in order to make the individual of type $\theta^{t-1}$ in expectation equally well of. This number is smaller than one because (i) the tax increase in period $t$ affects the individual in period $t$ with probability less than one and because of (ii) risk aversion.

To measure the welfare gain through this insurance role of income taxation, we ask the following question: If the government could increase the lump-sum transfer in period $t$ by a different amount for each $\theta^{t-1}$-type such that expected period $t$ utilities are unchanged for all $\theta^{t-1}$-types (i.e. by $\mathcal{CU}(\theta^{t-1};\theta)$ respectively), how much resources could the government save due to this insurance against income risk? From each individual of type $\theta^{t-1}$, the government obtains tax revenue of $\frac{1-F_t(\theta|\theta_{t-1})}{(1+r)^{t-1}}$ in present value terms. To hold utility constant for that individual from a period $t-1$ perspective, only $\frac{\mathcal{CU}(\theta^{t-1};\theta)}{(1+r)^{t-1}}$ of resources have to be spent (in present value terms). Adding up and integrating over all histories yields (24). $\mathcal{M}_t^I(\theta)$ reflects the gains from insurance for individuals in period $t-1$ against their period $t$ shocks. This insurance gain is measured in first period resources. The more pronounced labor income risk, conditional on $\theta^{t-1}$, and the stronger risk aversion, the larger is this insurance effect. It is simple to show that it is always positive.

Given that the lump-sum transfer in period $t$ is not increased by $\mathcal{CU}(\theta^{t-1};\theta)$ for type $\theta^{t-1}$ but by $\Delta(\theta)$ instead, expected period $t$ utility will not stay constant. Instead expected utility will be increased by

$$\mathcal{R}(\theta^{t-1};\theta) = \Delta(\theta) - \mathcal{CU}(\theta^{t-1};\theta)$$

in monetary terms for individuals of type $\theta^{t-1}$. $\mathcal{R}(\theta^{t-1};\theta)$ captures the redistributive element of this tax reform for period $t$. The government necessarily redistributes between different $\theta^{t-1}$-types also in period $t$. We now derive the welfare consequences of this implied redistribution. Recall that $\mathcal{R}(\theta^{t-1};\theta)$ measures the (possibly negative) expected utility increase of type $\theta^{t-1}$ in monetary terms. A marginal increase in consumption in period $t$ for individuals of type $\theta^{t-1}$ (and for each realization of the shock $\theta_t$) is valued

$$\beta^{t-1} \int_{\Theta} U'(\theta^{t-1},\theta_t)dF_t(\theta_t|\theta_{t-1})\tilde{h}_{t-1}(\theta^{t-1}) - \frac{\lambda}{(1+r)^{t-1}}h_{t-1}(\theta^{t-1})$$
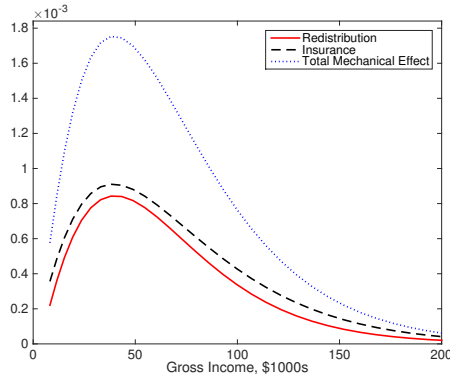
Figure 5: Insurance and Redistribution Part of Optimal Taxes

by the planner. Thus, aggregating over all types and weighing by $\mathcal{R}(\theta^{t-1}; \theta)$ yields:

$$\mathcal{M}_t^{R*}(\theta) = \int_{\Theta^{t-1}} \left( \frac{\lambda h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} - \frac{\tilde{h}_{t-1}(\theta^{t-1}) \int_{\theta_t} U'(\theta^{t-1}, \theta_t) dF_t(\theta_t|\theta_{t-1})}{(1+r)^{t-1}} \right)$$
$$\times \left( \mathcal{C}\mathcal{U}(\theta^{t-1}; \theta) - \Delta(\theta) \right) d\theta^{t-1}. \tag{27}$$

The last term $\Delta(\theta)$ can now be ignored because it is independent of $\theta^{t-1}$ and because it is possible to show that

$$\int_{\Theta^{t-1}} \left( \frac{\lambda h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} - \frac{\tilde{h}_{t-1}(\theta^{t-1}) \int_{\theta_t} U'(\theta^{t-1}, \theta_t) dF_t(\theta_t|\theta_{t-1})}{(1+r)^{t-1}} \right) d\theta^{t-1} = 0.$$

This follows from the expression for $\lambda$ which says that average period one social marginal utility is equal to $\lambda$ (see (41) in Appendix B.1.3) and the Euler equations. Combining this with (27) yields $\mathcal{M}_t^R(\theta)$ as defined in (25).

These arguments reveal that (25) indeed measures the welfare effect through redistribution between ex-ante heterogeneous agents. The $\mathcal{C}\mathcal{U}(\theta^{t-1}; \theta)$-term should be higher for high $\theta_{t-1}$-types because they are likely to draw a better shock in Period $t$. For redistributive Pareto-weights (i.e. $F(\theta_1) \leq \tilde{F}(\theta_1)$ for each $\theta_1$) the term $\mathcal{M}_t^R(\theta)$ should therefore be positive at every $\theta$. If Pareto weights are sufficiently strong in favor of high innate types, the welfare effect from redistribution can certainly be negative.

**Quantitative Illustration of Decomposition.** Figure 4.2 illustrates the insurance and redistribution components of the mechanical effect at the Utilitarian optimum as discussed in Section 3.4. We plot each component added up over all periods: $\sum_{t=1}^{3} \mathcal{M}_t(\theta)$, $\sum_{t=1}^{3} \mathcal{M}_t^I(\theta)$ and $\sum_{t=1}^{3} \mathcal{M}_t^R(\theta)$. By definition, $\mathcal{M}_1^I(\theta) = 0$, as in the first period taxes purely redistribute but do not provide any insurance value. All three functions are hump-shaped. Moreover, the insurance and redistribution co-move closely across the income distribution. Interestingly,

28

the insurance component is slightly bigger than the redistribution component for most of the income distribution. In the next section, we shed light on how the shape of the mechanical effect influences the shape of optimal tax rates.

## 4.3 Relation To The Seminal Results of The Static Optimal Tax Problem

We next turn to the question how optimal history-independent taxes compare to the policy prescriptions of the static model. Diamond (1998) investigates the main drivers of the shape of optimal tax rates in the static model. We start by comparing the dynamic case to the static one in the spirit of the Diamond paper. Next, we reinvestigate results for the optimal tax rate under the limit assumption that the right tail of the income distribution is Pareto distributed.

### 4.3.1 The Shape of Marginal Tax Rates: Dynamic versus Static

For a static environment, Diamond (1998) investigates conditions, which help to understand the shape of optimal taxes. In this section, we study the shape of optimal history-independent tax rates in dynamic environments and compare them to the static model. Recall that age-independent Pareto optimal marginal labor income tax rates are given by:

$$\frac{\mathcal{T}'(y(\theta))}{1 - \mathcal{T}'(y(\theta))} = \left(1 + \frac{1}{\varepsilon}\right) \frac{1}{\theta f^*(\theta)} \times \left[\sum_{t=1}^{T} \mathcal{M}_t(\theta) + \mathcal{S}^I(\theta)\right],$$

where $f^*(\theta), \mathcal{M}_t(\theta)$ and $\mathcal{S}^I(\theta)$ are defined by (17), (13) and (18). Inspired by Diamond (1998), we decompose our formula into three parts:

$$\frac{\mathcal{T}'(y(\theta))}{1 - \mathcal{T}'(y(\theta))} = \mathcal{A}(y(\theta)) \cdot \mathcal{B}(y(\theta)) \cdot \mathcal{C}(y(\theta))$$

where

$$\mathcal{A}(y(\theta)) = \left(1 + \frac{1}{\varepsilon}\right), \quad \mathcal{B}(y(\theta)) = \frac{\sum_{i=1}^{T} \mathcal{M}_i(\theta) + \mathcal{S}^I(\theta)}{1 - F^*(\theta)}, \quad \mathcal{C}(y(\theta)) = \frac{1 - F^*(\theta)}{\theta f^*(\theta)}.$$

with $F^*(\theta)$ as defined in (26). How does each term contribute to the optimal shape of taxes? A standard assumption in the literature is to work with a constant elasticity so that $\mathcal{A}$ is a constant. The term $\mathcal{C}$ is similar as in the static model and captures the share of individuals earning more than $y(\theta)$ over the share of individuals earning $y(\theta)$ multiplied by their productivity $\theta$. For the static case, the cross sectional skill distribution is relevant. Here, in a dynamic lifecycle perspective, it is the ratio of the adjusted lifecycle distribution $F^*(\theta)$ that matters. The role of $F^*(\theta)$ for optimal history-independent taxes is therefore equivalent to the role of the cross-sectional skill distribution for optimal taxes in a static problem.

Term $\mathcal{B}$ differs compared to the static approach for two reasons. First, it is extended by the savings effect $\mathcal{S}^I(\theta)$ that is not present in the static formula. Second, the mechanical effect $\sum_{i=1}^{T} \mathcal{M}_i(\theta)$ does not reflect the welfare effect of taxation through redistribution between ex-ante different agents only but also social insurance against wage risk – see Section 4.2. For ease of exposition, we define $\mathcal{B}_1(\theta) = \frac{\sum_{i=1}^{T} \mathcal{M}_i(\theta)}{1-F^*(\theta)}$ and $\mathcal{B}_2(\theta) = \frac{\mathcal{S}^I(\theta)}{1-F^*(\theta)}$.

We start by explaining the role of $\mathcal{B}_2(\theta)$, which measures the effect of the marginal labor income tax rate on savings tax revenue. Remember from (18) that this depends crucially on how labor tax rates at a given point change savings behavior. To fix ideas consider the following scenario: suppose there are only two periods and tax rates at income level $y$ are increased – this triggers two different kinds of income effects on savings by all agents with income above $y$ today. First, disposable income today is reduced, which reduces savings. Second, the tax increases savings because the expected tax burden in the future period is increased. The sign of $\mathcal{B}_2(\theta)$ depends on the relative strength of these effects. For our quantitative model – calibrated to match a realistic life cycle structure – we find that $\mathcal{B}_2(\theta)$ is positive up to incomes of about \$25,000, so that up to that income level the second effect dominates. Moreover, $\mathcal{B}_2(\theta)$ is monotonically decreasing. Taking together the new term $\mathcal{B}_2(\theta)$ is a force for higher marginal tax rates for low incomes, lower marginal tax rates for higher incomes and therefore contributes to decreasing marginal tax rates over the schedule. Quantitatively however, $\mathcal{B}_2(\theta)$ is extremely small in absolute value. When we relate it to $\mathcal{B}_1(\theta)$, the term $\mathcal{B}_2(\theta)$ is almost two orders of magnitudes smaller on average. This is consistent with the results form Section 4.1.3, where we found only minuscule differences in the optimal allocation if the savings effects, captured by $\mathcal{B}_2(\theta)$, are ignored. Whereas the savings term $\mathcal{B}_2(\theta)$ does hardly influence the shape of optimal tax rates, we now illustrate how $\mathcal{B}_1(\theta)$ can be very different in the static and the dynamic context. This crucially depends on the social welfare function and the taste for redistribution across different ability groups.

**The Role of the Social Welfare Function.** To discuss the role of the social welfare function, we start with a case where we choose Pareto weights $\tilde{f}_1(\theta_1)$ such that in a static world with cross sectional *cdf* $F_1(\theta)$ optimal marginal tax rates would be zero and everybody would consume her income. We refer to this social welfare function as 'laissez-faire' Pareto weights. Figure 6(a) shows that optimal marginal tax rates in the dynamic context differ dramatically from the zero tax rates in the static case. This is entirely driven by the insurance value of taxation in a dynamic setting. Tax rates start at about 25% then fall and finally converge to a level around 10%.

