University of Mannheim / Department of Economics

Working Paper Series

Efficient Labor and Capital Income Taxation over the Life Cycle

Sebastian Findeisen Dominik Sachs

Working Paper 14-17

May 2014

Efficient Labor and Capital Income Taxation over the Life Cycle*

Sebastian Findeisen
University of Mannheim

Dominik Sachs
CGS, University of Cologne

May 11, 2014

Abstract

This paper analyzes Pareto optimal taxation of labor and capital income in a life-cycle framework with private information and idiosyncratic risk. We focus on history-independent tax systems. We thereby complement the *Mirrlees* taxation literature, which has so far typically either characterized optimal history-dependent distortions or focused on static environments. For labor income taxes, we provide a novel decomposition of tax formulas into a *redistribution* and an *insurance* component. The latter is independent of redistributive motives and is determined by the degree of income risk and risk aversion. We show that the optimal linear capital tax rate is non-zero and derive a simple formula, which trades off redistributive and insurance benefits against the efficiency loss from savings distortions. Our quantitative results show that the insurance component contributes significantly to optimal labor tax rates. Optimal capital taxes are significant and yield sizable welfare gains.

JEL-**classification:** H21, H23

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

*Contact: findeisen@uni-mannheim.de, sachs@wiso.uni-koeln.de. We are grateful to Emmanuel Saez for valuable comments. We also thank Alan Auerbach, Spencer Bastani, Felix Bierbrauer, Friedrich Breyer, Carlos da Costa, Bas Jacobs, Emanuel Hansen, Leo Kaas, Fabrizio Zilibotti, Josef Zweimueller and seminar participants in Berkeley, Konstanz, Zurich, the SAET 2013 (Paris) and the IIPF 2013 (Taormina) for helpful suggestions and discussions. Sebastian Findeisen gratefully acknowledges the hospitality of Berkeley and the Center For Equitable Growth. We are thankful to Fatih Karahan and Serdar Ozkan who kindly made their estimates available to us. Dominik Sachs' research is funded by a post-doc fellowship of the Fritz-Thyssen Foundation and the Cologne Graduate School in Management, Economics and Social Sciences.

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. In this sense, we bridge the classical public finance approach with the recent *NDPF*. The big advantage of such an endeavor is that the instruments we characterize are within the realm of current tax practices. Further, by restricting the power of labor income taxes, we can investigate the desirability of capital taxation from a new and, arguably, more realistic angle.

Formally, let y_t be the income of an individual in period t (or, equivalently, at age t) and θ_t be the productivity in that period. As emphasized by the *NDPF* literature, in the second-best 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. Decentralizing such

[1] See Meghir and Pistaferri (2011) and Jappelli and Pistaferri (2010) for 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.

an allocation requires taxes that condition on the whole history of incomes $y^t = (y_1, y_2, ..., y_t)$.[3] *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 simple and 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 θ_t and not on accumulated wealth. For the allocation, this implies that income is solely a function of θ_t and not of θ^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. Further, this approach is flexible enough to allow the study of age-independent as well as age-dependent policies. The latter have attracted increasing interest lately (Weinzierl 2011, Bastani, Blomquist, and Micheletto 2013) and it has been argued that this approach "seems to have a good probability of leading to significant policy improvements" (Banks and Diamond 2011).

Theoretical Results. Our first theoretical result for optimal history-independent tax systems is the derivation of the formula for optimal marginal labor income tax rates and its decomposition into a redistribution and an insurance component. The natural comparison for our results is the seminal formula obtained by Diamond (1998) for the static model. We show that the forces determining the shape of optimal tax schedules in the static model also turn out to be important in the dynamic context. Tax rates are decreasing in the labor supply elasticity and the weighted mass of individuals whose labor supply is distorted. They are increasing in what is called the mechanical effects of income taxation. The mechanical effect $\mathcal{M}_t(\theta_t)$ of a marginal tax rate for individuals of type θ_t is defined as the welfare gain of taking one marginal dollar from all individuals with ability $\geq \theta_t$ in period (at age) t in the absence of behavioral responses (Saez 2001).

An important difference to the static (or dynamic deterministic) perspective is that the mechanical effect 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 novel 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. Based on these considerations, we show that the insurance value of taxation actually imposes meaningful lower bounds on Pareto optimal taxes and imposes a strong *necessary* condition for the Pareto effi-

[3] See Kocherlakota (2004, 2005) or Werning (2011).

ciency of age-dependent tax schedules. Thus, whereas Pareto efficiency alone imposes barely any restriction on income tax schedules in the static model (Werning 2007), it does so in dynamic stochastic environments.

Our second theoretical result for optimal tax design is to show that optimal linear capital taxes are different from zero and to characterize the forces underlying optimal capital taxation. We derive a formula for the optimal linear capital tax that follows a very simple and intuitive equity-efficiency relationship: the gains from redistributing wealth are traded-off against the negative incentive effect on the savings margin. Our results for the taxation of savings are different from the classical *NDPF* argument in favor of capital taxation, where savings are distorted in order to relax incentive constraints via an income effect on labor supply, as implied by the inverse Euler equation.[4] By contrast, we shut this channel down by assuming no-income effects, and show that with simple tax instruments savings distortions play a very different role. The case for a positive savings tax now arises, because it is an effective device to achieve redistribution and insurance across individuals. This motive is typically not present with history dependent labor taxes, as those can provide insurance by conditioning on the whole history of shocks, see Kocherlakota (2005).

Quantitative Results. We finally demonstrate how our first-order approach can be applied numerically to simulate optimal history-independent tax schedules for a three period economy. We base our calibration on recent estimates of income risk parameters, which are allowed to condition on age, providing a realistic life-cycle structure for the evolution of income risk (Karahan and Ozkan 2013). One goal of the quantitative exploration is to investigate how results change for different social welfare criteria *along* the (second-best) *Pareto frontier*. Our main case of interest is the Utilitarian one but we consider parts of the Pareto frontier ranging from the Rawlsian case to one, in which Pareto weights are chosen such that they imply exactly zero marginal tax rates and, therefore, no redistribution in a static Mirrlees economy. We refer to this set of Pareto weights as *laissez-faire* weights. This case makes the role of insurance particularly transparent. In contrast to the static economy, where they are zero by construction, optimal marginal labor income tax rates in the dynamic economy are significantly positive for all income levels.

We also investigate the pattern of optimal age-dependent taxes. These results illustrate the insurance value of taxation even more clearly. The insurance motive is increasing over the life cycle because the degree of income uncertainty is naturally increasing from an ex-ante perspective. This calls for increasing taxes over the life cycle. This insurance effect, however, is counteracted by a well known distributional effect. A well known result from the optimal tax literature is that the *hazard rates* of the income and skill distributions are both very informative statistics for the optimal pattern of marginal tax rates (Diamond 1998, Saez 2001). Let $F(\theta)$ be the distribution function of ability θ and $f(\theta)$ the related density. The higher the ratio $\frac{1-F(\theta)}{f(\theta)\theta}$,

[4] Golosov, Troshkin, and Tsyvinski (2013) also consider non-separable preferences and consider other arguments for savings distortions than the inverse Euler logic in a *NDPF* environment.

the higher is the optimal tax rate for this productivity level, ceteris paribus. In the calibration based on age-dependent income risk processes, these ratios are highest for the middle-aged, lowest for the old, and lie in-between for the young for most parts of the income support. Solely based on this "tagging" logic (Akerlof 1978), it seems desirable to tax the old the least and the middle-aged the most, since this would minimize labor supply distortions. Because of the insurance motive of taxation, however, labor income taxes for the old are higher than labor income taxes on the young. Our quantitative and theoretical results, hence, show that to understand the economic forces behind optimal age-dependent taxation and make policy statements about how taxes should change with age over the life cycle, one has to move beyond a pure "tagging on age" argument and take into account the insurance value of labor taxation.

Lastly, our quantitative results reveal that optimal capital tax rates are significantly positive and yield sizable welfare gains, as they provide an additional tool to the labor tax to achieve redistributive goals and provide insurance. Moreover, the desirability to tax savings is driven by the effectiveness of savings taxes on the old. The Utilitarian planner, for example, would like to place an almost zero tax rate on the savings of the young and then have savings taxes increase significantly over the life cycle. The reason is that wealth inequality is increasing over the life cycle, which in turn increases the redistributive desirability of capital taxation. Lastly, we conduct comparative static exercises with respect to risk-aversion and show that capital taxes and its welfare gains are increasing and convex in risk-aversion. The reason is that higher risk aversion (i) increases the gains from redistribution and insurance via capital taxes and implies (ii) that savings are less responsive to taxes.

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.[5] 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 more general case of persistent shocks.

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

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 optimal age-dependent labor income taxation.[6] 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. In addition, our paper is the first to study age-dependent capital income taxation, which we find to increase over the life cycle.

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.[7]

This paper is also related to Golosov, Tsyvinski, and Werquin (2013), 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. The main conceptual difference is that we focus on the full optimum instead of a tax reform approach and focus on simple history-independent and separable tax systems.

Finally, Piketty and Saez (2013) have recently derived optimal linear inheritance tax rates for a class of models with multiple generations. The main difference to the present paper is that we concentrate on the implications of life cycle instead of intergenerational considerations for capital taxes. Conesa, Kitao, and Krueger (2009), in tradition with the Ramsey approach to optimal taxation, study optimal labor and capital income taxes in a computational life cycle framework. While our approach shares some features from a Ramsey type of exercise, we allow labor income taxes to be an arbitrarily nonlinear function in the Mirrlees tradition and theoretically highlight the forces driving labor and capital taxation.

This paper is organized as follows. In Section 2, we state our formal framework and show how we make the optimal tax problem tractable. In Section 3, we derive our main results on optimal taxes in a two-period version of the model. We outline how results extend to the T-periods case in Section 4 and provide a quantitative assessment of optimal policies in Section 5. In Section 6 we conclude.

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

[7] 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.

2 The Formal Framework

2.1 Economic Environment

We consider a life cycle framework with T periods where individuals at a certain point in time t are characterized by their productivity $\theta_t \in \Theta = [\underline{\theta}, \overline{\theta}]$.[8] Further, we denote the history of shocks by $\theta^t = (\theta_1, \theta_2, \ldots, \theta_t)$. In each period, individuals make a savings and a labor supply decision. Flow utility is given by

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

where $U'' < 0$, $\Psi'' > 0$, c_t is consumption in period t and y is gross income in period t. $\frac{y_t}{\theta_t}$ 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 $U''(\theta^t)$.

Importantly, the functional form of U eliminates income effects on labor supply, while allowing for risk-aversion. This assumption is crucial for the tractability of the dynamic optimal tax problem.[9] Without income effects, individual labor supply is independent of the amount of assets. This simplifies the analysis heavily in two manners. First, there is an obvious ordering of the income vector. Second, the Hessian matrix of the individual choice problem (choosing labor supply and savings jointly) has an empty minor diagonal; taking into account optimal behavior of individuals is thus much simpler. 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 (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. There is no initial heterogeneity in assets and we assume that all agents start with zero assets.

In the following periods, productivities evolve stochastically over time according to a Markov process. The respective *cdf* is $F_t(\theta_t|\theta_{t-1})$. Further, denote by $h_t(\theta^t)$ the probability of history θ^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 consider a small open economy, so the interest on savings r is fixed. Further, we assume incomplete markets in a sense that individuals only have access to risk-free one period bonds.[10]

[8] For simplicity, but without loss of generality, we assume that the support of the distribution functions is constant.

[9] 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). In macroeconomics, this class of preferences has shown to be very useful in matching business cycle moments (Greenwood, Hercowitz, and Huffman 1988, Mendoza and Yue 2012).

[10] 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 $\sum_{t=s}^{T} y_t(\underline{\theta})$. Secondly, individuals can actually not borrow that much as repaying everything would yield a

2.2 Policy Instruments and Planner's Objective

We are interested in the Pareto efficient set of nonlinear labor income tax schedule and linear capital income tax rates that only condition on current income and potentially age. Thus, we are not solving for a second-best Pareto problem, where the government could condition policy instruments on all public information (typically the history of income and savings), but rather restrict the set of policy instruments in a Ramsey manner. However, our approach shares the feature of the Mirrlees approach that labor income taxes can be an arbitrarily nonlinear function of current income, see also the discussion at the end of this subsection. One could call it a third-best Pareto problem, where third-best refers to the restriction on policy instruments. In the remainder of this paper, solely the phrase Pareto optimal will be used.

We consider two scenarios. In the first, the government can condition labor income tax schedules $\mathcal{T}(\cdot)$ and capital tax rates τ on time t, so $\mathcal{T} = \{\mathcal{T}_t(\cdot)\}_{t=1,\dots,T}$ and $\tau = \{\tau_t\}_{t=1,\dots,T}$. This is equivalent to age-dependent taxation. In the second scenario, we study income tax functions, which are independent of time/age. In this case $\mathcal{T}$ is a single function and τ a scalar.