In Figure 6(b), we explore what contributes to this U-shape. First, note that in the dynamic model the $\mathcal{C}$ component is convexly decreasing. It becomes clear that the U-shape in the dynamic case comes from the interaction of $\mathcal{C}$ and $\mathcal{B}$. The shape of $\mathcal{B}$ actually contributes to increasing marginal tax rates, but this is offset for low incomes by the decrease in $\mathcal{C}$.
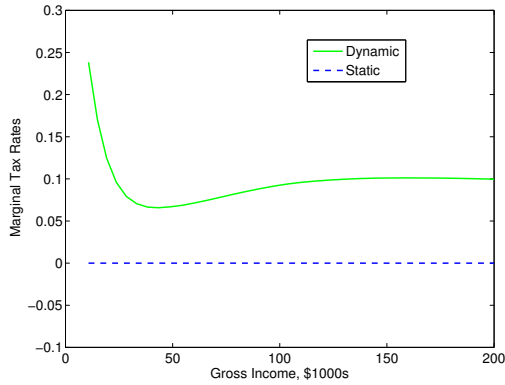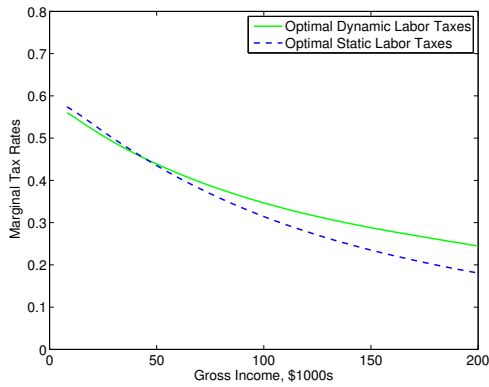
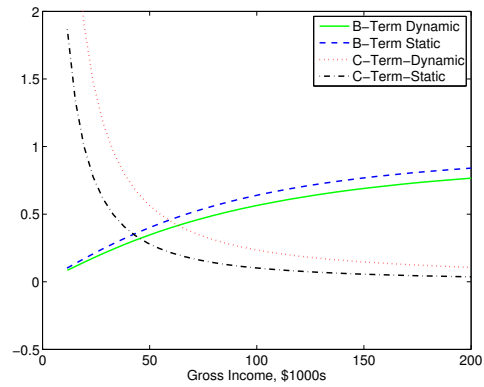(a) Marginal Labor Income Tax Rates
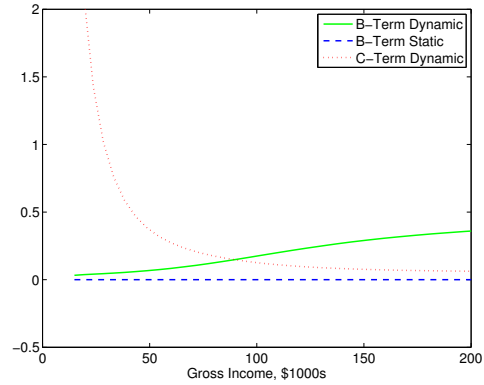
(b) B and C Terms

Figure 6: 'Laissez-Faire' Pareto Weights



(a) Marginal Labor Income Tax Rates

(b) B and C Terms

Figure 7: Utilitarian Pareto Weights

By contrast, for Pareto weights which lead to significant redistribution already in the static model, the shape of tax rates in the dynamic and static case become much more similar. We illustrate this for a Utilitarian social welfare function in Figures 7(a) and 7(b). The differences in optimal tax policies is much smaller. This is intuitive: as is well-known, a Utilitarian social welfare function in the static context can be interpreted as expected utility maximization before the ability type is revealed (behind a veil of ignorance). This implies that the static and the dynamic planner both share a similar desire for redistribution: the $\mathcal{B}$-terms are both concave and increasing.[29] Taxes are slightly higher in the dynamic case for higher income levels which is mainly driven by differences in the static and dynamic $\mathcal{C}$-term.

---

[29]The $\mathcal{B}$ term in the static case is slightly larger here, because marginal utility of the very low skilled is slightly higher in the static case compared to the dynamic case. This is driven by the fact that wages grow from the first to the second period so that individuals can smooth consumption and because shocks are not perfectly persistent, implying that some people are only low skilled for a limited amount of time, allowing them to self-insure.

Summing up, the shape and level of labor income tax rates can be very different taking either a dynamic or a static perspective. The difference is more pronounced when the social welfare function does not imply a lot of redistribution from a static perspective. Intuitively, in the dynamic model even without a taste for redistribution, there remains the insurance motive of taxation. The differences vanish, when the social welfare function already implies a lot of redistribution from a static perspective. In this section, we illustrated this insight by exploring 'laissez-faire' Pareto weights and the Utilitarian case. The intuition is more general, however – in unreported results, we have also compared other other points on the Pareto frontier, finding the same patterns.

### 4.3.2 The Top Tax Rate

We now provide a limit result for the optimal labor tax rate at the top of the income distribution. With our stochastic multi-period environment, this forms the counterpart to the famous Saez (2001) formula for the static model. We follow Saez and express the optimal tax formula in terms of income instead of skill distributions; formally, let $e_{t|t-1}(y_t|y_{t-1})$ denote the conditional distribution in period $t$. The standard formula assumes, in line with the data, that the right tail of the cross-sectional income distribution is Pareto distributed. In the dynamic model more assumptions are needed. One needs an assumption for the right tail for every conditional distribution $e_{t|t-1}(y_t|y_{t-1})$ in every time period $t$ and for any previous income level $y_{t-1}$. Given assumptions on the value of the Pareto parameter for each conditional distribution, it would then be possible to derive a top tax rate formula, depending on all Pareto parameters across the conditional distributions.

In what follows, instead of placing assumptions on every single conditional distribution $e_{t|t-1}(y_t|y_{t-1})$, we consider a scenario where Pareto parameters differ by age groups only. We do this for two reasons. First, there is no empirical guidance on how Pareto parameters differ across different conditional distributions, however, some recent papers have show how these parameters differ by age groups. Second, it simplifies the exposition while delivering the same intuition. Formally, we assume that at every age above a high-income threshold $y^H$, earnings are Pareto distributed with an age-dependent parameter $a_t$. We obtain the following top tax result for the dynamic economy:

**Proposition 4.** *Assume that welfare weights for top incomes in both periods converge to a constant $\overline{\mathcal{W}}$ across age groups and elasticities at the top converge to a constant $\epsilon$. Then the optimal top marginal tax rate above income level $y^H$ satisfies:*

$$\frac{\overline{\mathcal{T}'}}{1-\overline{\mathcal{T}'}} = \frac{\left(1-\overline{\mathcal{W}}\right)\sum_{t=1}^{T}\frac{1}{(1+r)^{t-1}}\frac{1}{a_t-1} + \overline{S}}{\epsilon \cdot \sum_{t=1}^{T}\frac{1}{(1+r)^{t-1}}\frac{a_t}{a_t-1}}, \tag{28}$$

*where $\overline{S} = \tau_s^I \sum_{t=2}^T \frac{1}{(1+r)^{t-2}} \int_{\Theta^{j-1}} \int_{\theta H}^{\overline{\theta}} \frac{\partial a_j(\theta^{j-1})}{\partial \mathcal{T}(y(\tilde{\theta}))} d\tilde{\theta} h_{j-1}(\theta^{j-1}) d\theta^{j-1}/y^H$ captures the fiscal externality on capital taxes, normalized by the Pareto threshold, and, as stated above the welfare weights converge:* $\int_{\Theta^{t-1}} \int_{\theta H}^{\overline{\theta}} \beta^{t-1} U'(\theta^{j-1}, \tilde{\theta}) \tilde{h}_{t-1}(\theta^{t-1}) f_t(\tilde{\theta}|\theta_{t-1}) d\tilde{\theta} d\theta^{j-1} = \overline{\mathcal{W}}.$

*Proof.* Mechanical effects by a small reform of $\tau$ are given by $\overline{M} = \sum_{t=1}^T (y_t^m - y^H) \frac{1}{(1+r)^{t-1}} \left(1 - \overline{\mathcal{W}}\right)$, where $y_t^m$ is average income above the threshold. Labor supply effects are $\overline{LS} = -\frac{\overline{\mathcal{T}'}}{1-\overline{\mathcal{T}'}} \epsilon \sum_{t=1}^T y_t^m \frac{1}{(1+r)^{t-1}}$. Using $\overline{LS} + \overline{M} + \overline{S} y^H = 0$ and $\frac{a_t}{a_{t-1}} = \frac{y^H}{y_t^m}$ gives the result. $\square$

The optimal top tax rate is decreasing in the Pareto parameters $a_t$. As in the static model, a higher $a$ reflects a thinner tail and less inequality driven by the top. Moreover, we see a savings term $\overline{S}$ as for the general tax formula discussed in the last section: also the top labor income tax rate may affect savings behavior, causing a fiscal externality. Relaxing the assumption of separate Pareto parameters, i.e. assuming $a_t = a$, it is then easy to show that:

$$\frac{\overline{\mathcal{T}'}}{1 - \overline{\mathcal{T}'}} = \frac{1 - \overline{\mathcal{W}} + \frac{(a-1)}{\left(1 + \frac{1}{1+r}\right)} \times \overline{S}}{\epsilon \cdot a}.$$

This formula is the famous top tax prescription from Saez, except for the presence of the savings term $\overline{S}$ weighted by $\frac{(a-1)}{\left(1 + \frac{1}{1+r}\right)}$ – the formula collapses to the static one if $\overline{S} = 0$. Numerically in line with the results of the previous section, we always have found a very small role of $\overline{S}$, hardly influencing optimal top rates.

Moreover, simple numerical exercises on formula (28) reveal that with heterogeneous Pareto tails across different distributions, the relatively thicker tails dominate optimal tax rates. Suppose, for example, we take two age groups and that the Pareto parameter of the young $a_1$ is around 3.5 and that of the old $a_2$ is 1.75 – this corresponds to the numbers found for the US for the age groups around 25 years (the young) and 50 years (the old) based on administrative data (Badel and Huggett 2014). Using formula (28), setting $r = \overline{S} = \overline{\mathcal{W}} = 0$ for simplicity and an elasticity parameter of 0.3, would yield an optimal top tax around 61%. In comparison the optimal tax rate would be around 65.5% if both tails were thick ($a_1 = a_2 = 1.75$) but only 48.8% if both tail were thin ($a_1 = a_2 = 3.5$). In this sense, the optimal tax formula allowing for heterogenous Pareto coefficients, strongly overweighs those distributions with a thicker tail. This result is robust when we consider more than just two age groups.

# 5 Conclusion

This paper analyzes Pareto optimal nonlinear taxation of annual labor income as well as linear taxation of capital in a framework with heterogenous agents whose skills evolve stochastically over time. By focusing on preferences without income effects on labor supply, we developed a first-order approach to make this problem tractable for a continuous type space. The paper can be seen as providing a link between the static optimal taxation literature and dynamic

public finance models: whereas we explicitly take into account dynamics and idiosyncratic uncertainty, we optimize over simple tax functions instead of looking at optimal history-dependent distortions.

We have provided a detailed welfare accounting in order to quantify the welfare losses from the different modifications as compared to the constrained optimal allocations. We found that the following restrictions on labor distortions yield welfare losses of similar size: from history dependence to age dependence, from age dependence to age independence and finally from age independence to linear (and age-independent) labor income taxes.

We also characterized properties of optimal simple tax schedules, both theoretically and quantitatively. The most important difference to the static perspective on optimal income taxation is the additional insurance motive of taxation which provides a meaningful lower bound on taxation. Further, we have quantitatively shown that capital taxation is desirable and the level of the optimal capital tax rate is decreasing the degree of progressivity of the labor income tax schedule.

It is likely that our method to study simple history-independent tax instruments in dynamic environments can also be applied more broadly in other contexts.[30] For example, we have left out an explicit role for retirement savings and different sources of capital income, potentially with stochastic returns. Shourideh (2013) investigates such a model with risk-return trade-offs for different kinds of capital. We have focused on labor supply incentives along the intensive margin and have neglected labor market participation decisions. Jacquet, Lehmann, and Van der Linden (2013) provide a state of the art treatment of a static Mirrlees model with labor market participation decisions in addition to intensive labor supply decisions. Incorporating such realistic features into the life cycle framework with labor income risk and taxation seems to be a fruitful and promising avenue for future research.

Finally, in future research one would like to combine the analysis of history-dependent and history-independent instruments. For example, by restricting certain parts of the redistribution & insurance system to be history independent or by restricting the degree of complexity of history dependence.

---

[30]In Findeisen and Sachs (2013), we use this approach to study optimal education-independent income taxes.

# References

AIYAGARI, S. R. (1994): "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 109(3), 659–684.

——— (1995): "Optimal capital income taxation with incomplete markets, borrowing constraints, and constant discounting," *Journal of Political Economy*, pp. 1158–1175.

ALBANESI, S., AND C. SLEET (2006): "Dynamic optimal taxation with private information," *The Review of Economic Studies*, 73(1), 1.

BADEL, A., AND M. HUGGETT (2014): "Taxing top earners: a human capital perspective," *Federal Reserve Bank of St. Louis Working Paper Series*, (2014-017).

BASTANI, S., S. BLOMQUIST, AND L. MICHELETTO (2013): "The welfare gains of age related optimal income taxation," *International Economic Review*, 54, 1219–1249.

BEST, M., AND H. KLEVEN (2013): "Optimal Income Taxation with Career Effects of Work Effort," *Working Paper, London School of Economics*.

BLOMQUIST, S., AND L. MICHELETTO (2008): "Age-related Optimal Income Taxation*," *The Scandinavian Journal of Economics*, 110(1), 45–71.

BOADWAY, R., AND M. SATO (2012): "Optimal Income Taxation with Uncertain Earnings: A Synthesis," *Working Paper*.

CHETTY, R. (2012): "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply," *Econometrica*, 80(3), 969–1018.

CONESA, J. C., S. KITAO, AND D. KRUEGER (2009): "Taxing Capital? Not a Bad Idea after All!," *The American Economic Review*, pp. 25–48.

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

FARHI, E., C. SLEET, I. WERNING, AND S. YELTEKIN (2013): "Nonlinear Capital Taxation Without Commitment," *The Review of Economic Studies*.

FARHI, E., AND I. WERNING (2013): "Insurance and Taxation over the Life Cycle," *The Review of Economic Studies*.

FINDEISEN, S., AND D. SACHS (2013): "Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans," *Working Paper*.

——— (2014): "Efficient Labor and Capital Income Taxation over the Life Cycle," *Working Paper*.

GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70(3), 569–587.

GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2013): "Redistribution and Social Insurance," *NBER Working Paper No. 17642*.

GOLOSOV, M., AND A. TSYVINSKI (2007): "Optimal taxation with endogenous insurance markets," *Quarterly Journal of Economics*, 122(2), 1769–1787.

GOLOSOV, M., A. TSYVINSKI, AND N. WERQUIN (2014): "A Variational Approach to the Analysis of Tax Systems," *Working Paper*.

GRUBER, J., AND E. SAEZ (2002): "The elasticity of taxable income: evidence and implications," *Journal of Public Economics*, 84(1), 1–32.

GUNER, N., R. KAYGUSUZ, AND G. VENTURA (2013): "Income Taxation of U.S. Households: Facts and Parametric Estimates," *Review of Economic Dynamics (forthcoming)*.

HAMMOND, P. J. (1987): "Markets as constraints: multilateral incentive compatibility in continuum economies," *The Review of Economic Studies*, 54(3), 399–412.

JACOBS, B., AND D. SCHINDLER (2012): "On the desirability of taxing capital income in optimal social insurance," *Journal of Public Economics*, 96(9-10), 853–868.

JACQUET, L., E. LEHMANN, AND B. VAN DER LINDEN (2013): "Optimal redistributive taxation with both extensive and intensive responses," *Journal of Economic Theory*, 148(5), 1770–1805.

JAPPELLI, T., AND L. PISTAFERRI (2010): "The Consumption Response to Income Changes," *Annual Review of Economics*, 2(1), 479–506.

KAPICKA, M. (2006): "Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping," *Review of Economic Dynamics*, 9(4), 612–639.

———— (2013): "Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach," *The Review of Economic Studies*, forthcoming.

KAPICKA, M., AND J. NEIRA (2013): "Optimal taxation in a life-cycle economy with endogenous human capital formation," *Working paper, UC Santa Barbara*.

KLEVEN, H. J., AND E. A. SCHULTZ (2013): "Estimating Taxable Income Responses using Danish Tax Reforms," *American Economic Journal:Economic Policy*, forthcoming.

KOCHERLAKOTA, N. R. (2004): "Wedges and taxes," *The American Economic Review*, 94(2), 109–113.

MEGHIR, C., AND L. PISTAFERRI (2011): "Earnings, consumption and life cycle choices," *Handbook of Labor Economics*, 4, 773–854.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38(2), 175–208.

PAVAN, A., I. SEGAL, AND J. TOIKKA (2013): "Dynamic Mechanism Design: A Myersonian Approach," *Econometrica*, forthcoming.

PIKETTY, T. (1997): "La redistribution fiscale face au chomage," *Revue française d'économie*, 12(1), 157–201.

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

SALANIÉ, B. (2003): *The Economics of Taxation.* The MIT Press.

SHOURIDEH, A. (2013): "Optimal Taxation of Wealthy Individuals," *Working Paper, Wharton School.*

STANTCHEVA, S. (2013): "Optimal Taxation and Human Capital Policies Over the Lifecycle," *Working Paper.*

WEINZIERL, M. (2011): "The Surprising Power of Age-Dependent Taxes," *The Review of Economic Studies*, 78(4), 1490–1518.

WERNING, I. (2011): "Nonlinear Capital Taxation," *Working Paper, MIT.*

# A    Derivation of History-Dependent Formulas

## A.1    Full Optimum

For the history-dependent problem, define $v_t(\theta^t) = U\left(c_t(\theta^t) - \Psi\left(\frac{y_t(\theta^t)}{\theta_t}\right)\right)$ and $\Gamma(v_t) = c_t - \Psi\left(\frac{y_t}{\theta_t}\right)$. Define $V_t(\theta^t) = U(c_t(\theta^t) - \Psi\left(\frac{y_t(\theta^t)}{\theta_t}\right) + \sum_{i=t+1}^{T} \beta^{i-1} \int_{\Theta^i(\theta^t)} \left(U(c_i(\theta^i) - \Psi\left(\frac{y_i(\theta^i)}{\theta_i}\right)\right) h_i(\theta^i) d\theta^i$, where $\Theta^i(\theta^t)$ is the set of histories at $i$ that contains $\theta^t$.

To handle (1), we apply a first-order approach (Golosov, Troshkin, and Tsyvinski 2013, Farhi and Werning 2013), which yields the following envelope conditions:

$$\frac{\partial V(\theta^{t-1}, \theta_t)}{\partial \theta_t} = U_t' \Psi' \frac{y_t(\theta^t)}{\theta_t^2} + \beta \int_{\Theta_{t+1}} V_{t+1}\left(\theta^{t-1}, \theta_t, \theta_{t+1}\right) \frac{\partial f_{t+1}(\theta_{t+1})}{\partial \theta_t} d\theta_{t+1}.$$

We assume that the second order conditions are fulfilled in the following. For a discussion of sufficient-conditions for the first-order approach to be valid, see Golosov, Troshkin, and Tsyvinski (2013). Setting up a Lagrangian and applying integration by parts yields

$$\max_{v_t(\theta^t), y_t(\theta^t)} \mathcal{L} = \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^t} v_t(\theta^t) d\tilde{h}(\theta^t) d\theta^t - \sum_{t=1}^{T} \int_{\Theta^t} \eta_t'(\theta^t) V_t(\theta^t) d\theta^t - \sum_{t=1}^{T} \int_{\Theta^t} \eta_t(\theta^t) U_t' \Psi' \frac{y_t(\theta^t)}{\theta_t^2} d\theta^t$$

$$- \sum_{t=1}^{T} \int_{\Theta^t} \eta_t(\theta^t) \beta \int_{\Theta_{t+1}} V_{t+1}\left(\theta^t, \theta_{t+1}\right) \frac{\partial f_{t+1}(\theta_{t+1})}{\partial \theta_t} d\theta_{t+1}$$

$$+ \lambda \sum_{t=1}^{T} \frac{1}{R^{t-1}} \int_{\Theta^t} \left(y_t(\theta^t) - \Gamma(v_t(\theta^t)) + \Psi\left(\frac{y_t(\theta^t)}{\theta_t}\right)\right) h(\theta^t) d\theta^t,$$

where $\eta_t(\theta^t)$ is the Lagrangian multiplier for the envelope condition for history $\theta^t$. The first-order condition for $v_t(\theta^t)$ reads as

$$\beta^{t-1} \tilde{h}(\theta^t) - \lambda \frac{1}{R^{t-1}} \frac{1}{U_t'(\theta^t)} h(\theta^t) - \sum_{j=1}^{t} \eta_j'(\theta^j) \beta^{t-j} \prod_{i=j+1}^{t} f_i(\theta_i | \theta_{i-1})$$

$$- \eta_t(\theta^t) \gamma_t(\theta^t) \frac{1}{\theta_t} - \sum_{j=1}^{t-1} \eta_j(\theta^j) \beta^{t-j} \frac{\partial f_{j+1}}{\partial \theta_j} \prod_{i=j+2}^{t} f_i(\theta_i | \theta_{i-1}) = 0. \tag{29}$$

Applying this first-order condition also for period $t-1$, solving for $\eta_{t-1}'(\theta^{t-1})$, inserting into (29) and using $\beta R = 1$ yields

$$\lambda \frac{1}{R^{t-1}} h(\theta^t) \left(\frac{1}{U_{t-1}'} - \frac{1}{U_t'}\right) - \eta_t' - \eta_{t-1} \beta \frac{\partial f_t}{\partial \theta_{t-1}} - \eta_t \gamma_t \frac{1}{\theta_t} + \beta f_t \eta_{t-1} \gamma_{t-1} \frac{1}{\theta_{t-1}} = 0 \tag{30}$$

Solving this differential equation, we obtain:

$$\eta_t(\theta_t) = \int_{\theta_t}^{\overline{\theta}} \exp\left(\int_{\theta}^{x} \gamma(s) \frac{1}{s} ds\right) \left\{ \lambda \frac{1}{R^{t-1}} h(\theta^{t-1}, x) \left(\frac{1}{U_t'(\theta^{t-1}, x)} - \frac{1}{U_{t-1}'(\theta^{t-1})}\right) \right.$$

$$\left. + \beta \eta_{t-1} \frac{\partial f_t(x | \theta_{t-1})}{\partial \theta_{t-1}} - \beta f_t(x | \theta_{t-1}) \eta_{t-1} \gamma_{t-1} \frac{1}{\theta_{t-1}} \right\} dx.$$

Next, it is easy to show that the first-order condition for $y_t(\theta^t)$ yields

$$\frac{\tau_{Lt}(\theta^t)}{1 - \tau_{Lt}(\theta^t)} = \left(1 + \frac{1}{\varepsilon}\right) \frac{U'(\theta^t)\eta_t(\theta^t)}{\lambda\left(\frac{1}{R}\right)^{t-1}\theta_t h(\theta^t)}. \tag{31}$$