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 weights for individuals with certain histories.

In the remainder of this paper, we use the notions wealth, savings or capital for a_t interchangeably. Also note that the way we define τ_t, it is a stock tax not a flow tax. However, there is always a one to one mapping between such a stock tax and a tax on capital income, i.e. ra_t. Thus, there is no loss of generality in the way we defined τ_t. In the following, we use the notions capital taxes, wealth taxes and capital income taxes interchangeably. Further note that the way we define capital taxes implies that borrowing is subsidized at the same rate as saving is taxed.[11]

Relation to Ramsey and New Dynamic Public Finance. The tax problem we look at lies between the classical Ramsey approach and the *NDPF* approach. The Ramsey equivalent to our problem would be a parametric restriction on $\mathcal{T}$, e.g., a linearity restriction. In that sense, the allocations that can be attained via a Ramsey approach are a subset of the allocations, that we can attain with our policy instruments. In a *NDPF* approach, policy instruments are normally only restricted by informational asymmetries. Therefore, the the set of allocations that we can attain with our policy instruments is a subset of the allocations that can be reached via a *NDPF* approach.

negative number in $U(\cdot)$ because of the disutility of labor. Therefore, in the absence of taxes, the maximal amount of debt is $\sum_{t=s}^{T}\left(y_t(\underline{\theta}) - \Psi\left(\frac{y_t(\underline{\theta})}{\underline{\theta}}\right)\right)$. For CRRA preferences, e.g., this constraint would never be binding as the marginal utility of consumption would then be ∞ in the worst case scenario.

[11] One could impose a zero subsidy (tax) on borrowing as an additional constraint. This would make notation more burdensome without changing the main results on the desirability of non-zero capital taxes.

2.3 Individual Problem Given Taxes

In the remainder of this section, for simplicity, we always use notation that refers to age-dependent taxation, i.e. $\mathcal{T}_t$ and τ_t. For the age-independent tax problem we would only have to drop the subscript t. Each period, individuals make a work and savings decision. Formally, the recursive problem of individuals given age-dependent taxes $\{\mathcal{T}_j, \tau_j\}_{j=t,...,T}$ reads as:

$$\begin{aligned} V_t(\theta_t, a_t, \{\mathcal{T}_j, \tau_j\}_{j=t,...,T}) = \max_{a_{t+1}, y_t} \; & U\left(y_t - \mathcal{T}_t(y_t) + (1+r)(1-\tau_t)a_t - a_{t+1} - \Psi\left(\frac{y_t}{\theta_t}\right)\right) \\ & + E_t\left[V_{t+1}(\theta_{t+1}, a_{t+1}, \{\mathcal{T}_j, \tau_j\}_{j=t+1,...,T})\right], \end{aligned} \tag{1}$$

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

Lemma 1. *The optimal gross income y_t that solves (1) is independent of assets, capital taxes $\{\tau\}_{j=t,...,T}$ 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 will greatly simplify the optimal tax analysis. For the resulting allocation, this implies that y_t is only a function of θ_t and not of θ^t. The savings decision of individuals, in contrast, will depend on all state variables: $a_{t+1}(\theta_t, a_t, \{\mathcal{T}_j, \tau_j\}_{j=t,...,T})$. Recursively inserting, one can also write $a_{t+1}(\theta^t, \{\mathcal{T}_j, \tau_j\}_{j=1,...,T})$. For the resulting allocation, this implies that assets are a function of the history of shocks: $a_{t+1}(\theta^t)$.

2.4 The Social Planner's Problem

The age-dependent tax problem of the social planner reads as:

$$\max_{\{\mathcal{T}_j, \tau_j\}_{j=t,...,T}} \int_\Theta V_1(\theta_1, 0, \{\mathcal{T}_j, \tau_j\}_{j=1,...,T}) d\tilde{F}_1(\theta_1) \tag{2}$$

where $V_1(\theta_1, 0, \mathcal{T}, \tau)$ is the solution to (1) for each θ_1 and subject to an intertemporal budget constraint:

$$\begin{aligned} & \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^t} \mathcal{T}_t(y_t(\theta_t)) h_t(\theta^t) d\theta^t \\ & + \sum_{t=2}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \tau_t (1+r) a_t(\theta^{t-1}, \{\mathcal{T}_j, \tau_j\}_{j=1,...,T}) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \geq \mathcal{R} \end{aligned} \tag{3}$$

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

Constraint (1) makes the solution of the problem with Lagrangian methods nontrivial. In the following subsection, we argue that (1) can be replaced by a set of first-order conditions for a_{t+1} and y_t and a monotonicity condition on y_t that is well known from the static Mirrlees literature.

2.5 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 $a_t(\theta^{t-1}, \{\mathcal{T}_j, \tau_j\}_{j=1,\dots,T})$.

We now want to show how (1) can be replaced by two first-order conditions and a monotonicity constraint. The set of first-order conditions for the individual problem (1) are standard. For the labor supply decision, we have $\forall t$ and $\forall \theta_t \in \Theta$:

$$1 - \mathcal{T}_t'(y_t(\theta_t) = \Psi'\left(\frac{y_t(\theta_t}{\theta_t}\right)\frac{1}{\theta_t}. \tag{4}$$

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

$$\begin{aligned} &U'\left(y_t(\theta_t) - \mathcal{T}_t(y_t(\theta_t)) - a_{t+1}(\theta^t) + (1-\tau_t)a_t(\theta^{t-1}) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right) \\ =&\beta(1+r)(1-\tau_t)\int_\Theta U'\Bigg(y_{t+1}(\theta_{t+1}) - \mathcal{T}_{t+1}(y_{t+1}(\theta_{t+1})) - a_{t+2}(\theta^t, \theta_{t+1}) \\ &+ (1-\tau_{t+1})a_{t+1}(\theta^t) - \Psi\left(\frac{y_{t+1}(\theta_{t+1})}{\theta_{t+1}}\right)\Bigg) dF_{t+1}(\theta_{t+1}|\theta_t). \end{aligned} \tag{5}$$

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 (4) and (5) to represent a maximum, only the second derivatives of the value function with respect to labor supply and consumption have to be ≤ 0. For labor supply, a familiar argument from the standard Mirrlees model implies that this holds if and only if[12]

$$y_t'(\theta_t) \geq 0. \tag{6}$$

The second-order condition for savings is always fulfilled due to concavity of the utility function. Hence, (4) and (5) represent a maximum whenever $y_t'(\theta_t) \geq 0$. As $y_t'(\theta_t) \geq 0$ even implies global concavity, (4) and (5) even represent a global maximum if $y_t'(\theta_t) \geq 0$ holds. These considerations yield the following lemma:

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

Lemma 2. *Instead choosing* $\{\mathcal{T}_t, \tau_t\}_{t=1,\ldots,T}$ *to maximize (2) subject to (3) and (1), the planner can also choose* $\{\mathcal{T}_t, \tau_t, \{y_t(\theta_t)\}_{\theta_t \in \Theta}, \{a_t(\theta^{t-1})\}_{\theta^{t-1} \in \Theta^{t-1}}\}_{t=1,\ldots,T}$ *subject to (3), (4), (5) and (6).*

Incorporating (4) into a Lagrangian, however, is still problematic as it contains $\mathcal{T}_t'$, i.e. the derivative of the function with respect to which we want to maximize. To tackle this problem, we make use of the following derivative

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

Inserting (4) into this derivative yields:

$$\frac{\partial\left(y_t(\theta_t) - \mathcal{T}_t(y_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)}{\partial \theta_t} = \Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right)\frac{y_t(\theta_t)}{\theta_t^2}. \tag{7}$$

Thus, (7) implies (4). We can therefore slightly rephrase Lemma 2:

Lemma 3. *Instead choosing* $\{\mathcal{T}_t, \tau_t\}_{t=1,\ldots,T}$ *to maximize (2) subject to (3) and (1), the planner can also choose* $\{\mathcal{T}_t, \tau_t, \{y_t(\theta_t)\}_{\theta_t \in \Theta}, \{a_t(\theta^{t-1})\}_{\theta^{t-1} \in \Theta^{t-1}}\}_{t=1,\ldots,T}$ *subject to (3), (7), (5) and (6).*

As we show in Appendix A.1, (7) and (5) can be incorporated into a Lagrangian or optimal control problem. Further, when solving for optimal policies, we do not incorporate the monotonicity constraint (6) into the Lagrangian as is standard practice in the optimal tax literature. In the numerical simulations we will ex-post check whether the monotonicity condition is fulfilled or not. The Lagrangian and all first-order conditions for the age-dependent and age-independent case are stated in Appendices A.1 and B.1 respectively.

3 Pareto Optimal Taxes – The Two-Period Case

We start by deriving and discussing optimal taxes in a two-period model, so $T = 2$. The reason is that optimal taxes in the two periods case are much easier to interpret and basically convey all the economic intuition. Building on our results we then briefly discuss formulas for the general case in Section 4. For pedagogical reasons, we begin with the case, in which the government does not tax savings. We then build on that case and move to Pareto optima with capital taxation.

3.1 Labor Income Taxation in the Absence of Savings Taxes

We start by characterizing optimal labor income taxes for the case that wealth taxes are constrained to be zero. This case is an important benchmark, where it is particularly transparent to elaborate the insurance motive of labor income taxes in addition to the redistribution motive. In addition, this benchmark serves as a helpful starting point for Section 3.2 where we discuss the optimality of non-zero capital taxes.

3.1.1 Optimal Labor Income Taxes

We first consider the case of age-dependent taxation; we refer to taxes in period $t = 1$ ($t = 2$) as taxes for the young (old).

Proposition 1. *Pareto optimal marginal labor income tax rates for the young satisfy:*

$$\frac{\mathcal{T}_1'(y_1(\theta_1))}{1-\mathcal{T}_1'(y_1(\theta_1))} = \left(1+\frac{1}{\varepsilon(\theta_1)}\right)\frac{1}{\lambda\theta_1 f_1(\theta_1)} \times \mathcal{M}_1(\theta_1) \tag{8}$$

where

$$\mathcal{M}_1(\theta_1) = \int_{\theta_1}^{\overline{\theta}} \left(\lambda - U'(\tilde{\theta}_1)\frac{\tilde{f}_1(\tilde{\theta}_1)}{f_1(\tilde{\theta}_1)}\right) dF_1(\tilde{\theta}_1)$$

and $\varepsilon(\theta_t)$ denotes the elasticity of taxable labor income w.r.t. to the net-of-tax-rate in period t, which is defined by: $\varepsilon(\theta_t) = \frac{\partial y_t(\theta_t)}{\partial(1-\mathcal{T}_t'(y_t(\theta_t)))}\frac{1-\mathcal{T}_t'(y_t(\theta_t))}{y_t(\theta_t)}$.

λ is the marginal value of public funds and given by

$$\lambda = \int_{\Theta} U'(\theta_1) d\tilde{F}_1(\theta_1). \tag{9}$$

Further, the distortion at the top and bottom of the income distribution is zero, i.e. $\mathcal{T}_1'(y_1(\underline{\theta})) = \mathcal{T}_1'(y_1(\overline{\theta})) = 0$.

Proof. See Appendix A.2.4. □

This formula is very similar to the static model – see Diamond (1998).[13] Note that in general the optimal tax formula in Propositions 1 can intuitively be derived by tax perturbation methods, where an infinitesimal increase of the marginal tax rate $\Delta T'$ at an infinitesimal income interval with length $\Delta y_1(\theta_1)$ around income level $y_1(\theta_1)$ is considered. The mechanical welfare gain from taking money from all young individuals with income $> y_1(\theta_1)$ is given by $\mathcal{M}_1(\theta_1) \times \Delta T'\Delta y_1(\theta_1)$, which will henceforth be called the *mechanical effect* (Saez 2001). It depends on the redistributive preferences, i.e. the Pareto weights, of the planner, the degree of risk aversion as well as on the share of individuals with income $> y_1(\theta_1)$. The planner trades off this *mechanical effect* against a loss in tax revenue which is induced by lower labor supply of individuals of type θ_1; as shown by Piketty (1997), this term is equivalent to $LS(\theta_1) = \lambda\frac{\mathcal{T}_1'(y_1(\theta_1))}{1-\mathcal{T}_1'(y_1(\theta_1))}\theta_1 f_1(\theta_1)\frac{\varepsilon(\theta_1)}{\varepsilon(\theta_1)+1} \times \Delta T'\Delta y_1(\theta_1)$. Setting $LS(\theta_1) = \mathcal{M}_1(\theta_1) \times \Delta T'\Delta y_1(\theta_1)$ yields formula (8). Note that even though savings will also adjust as a consequence of the small tax perturbation, this has no consequences for welfare via individual utilities by the envelope theorem. In the presence of capital taxes, this will no longer be true as we will discuss in Section 3.3.