We obtain

$$\begin{aligned}
\frac{\tau_{Lt}(\theta^t)}{1 - \tau_{Lt}(\theta^t)} = &\left(1 + \frac{1}{\varepsilon}\right) \frac{1}{\lambda\left(\frac{1}{R}\right)^{t-1}\theta_t h(\theta^t)} \int_{\theta_t}^{\overline{\theta}} \exp\left(\int_{\theta}^{x} \gamma(s)\frac{1}{s}ds\right) \\
& U'(\theta^t)\Bigg\{\lambda\frac{1}{R^{t-1}}h(\theta^{t-1},x)\left(\frac{1}{U'_t(\theta^{t-1},x)} - \frac{1}{U'_{t-1}(\theta^{t-1})}\right) \\
& + \beta\eta_{t-1}\frac{\partial f_t(x|\theta_{t-1})}{\partial \theta_{t-1}} - \beta f_t(x|\theta_{t-1})\eta_{t-1}\gamma_{t-1}\frac{1}{\theta_{t-1}}\Bigg\}dx.
\end{aligned} \tag{32}$$

Using (31) for period $t - 1$, solving for $\eta_{t-1}(\theta^{t-1})$ and inserting yields:

$$\begin{aligned}
\frac{\tau_{Lt}(\theta^t)}{1 - \tau_{Lt}(\theta^t)} = &\left(1 + \frac{1}{\varepsilon}\right) \frac{1}{\theta_t f_t(\theta_t|\theta_{t-1})} \int_{\theta_t}^{\overline{\theta}} \exp\left(\int_{\theta}^{x} \gamma(s)\frac{1}{s}ds\right) \\
& U'(\theta^t)\Bigg\{ f_t(x|\theta_{t-1})\left(\frac{1}{U'_t(\theta^{t-1},x)} - \frac{1}{U'_{t-1}(\theta^{t-1})}\right) \\
& + \frac{\tau_{Lt-1}}{1 - \tau_{Lt-1}}\left(1 + \frac{1}{\varepsilon}\right)^{-1}\theta_{t-1}\frac{1}{U'_{t-1}}\frac{\partial f_t(x|\theta_{t-1})}{\partial \theta_{t-1}} \\
& - \frac{\tau_{Lt-1}}{1 - \tau_{Lt-1}}\left(1 + \frac{1}{\varepsilon}\right)^{-1}\theta_{t-1}\frac{1}{U'_{t-1}}f_t(x|\theta_{t-1})\gamma_{t-1}\frac{1}{\theta_{t-1}}\Bigg\}dx
\end{aligned} \tag{33}$$

after some slight simplifications.

### A.1.1 Relation to the Previous Literature

In this appendix we show how to transform our formula into that of Golosov, Troshkin, and Tsyvinski (2013). For this purpose, define $\lambda_t(\theta^t)$ (equivalent to $\lambda_{1,t}$):

$$\lambda_t(\theta^t) = \frac{\tau_{Lt}(\theta^t)}{1 - \tau_{Lt}(\theta^t)}\frac{1}{U'(\theta^t)}\left(1 + \frac{1}{\varepsilon}\right)^{-1} f_t(\theta_t|\theta_{t-1})\theta_t\gamma_t(\theta^t)\frac{1}{\theta_t} + \frac{f_t(\theta_t|\theta_{t-1})}{U'(\theta^t)}.$$

Using this definition, we can rewrite (33) as:

$$\begin{aligned}
\frac{\tau_{Lt}(\theta^t)}{1 - \tau_{Lt}(\theta^t)} = &\left(1 + \frac{1}{\varepsilon}\right) \frac{1}{\theta_t f_t(\theta_t|\theta_{t-1})} \int_{\theta_t}^{\overline{\theta}} \exp\left(\int_{\theta}^{x} \gamma(s)\frac{1}{s}ds\right) U'(\theta^t)f_t(x|\theta_{t-1})\left(\frac{1}{U'_t(\theta^{t-1},x)} - \lambda_{t-1}(\theta^{t-1})\right) \\
& + \frac{\tau_{Lt-1}}{1 - \tau_{Lt-1}}\left(1 + \frac{1}{\varepsilon}\right)^{-1}\frac{\theta_{t-1}}{\theta_t}\frac{1}{f_t(\theta_t|\theta_{t-1})}\frac{1}{U'_{t-1}(\theta^{t-1})}\int_{\theta_t}^{\overline{\theta}}\exp\left(\int_{\theta}^{x}\gamma(s)\frac{1}{s}ds\right)\frac{\partial f_t(x|\theta_{t-1})}{\partial \theta_{t-1}}dx
\end{aligned}$$

which is the expression of Golosov, Troshkin, and Tsyvinski (2013) for GHH preferences at the beginning of their Appendix A4. To obtain an expression for $\lambda_t(\theta^t)$, one has to use the transversality condition, i.e. $\eta_t(\theta^{t-1},\underline{\theta}) = 0$ and solve for $\lambda_t(\theta^t)$.

## A.2 Linear Capital Wedges

The problem is only slightly altered as compared to Appendix A.1 in that (2) has to be considered as well. The Lagrangian now reads as:

$$
\max_{v_t(\theta^t), y_t(\theta^t), \tau_s} \mathcal{L} = \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^t} v_t(\theta^t) d\tilde{h}(\theta^t) d\theta^t - \sum_{t=1}^{T} \int_{\Theta^t} \eta_t'(\theta^t) V_t(\theta^t) d\theta^t - \sum_{t=1}^{T} \int_{\Theta^t} \eta_t(\theta^t) U_t' \Psi' \frac{y_t(\theta^t)}{\theta_t^2} d\theta^t
$$

$$
- \sum_{t=1}^{T} \int_{\Theta^t} \eta_t(\theta^t) \beta \int_{\Theta_{t+1}} V_{t+1}\left(\theta^t, \theta_{t+1}\right) \frac{\partial f_{t+1}(\theta_{t+1})}{\partial \theta_t} d\theta_{t+1}
$$

$$
+ \lambda \sum_{t=1}^{T} \frac{1}{R^{t-1}} \int_{\Theta^t} \left( y_t(\theta^t) - \Gamma(v_t(\theta^t)) + \Psi\left(\frac{y_t(\theta^t)}{\theta_t}\right) \right) h(\theta^t) d\theta^t
$$

$$
+ \sum_{t=1}^{T-1} \int_{\Theta^t} \mu_t(\theta^t) \left( -U_t'(\theta^t) + (1-\tau_s) \int_{\Theta_{t+1}} U_{t+1}'(\theta^t, x) dF_{t+1}(x|\theta_t) \right).
$$

The FOC for $v_t(\theta^t)$ now reads as

$$
\beta^{t-1} \tilde{h}(\theta^t) - \lambda \frac{1}{R^{t-1}} \frac{1}{U_t'(\theta^t)} h(\theta^t) - \sum_{j=1}^{t} \eta_j'(\theta^j) \beta^{t-j} \prod_{i=j+1}^{t} f_i(\theta_i|\theta_{i-1})
$$

$$
- \eta_t(\theta^t) \gamma_t(\theta^t) \frac{1}{\theta_t} - \sum_{j=1}^{t-1} \eta_j(\theta^j) \beta^{t-j} \frac{\partial f_{j+1}}{\partial \theta_j} \prod_{i=j+2}^{t} f_i(\theta_i|\theta_{i-1})
$$

$$
- \mu_t(\theta^t) \frac{U_t''}{U_t'} + \mu_{t-1}(\theta^{t-1})(1-\tau_s) \frac{U_t''}{U_t'} f_t. = 0 \tag{34}
$$

Applying this first-order condition also for period $t-1$, solving for $\eta_{t-1}'(\theta^{t-1})$, inserting into (34) and using $\beta R = 1$ yields

$$
0 = \lambda \frac{1}{R^{t-1}} h(\theta^t) \left( \frac{1}{U_{t-1}'} - \frac{1}{U_t'} \right) - \eta_t' - \eta_{t-1} \beta \frac{\partial f_t}{\partial \theta_{t-1}} - \eta_t \gamma_t \frac{1}{\theta_t} + \beta f_t \eta_{t-1} \gamma_{t-1} \frac{1}{\theta_{t-1}}
$$

$$
- \mu_t(\theta^t) \frac{U_t''}{U_t'} + \mu_{t-1}(\theta^{t-1})(1-\tau_s) \frac{U_t''}{U_t'} f_t - \beta f_t \left( -\mu_{t-1}(\theta^{t-1}) \frac{U_{t-1}''}{U_{t-1}'} + \mu_{t-2}(\theta^{t-2})(1-\tau_s) \frac{U_{t-1}''}{U_{t-1}'} f_{t-1} \right). \tag{35}
$$

Solving the differential equation, we obtain:

$$
\eta_t(\theta_t) = \int_{\theta_t}^{\bar{\theta}} \exp\left( \int_{\underline{\theta}}^{x} \gamma(s) \frac{1}{s} ds \right) \left\{ \lambda \frac{1}{R^{t-1}} h(\theta^{t-1}, x) \left( \frac{1}{U_t'(\theta^{t-1}, x)} - \frac{1}{U_{t-1}'(\theta^{t-1})} \right) \right.
$$

$$
+ \beta \eta_{t-1} \frac{\partial f_t(x|\theta_{t-1})}{\partial \theta_{t-1}} - \beta f_t(x|\theta_{t-1}) \eta_{t-1} \gamma_{t-1} \frac{1}{\theta_{t-1}} + \mu_t(\theta^{t-1}, x) \frac{U_t''}{U_t'} - \mu_{t-1}(\theta^{t-1}) \beta (1+r)(1-\tau) \frac{U_t''}{U_t'} f_t
$$

$$
\left. + \beta f_t \left( -\mu_{t-1}(\theta^{t-1}) \frac{U_{t-1}''}{U_{t-1}'} + \mu_{t-2}(\theta^{t-2}) \beta (1+r)(1-\tau) \frac{U_{t-1}''}{U_{t-1}'} f_{t-1} \right) \right\} dx
$$

It is easy to show that (31) also holds in this case. Further using the same algebraic manipulations as from (32) to (33) yields:

$$\frac{\tau_{Lt}(\theta^t)}{1-\tau_{Lt}(\theta^t)} = \left(1+\frac{1}{\varepsilon}\right)\frac{1}{\lambda\left(\frac{1}{R}\right)^{t-1}\theta_t h(\theta^t)}\int_{\theta_t}^{\overline{\theta}}\exp\left(\int_{\theta}^x \gamma(s)\frac{1}{s}ds\right)$$

$$U'(\theta^t)\left\{\lambda\frac{1}{R^{t-1}}h(\theta^{t-1},x)\left(\frac{1}{U_t'(\theta^{t-1},x)}-\frac{1}{U_{t-1}'(\theta^{t-1})}\right)\right.$$

$$+\frac{\tau_{t-1}}{1-\tau_{t-1}}\left(1+\frac{1}{\varepsilon}\right)^{-1}\lambda\beta^{t-1}\theta_{t-1}\frac{h_{t-1}(\theta^{t-1})}{U_{-1}'}\frac{\partial f_t(x|\theta_{t-1})}{\partial\theta_{t-1}}$$

$$-\frac{\tau_{t-1}}{1-\tau_{t-1}}\left(1+\frac{1}{\varepsilon}\right)^{-1}\lambda\beta^{t-1}\theta_{t-1}\frac{h_{t-1}(\theta^{t-1})}{U_{-1}'}f_t(x|\theta_{t-1})\gamma_{t-1}\frac{1}{\theta_{t-1}}$$

$$+\mu_t(\theta^{t-1},x)\frac{U_t''}{U_t'}-\mu_{t-1}(\theta^{t-1})(1-\tau_s)\frac{U_t''}{U_t'}f_t$$

$$\left.+\beta f_t\left(-\mu_{t-1}(\theta^{t-1})\frac{U_{t-1}''}{U_{t-1}'}+\mu_{t-2}(\theta^{t-2})(1-\tau_s)\frac{U_{t-1}''}{U_{t-1}'}f_{t-1}\right)\right\}dx.$$

# B  History-Independent Taxes

## B.1  Lagrangian, First-Order Conditions and Multipliers

### B.1.1  The Lagrangian

Here, we state Lagrangian for the problem as stated in Proposition 1. The age-independent counterpart is basically equivalent; the subscript $t$ would have to dropped for $y_t, M_t$ and $\eta_t$.

$$\mathcal{L} = \sum_{t=1}^T \beta^{t-1}\int_{\Theta^t}U\Bigg(M_t(\theta_t)-a_{t+1}(\theta^t)$$

$$+(1-\tau_s)(1+r)a_t(\theta^{t-1})-\Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\Bigg)\tilde{h}_t(\theta^t)d\theta^t$$

$$+\lambda\sum_{t=1}^T\frac{1}{(1+r)^{t-1}}\int_{\Theta^{t-1}}\int_\Theta y_t(\theta_t)-M(\theta_t)+\tau_s(1+r)a_t(\theta^{t-1})dF_t(\theta_t|\theta_{t-1})h_{t-1}(\theta^{t-1})d\theta^{t-1}$$

$$+\sum_{t=1}^{T-1}\int_{\Theta^t}\mu_t(\theta^t)\Bigg[U'\Bigg(M_t(\theta_t)-a_{t+1}(\theta^t)+(1-\tau_s)(1+r)a_t(\theta^{t-1})-\Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\Bigg)$$

$$-\beta(1+r)(1-\tau_s)\int_\Theta U'\Bigg(M_{t+1}(\theta_{t+1})-a_{t+2}(\theta^t,\theta_{t+1})$$

$$+(1-\tau_s)(1+r)a_{t+1}(\theta^t)-\Psi\left(\frac{y_{t+1}(\theta_{t+1})}{\theta_{t+1}}\right)\Bigg)dF_t(\theta_{t+1}|\theta_t)\Bigg]d\theta^t$$

$$+\sum_{t=1}^T\int_\Theta\eta_t(\theta_t)\frac{\partial\left(M_t(\theta_t)-\Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)}{\partial\theta_t}d\theta_t-\sum_{t=1}^T\int_\Theta\eta_t(\theta_t)\Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right)\frac{y_t(\theta_t)}{\theta_t^2}d\theta_t.$$

Partially integrating $\int_\Theta\eta_t(\theta_t)\frac{\partial\left(M_t(\theta_t)-\Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)}{\partial\theta_t}d\theta_t$ yields

$$\eta_t(\overline{\theta})\left(M_t(\overline{\theta})-\Psi\left(\frac{y_t(\overline{\theta})}{\overline{\theta}}\right)\right)-\eta_t(\underline{\theta})\left(M_t(\underline{\theta})-\Psi\left(\frac{y_t(\underline{\theta})}{\underline{\theta}}\right)\right)-\int_\Theta\eta_t'(\theta_t)\left(M_t(\theta_t)-\Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)d\theta_t,$$

which can then be replaced yielding

$$\mathcal{L} = \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^t} U\Bigg( M_t(\theta_t) - a_{t+1}(\theta^t)$$

$$+ (1-\tau)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right) \Bigg) \tilde{h}_t(\theta^t)d\theta^t$$

$$+ \lambda \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\Theta} y_t(\theta_t) - M(\theta_t) + \tau(1+r)a_t(\theta^{t-1})dF_t(\theta_t|\theta_{t-1})h_{t-1}(\theta^{t-1})d\theta^{t-1}$$

$$+ \sum_{t=1}^{T-1} \int_{\Theta^t} \mu_t(\theta^t)\Bigg[ U'\left( M_t(\theta_t) - a_{t+1}(\theta^t) + (1-\tau)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)$$

$$- \beta(1+r)(1-\tau_{t+1}) \int_{\Theta} U'\Bigg( M_{t+1}(\theta_{t+1}) - a_{t+2}(\theta^t, \theta_{t+1})$$

$$+ (1-\tau_{t+1})(1+r)a_{t+1}(\theta^t) - \Psi\left(\frac{y_{t+1}(\theta_{t+1})}{\theta_{t+1}}\right) \Bigg) dF_t(\theta_{t+1}|\theta_t)\Bigg] d\theta^t$$

$$- \sum_{t=1}^{T} \int_{\Theta} \eta_t'(\theta_t)\left( M_t(\theta_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)d\theta_t - \sum_{t=1}^{T} \int_{\Theta} \eta_t(\theta_t)\Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right)\frac{y_t(\theta_t)}{\theta_t^2}d\theta_t.$$

$$+ \sum_{t=1}^{T} \eta_t(\bar{\theta})\left( M_t(\bar{\theta}) - \Psi\left(\frac{y_t(\bar{\theta})}{\bar{\theta}}\right)\right) - \eta_t(\underline{\theta})\left( M_t(\underline{\theta}) - \Psi\left(\frac{y_t(\underline{\theta})}{\underline{\theta}}\right)\right).$$

### B.1.2 First-Order Conditions

The derivatives with respect to the endpoint conditions yield $\forall t:\ \eta_t(\bar{\theta}) = \eta_t(\underline{\theta}) = 0$ (equivalently, $\eta(\bar{\theta}) = \eta(\underline{\theta}) = 0$ for the age-independent case). The first-order conditions read as

$$\frac{\partial \mathcal{L}}{\partial M_s(\theta_s)} = - \frac{\lambda}{(1+r)^{s-1}} \int_{\Theta^{s-1}} f_s(\theta_s|\theta_{s-1})h_{s-1}(\theta^{s-1})d\theta^{s-1}$$

$$+ \beta^{s-1} \int_{\Theta^{s-1}} U'(\theta^{s-1}, \theta_s)f_s(\theta_s|\theta_{s-1})\tilde{h}_{s-1}(\theta^{s-1})d\theta^{s-1}$$

$$+ \int_{\Theta^{s-1}} \mu_s(\theta^{s-1}, \theta_s)U''(\theta^{s-1}, \theta_s)d\theta^{s-1}$$

$$- (1-\tau) \int_{\Theta^{s-1}} \mu_{s-1}(\theta^{s-1})U''(\theta^{s-1}, \theta_s)f_s(\theta_s|\theta_{s-1})d\theta^{s-1}$$

$$- \eta_s'(\theta_s) = 0. \tag{36}$$

For the age-independent case, one basically would have the summation over $t = 1$ to $T$ for the first two lines, the summation over $t = 1$ to $T - 1$ for the third line, the summation over $t = 2$ to $T$ for the fourth line (apart form dropping subscripts for $y_t, M_t, \eta_t', \eta_t$). The same argument applies for the first-order condition for gross income

$$\frac{\partial \mathcal{L}}{\partial y_s(\theta_s)} = \frac{\lambda}{(1+r)^{s-1}} \int_{\Theta^{s-1}} f_s(\theta_s|\theta_{s-1}) h_{s-1}(\theta^{s-1}) d\theta^{s-1}$$

$$- \beta^{s-1} \int_{\Theta^{s-1}} U'(\theta^{s-1}, \theta_s) \Psi' \left( \frac{y_s(\theta_s)}{\theta_s} \right) \frac{1}{\theta_s} f_s(\theta_s|\theta_{s-1}) \tilde{h}_{s-1}(\theta^{s-1}) d\theta^{s-1}$$

$$+ \int_{\Theta^{s-1}} \mu_s(\theta^{s-1}, \theta_s) U''(\theta^{s-1}, \theta_s) \Psi' \left( \frac{y_s(\theta_s)}{\theta_s} \right) \frac{1}{\theta_s} d\theta^{s-1}$$

$$- \beta(1+r)(1-\tau) \int_{\Theta^{s-1}} \mu_{s-1}(\theta^{s-1}) U''(\theta^{s-1}, \theta_s) \Psi' \left( \frac{y_s(\theta_s)}{\theta_s} \right) \frac{1}{\theta_s} f_s(\theta_s|\theta_{s-1}) d\theta^{s-1}$$

$$- \eta_s'(\theta_s) \Psi' \left( \frac{y_s(\theta_s)}{\theta_s} \right) \frac{1}{\theta_s} - \eta_s(\theta_s) \left( \Psi' \left( \frac{y_s(\theta_s)}{\theta_s} \right) \frac{1}{\theta_s^2} + \Psi'' \left( \frac{y_s(\theta_s)}{\theta_s} \right) \frac{y_s(\theta_s)}{\theta_s^2} \right) = 0. \qquad (37)$$

The following first-order conditions are equivalent in both cases (apart form dropping subscripts for $y_t$, $M_t$, $\eta_t'$, $\eta_t$):

$$\frac{\partial \mathcal{L}}{\partial a_{s+1}(\theta^s)} = \frac{\lambda}{(1+r)^{s-1}} \tau h_s(\theta^s) - \mu_s(\theta^s) U''(\theta^s)$$

$$- (1-\tau)^2 \beta (1+r)^2 \mu_s(\theta^s) \int_\Theta U''(\theta^s, \theta_{s+1}) dF_{s+1}(\theta_{s+1}|\theta_s)$$

$$+ (1-\tau)\beta(1+r)\mu_{s-1}(\theta^{s-1}) U''(\theta^s) f_s(\theta_s|\theta_{s-1})$$

$$+ (1-\tau)(1+r) \int_\Theta \mu_{s+1}(\theta^s, \theta_{s+1}) U''(\theta^s, \theta_{s+1}) d\theta_{s+1} = 0. \qquad (38)$$

$$\frac{\partial \mathcal{L}}{\partial \tau} = \sum_{t=2}^T \frac{\lambda}{(1+r)^{t-2}} \int_{\Theta^{t-1}} a_t(\theta^{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}$$

$$- \sum_{t=2}^T \beta^{t-1}(1+r) \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \int_\Theta U'(\theta^{t-1}, \theta_t) dF_t(\theta_t|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1}$$

$$- \sum_{t=1}^{T-1} (1+r) \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \int_\Theta \mu_t(\theta^{t-1}, \theta_t) U''(\theta^{t-1}, \theta_t) d\theta_t d\theta^{t-1}$$

$$+ \sum_{t=2}^T \beta(1+r)^2(1-\tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_\Theta U''(\theta^{t-1}, \theta_t) a_t(\theta^{t-1}) dF_t(\theta_t|\theta_{t-1}) d\theta^{t-1}$$

$$+ \sum_{t=2}^T \beta(1+r) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_\Theta U'(\theta^{t-1}, \theta_t) dF_t(\theta_t|\theta_{t-1}) d\theta^{t-1} = 0. \qquad (39)$$

### B.1.3 Multiplier Functions

First, we derive an expression for the marginal value of public funds $\lambda$. Using the transversality condition $\eta_s(\underline{\theta}) = 0$ yields:

$$\lambda = (1+r)^{s-1} \frac{1}{\int_{\Theta^{s-1}} f_s(\theta_s|\theta_{s-1}) h_{s-1}(\theta^{s-1}) d\theta^{s-1}} \times$$

$$\left\{ \beta^{s-1} \int_{\Theta^{s-1}} U'(\theta^{s-1}, \theta_s) f_s(\theta_s|\theta_{s-1}) \tilde{h}_{s-1}(\theta^{s-1}) d\theta^{s-1} \right.$$

$$+ \int_{\Theta^{s-1}} \mu_s(\theta^{s-1}, \theta_s) U''(\theta^{s-1}, \theta_s) d\theta^{s-1}$$

$$\left. - (1-\tau_s) \int_{\Theta^{s-1}} \mu_{s-1}(\theta^{s-1}) U''(\theta^{s-1}, \theta_s) f_s(\theta_s|\theta_{s-1}) d\theta^{s-1} \right\} \qquad (40)$$