The fact that the marginal value of public funds equals the average social marginal utility – as reflected by (9) – is a well known result from the static Mirrlees literature without income effects, that generalizes to our dynamic setting if wealth taxes are constrained to be zero. We now turn to optimal labor income taxes for the old.

[13]Relatedly, for the case of *history-dependent* taxes, Golosov, Troshkin, and Tsyvinski (2013) also find that optimal labor wedges in the initial period follow the same forces as in the static model.

Proposition 2. *Pareto optimal marginal labor income tax rates for the old satisfy:*

$$\frac{\mathcal{T}_2'(y_2(\theta_2))}{1-\mathcal{T}_2'(y_2(\theta_2))} = \left(1+\frac{1}{\varepsilon(\theta_2)}\right)\frac{1}{\frac{\lambda}{1+r}\theta_2\int_\Theta f_2(\theta_2|\theta_1)dF_1(\theta_1)} \times \mathcal{M}_2(\theta_2),$$

where

$$\mathcal{M}_2(\theta_2) = \int_\Theta \int_{\theta_2}^{\overline{\theta}} \left(\frac{\lambda}{1+r} - \beta U'(\theta_1,\tilde{\theta}_2)\frac{\tilde{f}_1(\theta_1)}{f_1(\theta_1)}\right) dF_2(\tilde{\theta}_2|\theta_1)dF_1(\theta_1).$$

Further, the distortion at the top and bottom of the income distribution is zero, i.e. $\mathcal{T}_2'(y_2(\underline{\theta})) = \mathcal{T}_2'(y_2(\overline{\theta})) = 0$.

Proof. See Appendix A.2.4. □

The interpretation is very similar to the labor income taxes for the young. An interesting difference is that the mechanical effect can now be decomposed into an insurance and a redistribution term. We will elaborate on this decomposition in Section 3.1.2. Finally notice that age-dependent transfers are not uniquely pinned down. This follows by simple Ricardian equivalence arguments. Perhaps more surprisingly, we will show that this result also extends to the case where wealth is taxed in addition to labor income in Section 3.3.

Finally, the following proposition, summarizes our results for optimal age-independent marginal tax rates.

Proposition 3. *Age-independent Pareto optimal marginal labor income tax rates are given by:*

$$\frac{\mathcal{T}'(y(\theta))}{1-\mathcal{T}'(y(\theta))} = \left(1+\frac{1}{\varepsilon(\theta)}\right)\frac{1}{\lambda\theta\left(f_1(\theta)+\frac{1}{1+r}\int_\Theta f_2(\theta|\theta_1)dF_1(\theta_1)\right)} \times \left[\sum_{i=1}^{2}\mathcal{M}_i(\theta)\right] \tag{10}$$

and λ *is defined as in (9). Further, the distortion at the top and bottom of the income distribution is zero, i.e.* $\mathcal{T}'(y(\underline{\theta})) = \mathcal{T}'(y(\overline{\theta})) = 0$.

Proof. See Appendix B.2.3. □

The difference here is that here labor supply incentives of the young *and* old are traded off against the mechanical effect for young *and* old. Trivially, age-independent taxes by definition lead to lower welfare as the trade-off cannot be optimally solved for each age group separately.

In this section we derived simple labor income taxes, which are allowed to condition on current income and (potentially) age only. We have shown that the formulas and their underlying intuition closely resemble their static model counterpart. Despite this close connection between the static and dynamic model, we now show how optimal taxes in the dynamic, risky environment can be decomposed into an insurance and a redistribution part. In a dynamic model, these concepts have a meaningful distinction, as becomes clear in the next section.

3.1.2 Insurance versus Redistribution and Pareto Optimality

We now show how the role of income taxes in a dynamic and risky economy can be meaningfully decomposed into two parts: insurance and redistribution.[14] We proceed by showing it for the case of age-dependent taxes on the old, but the general insights also hold true for age-independent taxation. For this purpose, we slightly reinterpret the classical tax perturbation. Assume that $\mathcal{T}_2'(y_2(\theta_2))$ is increased in a way that everyone with old age income above $y_2(\theta_2)$ now pays one more dollar of labor income tax. Then, assume that the additional tax revenue generated by this increase is redistributed in a lump sum fashion in the second period; this reinterpretation is harmless as this lump sum increase has no first-order impact on welfare via the implied responses of savings behavior.[15]

Note that due to the reform, it is likely that the expected life-time utility of some θ_1-types is increased, while the utility of others is decreased. The reform typically creates winners and losers.[16] We denote the uniform lump sum increase in period two by $\Delta L(\theta_2)$; its value is given by $\int_\Theta \int_{\theta_2}^\infty dF_2(\tilde{\theta}_2|\theta_1)dF_1(\theta_1)$. For each type θ_1, we now introduce a 'constant utility term' $\mathcal{CU}(\theta_1,\theta_2)$ and a 'redistribution term' $\mathcal{R}(\theta_1,\theta_2)$. The constant utility term is defined as follows:

$$\mathcal{CU}(\theta_1,\theta_2) = \frac{\int_{\theta_2}^{\overline{\theta}} U'(\theta_1,\tilde{\theta}_2)dF_2(\tilde{\theta}_2|\theta_1)}{\int_{\underline{\theta}}^{\overline{\theta}} U'(\theta_1,\theta_2)dF_2(\theta_2|\theta_1)}.$$

The numerator captures the (expected) utility loss in period two due to the tax increase. Dividing it by expected marginal utility in period two says by how much consumption had to be increased in period two, in every state of the world, in order to make the individual of type θ_1 in expectation equally well of. This number is smaller than one because (i) the tax increase in period 2 affects the individual in period 2 with probability less than one and because of (ii) risk aversion. As we will argue below, this term is useful to measure the insurance gain from taxation.

We define:

$$\mathcal{R}(\theta_1,\theta_2) = \Delta L(\theta_2) - \mathcal{CU}(\theta_1,\theta_2).$$

The term $\mathcal{R}(\theta_1,\theta_2)$ captures redistribution. It is positive (negative) if expected utility for that θ_1-type increases (decreases) because the increase in the lump sum transfer $\Delta L(\theta_2)$ is larger (smaller) than $\mathcal{CU}(\theta_1,\theta_2)$.

[14] 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.

[15] 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 A.2.5.

[16] This does not have to be the case. There could also be solely loser or solely winners. See the discussion after Proposition 4.

Given this decomposition, we now determine the overall welfare consequences of the two effects. We start with the insurance effect and ask the following question: If the government could increase the lump-sum transfer in period two differently for each θ_1-type such that expected utilities are unchanged for all θ_1-types (i.e. by $\mathcal{CU}(\theta_1, \theta_2)$ respectively), how much resources could the government save due to this insurance against income risk? Since $\mathcal{CU}(\theta_1, \theta_2)$ by definition does not change expected lifetime utilities it has no direct impact on welfare. However, this reform has a resource savings effect. From each individual of type θ_1, the government obtains tax revenue of $\frac{1-F_2(\theta_2|\theta_1)}{1+r}$ in present value terms. To hold utility constant for that individual, only $\frac{\mathcal{CU}(\theta_1,\theta_2)}{1+r}$ of resources have to be spend (in present value terms). Formally, we write:

$$\mathcal{M}_2^I(\theta_2) = \frac{\lambda}{1+r} \int_\Theta \left[(1 - F_2(\theta_2|\theta_1)) - \mathcal{CU}(\theta_1, \theta_2) \right] dF_1(\theta_1) > 0. \tag{11}$$

The more pronounced labor income risk, conditional on θ_1, and the stronger risk aversion, the larger is this insurance effect. Note, that it is – as one would expect – always positive.

However, the government cannot increase the lump-sum transfer differently for individuals with different income histories. Thus, the government necessarily effectively redistributes between different θ_1-types. We now derive the welfare consequences of this implied redistribution. Note that $R(\theta_1, \theta_2)$ measures the (possibly negative) expected utility increase of type θ_1 in monetary terms. This increase is valued $\beta \int_\Theta U'(\theta_1, \theta_2) dF_2(\theta_2|\theta_1) \tilde{f}_1(\theta_1) - \frac{\lambda}{1+r} f_1(\theta_1)$ by the planner for each type θ_1. Thus, aggregating over all types and weighing by $R(\theta_1, \theta_2)$, yields:

$$\mathcal{M}_2^R(\theta_2) = \int_\Theta \left(\frac{\lambda}{1+r} - \beta \int_{\theta_2} U'(R_1(\theta_1, \theta_2)) dF_2(\theta_2|\theta_1) \frac{\tilde{f}_1(\theta_1)}{f_1(\theta_1)} \right) \times \left(\mathcal{CU}(\theta_1, \theta_2) - \Delta L(\theta_2) \right) dF_1(\theta_1).$$

The last term $\Delta L(\theta_2)$ can now be ignored since it is independent of θ_1 and it is possible to show that

$$\int_{\theta_1} \left(\frac{\lambda}{1+r} - \beta \int_\Theta U'(\theta_1, \theta_2) dF_2(\theta_2|\theta_1) \frac{\tilde{f}_1(\theta_1)}{f_1(\theta_1)} \right) dF_1(\theta_1) = 0$$

based on the (9) and the Euler equation for each θ_1-type. Thus, the following remains:

$$\mathcal{M}_2^R(\theta_2) = \int_\Theta \left(\frac{\lambda}{1+r} - \beta \int_\Theta U'(\theta_1, \theta_2) dF_2(\theta_2|\theta_1) \frac{\tilde{f}_1(\theta_1)}{f_1(\theta_1)} \right) \times \mathcal{CU}(\theta_1, \theta_2) dF_1(\theta_1). \tag{12}$$

The $\mathcal{CU}(\theta_1, \theta_2)$-term should be higher for high θ_1-types typically because they are likely to have a better shock in Period 2. Therefore, for redistributive Pareto-weights, the term $\mathcal{M}_2^R(\theta_2)$ should be positive for each θ_2. If Pareto weights are sufficiently strong in favor of high innate types, however, the welfare effect can be negative. We will get back to this in our numerical simulations. Note that this decomposition is defined exactly in such a way that:

$$\mathcal{M}_2(\theta_2) = \mathcal{M}_2^I(\theta_2) + \mathcal{M}_2^R(\theta_2).$$

Finally, the decomposition of the mechanical effect also applies for the case of age-independent taxes since $M_2(\theta_2)$ also shows up in (10).

Two Pareto Tests. Instead of characterizing optimal tax schedules for given Pareto-weights, our results from above can fruitfully be applied to derive a test for whether a given tax function is Pareto optimal.

Werning (2007) has pioneered this approach for the static Mirrlees model.[17] As he shows Pareto-efficiency alone is a rather weak condition on static tax schedules. By contrast, in dynamic environments efficiency alone places stronger bounds because of the insurance property of taxation. We focus on deriving a lower bound on income taxes for the old; for the young such a lower bound naturally does not exist as there is no insurance motive:[18]

Proposition 4. *A marginal tax rate at income level $y(\theta_2)$ is inefficiently low whenever:*

$$\int_\Theta (1 - F_2(\theta_2|\tilde{\theta}_1))dF_1(\tilde{\theta}_1) - \frac{T_2'(y_2(\theta_2))}{1 - T_2'(y_2(\theta_2))}\frac{\epsilon(\theta_2)}{\epsilon(\theta_2)+1}\theta_2 \int_\Theta f_2(\theta_2|\tilde{\theta}_1)dF_1(\tilde{\theta}_1) > \mathcal{CU}(\theta_1, \theta_2)\ \forall\ \theta_1. \tag{13}$$

We omit a formal proof and rather provide a heuristic one in the following lines. The LHS of (13) captures the change in tax revenue due to a marginal increase of $\mathcal{T}_2'(y_2(\theta_2))$ such that absolute tax payments for individuals with old age income $> y_2(\theta_2)$ is increased by one dollar. The first term captures the mechanical gain. The second captures the loss due to behavioral responses. Whenever the LHS is smaller than zero, the marginal tax rate is above its Laffer bound.[19] More interestingly in a dynamic environment, whenever the LHS is larger than $\mathcal{CU}(\theta_1, \theta_2)$ for each θ_1, the marginal tax rate $\mathcal{T}_2'(y_2(\theta_2))$ is inefficiently low and an increase of it combined with an increase in the lump sum transfer can make everybody better off. Why? As the LHS is the gain in tax revenue, it is also the increase in the lump sum in period 2. For example, assume that this increase is 5 and that the dollar amount that can compensate for the second period tax increase is below 5 for everybody: this tax reform makes everyone better off – even though it redistributes between ex-ante types. This condition can be tested for a given elasticity and risk-aversion parameter, after backing out the skill distributions. In addition, knowledge about income and consumption or savings would be required.