43

for the age-dependent case. For the age-independent case it reads as

$$\lambda = \sum_{s=1}^{T} (1+r)^{s-1} \frac{1}{\int_{\Theta^{s-1}} f_s(\theta_s|\theta_{s-1}) h_{s-1}(\theta^{s-1}) d\theta^{s-1}} \times$$

$$\left\{ \beta^{s-1} \int_{\Theta^{s-1}} U'(\theta^{s-1}, \theta_s) f_s(\theta_s|\theta_{s-1}) \tilde{h}_{s-1}(\theta^{s-1}) d\theta^{s-1} \right.$$

$$+ \int_{\Theta^{s-1}} \mu_s(\theta^{s-1}, \theta_s) U''(\theta^{s-1}, \theta_s) d\theta^{s-1}$$

$$\left. - (1-\tau_s) \int_{\Theta^{s-1}} \mu_{s-1}(\theta^{s-1}) U''(\theta^{s-1}, \theta_s) f_s(\theta_s|\theta_{s-1}) d\theta^{s-1} \right\} \tag{41}$$

We next derive expressions for the Lagrangian multiplier function(s) on the envelope condition(s). For the age-dependent case, use (36) to obtain

$$\eta_s(\theta_s) = \frac{\lambda}{(1+r)^{s-1}} \int_{\Theta^{s-1}} \int_{\theta_s}^{\overline{\theta}} dF_s(\tilde{\theta}_s|\theta_{s-1}) h_{s-1}(\theta^{s-1}) d\theta^{s-1}$$

$$- \beta^{s-1} \int_{\Theta^{s-1}} \int_{\theta_s}^{\overline{\theta}} U'(R_s(\theta^{s-1}, \theta_s)) dF_s(\tilde{\theta}_s|\theta_{s-1}) \tilde{h}_{s-1}(\theta^{s-1}) d\theta^{s-1}$$

$$- \int_{\Theta^{s-1}} \int_{\theta_s}^{\overline{\theta}} \mu_s(\theta^{s-1}, \theta_s) U''(R_s(\theta^{s-1}, \theta_s)) d\tilde{\theta}_s d\theta^{s-1}$$

$$+ \beta(1+r)(1-\tau_s) \int_{\Theta^{s-1}} \mu_{s-1}(\theta^{s-1}) \int_{\theta_s}^{\overline{\theta}} U''(R_s(\theta^{s-1}, \theta_s)) dF_s(\tilde{\theta}_s|\theta_{s-1}) d\theta^{s-1}. \tag{42}$$

For the age-independent case, it reads as

$$\eta(\theta) = \sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\theta}^{\overline{\theta}} dF_t(\tilde{\theta}|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}$$

$$- \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^{t-1}} \int_{\theta}^{\overline{\theta}} U'(\theta^{t-1}, \theta_t) dF_t(\tilde{\theta}|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1}$$

$$- \sum_{t=1}^{T} \int_{\Theta^{t-1}} \int_{\theta}^{\overline{\theta}} \mu_t(\theta^{t-1}, \tilde{\theta}) U''(\theta^{t-1}, \tilde{\theta}) d\tilde{\theta} d\theta^{t-1}$$

$$+ \sum_{t=1}^{T} (1-\tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta}^{\overline{\theta}} U''(\theta^{t-1}, \tilde{\theta}) dF_t(\tilde{\theta}|\theta_{t-1}) d\theta^{t-1}. \tag{43}$$

Next, we derive an expression for $\mu_t(\theta^t)$. The notation refers to the age-dependent and the age-independent case are equivalent. Use (38) to obtain, with $SOC_s(\theta^s)$ being the second-order condition for savings from the individuals problem:

$$\mu_s(\theta^s) = \frac{\frac{\lambda}{(1+r)^{s-1}} \tau h(\theta^{s-1}) + (1-\tau_s)\beta(1+r)\mu_{s-1}(\theta^{s-1}) U''(\theta^s) f_s(\theta_s|\theta_{s-1})}{SOC_s(\theta^s)}$$

$$+ \frac{(1-\tau)(1+r) \int_{\Theta} \mu_{s+1}(\theta^s, \theta_{s+1}) U''(\theta^s, \theta_{s+1}) d\theta_{s+1}}{SOC_s(\theta^s)}. \tag{44}$$

Therefore, we define some terms that make notation less burdensome:

$$A_s(\theta^s) = \frac{\frac{\lambda}{(1+r)^{s-1}}\tau h(\theta^{s-1})}{SOC_s}$$

$$B_s(\theta^s) = \frac{(1-\tau)U''(\theta^s)\,f_s(\theta_s|\theta_{s-1})}{SOC_s}$$

$$C_s(\theta^s,\theta_{s+1}) = \frac{(1-\tau)(1+r)U''(\theta^s,\theta_{s+1})}{SOC_s},$$

then, we can rewrite (44) as

$$\mu_s(\theta^s) = A_s(\theta^s) + B_s(\theta^s)\mu_{s-1}(\theta^{s-1}) + \int_\Theta C_s(\theta^s,\theta_{s+1})\mu_{s+1}(\theta^s,\theta_{s+1})d\theta_{s+1}.$$

Or, more concretely for $s = T-2$:

$$\begin{aligned}\mu_{T-2}(\theta^{T-2}) =& A_{T-2}(\theta^{T-2}) + B_{T-2}(\theta^{T-2})\mu_{T-3}(\theta^{T-3}) \\ &+ \int_\Theta C_{T-2}(\theta^{T-2},\theta_{T-1})\mu_{T-1}(\theta^{T-2},\theta_{T-1})d\theta_{T-1}.\end{aligned} \quad (45)$$

For $s = T-1$, we get:

$$\mu_{T-1}(\theta^{T-1}) = A_{T-1}(\theta^{T-1}) + B_{T-1}(\theta^{T-1})\mu_{T-2}(\theta^{T-2}). \quad (46)$$

Now insert (46) into (45). Omitting arguments, this yields:

$$\mu_{T-2} = \frac{A_{T-2} + B_{T-2}\mu_{T-3} + \int_\Theta C_{T-2}A_{T-1}d\theta_{T-1}}{1 - \int_\Theta C_{T-2}(\theta^{T-2},\theta_{T-1})B_{T-1}(\theta^{T-1})d\theta_{T-1}}.$$

Now insert this into $\mu_{T-3}$

$$\mu_{T-3} = A_{T-3} + B_{T-3}\mu_{T-4} + \int_\Theta C_{T-3}\frac{A_{T-2} + B_{T-2}\mu_{T-3} + \int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1 - \int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}, \quad (47)$$

yielding

$$\mu_{T-3} = \frac{A_{T-3} + B_{T-3}\mu_{T-4} + \int_\Theta C_{T-3}\frac{A_{T-2}+\int_\Theta C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}}{1 - \int_\Theta \frac{C_{T-3}B_{T-2}}{1-\int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}}. \quad (48)$$

Now insert this into $\mu_{T-4}$

$$\begin{aligned}\mu_{T-4} =& A_{T-4} + B_{T-4}\mu_{T-5} \\ &+ \int_\Theta C_{T-4}\frac{A_{T-3} + B_{T-3}\mu_{T-4} + \int_\Theta C_{T-3}\frac{A_{T-2}+\int_\Theta C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}}{1 - \int_\Theta \frac{C_{T-3}B_{T-2}}{1-\int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}}d\theta_{T-3}.\end{aligned} \quad (49)$$

Rewrite to obtain

$$\mu_{T-4} = \left[1 - \int_{\Theta} C_{T-4}B_{T-3}\left[1 - \int_{\theta_{T-2}} C_{T-3}B_{T-2}\left[1 - \int_{\Theta} C_2 B_{T-1}d\theta_{T-1}\right]^{-1}\right]^{-1}\right]^{-1}$$

$$\left(A_{T-4} + B_{T-4}\mu_{T-5}\right.$$

$$\left. + \int_{\Theta} C_{T-4}\frac{A_{T-3} + \int_{\Theta} C_{T-3}\frac{A_{T-2}+\int_{\Theta} C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}}{1 - \int_{\Theta}\frac{C_{T-3}B_{T-2}}{1-\int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}}}d\theta_{T-3}\right). \tag{50}$$

Finally, calculate $\mu_{T-5}$, after which the pattern should become clear.

$$\mu_{T-5} = \left[1 - \int_{\Theta} C_{T-5}B_{T-4}\left[...\left[1 - \int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}\right]^{-1}....\right]^{-1}\right]^{-1}$$

$$\left(A_{T-5} + B_{T-5}\mu_{T-6}\right.$$

$$\left. + \int_{\Theta} C_{T-5}\frac{A_{T-4} + \int_{\Theta} C_{T-4}\frac{A_{T-3}+\int_{\Theta} C_{T-3}\frac{A_{T-2}+\int_{\Theta} C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}}{1-\int_{\Theta}\frac{C_{T-3}B_{T-2}}{1-\int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}}}}{1 - \int_{\Theta}\frac{C_{T-4}B_{T-3}}{1-\int_{\Theta}\frac{C_{T-3}B_{T-2}}{1-\int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}}}}\right). \tag{51}$$

Now define

$$D_s = \left[1 - \int_{\Theta} C_s B_{s+1}\left[1 - \int_{\Theta} C_{s+1}B_{s+2}\left[...\left[1 - \int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}\right]^{-1}....\right]^{-1}d\theta_{s+2}\right]^{-1}d\theta_{s+1}\right]^{-1}.$$

Using this definition, we can write $\mu_{T-5}$ as

$$\mu_{T-5} = \frac{A_{T-5} + B_{T-5}\mu_{T-6} + \int_{\Theta} C_{T-5}\frac{A_{T-4}+\int_{\theta_{T-3}} C_{T-4}\frac{A_{T-3}+\int_{\Theta} C_{T-3}\frac{A_{T-2}+\int_{\Theta} C_{T-2}A_{T-1}}{D_{T-2}}}{D_{T-3}}}{D_{T-4}}}{D_{T-5}}. \tag{52}$$

It now turns out helpful to make another definition:

$$E_s = \int_{\Theta} C_s \frac{A_{s+1}\int_{\Theta} C_{s+1}\frac{A_{s+2}+\int_{\Theta} C_{s+2}\frac{A_{s+3}+\int_{\Theta} C_{s+3}\frac{A_{s+4}\cdots\cdots}{D_{s+4}}}{D_{s+3}}}{D_{s+2}}}{D_{s+1}}.$$

Then we can write $\mu_{T-5}$ as

$$\mu_{T-5} = \frac{A_{T-5} + B_{T-5}\mu_{T-6} + E_{T-5}}{D_{T-5}}.$$

In genereal, we thus obtain:

$$\mu_s = \frac{A_s + B_s\mu_{s-1} + E_s}{D_s}.$$

For the second period, we obtain

$$\mu_2 = \frac{A_2 + B_2\mu_1 + E_2}{D_2}. \tag{53}$$

and get

$$\mu_1 = \frac{A_1 + E_1}{D_1}. \tag{54}$$

Now we can recursively calculate all other $\mu_t$ for $t = 2, ..., T$.

In equation (54) one can see that the $\mu_1(\theta_1) = 0$ if savings taxes are zero. Recursive calculation reveals that all $\mu_t$ are equal to zero.