In addition, we can also derive a strong *necessary* condition on Pareto optimality for joint design of taxes on the young on the old. This is again related to insurance role of taxes. To derive this test intuitively, assume an increase of $\mathcal{T}_2'(y_2(\theta_2))$ by an infinitesimal amount, while decreasing the whole schedule $\mathcal{T}_1$ in a way that expected utilities of all θ_1-types stay constant. Whereas this reform (by definition) has no direct effect on welfare via individual utilities, it has resource effects via the change in labor supply in both periods. Whenever the reform leads

[17]Scheuer (2013) derives a test for an occupational choice model with entrepreneurs and workers.

[18]A lower bound for age-independent taxes can be obtained similarly. Of course, it is slightly weaker.

[19] Lorenz and Sachs (2012) derive such a condition for a static Mirrlees model with intensive and extensive labor supply responses.

to an increase in resources it yields a Pareto improvement. Whenever it leads to a decrease of resources, a reform into the opposing direction yields a Pareto improvement. Thus, we obtain the following necessary condition for the Pareto efficiency of age-dependent taxes:

Proposition 5. *A sequence of age-dependent labor income tax functions* $\{\mathcal{T}_1, \mathcal{T}_2\}$ *is second-best Pareto efficient if and only if*

$$\begin{aligned}&\int_\Theta \int_{\theta_2}^{\bar{\theta}} \left(1 - \mathcal{CU}(\theta_1, \tilde{\theta}_2)\right) dF_2(\tilde{\theta}_2|\theta_1) dF_1(\theta_1)\\ &- \frac{\mathcal{T}_2'(y_2(\theta_2))}{1-\mathcal{T}_2'(y_2(\theta_2))}\theta_2 \frac{\varepsilon(\theta_2)}{1+\varepsilon(\theta_2)} \int_\Theta f_2(\theta_2|\theta_1) dF_1(\theta_1)\\ &+ \int_\Theta \frac{\mathcal{T}_1'(y_1(\theta_1))}{1-\mathcal{T}_1'(y_1(\theta_1))}\theta_1 \frac{\varepsilon(\theta_1)}{1+\varepsilon(\theta_1)} \frac{\partial \frac{\int_{\theta_2}^{\bar{\theta}} U'(\theta_1,\tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{U'(\theta_1)}}{\partial \theta_1} dF_1(\theta_1) = 0.\end{aligned}$$

Proof. See Appendix A.2.6 □

Note that in the above formula, Pareto weights do not show up. The first line captures the insurance gains, whereas the second and third reflect labor supply responses for both periods. Summing up, our analysis shows that Pareto efficiency is a much stronger criterion in dynamic, risky environments in comparison to the static tax model.

3.2 On the Optimality of Capital Taxes

If the government can tax savings, it will generally set a non-zero savings tax. The intuition is that, starting from zero, a small change in the savings tax will always trigger a first order redistributive gain, from individuals with higher savings to ones with lower savings. If the government is interested in redistributing into the other direction, the same argument goes through for a small decrease in the tax rate. Going against this, is an efficiency loss; this loss will, however, be of second-order, starting from a zero tax.[20] We will now build on this intuition and derive a formula for the optimal linear tax rate.

In the two period model there is no difference between age-dependent and age-independent capital taxation because young agents start with zero wealth and capital is only taxed in period two. We will use a perturbation argument for pedagogical reasons, looking at a small change in the capital tax rate $d\tau$. In Appendix A.3.2, we provide a derivation of the formula with Lagrangian methods .

The small change will increase government's revenue in present value terms by

$$\int_\Theta a_2(\theta_1) dF_1(\theta_1) \tag{14}$$

[20]See equation (17). Evaluated at $\tau = 0$ it is 0.

and decrease utility of individuals, which is valued (in terms of public funds) by

$$-\int_{\Theta} a_2(\theta_1)\beta(1+r)\int_{\underline{\theta}_2}^{\overline{\theta}} \frac{U'(\theta_1,\tilde{\theta}_2)}{\lambda} dF_2(\tilde{\theta}_2|\theta_1)d\tilde{F}_1(\theta_1). \tag{15}$$

It also discourages savings and thereby decreases tax revenue from the capital income tax. This effect on public funds (in present value terms) is given by:

$$\tau\int_{\Theta}\frac{\partial a_2(\theta_1)}{\partial\tau}dF_1(\theta_1). \tag{16}$$

Again, note that for $\tau = 0$ this effect is of second order indicating that increasing or decreasing τ from zero has no first-order incentive costs and a non-zero capital tax is desirable whenever $(14)|_{\tau=0} + (15)|_{\tau=0} \neq 0$.

For $\tau \neq 0$,however, it holds that $(16) \neq 0$. Note that $\frac{\partial a_2(\theta_1)}{\partial\tau}$ is in general of ambiguous sign. In fact, the total behavioral response could be decomposed into an income and into a substitution effect. The substitution effect calls for lower savings due to an increase in τ. The income effect is of opposite sign: the higher τ, the less after tax wealth one has tomorrow (for a given amount of savings), which makes one save more. For expositional reasons, we now rewrite (16) as

$$-\tau\int_{\Theta}\zeta_{a_2,1-\tau}(\theta_1)\frac{a_2(\theta_1)}{1-\tau}dF_1(\theta_1), \tag{17}$$

where $\zeta_{a_2,1-\tau}(\theta_1)$ is the uncompensated elasticity of savings with respect to the net of tax rate $1-\tau$. If $\frac{\partial a_2(\theta_1)}{\partial(1-\tau)} > (<)0$, $\zeta_{a_2,1-\tau}(\theta_1)$ is positive if $a_2(\theta_1) > (<)\ 0$.

The absence of income effects on labor implies that labor supply does not change in response to a capital tax increase.[21] Optimality of τ then requires $(14) + (15) + (17) = 0$, which yields the following result, assuming $\beta(1+r) = 1$ for simplicity in the exposition:

Proposition 6. *The optimal linear capital tax rate τ satisfies*

$$\frac{\tau}{1-\tau} = \frac{\int_{\Theta} a_2(\theta_1)\left[f_1(\theta_1) - \int_{\Theta}\frac{U'(\theta_1,\tilde{\theta}_2)\tilde{f}_1(\theta_1)}{\lambda}dF_2(\tilde{\theta}_2|\theta_1)\right]d\theta_1}{\int_{\Theta} a_2(\theta_1)\zeta_{a_2,1-\tau}(\theta_1)dF_1(\theta_1)}. \tag{18}$$

Proof. See Appendix A.3.2. □

Interpretation. The optimal taxation of capital follows a very simple and intuitive equity-efficiency trade-off as is standard in the public finance literature. Capturing efficiency arguments, the tax rate is decreasing in the weighted elasticity of savings with respect to the net of tax rate $1-\tau$. Capturing equity arguments, it is higher, the more the government values

[21] With income effects there would be two additional effects. If leisure is a normal good, labor supply in period two would increase which would raise additional tax revenue. Labor supply in period 1, in contrast, would decrease. Whether the presence of income effects makes the case for capital taxes more likely depends on which of these effects dominates.

redistribution from high savers to low savers, or in other words from the wealthy to the non-wealthy. The numerator is similar to the mechanical effects as defined in Propositions 2 and 1. Without the term $a_2(\theta_1)$ the numerator in (18) would say by how much welfare (in terms of public funds) increases if one dollar is redistributed in period two from all individuals to the government. With the term $a_2(\theta_1)$, however, this effect is weighted for each θ_1 by the amount of savings. For reasonable conditional distribution functions $F_2(\theta_2|\theta_1)$, one can expect $a_2'(\theta_1) > 0$ implying that individuals with higher innate ability θ_1 save more and therefore are affected stronger from the increase in the capital income tax rate τ. If in addition Pareto weights are such that the government wants to redistribute from high innate types to low innate types, the numerator of (18) is positive, yielding positive capital taxes. For a given set of Pareto weights, capital taxes are then increasing in wealth inequality. Our numerical results in Section 5 will confirm this intuition.

Relation To Previous Public Finance Literature. Our results for the taxation of savings are different from other recent arguments in favor of capital taxation. In the *NDPF*-literature, savings are distorted in order to relax incentive constraints via an income effect on labor supply. They are not used to redistribute and are only used to provide labor incentives (see Kocherlakota (2005)).[22]

The reason is that, capital taxes are superfluous for insurance/redistribution since history dependent labor taxes can condition on the whole history of shocks θ^t and are sufficient for all desired insurance/redistribution by the planner.

As we show, with simpler and more realistic policy instruments like standard income taxes or age-dependent income taxes, capital taxes can provide additional redistribution.

A similar logic as in the *NDPF* arises in Blomquist and Micheletto (2008), who consider age-dependent nonlinear taxes in a two period model with ex-ante homogenous agents, which can end up as a high or low skilled agent in period 2. Savings are taxed to relax the incentive constraint in period 2 due to an income effect on labor supply. Savings are not taxed for redistributive issues because ex-ante homogeneous all save the same amount.[23]

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. In their framework, a positive capital tax also provides insurance against idiosyncratic risk. In addition, their timing assumptions are also different in that individuals make consumption and labor supply decisions before their shock realizes.

Developing a novel dynamic tax reform approach, Golosov, Tsyvinski, and Werquin (2013) look at the welfare effects of an increase of a linear capital tax rate starting from any given tax

[22] In Kocherlakota (2005), optimal wealth taxes are zero on average and raise no aggregate tax revenue.

[23] Bastani, Blomquist, and Micheletto (2013) numerically elaborate a similar discrete type model with ex-ante heterogeneity and raise a similar argument for taxing savings in order to relax incentive constraints. Their focus is more on the welfare gains from age-dependence and they do connect the desirability of capital taxation to wealth inequality, risk aversion and the elasticity of savings.

system and obtain a formula which is similar to (18). Finally, Piketty and Saez (2013) derive a formula for the optimal linear inheritance tax in an overlapping generations framework. As for the formulas presented in this paper, equity-efficiency considerations are key to understand optimal bequest taxation.

3.3 Labor Income Taxes in the Presence of Capital Taxes

We now move on to labor taxes in the presence of capital taxation. The following proposition summarizes our main result.

Proposition 7. *In the presence of wealth taxes, Pareto optimal marginal labor income tax rates for the young (t=1) and old (t=2) are given by*

$$\frac{\mathcal{T}_t'(y_t(\theta_t))}{1-\mathcal{T}_t'(y_t(\theta_t))} = \left(1+\frac{1}{\varepsilon(\theta_t)}\right)\frac{\eta_t(\theta_t)}{\lambda\frac{1}{(1+r)^{t-1}}\theta_t g_t(\theta_t)},$$

where

$$\eta_1(\theta_1) = \mathcal{M}_1(\theta_1) - \int_{\theta_1}^{\bar{\theta}} \mu_1(\theta_1)U''(\theta_1)d\tilde{\theta}_1$$

$$\eta_2(\theta_2) = \mathcal{M}_2(\theta_2) + \beta(1+r)(1-\tau)\int_{\Theta}\mu_1(\theta_1)\int_{\theta_2}^{\bar{\theta}} U''(\theta_1,\theta_2)dF_2(\tilde{\theta}_2|\theta_1)d\theta_1.$$

and

$$\lambda = \int_{\Theta} U'(\theta_1)d\tilde{F}_1(\theta_1) - \int_{\Theta}\mu_1(\theta_1)U''(\theta_1)d\theta_1. \tag{19}$$

$\mu_1(\theta_1)$ is the Lagrangian multiplier function on the Euler equation (5) and $g_1(\theta_1) = f_1(\theta_1)$ and $g_2(\theta_2) = \int_{\theta_1} f_2(\theta_2|\theta_1)dF_1(\theta_1)$.

Proof. See Appendix A.2.2. □

In addition to the terms in Propositions 1 and 2, additional terms show up. $\mu(\theta_1)$ denotes the Lagrangian multiplier on a type θ_1. These additional terms capture the impact of changes in savings behavior on capital tax revenue. In Appendix A.2.3, we show how these expressions can be written as behavioral responses of savings, as in the next proposition.

Proposition 8. *Pareto optimal marginal labor income tax rates for the young/old ($t = 1,2$) satisfy:*

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

where the savings tax effects equal

$$\mathcal{S}_1(\theta_1) = \lambda\tau\int_{\theta_1}^{\bar{\theta}_1}\frac{\partial a_2(\tilde{\theta}_1)}{\partial\mathcal{T}_1'(y_1(\tilde{\theta}_1))}dF_1(\tilde{\theta}_1)$$

and

$$\mathcal{S}_2(\theta_2) = \lambda\tau(1+r)\int_\Theta \frac{\partial a_2(\theta_1)}{\partial \mathcal{T}_2'(y_2(\theta_2))} dF_1(\theta_1).$$

The marginal value of public funds is defined by:

$$\lambda = \int_\Theta U'(\theta_1) d\tilde{F}_1(\theta_1) + \lambda\tau \int_\Theta \frac{\partial a_2(\theta_1)}{\partial \mathcal{T}_1(0)} dF_1(\theta_1).$$

Proof. See Appendix A.2.3. □

Comparing the optimal tax formula to the one from Proposition 1, the additional term $\mathcal{S}_i(\theta_i)$ appears. We call $\mathcal{S}_i(\theta_i)$ the savings tax effect. Such a term has been discovered by Golosov, Tsyvinski, and Werquin (2013) in a dynamic tax reform setting and by Blomquist and Micheletto (2008) in an optimal age-dependent tax setting. The effect is proportional to the savings tax τ. If capital taxes τ are different from zero, changes in savings caused by labor tax changes have a first-order effect on the government's budget. As can be seen, $\mathcal{S}_i(\theta_i)$ differs across age groups. Higher labor taxes in the first period will unambiguously induce individuals to save less and therefore $\mathcal{S}_1(\theta_1) < 0$ if $\tau > (<)0$. This reduces capital tax revenue and is a force towards lower labor taxes on the young. In contrast, higher labor taxes on the old will increase savings and therefore $\mathcal{S}_2(\theta_1) > 0$ if $\tau > 0$. Finally, the marginal value of public funds λ now also takes into account how marginally increasing consumption for everybody in period 1 has an impact on the government budget via implied savings responses.

In the age-independent case, it is then possible to show that the following proposition characterizes optimal tax rates:

Proposition 9. *Age-independent Pareto optimal marginal labor income tax rates are given by:*

$$\frac{\mathcal{T}'(y(\theta))}{1-\mathcal{T}'(y(\theta))} = \left(1+\frac{1}{\varepsilon(\theta)}\right) \frac{1}{\lambda\theta(f_1(\theta) + \frac{1}{1+r}\int_\Theta f_2(\theta|\theta_1)dF_1(\theta_1))} \times \left[\sum_{i=1}^{2} \mathcal{M}_i(\theta) + \mathcal{S}_i(\theta)\right].$$

Proof. See Appendix B.2.2. □

The effect of labor taxes on savings will in general be ambiguous – as our discussion on age-dependent labor taxes highlighted, higher taxes on the young reduce savings by an income effect but higher taxes on the old also increase savings as individuals anticipate higher taxes later in life. Determining the sign of $\mathcal{S}_1(\theta) + \mathcal{S}_2(\theta)$ is, hence, a quantitative question.

Finally, also in the presence of wealth taxes, the lump sum elements of the tax code are indeterminate.

Proposition 10. *Lump sum elements of the labor income tax functions are indeterminate also with savings taxes.*

Proof. See Appendix A.4 □

Finally, note that the mechanical effect can also be decomposed into an insurance and a redistribution component as in equations (11) and (12). The heuristic derivation via a tax reform argument is a bit more elaborate in the presence of wealth taxes. See Appendix A.2.5 for such a derivation.

4 The T-Period Case

We now present the generalization to the T-period case. Since most of the economic intuition can be well-understood from the two period case, we keep the discussion brief. Also for brevity and to avoid repetition, we representatively focus on age-dependent labor taxes and assume that capital taxation is possible and age-dependent. First, using the first-order approach described in Section 2, the following formula characterizes the optimal age-dependent marginal tax rate in period t, see Appendix A.2.1:

$$\frac{\mathcal{T}_t'(y_t(\theta_t))}{1-\mathcal{T}_t'(y_t(\theta_t))} = \left(1+\frac{1}{\varepsilon(\theta_t)}\right) \frac{\eta_t(\theta_t)}{\lambda \frac{1}{(1+r)^{t-1}} \theta_t \int_{\Theta^{t-1}} f_t(\theta_t|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}}, \tag{20}$$

where

$$\begin{aligned}\eta_t(\theta_t) = M_t(\theta_t) &- \int_{\Theta^{t-1}} \int_{\theta_t}^{\overline{\theta}} \mu_t(\theta^{t-1},\theta_s) U''(\theta^{t-1},\theta_s) d\tilde{\theta}_t d\theta^{t-1} \\ &+ \beta(1+r)(1-\tau_t) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta_t}^{\overline{\theta}} U''(\theta^{t-1},\theta_t) dF_t(\tilde{\theta}_t|\theta_{t-1}) d\theta^{t-1}.\end{aligned}$$

The mechanical effect is defined similarly as in the two-period case:

$$\begin{aligned}M_t(\theta_t) =& \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\theta_t}^{\overline{\theta}} dF_t(\tilde{\theta}_t|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \\ &- \beta^{t-1} \int_{\Theta^{t-1}} \int_{\theta_t}^{\overline{\theta}} U'(\theta^{t-1},\theta_t) dF_t(\tilde{\theta}_t|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1}.\end{aligned}$$

$\mu_t(\theta^t)$ is the Lagrangian multiplier on the Euler equation of individuals with history θ^t. Note that $\mu_0(\theta^0) = \mu_T(\theta^T) = 0$ for all θ^0 and θ^T. Analytical expressions for these multiplier functions are provided in Appendix A.1.3.

Insurance versus Redistribution: The mechanical effect can again be decomposed into an insurance and a redistribution effect. The insurance effect is given by

$$M_t^I(\theta_t) = \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \left(1 - F_t(\theta_t|\theta_{t-1}) - CU(\theta^{t-1},\theta_t)\right) h_{t-1}(\theta^{t-1}) d\theta^{t-1}$$

where

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

The redistribution effect is given by

$$M_t^R(\theta_t) = \int_{\Theta^{t-1}} \left(\frac{\lambda}{(1+r)^{t-1}} - \beta^{t-1} \int_{\Theta} U'(\theta^{t-1}, \theta_t) dF_t(\theta_t|\theta_{t-1}) \right) CU(\theta^{t-1}, \theta_t) h_{t-1}(\theta^{t-1}) d\theta^{t-1}. \tag{21}$$

For a formal derivation of this decomposition via a tax perturbation argument see Appendix A.2.5.

Savings Tax Effects. As in the two period model, the μ-terms capture savings adjustments of individuals and the fiscal effects of these adjustments have on the government budget. Alternatively, optimal marginal tax rates can also be expressed directly as functions of these savings responses. This could be derived by maximizing the planner objective subject to constraint (1) directly and invoking the envelope theorem, without making use of (5) and (7). All indirect effects then cancel out, and what remains are the direct (mechanical) effects of labor taxes on welfare as well as the effects on the government budget constraint (see Saez (2001) and Golosov, Tsyvinski, and Werquin (2013)). Based on these perturbation arguments, one can write the formula for optimal marginal tax rates as:

$$\frac{\mathcal{T}_t'(y_t(\theta_t))}{1 - \mathcal{T}_t'(y_t(\theta_t))} = \left(1 + \frac{1}{\varepsilon(\theta_t)}\right) \frac{1}{\frac{\lambda}{(1+r)^{t-1}} \theta_t f_t(\theta_t)} \times \left[\mathcal{M}_t(\theta_t) + \mathcal{S}_t(\theta_t)\right], \tag{22}$$

where

$$\mathcal{S}_t(\theta_t) = \lambda \sum_{j=1}^{T} \tau_j (1+r)^{j-1} \int_{\Theta^{j-1}} \frac{\partial a_j(\theta^{j-1})}{\partial \mathcal{T}_t'(y_t(\theta_t))} h_j(\theta^j) d\theta^j.$$

The difference to the two period model is just that labor taxes influence savings at all points in time. While the formula (22) provides a superior economic intuition compared to (20), it is very hard to get analytical expressions for responses of savings with respect to taxes. For the three period case, we analytically derive formulas for these savings responses and demonstrate the equivalence between (20) and (22), see Appendix A.2.7. This formal proof is already quite involved, as 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 big advantage of formula (20) is therefore that it can be more readily used for numerical simulations.

Capital taxes: For optimal capital taxes, analogously to labor taxes, the following condition characterizes Pareto optima as a function of the multiplier functions of the Euler equations, see Appendix A.3.1:

$$
\begin{aligned}
(1-\tau_t) = & \frac{1}{\beta(1+r)^2\mu_{t-1}(\theta^{t-1})\int_\Theta U''(\theta^{t-1},\theta_t)a_t(\theta^{t-1})dF_t(\theta_t|\theta_{t-1})d\theta^{t-1}} \\
& \Bigg[-\frac{\lambda}{(1+r)^{t-2}}\int_{\Theta^{t-1}} a_t(\theta^{t-1})h_{t-1}(\theta^{t-1})d\theta^{t-1} \\
& + \beta^{t-1}(1+r)\int_{\Theta^{t-1}} a_t(\theta^{t-1})\int_\Theta U'(\theta^{t-1},\theta_t)dF_t(\theta_t|\theta_{t-1})\tilde{h}_{t-1}(\theta^{t-1})d\theta^{t-1} \\
& + (1+r)\int_{\Theta^{t-1}} a_t(\theta^{t-1})\int_\Theta \mu_t(\theta^{t-1},\theta_t)U''(\theta^{t-1},\theta_t)d\theta_t d\theta^{t-1} \\
& - \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1})\int_\Theta U'(\theta^{t-1},\theta_t)dF_t(\theta_t|\theta_{t-1})d\theta^{t-1}\Bigg],
\end{aligned} \tag{23}
$$

We will use the above formula for our computations. It is equivalent to the optimal tax formula expressed in terms of behavioral responses, expressed here with $\beta(1+r)=1$:

$$
\begin{aligned}
\frac{\tau_t}{1-\tau_t} = & \frac{\int_{\Theta^{t-1}} a_t(\theta^{t-1})\left[h_{t-1}(\theta^{t-1}) - \int_\Theta \frac{U'(\theta^{t-1},\tilde{\theta}_t)\tilde{h}_{t-1}(\theta^{t-1})}{\lambda}dF_t(\tilde{\theta}_t|\theta_{t-1})\right]d\theta^{t-1}}{\int_{\Theta^{t-1}} a_t(\theta^{t-1})\zeta_{a_t,1-\tau_t}(\theta^{t-1})h_{t-1}(\theta^{t-1})d\theta^{t-1}} \\
& + \frac{\sum_{j=2,j\neq t}^{T}\tau_t\int_{\Theta^{t-1}}\frac{\partial a_t(\theta^{t-1})}{\partial \tau_t}h_{t-1}(\theta^{t-1})d\theta^{t-1}}{\int_{\Theta^{t-1}} a_t(\theta^{t-1})\zeta_{a_t,1-\tau_t}(\theta^{t-1})h_{t-1}(\theta^{t-1})d\theta^{t-1}}.
\end{aligned}
$$

As in the two period model, the numerator in the first line captures the redistribution effect of savings taxes, from the wealthy to individuals with low assets. In the T period model, the numerator in the second line captures that age-dependent savings taxes will have cross-price effects on savings in other periods, which in turn will have an effect on the government budget constraint. This effect was naturally not present in the 2-period case. Note that in line with the general idea behind a first-order approach, we check numerically whether the first-order conditions behind the optimal capital tax rate are not only necessary, but also sufficient.

5 A Numerical Exploration

We now simulate optimal policies for a $T=3$ period economy. The formulas derived in Section 4 will be the basis for our numerical simulation. In Section 5.1 we explain our parameterization. In the following subsections, we present the results for optimal taxes.