## B.2 Labor Income Taxes

### B.2.1 Labor Income Tax Formula

Dividing (37) by $\Psi' \frac{1}{\theta_s}$ and adding (36) yields a formula for the age-dependent marginal tax rate:

$$\frac{\mathcal{T}'_s(y_s(\theta_s))}{1 - \mathcal{T}'_s(y_s(\theta_s))} = \left(1 + \frac{1}{\varepsilon(\theta_s)}\right) \frac{\eta_s(\theta_s)}{\lambda \frac{1}{(1+r)^{s-1}} \theta_s \int_{\Theta^{s-1}} f_s(\theta_s|\theta_{s-1}) h_{s-1}(\theta^{s-1}) d\theta^{s-1}}. \tag{55}$$

Dividing the age-independent equivalent to (37) by $\Psi' \frac{1}{\theta}$ and adding the age-independent equivalent to (36) yields a formula for the age-independent marginal tax rate:

$$\frac{\mathcal{T}'(y(\theta))}{1 - \mathcal{T}'(y(\theta))} = \left(1 + \frac{1}{\varepsilon(\theta)}\right) \frac{\eta(\theta)}{\lambda \theta \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\Theta} f_t(\theta_t|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}}. \tag{56}$$

### B.2.2 Relation Between $\mu$-Formula and Savings Tax Effect formula in a Three-Period Economy

We now derive the savings responses in $T = 3$ period economy.

**Formulas for Savings Responses** We now derive a formula for the savings responses due to an increase of the marginal tax rate $\mathcal{T}'_2(y_2(\theta_2))$.

To get a formula for the savings responses $\frac{\partial a_3(\theta^2)}{\partial \mathcal{T}'_2(y_2(\theta_2))}$ and $\frac{\partial a_2(\theta_1)}{\partial \mathcal{T}'_2(y_2(\theta_2))}$ is not simple because the savings choices are made at different points in time (and for different shock realizations). Further, these savings choices at different points in time are interlinked with each other. The way we proceed is the following: We look at agents with certain histories as agents that decide independently from each other. E.g., we look at the savings adjustment of an agent with history $\theta^2 = (\theta_1, \theta_2)$ who ignores that his savings change will also change savings behavior of agents of type $\theta_1$, i.e. himself one period ago. We will then, in a second step, ask how type $\theta_1$ will react to the savings change of type $\theta^2$. But of course this reaction of type $\theta_1$ will cause a reaction of type $\theta^2$ again and so on and so forth. This will thus lead to an infinite adjustment of savings in all periods. As we show now, thinking about this like that can yield formulas for change of equilibrium savings behavior. We therefore proceed step by step and look at the adjustments in the first round of this game between the agents at different points in time, then at the second round and so on and so forth.

**First Round Effects:** In period 2, individuals with income higher than $y_2(\theta_2)$ will change savings as their net income is decreased. The formula for this savings change can be obtained by implicitly differentiating the respective Euler equation:

$$a'_3(\theta^2) \equiv -\frac{U''(\theta^2)}{SOC_2(\theta^2)}, \tag{57}$$

where $SOC_t(\theta^t)$ is the second-order condition for savings of an agent with history $\theta^t$. But also individuals in period one will react to the small tax reform. The increase in taxes tomorrow will make them save more today. To obtain the respective formula, implicitly differentiate the respective Euler equation and obtain:

$$A_1(\theta_1) \equiv \frac{\beta(1+r)(1-\tau) \int_{\theta_2}^{\overline{\theta}} U''(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{SOC_1(\theta_1)}. \tag{58}$$

Note that both of these tax changes are of hypothetical nature as they are computed as if the savings decisions were not interlinked. However, they are connected and thus (57) and (58) will trigger second-round effects.

**Second Round Effects:** The savings adjustment of period 2, (57), will make individuals in period 1 adjust savings:

$$A_2(\theta_1) \equiv \frac{\beta(1+r)(1-\tau) \int_{\theta_2}^{\overline{\theta}} U''(\theta_1, \tilde{\theta}_2) a_3'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{SOC_1(\theta_1)}. \tag{59}$$

The first round savings adjustment in period 1, captured by (58), will trigger savings adjustments in period 2 for all $\theta_2$:

$$\frac{U''(\theta_1, \theta_2)(1+r)(1-\tau)A_1(\theta_1)}{SOC_2(\theta^2)} = a_3'(\theta_1, \theta_2)(1+r)(1-\tau)A_1(\theta_1). \tag{60}$$

**Third Round Effects:** The savings adjustment of period 2 (60) will now trigger savings adjustments in period 1

$$\frac{\beta(1+r)^2(1-\tau)^2 \int_{\Theta} U''(\theta_1, \theta_2) a_3' A_1(\theta_1) dF_2(\theta_2|\theta_1)}{SOC_1(\theta_1)} = A_1(\theta_1)A_3(\theta_1), \tag{61}$$

where $A_3(\theta_1)$ is defined such that the equal sign holds.

The savings adjustment of period 1 (59) will now trigger savings adjustments in period 2

$$a_3'(\theta_1, \theta_2)(1+r)(1-\tau)A_2(\theta_1). \tag{62}$$

**Fourth Round Effects:** The savings adjustment of period 2 (62) will now trigger savings adjustments in period 1

$$\frac{(1+r)(1-\tau)^2 \int_{\Theta} U''(\theta_1, \theta_2) a_3' A_2(\theta_1) dF_2(\theta_2|\theta_1)}{SOC_1(\theta_1)} = A_2(\theta_1)A_3(\theta_1) \tag{63}$$

The savings adjustment of period 1 (63) will now trigger savings adjustments in period 2

$$a_3'(\theta_1, \theta_2)(1+r)(1-\tau)A_1(\theta_1)A_3(\theta_1) \tag{64}$$

One could now repeat this until infinity. It is easy to show that savings responses in period 2 for all types with income lower than $y_2(\theta_2)$ add up to

$$\frac{\partial a_3(\theta^2)}{\partial \mathcal{T}_2'(y_2(\theta_2))} = a_3'(1+r)(1-\tau)(A_1 + A_2) \sum_{i=0}^{\infty} A_3^i. \tag{65}$$

and for all types with income higher than $y_2(\theta_2)$ add up to

$$\frac{\partial a_3(\theta^2)}{\partial \mathcal{T}_2'(y_2(\theta_2))} = a_3' + a_3'(1+r)(1-\tau)(A_1 + A_2) \sum_{i=0}^{\infty} A_3^i. \tag{66}$$

Further, period 1 savings adjust according to:

$$\frac{\partial a_2(\theta_1)}{\partial \mathcal{T}_2'(y_2(\theta_2))} = (A_1 + A_2) \sum_{i=0}^{\infty} A_3^i. \tag{67}$$

**Relation to $\mu$-formulas**   To show the relationship between the $\mu$-formula and the one with savings-tax effects, we have to show that

$$- \int_{\Theta} \int_{\theta_2}^{\overline{\theta}} \mu_2(\theta_1, \tilde{\theta}_2) U''(\theta_1, \tilde{\theta}_2) d\tilde{\theta}_2 d\theta_1$$

$$+ \beta(1+r)(1-\tau) \int_{\Theta} \mu_1(\theta_1) \int_{\theta_2}^{\overline{\theta}} U''(\theta_1, \tilde{\theta}_2)) dF_2(\tilde{\theta}_2|\theta_1) d\theta_1 \tag{68}$$

which is the term appearing in the optimal tax formula in Appendix B.2.1 is equal to

$$\frac{\lambda\tau}{1+r} \int_{\Theta} \left( (A_1(\theta_1) + A_2(\theta_1)) \sum_{i=0}^{\infty} A_3(\theta_1)^i \right) dF_1(\theta_1) + \tau \frac{\lambda}{(1+r)^2} \left( a_3' + a_3'(1+r)(1-\tau)(A_1 + A_2) \sum_{i=0}^{\infty} A_3^i \right).$$

Using (54) and evaluating for $T = 3$, yields

$$\mu_1(\theta_1) = \frac{\lambda \frac{(1-\tau) \int_{\Theta} \tau a_3' dF_2(\theta_2|\theta_1) f_1(\theta_1)}{SOC_1} + \lambda \frac{f_1(\theta_1)\tau}{SOC_1}}{1 - \frac{(1-\tau)^2(1+r)^2 \int_{\Theta} a_3' U''(\theta_1,\theta_2) dF_2(\theta_2|\theta_1)}{SOC_1}}$$

and then inserting into (53), yields

$$\mu_2(\theta_1, \theta_2) = \mu_1(\theta_1)\beta(1-\tau)(1+r)\frac{U''(\theta_1, \theta_2) f_2(\theta_2|\theta_1)}{SOC_2} + \frac{\lambda}{1+r} \frac{h_2(\theta^2)\tau}{SOC_2} \tag{69}$$

Inserting these two terms into (68) yields

$$- \frac{\lambda\tau}{1+r} \int_{\Theta} \int_{\theta_2}^{\overline{\theta}} \frac{U''(\theta_1, \tilde{\theta}_2)}{SOC_2} dF_2(\tilde{\theta}_2|\theta_1) dF_1(\theta_1)$$

$$+ \beta(1+r)(1-\tau) \int_{\Theta} \mu_1(\theta_1) \int_{\theta_2}^{\overline{\theta}} U''(\theta_1, \tilde{\theta}_2) \left( 1 - \frac{U''(\theta_1, \tilde{\theta}_2)}{SOC_2} \right) dF_2(\tilde{\theta}_2|\theta_1) d\theta_1 \tag{70}$$

$$\frac{\lambda\tau}{1+r} \int_{\Theta} \int_{\theta_2}^{\overline{\theta}} a_3'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1) dF_1(\theta_1)$$

$$+ \lambda\tau\beta(1+r)(1-\tau) \int_{\Theta} \left( \sum_{i=0}^{\infty} A_3^i \right) \int_{\theta_2}^{\overline{\theta}} \frac{U''(\theta_1, \tilde{\theta}_2)}{SOC_1} (1 + a_3') dF_2(\tilde{\theta}_2|\theta_1) dF_1(\theta_1)$$

$$+ \lambda\tau\beta(1+r)(1-\tau)$$

$$\int_{\Theta} \left( (1-\tau) \sum_{i=0}^{\infty} A_3^i \right) \left( \int_{\theta_2} a_3' dF_2(\theta_2|\theta_1) \right) \int_{\theta_2}^{\overline{\theta}} \frac{U''(\theta_1, \tilde{\theta}_2)}{SOC_1} (1 + a_3') dF_2(\tilde{\theta}_2|\theta_1) dF_1(\theta_1) \tag{71}$$

Using the definitions of $A_1$ and $A_2$, this can be rewritten as

$$\frac{\lambda\tau}{1+r}\int_\Theta\int_{\theta_2}^{\overline{\theta}}a_3'(\theta_1,\tilde\theta_2)dF_2(\tilde\theta_2|\theta_1)dF_1(\theta_1)$$

$$+\lambda\tau\int_\Theta\left(\sum_{i=0}^{\infty}A_3(\theta_1)^i\right)(A_1(\theta_1)+A_2(\theta_1))dF_1(\theta_1)$$

$$+\lambda\tau\int_\Theta\left(\sum_{i=0}^{\infty}A_3(\theta_1)^i\right)(A_1(\theta_1)+A_2(\theta_1))\int_\Theta(1-\tau)a_3'(\theta_1,\theta_2)dF_2(\theta_2|\theta_1)dF_1(\theta_1)\qquad(72)$$

which completes the proof. The formulas for the optimal marginal labor income tax rates in period 1 and 2 read as:

$$\frac{\mathcal{T}_1'(y_1(\theta_1))}{1-\mathcal{T}_1'(y_1(\theta_1))}=\left(\frac{1}{\varepsilon}+1\right)\frac{1}{\theta_1\lambda f_1(\theta_1)}$$
$$\left[\lambda\int_{\theta_1}^{\overline{\theta}}dF_1(\tilde\theta_1)-\int_{\theta_1}^{\overline{\theta}}U'(\tilde\theta_1)d\tilde F_1(\tilde\theta_1)-\int_{\theta_1}^{\overline{\theta}}\mu_1(\tilde\theta_1)U''(R_1(\tilde\theta_1))d\tilde\theta_1\right]$$

and in period 3 it reads as

$$\frac{\mathcal{T}_3'(y_3(\theta_3))}{1-\mathcal{T}_3'(y_3(\theta_3))}=\left(\frac{1}{\varepsilon}+1\right)\frac{1}{\theta_3\lambda\int_{\theta^2}f_3(\theta_3|\theta_2)h_2(\theta^2)d\theta^2}$$
$$\times\left[\frac{\lambda}{(1+r)^2}\int_{\theta^2}\int_{\theta_3}^{\overline{\theta}}dF_3(\tilde\theta_3|\theta_2)h_2(\theta^2)d\theta^2-\beta^2\int_{\theta^2}\int_{\theta_3}^{\overline{\theta}}U'(\theta^2,\tilde\theta_3)dF_3(\tilde\theta_3|\theta_2)\tilde h_2(\theta^2)d\theta^2\right.$$
$$\left.+(1-\tau)\int_{\theta^2}\mu_2(\theta^2)\int_{\theta_3}^{\overline{\theta}}U''(\theta^2,\tilde\theta_3)dF_3(\tilde\theta_3|\theta_2)d\theta^2\right]$$

Using the definition of $\mu_1$ and $\mu_2$ from above, one can show that the '$\mu$-terms' in these optimal tax formulas reflect the fiscal externalities from the savings adjustment.