5.1 Inequality over the Life Cycle and Parameters

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 parameteri-

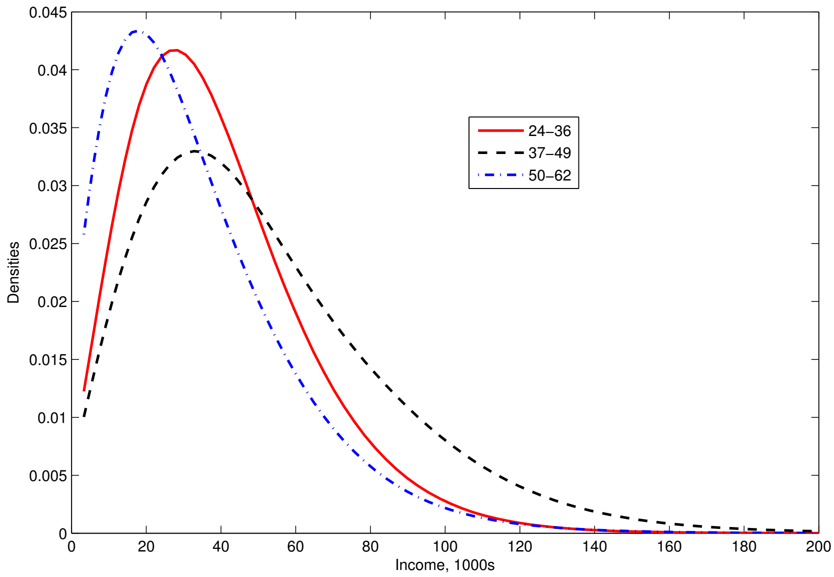


Figure 1: Income Distribution for Each Age Group

zation 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[24], 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 1 shows the three cross-sectional income distributions for each age group. 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 2 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. The roles of both persistence and risk for earnings become very clear from

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

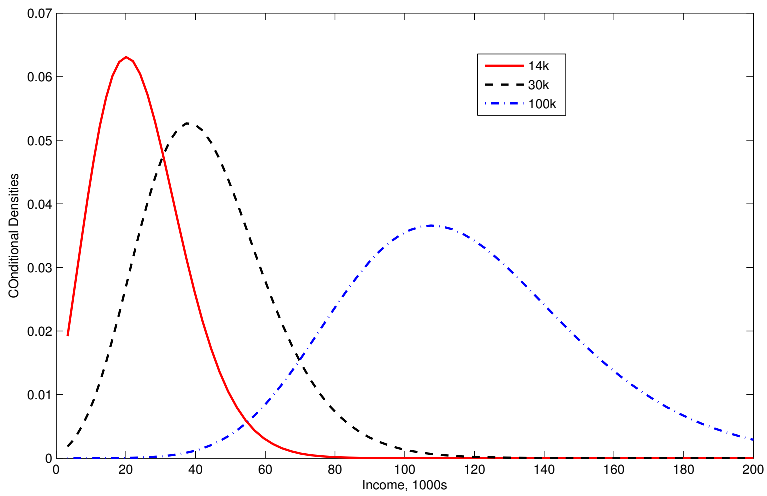


Figure 2: Conditional Income Distributions, Middle Aged Workers

this picture. To complete the parametrization of the model, we calibrate all conditional skill distributions from their income counterparts, as pioneered by Saez (2001).[25]

We assume that the utility function is of the form

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

For the benchmark, we set $\varepsilon = 3$, implying a labor supply elasticity of 0.33 (Chetty 2012) and set $\gamma = 1.5$ (Chetty 2006).[26]. The annual interest rate is 3% so in our simulations, we set the interest rate to $(1.03)^{13}$ and adjust the discount factor such that $\beta(1 + r) = 1$.

5.2 Results in the Benchmark Case

We present results for a Utilitarian social welfare function. We calculate optimal policies for four cases: age-dependent taxes are available or not and, for both cases, wealth/capital income taxes are available or not.

Optimal Labor Tax Schedules. In Figure 3, we plot optimal marginal labor income tax rates, both age-dependent and age-independent, for the case when wealth taxation is available to the government. First, all marginal tax rates are decreasing over the income distribution, reflecting that the income distributions have a log-normal shape. This marginal tax rate re-

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

[26] Our most important qualitative results are not sensitive to the choice of the labor supply elasticity. We conduct sensitivity analysis w.r.t. risk-aversion below.

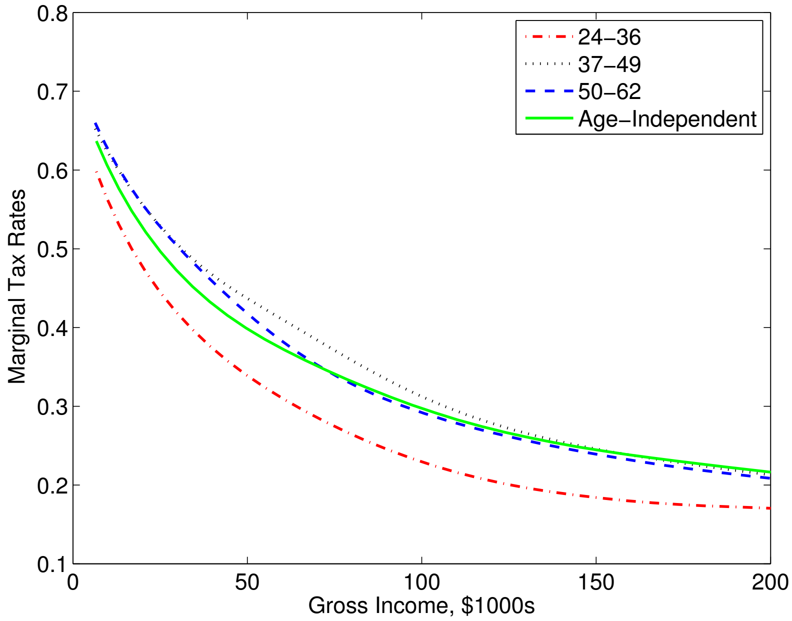


Figure 3: Optimal Utilitarian Marginal Tax Rates

gressivity is well-understood from the static literature on optimal income taxation (Diamond 1998).[27]

Second, labor income taxes are lowest for the *young*. The other tax schedules for the middle, old and the age-independent one 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 reflects a hazard-rate argument. A well known result in optimal nonlinear taxation is that the hazard rates of the income and skill distributions are both very informative statistics for the optimal pattern of marginal tax rates (Diamond 1998, Saez 2001). 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, see Figure 4.

Solely based on this reasoning, it seems desirable to tax the old the least and the middle-aged the most, since this would minimize labor supply distortions. But, as we described in theory part of the paper, in a dynamic and risky economy, age-dependent taxation has the additional power to provide insurance against income shocks. This 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 results proof that, in a world with income risk, the gains and forces behind age-dependency go beyond a pure hazard-rate logic. In the following, we illustrate this exploiting our novel decomposition of the mechanical effect into an insurance and a redistribution component.

[27]Notice that our income distributions have no Pareto tails.

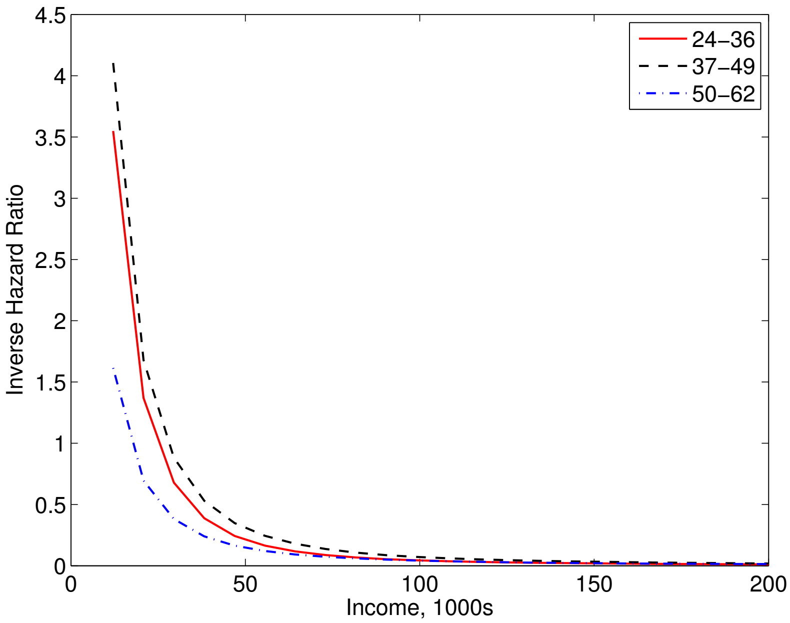


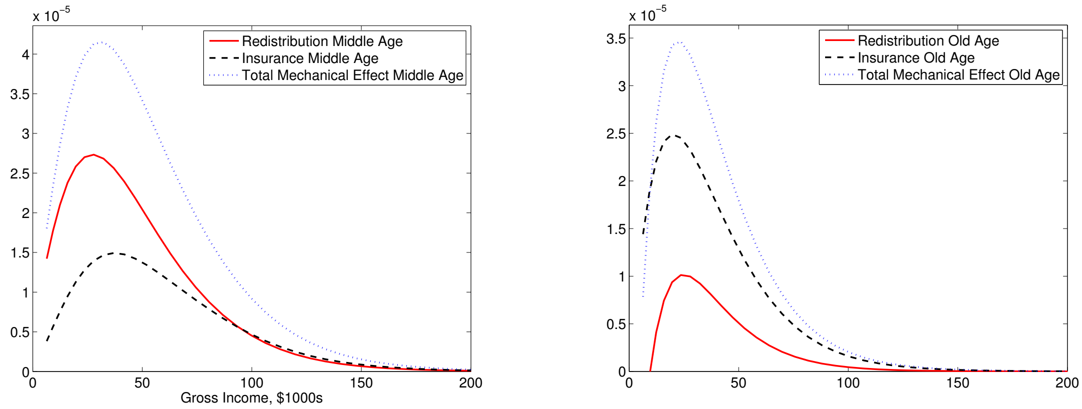
Figure 4: Ratio $\frac{1-F}{f\theta}$

Insurance Versus Redistribution. Figure 5 displays the insurance and redistribution components of the mechanical effect (equations (11) and (12) and their counterparts for period three.) For the middle-aged, the redistribution part is more important than the insurance part. In contrast, for the old, the insurance part dominates the redistribution part. This helps to understand that the insurance motive drives up taxes on the old.[28] Taxes on the young have no insurance value by definition, explaining why the old are taxed heavier than the young. Concerning the difference between the middle-aged and the old, the hazard-rate effect and the insurance effect almost cancel each other out. As taxes on the middle-aged are still higher, the hazard-rate effects is a bit more powerful in this case.

Savings Taxes. The optimal age-independent wealth tax is 4.73%. Optimal age-dependent taxation implies a wealth tax of 0.06% for the middle-aged and of 12.42% for the old. We rephrase these results in terms of a capital income tax.[29] In the age-independent case the optimal capital income tax is 14.82%. For age-dependent taxes the numbers are 0,18% for the middle-aged and 38.93% for the old. The first lesson is that the desire to tax capital in the age-independent case, is almost exclusively driven by the desire to tax the old. This can be understood with the help of our formulas for optimal capital taxes provided in the theory section. Wealth inequality is increasing over the life cycle; the standard deviation of wealth is twice as high for the old. In addition, note that there is also an interaction between capital

[28] Note that for the lowest income levels, the mechanical effect is negative. This follows from the combination of a positive savings effect and no distortion at the bottom.

[29] The simple transformation is: $\tau^r = \tau^w \frac{1+r}{r}$, where τ^w is the wealth tax. The capital income tax implements the same allocation as the wealth tax.



(a) Middle-Aged

(b) Old

Figure 5: Insurance and Redistribution Part of Optimal Taxes

taxes in period two and three. Given high capital taxes in period three, the government has the incentivize to tax period two capital less heavily because this increases savings from period two to three, which in turn yields tax revenue at the margin.[30]

Varying the yearly interest rate has only very small effects on the optimal wealth tax as long as we adjust the discount factor such that $\beta(1+r) = 1$. This implies in turn that capital income taxes are higher (lower) if interest rates are lower (higher).

The Effect of Savings Taxes on Labor Taxes. An interesting question is to what extent the presence of capital taxes influences optimal labor income taxes, or in other words how large the savings effects in the optimal labor income tax formulas turn out quantitatively. Figure 6 shows optimal taxes in the presence of a capital tax and for the case where we constrained capital taxes to be zero. Age-independent taxes are almost equivalent in both cases; the solid green line and the dotted green line are almost indistinguishable. For age-dependent taxes, we find that the presence of positive capital taxes leads to slightly higher labor income taxes on the old and slightly lower labor income taxes on the young and middle-aged. This is in line with our interpretations of the theoretical formulas in Section 3 and the findings by Golosov, Tsyvinski, and Werquin (2013). Summing up, however, we only find small quantitative effect of the savings tax effect on optimal labor taxes.

5.3 Welfare Gains, Comparative Statics and Risk Aversion

We now construct comparative static exercises changing the CRRA coefficient. We present results for optimal savings tax rates and the welfare gains from being able to tax capital. As it turns out, the conclusions regarding optimal labor tax schedules remain largely unaffected, so we refrain from showing them here.

[30] Age-dependent capital taxes may be ineffective if arbitrage is possible. In this case, the results on age-dependent capital taxes still give guidance on what drives age-independent capital taxation.

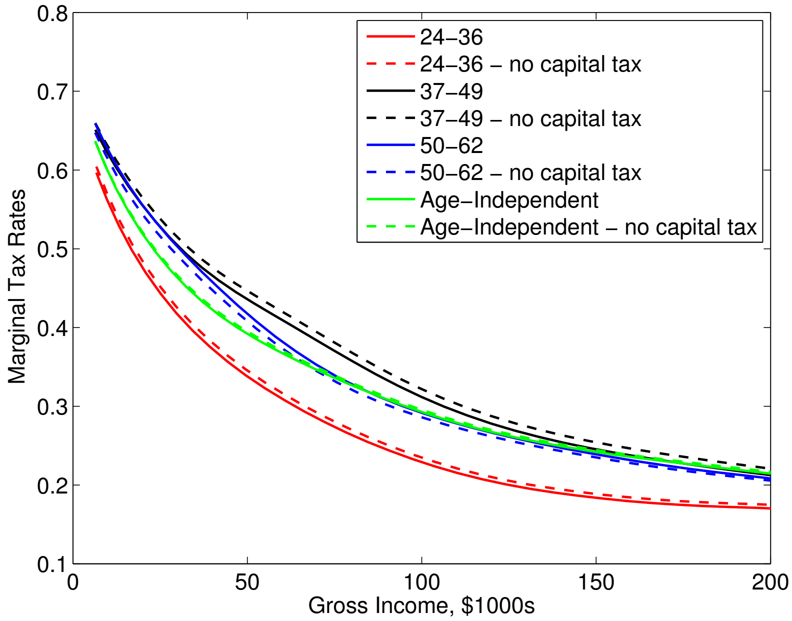


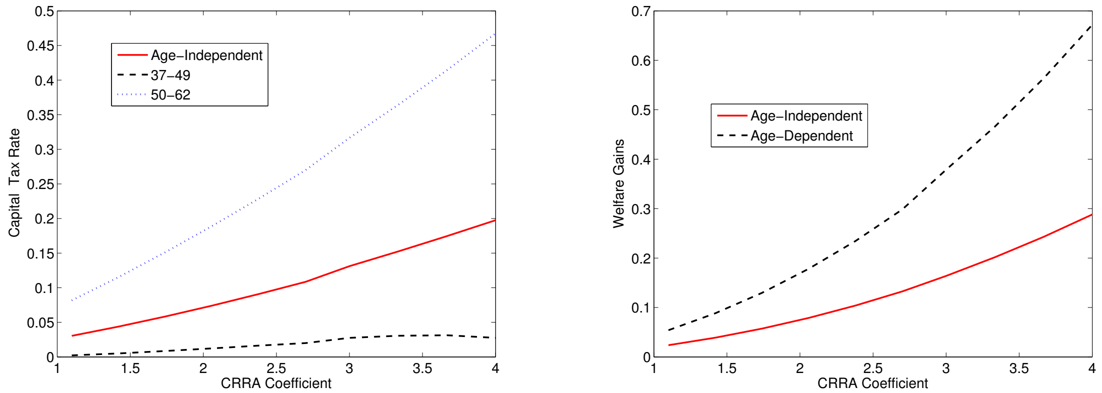
Figure 6: Illustration of Savings Effect

Capital Taxes and Risk-Aversion. Panel A of Figure 7 shows that capital tax rates (in the graph illustrated as taxes on wealth) are increasing in γ.[31] Next, notice that the optimal age-independent and the tax rate for the old are sensitive to the choice of the parameter. The tax rate for the middle-aged remains relatively stable and comparably small. This reinforces the idea the desire to tax capital, is mainly driven by the desire to tax the old. Interestingly, the slope is almost linear for all three functions. Finally, notice that for all levels of risk-aversion, age-independent tax rates are positive and significant, starting from a lower bound of about 0.03.

Welfare Gains of Capital Taxation. Panel B of Figure 7 plots the welfare gains of being able to tax capital, using the standard measure of a percentage consumption increase. For each value of γ we compute four optimal allocations, allowing the government to condition on age or not, and tax capital or not. We then compute the welfare gains from capital taxation under both scenarios.

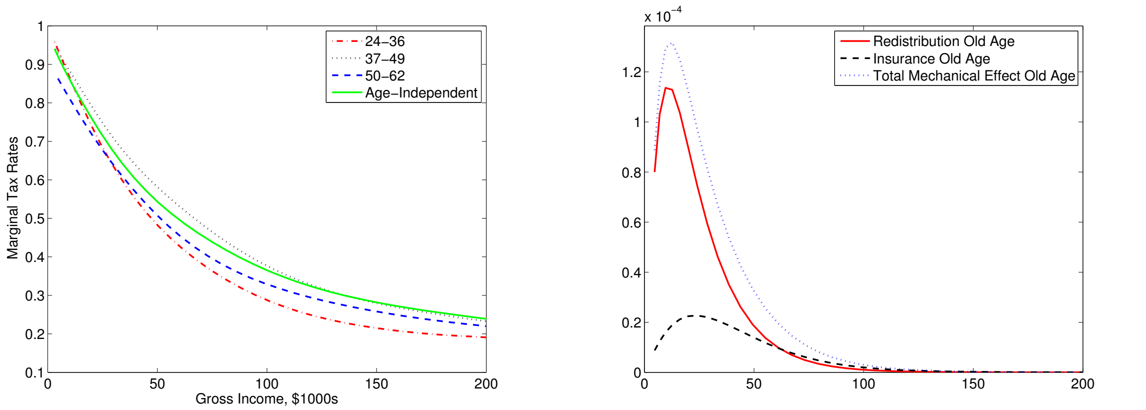
The welfare gains are sizable in all cases. They range from a lower bound of about 0.02% (age-independence, $\gamma \approx 1$) to up to 0.7% (age-dependence, $\gamma \approx 4$). The gains are larger for age-dependence, in line with the previous results that capital tax rates for the old are much higher than for the middle-aged. Finally note that the welfare gains are convexly increasing in γ, with steeper increases in the case of age-dependent taxation.

[31] Note that γ controls risk-aversion as well the intertemporal elasticity of substitution; both can concenptually influence the optimal capital tax rate.



(a) Risk-Aversion and Capital Taxes (b) Welfare Gains of Capital Taxation

Figure 7: Capital Income Taxes



(a) Marginal Labor Income Tax Rates (b) Decomposition of Mechanical Effect: Period 3

Figure 8: Rawlsian Social Welfare Function

5.4 Exploring the Pareto Frontier

In this subsection, we also explore other points on the Pareto-frontier. We first start with an often used social welfare function – the Rawlsian one. Afterwards we look at a case, which is rarely considered in the literature. We investigate Pareto weights that are less redistributive than Utilitarian weights; in fact we will consider Pareto weights that would yield the laissez-faire outcome with zero tax rates in a related static economy.

Rawlsian Optimum. The optimal labor income tax rates are plotted in Figure 8(a). The shape looks similar as in the Utilitarian optimum, however, taxes in general are – not surprisingly – larger. Figure 8(b) shows the decomposition of the mechanical effect for the old. As can be seen, redistribution here plays a much stronger role as compared to insurance. Whereas the insurance effect is of similar amount in absolute terms as in the Utilitarian case (see Figure 5(b)), the gains from redistribution are significantly larger which drives up marginal tax rates.

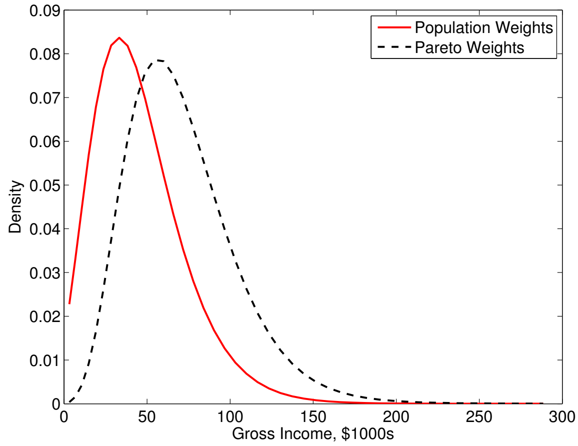
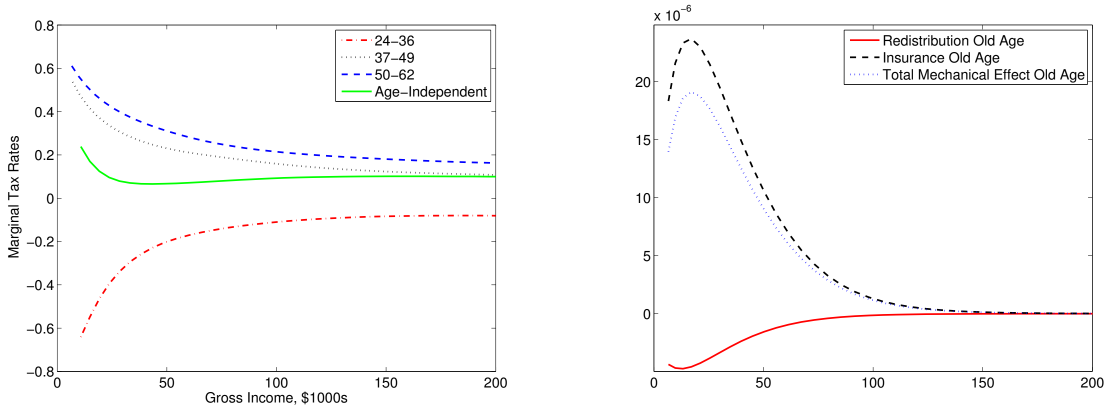


Figure 9: Illustration 'Laissez-Faire' Weights

For the optimal age-independent wealth tax, we obtain a rate of 4.3%. One might wonder why this number is lower than in the Utilitarian case. The reason is that in the Rawlsian equilibrium, rich individuals save less, because more of their labor income is already taxed. For age-dependent wealth taxes, we obtain the surprising result that the savings decision of the young individuals is subsidized because the wealth tax for the middle age individuals is -2.56%. To understand the intuition behind this result, it is necessary to take the wealth tax rate for the old into account which stands at 14.78%. Subsidizing young individuals at the margin also induces higher savings later in their life cycle which are then taxed heavily.

Laissez-Faire Weights. In a framework with heterogeneous agents, there is no correct or wrong normative objective. Typically the literature focuses on the Utilitarian and Rawlsian objective, or intermediate cases. We leave this path and instead also ask the following question: To what extent can redistributive taxation be grounded on the idea of social insurance? We therefore make the following thought experiment: We consider a static economy where productivities are distributed as in the first period of our dynamic economy. We then consider a static Mirrlees problem and back out the Pareto weights that would yield the laissez-faire equilibrium as the optimum, i.e. the Pareto weights that would lead to zero redistribution in a static economy. The respective Pareto weights are illustrated in Figure 9; richer individuals obtain a higher weight as compared to their population share.

Figure 10(a) contains the results for optimal marginal tax rates. In case the government is restricted to use age-independent labor income taxes, the government imposes marginal tax rates that are significantly larger than zero, which can be attributed to the insurance value of taxation. With age-dependent instruments, taxes rates are large in period two and three of the life cycle; this is clearly driven by the insurance effect of taxation. See Figure 10(b) for the decomposition of the mechanical effect in period three. The insurance effect is of the same size as in the Utilitarian and the Rawlsian cases, compare Figures 5(b) and 8(b). However, the fact that this insurance also implies redistribution from ex-ante high to



(a) Marginal Labor Income Tax Rates (b) Decomposition of Mechanical Effect: Period 3

Figure 10: 'Laissez-Faire' Pareto Weights

ex-ante to low types, is valued negatively by the 'Laissez-Faire' planner. To counteract this undesired redistribution, the planner therefore sets negative marginal tax rates in the first period throughout the whole income distribution. With age-dependent taxes, the planner can provide (i) insurance for individuals which is in general efficient independently from Pareto weights and (ii) avoid (too much) redistribution from ex-ante rich to ex-ante poor by counteracting (i) via negative marginal tax rates in period one. For this reason, welfare gains from age-dependence are much higher in this case (about 2% of life-time consumption) compared to the Utilitarian (0.1%) and the Rawlsian (0.5%). Finally, the age-independent wealth tax is -0.26% and the age-dependent wealth tax increases from -11.49% to 7.80% over the life cycle.

6 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. This method can be used to study age-dependent and age-independent taxes. 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 missing 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 also shown that in the presence of simple labor income taxes on the one hand and uninsurable idiosyncratic risk on the other hand, optimal capital taxes are typically non-zero. Our formula for the optimal capital tax highlights a classical equity-efficiency trade-off. This implies that capital taxes increase the redistributive power of the government. In our quantitative section, we have shown that the use of capital taxes yields significant welfare gains.

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.[32] 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 further 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.

[32]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.

AKERLOF, G. (1978): "The economics of 'tagging' as applied to the optimal income tax, welfare programs, and manpower planning," *The American Economic Review*, 68(1), 8–19.

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

ATKINSON, A., AND J. STIGLITZ (1976): "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, 6(1-2), 55–75.

BANKS, J., AND P. DIAMOND (2011): "The Base for Direct Taxation," in *Dimensions of Tax Design. The Mirrlees Review*, chap. 6, pp. 548–648. Oxford University Press.

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. (2006): "A new method of estimating risk aversion," *The American Economic Review*, 96(5), 1821–1834.

——— (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., 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.*

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., A. Tsyvinski, and N. Werquin (2013): "Dynamic Tax Reforms," *Working Paper.*

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988): "Investment, capacity utilization, and the real business cycle," *The American Economic Review*, pp. 402–417.

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

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.*

Karahan, F., and S. Ozkan (2013): "On the persistence of income shocks over the life cycle: Evidence, theory, and implications," *Review of Economic Dynamics*, 26(3), 452–476.

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.

Kocherlakota, N. R. (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73(5), 1587–1621.