### B.2.3  Income Effect on Savings

Consider the savings effect for history-independent taxes, defined by:

$$\mathcal{S}^I(\theta)=\tau_s^I\sum_{j=2}^{T}(1+r)^{j-2}\int_{\Theta^{j-1}}\int_\theta^{\overline{\theta}}\frac{\partial a_j(\theta^{j-1})}{\partial\mathcal{T}(y(\tilde\theta))}d\tilde\theta h_{j-1}(\theta^{j-1})d\theta^{j-1}.$$

We can now define income effects, similar to Saez (2001) for labor income or Golosov, Tsyvinski, and Werquin (2014) for savings and labor income.

$$\eta^I(\theta^{j-1},\theta)(1-\tau_s^I)=\frac{\partial a_j(\theta^{j-1})}{\partial\mathcal{T}(y(\theta))}.$$

If we write the average income effect as $\int_{\Theta^{j-1}}\int_\theta^{\overline{\theta}}\frac{\partial a_j(\theta^{j-1})}{\partial\mathcal{T}(y(\tilde\theta))}d\tilde\theta h_{j-1}(\theta^{j-1})d\theta^{j-1}=\bar\eta_j^I(\theta)(1-\tau_s^I)$, we obtain:

$$\mathcal{S}^I(\theta)=\frac{\tau_s^I}{1-\tau_s^I}\sum_{j=2}^{T}(1+r)^{j-2}\cdot\bar\eta_j^I(\theta).$$

Two important aspects make $\bar{\eta}_j^I(\theta)$ different from a standard income effect in a static model. First, with age-independent taxation a change in labor taxes for some income level affects an individual's income (potentially) in multiple periods. With age-dependent taxation $\eta_t^A(\theta^{j-1}, \theta)(1 - \tau_s^I) = \frac{\partial a_j(\theta^{j-1})}{\partial \mathcal{T}_t(y_t(\theta_t))}$ has opposing signs for all savings decisions decided upon before/in the time period of the tax increase ($j \leq t$) and all periods after the tax increase ($j > t$). In both the age-dependent and the age-independent case, the sign of $\mathcal{S}^I(\theta)$ and $\mathcal{S}_t^A(\theta)$ is hence ambiguous. Second with idiosyncratic uncertainty changes in taxes also affect the riskiness of the income stream and hence the need for precautionary savings.

## B.3 Decomposition of the Mechanical Effect in the Presence of Capital Taxes

In this appendix we show that the decomposition of the mechanical effect as derived in Section 4.2 also holds in the presence of a non-zero capital tax rate $\tau_s$. Again, we consider a reform, where the marginal tax rate for individuals with income $y(\theta)$ is increased such that all individuals with income $y > y(\theta)$ will pay exactly one more dollar of income taxes. Besides the labor supply effect on welfare, this affects welfare by $\sum_{t=1}^{T} M_t(\theta) + \mathcal{S}^I(\theta)$. We now slightly reinterpret the tax perturbation to intuitively derive that the effect on welfare can also be written as $\sum_{t=1}^{T} M_t^R(\theta) + M_t^I(\theta) + \mathcal{S}^I(\theta)$, where $M^I(\theta)$ is as defined in (24) and $M_t^R(\theta)$ is as defined in (25).

For the reinterpretation, we assume that the additional labor income tax revenue $\Delta(\theta)$ is then redistributed lump-sum.

In the presence of a non-zero capital tax $\tau_s$, one also has to take into account that the increase in the lump sum element by $\Delta(\theta)$ (as defined in (4.2)) causes savings responses that affect welfare. Formally this effect on welfare through the change in the Euler equations is given by:

$$
S^*(\theta) = \Delta(\theta) \Bigg\{ \sum_{t=1}^{T-1} \int_{\Theta^{t-1}} \int_\theta \mu_t(\theta^{t-1}, \theta_t) U''(\theta^{t-1}, \theta_t) d\theta_t d\theta^{t-1}
$$
$$
- \sum_{t=2}^{T} (1 - \tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_\Theta U''(\theta^{t-1}, \theta_t) dF_t(\theta_t | \theta_{t-1}) d\theta^{t-1} \Bigg\}. \tag{73}
$$

Equations (24) and (27) can be derived in the same way as in the main text. However, now we do not have $M_t^{R*}(\theta) = M_t^R(\theta)$ but instead $\sum_{t=1}^{T} M_t^{R*}(\theta) + S^*(\theta) = \sum_{t=1}^{T} M_t^R(\theta)$:

$$
\sum_{t=1}^{T} M_t^{R*}(\theta_t) + S^*(\theta) = \int_\theta^{\overline{\theta}} \lambda f_1(\theta_1) - \tilde{f}_1(\theta_1) U'(\theta_1) d\theta_1
$$
$$
+ \sum_{t=2}^{T} \int_{\Theta^{t-1}} \left( \frac{\lambda h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} - \frac{\tilde{h}_{t-1}(\theta^{t-1}) \int_{\theta_t} U'(\theta^{t-1}, \theta_t)) dF_t(\theta_t | \theta_{t-1})}{(1+r)^{t-1}} \right) \times \mathcal{CU}(\theta^{t-1}; \theta) d\theta^{t-1}
$$
$$
- \Delta(\theta) \sum_{t=1}^{T} \int_{\Theta^{t-1}} \left( \frac{\lambda h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} - \frac{\tilde{h}_{t-1}(\theta^{t-1}) \int_{\theta_t} U'(\theta^{t-1}, \theta_t)) dF_t(\theta_t | \theta_{t-1})}{(1+r)^{t-1}} \right) d\theta^{t-1} + S^*(\theta). \tag{74}
$$

which is equal to $\sum_{t=1}^{T} M_t^R(\theta)$ because the last line is equal to zero because of the transversality condition $\eta(\underline{\theta}) = 0$.

# C    Details on Calibration

We use the empirical model from Karahan and Ozkan (2013), who estimate their model using PSID-data. $y_{h,t}^i$ denotes log income of individual $i$ at age $h$ in period $t$. To obtain residual log incomes $\tilde{y}_{h,t}^i$, the authors regress log earnings on some observables (age and education):

$$y_{h,t}^i = f(X_a^i; \theta_t) + \tilde{y}_{h,t}^i,$$

where $f(X_a^i)$ is a function of the observable characteristics. Residual income is then decomposed into a fixed effect ($\alpha^i$), an AR(1) component ($z_{h,t}^i$) and a transitory component ($\phi_t \epsilon_h^i$):

$$\tilde{y}_{h,t}^i = \alpha^i + z_{h,t}^i + \phi_t \epsilon_{h,t}^i,$$

where the AR(1) process is given by

$$z_{h,t}^i = \rho_{h-1} z_{h-1,t-1}^i + \pi_t \eta_h^i,$$

and where the error term $\eta_h^i$ captures persistent shocks, $\pi_t$ is a time dependent loading factor and $\rho_{h-1}$ measures the persistence of these shocks.

Based on non-parametric estimates, Karahan and Ozkan (2013) divide individuals into three age groups: 24-33 (young), 34-52 (middle age) and 53-60 (old). In the following, we list the values they obtain for the different parameters, where the indices $Y, M, O$ correspond to the three age groups from their paper.

**Age-dependent parameters:**

- Persistence parameters: $\rho_Y = 0.88$, $\rho_M = 0.97$ and $\rho_O = 0.96$,

- Variances of the persistent error terms: $\sigma_{\eta,Y}^2 = 0.027$, $\sigma_{\eta,M}^2 = 0.013$ and $\sigma_{\eta,O}^2 = 0.026$

- Variances of the transitory shock: $\sigma_{\epsilon,Y}^2 = 0.056$, $\sigma_{\epsilon,M}^2 = 0.059$ and $\sigma_{\epsilon,O}^2 = 0.068$

**Age-independent parameters:**

- Variance of individual fixed effect: $\sigma_\alpha^2 = 0.0707$

- Variance of $z_1$ (i.e. the starting value of the persistence term): $\sigma_z^2 = 0.0767$

**Time-dependent parameters:**

- As we consider only one cohort, we assume the time dependent loading factors $\pi_t$ and $\phi_t$ to be constant. Indeed, we set them to $\pi = 1.1253$ and $\phi = 1.1115$ which corresponds to the values from 1996 as they lie in the middle of all estimates for the years from 1968-1997.

**Parameters in $f(X_a^i; \theta_t)$:**

- The function takes the form of a 3rd order polynomial in age. The coefficients are 0.0539713 for $age$, -0.153567 for $(age/10)^2$ and 0.0111291 for $(age/10)^3$.

- As Karahan and Ozkan (2013), we distinguish three education groups: individuals without high school degree, high school graduates and college graduates. The education dummies take on the values 9.570346, 9.91647 and 10.26789 respectively.

Based on all these parameters, one can now simulate the evolution of the earnings distribution. We simulated millions of lives such that a law of large numbers applies. For each simulated life, we then have the income for each year, which allows us to calculate the average income of one individual for all three parts of his life. For our simulations these are the age groups 24-36, 37-49 and 50-62 – see main text. We set the initial share of non
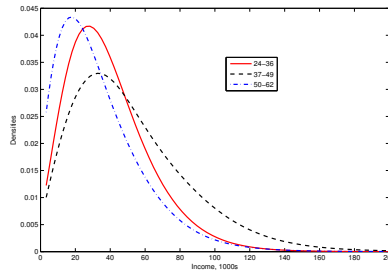
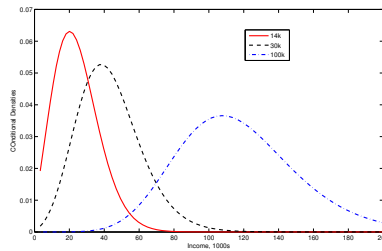Figure 8: Income Distribution for Each Age Group



Figure 9: Conditional Income Distributions, Middle Aged Workers

high-school graduates to 0.15, for high-school graduates to 0.60 and for college graduates to 0.15. This matches well US numbers – see, for example, the NLSY97.

We next discretize the earnings distribution. Thus for each simulated life, we then have 3 grid points; one for each period. With a standard kernel smoother (bandwidth of $2,500), we then smoothed the unconditional earnings distributions over this grid space as well as the conditional earnings distributions and therefore the transition probabilities. The final step was then to calibrate the skill distributions from the earnings distributions, as is commonly done (Saez 2001).

It becomes clear how inequality evolves over the life cycle. In the middle age group there are much more people with high incomes relative to the young and old. The income distribution first fans out, going from young to middle, and then compresses again in the last part of the life cycle. Figure 9 shows three conditional income distributions for the middle age-group, conditioning on earnings of $14,000, $30,000 and $100,000 in the previous period respectively.