Lorenz, N., and D. Sachs (2012): "Optimal Participation Taxes and Efficient Transfer Phase-Out," *Working Paper, University of Konstanz.*

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

Mendoza, E. G., and V. Z. Yue (2012): "A general equilibrium model of sovereign default and business cycles," *The Quarterly Journal of Economics*, 127(2), 889–946.

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.

Piketty, T., and E. Saez (2013): "A Theory of Optimal Inheritance Taxation," *Econometrica*, forthcoming.

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.

Scheuer, F. (2013): "Entrepreneurial Taxation with Endogenous Entry," *American Economic Journal: Policy*, forthcoming.

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. (2007): "Pareto efficient income taxation," *Massachusetts Institute of Technology, Mimeo.*

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

A Age-Dependent Taxes

A.1 Lagrangian, First-Order Conditions and Multipliers

A.1.1 The Lagrangian

Define $M_t(\theta_t) = y_t(\theta_t) - \mathcal{T}_t(y_t(\theta_t))$. Then, the Lagrangian reads as

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

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

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

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

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

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

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

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

$$
\begin{aligned}
\mathcal{L} = &\sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^t} U\Bigg(M_t(\theta_t) - a_{t+1}(\theta^t) \\
&+ (1-\tau_t)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\Bigg)\tilde{h}_t(\theta^t)d\theta^t \\
&+ \lambda \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\Theta} y_t(\theta_t) - M(\theta_t) + \tau_t(1+r)a_t(\theta^{t-1})dF_t(\theta_t|\theta_{t-1})h_{t-1}(\theta^{t-1})d\theta^{t-1} \\
&+ \sum_{t=1}^{T-1} \int_{\Theta^t} \mu_t(\theta^t)\Bigg[U'\left(M_t(\theta_t) - a_{t+1}(\theta^t) + (1-\tau_t)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right) \\
&- \beta(1+r)(1-\tau_{t+1}) \int_{\Theta} U'\Bigg(M_{t+1}(\theta_{t+1}) - a_{t+2}(\theta^t, \theta_{t+1}) \\
&+ (1-\tau_{t+1})(1+r)a_{t+1}(\theta^t) - \Psi\left(\frac{y_{t+1}(\theta_{t+1})}{\theta_{t+1}}\right)\Bigg)dF_t(\theta_{t+1}|\theta_t)\Bigg]d\theta^t \\
&- \sum_{t=1}^{T} \int_{\Theta} \eta_t'(\theta_t)\left(M_t(\theta_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)d\theta_t - \sum_{t=1}^{T}\int_{\Theta} \eta_t(\theta_t)\Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right)\frac{y_t(\theta_t)}{\theta_t^2}d\theta_t. \\
&+ \sum_{t=1}^{T} \eta_t(\overline{\theta})\left(M_t(\overline{\theta}) - \Psi\left(\frac{y_t(\overline{\theta})}{\overline{\theta}}\right)\right) - \eta_t(\underline{\theta})\left(M_t(\underline{\theta}) - \Psi\left(\frac{y_t(\underline{\theta})}{\underline{\theta}}\right)\right).
\end{aligned}
$$

A.1.2 First-Order Conditions

The derivatives with respect to the endpoint conditions yield $\forall t:\ \eta_t(\overline{\theta}) = \eta_t(\underline{\theta}) = 0$. The first-order conditions read as

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial M_s(\theta_s)} = &- \frac{\lambda}{(1+r)^{s-1}} \int_{\Theta^{s-1}} f_s(\theta_s|\theta_{s-1})h_{s-1}(\theta^{s-1})d\theta^{s-1} \\
&+ \beta^{s-1}\int_{\Theta^{s-1}} U'(\theta^{s-1},\theta_s)f_s(\theta_s|\theta_{s-1})\tilde{h}_{s-1}(\theta^{s-1})d\theta^{s-1} \\
&+ \int_{\Theta^{s-1}} \mu_s(\theta^{s-1},\theta_s)U''(\theta^{s-1},\theta_s)d\theta^{s-1} \\
&- \beta(1+r)(1-\tau_s)\int_{\Theta^{s-1}} \mu_{s-1}(\theta^{s-1})U''(\theta^{s-1},\theta_s)f_s(\theta_s|\theta_{s-1})d\theta^{s-1} \\
&- \eta_s'(\theta_s) = 0
\end{aligned} \tag{24}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial y_s(\theta_s)} = &\frac{\lambda}{(1+r)^{s-1}} \int_{\Theta^{s-1}} f_s(\theta_s|\theta_{s-1})h_{s-1}(\theta^{s-1})d\theta^{s-1} \\
&- \beta^{s-1}\int_{\Theta^{s-1}} U'(\theta^{s-1},\theta_s)\Psi'\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{1}{\theta_s}f_s(\theta_s|\theta_{s-1})\tilde{h}_{s-1}(\theta^{s-1})d\theta^{s-1} \\
&+ \int_{\Theta^{s-1}} \mu_s(\theta^{s-1},\theta_s)U''(\theta^{s-1},\theta_s)\Psi'\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{1}{\theta_s}d\theta^{s-1} \\
&- \beta(1+r)(1-\tau_s)\int_{\Theta^{s-1}} \mu_{s-1}(\theta^{s-1})U''(\theta^{s-1},\theta_s)\Psi'\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{1}{\theta_s}f_s(\theta_s|\theta_{s-1})d\theta^{s-1} \\
&- \eta_s'(\theta_s)\Psi'\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{1}{\theta_s} - \eta_s(\theta_s)\left(\Psi'\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{1}{\theta_s^2} + \Psi''\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{y_s(\theta_s)}{\theta_s^2}\right) = 0
\end{aligned} \tag{25}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_{s+1}(\theta^s)} =& \frac{\lambda}{(1+r)^{s-1}}\tau_{s+1}h_s(\theta^s) - \mu_s(\theta^s)U''(\theta^s) \\
&- (1-\tau_{s+1})^2\beta(1+r)^2\mu_s(\theta^s)\int_\Theta U''(\theta^s,\theta_{s+1})dF_{s+1}(\theta_{s+1}|\theta_s) \\
&+ (1-\tau_s)\beta(1+r)\mu_{s-1}(\theta^{s-1})U''(\theta^s)\, f_s(\theta_s|\theta_{s-1}) \\
&+ (1-\tau_{s+1})(1+r)\int_\Theta \mu_{s+1}(\theta^s,\theta_{s+1})U''(\theta^s,\theta_{s+1})d\theta_{s+1} = 0
\end{aligned} \tag{26}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \tau_s} =& \frac{\lambda}{(1+r)^{s-2}}\int_{\Theta^{s-1}} a_t(\theta^{s-1})h_{s-1}(\theta^{s-1})d\theta^{s-1} \\
&- \beta^{s-1}(1+r)\int_{\Theta^{s-1}} a_s(\theta^{s-1})\int_\Theta U'(\theta^{s-1},\theta_s)dF_s(\theta_s|\theta_{s-1})\tilde{h}_{s-1}(\theta^{s-1})d\theta^{s-1} \\
&- (1+r)\int_{\Theta^{s-1}} a_s(\theta^{s-1})\int_\Theta \mu_s(\theta^{s-1},\theta_s)U''(\theta^{s-1},\theta_s)d\theta_s d\theta^{s-1} \\
&+ \beta(1+r)^2(1-\tau_s)\int_{\Theta^{s-1}} \mu_{s-1}(\theta^{s-1})\int_\Theta U''(\theta^{s-1},\theta_s)a_s(\theta^{s-1})dF_s(\theta_s|\theta_{s-1})d\theta^{s-1} \\
&+ \beta(1+r)\int_{\Theta^{s-1}} \mu_{s-1}(\theta^{s-1})\int_\Theta U'(\theta^{s-1},\theta_s)dF_s(\theta_s|\theta_{s-1})d\theta^{s-1} = 0
\end{aligned} \tag{27}
$$

A.1.3 Multiplier Functions

Use (24) to obtain

$$
\begin{aligned}
\eta_s(\theta_s) =& \frac{\lambda}{(1+r)^{s-1}}\int_{\Theta^{s-1}}\int_{\theta_s}^{\bar{\theta}} dF_s(\tilde{\theta}_s|\theta_{s-1})h_{s-1}(\theta^{s-1})d\theta^{s-1} \\
&- \beta^{s-1}\int_{\Theta^{s-1}}\int_{\theta_s}^{\bar{\theta}} U'(R_s(\theta^{s-1},\theta_s))dF_s(\tilde{\theta}_s|\theta_{s-1})\tilde{h}_{s-1}(\theta^{s-1})d\theta^{s-1} \\
&- \int_{\Theta^{s-1}}\int_{\theta_s}^{\bar{\theta}} \mu_s(\theta^{s-1},\theta_s)U''(R_s(\theta^{s-1},\theta_s))d\tilde{\theta}_s d\theta^{s-1} \\
&+ \beta(1+r)(1-\tau_s)\int_{\Theta^{s-1}} \mu_{s-1}(\theta^{s-1})\int_{\theta_s}^{\bar{\theta}} U''(R_s(\theta^{s-1},\theta_s))dF_s(\tilde{\theta}_s|\theta_{s-1})d\theta^{s-1}.
\end{aligned} \tag{28}
$$

Next, we derive μ_t. Use (26) to obtain, with $SOC_s(\theta^s)$ being the second-order condition for savings from the individuals problem:

$$
\begin{aligned}
\mu_s(\theta^s) =& \frac{\frac{\lambda}{(1+r)^{s-1}}\tau_{s+1}h(\theta^{s-1}) + (1-\tau_s)\beta(1+r)\mu_{s-1}(\theta^{s-1})U''(\theta^s)\, f_s(\theta_s|\theta_{s-1})}{SOC_s(\theta^s)} \\
&+ \frac{(1-\tau_{s+1})(1+r)\int_\Theta \mu_{s+1}(\theta^s,\theta_{s+1})U''(\theta^s,\theta_{s+1})d\theta_{s+1}}{SOC_s(\theta^s)}.
\end{aligned} \tag{29}
$$

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

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

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

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

then, we can rewrite (29) 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 $s = T-2$:

$$\begin{aligned}\mu_{T-2}(\theta^{T-2}) =& A_{T-2}(\theta^{T-2}) + B_{T-2}(\theta^{T-2})\mu_{T-3}(\theta^{T-3}) \\ &+ \int_\Theta C_{T-2}(\theta^{T-2}, \theta_{T-1})\mu_{T-1}(\theta^{T-2}, \theta_{T-1})d\theta_{T-1}. \end{aligned} \tag{30}$$

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

$$\mu_{T-1}(\theta^{T-1}) = A_{T-1}(\theta^{T-1}) + B_{T-1}(\theta^{T-1})\mu_{T-2}(\theta^{T-2}). \tag{31}$$

Now insert (31) into (30). 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 μ_{T-3}

$$\mu_{T-3} = A_{T-3} + B_{T-3}\mu_{T-4} + \int_\Theta C_{T-3}\frac{A_{T-2} + B_{T-2}\mu_{T-3} + \int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1 - \int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}, \tag{32}$$

yielding

$$\mu_{T-3} = \frac{A_{T-3} + B_{T-3}\mu_{T-4} + \int_\Theta C_{T-3}\frac{A_{T-2} + \int_\Theta C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}}{1 - \int_\Theta \frac{C_{T-3}B_{T-2}}{1-\int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}}. \tag{33}$$

Now insert this into μ_{T-4}

$$\begin{aligned}\mu_{T-4} =& A_{T-4} + B_{T-4}\mu_{T-5} \\ &+ \int_\Theta C_{T-4}\frac{A_{T-3} + B_{T-3}\mu_{T-4} + \int_\Theta C_{T-3}\frac{A_{T-2} + \int_\Theta C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}}{1 - \int_\Theta \frac{C_{T-3}B_{T-2}}{1-\int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}}d\theta_{T-3}. \end{aligned} \tag{34}$$

Rewrite to obtain

$$\mu_{T-4} = \left[1 - \int_{\Theta} C_{T-4} B_{T-3} \left[1 - \int_{\theta_{T-2}} C_{T-3} B_{T-2} \left[1 - \int_{\Theta} C_2 B_{T-1} d\theta_{T-1}\right]^{-1}\right]^{-1}\right]^{-1}$$
$$\Bigg(A_{T-4} + B_{T-4}\mu_{T-5}$$
$$+ \int_{\Theta} C_{T-4} \frac{A_{T-3} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + \int_{\Theta} C_{T-2} A_{T-1} d\theta_{T-1}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}} d\theta_{T-2}}{1 - \int_{\Theta} \frac{C_{T-3} B_{T-2}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}}} d\theta_{T-3} \Bigg). \tag{35}$$

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

$$\mu_{T-5} = \left[1 - \int_{\Theta} C_{T-5} B_{T-4} \left[... \left[1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}\right]^{-1} .... \right]^{-1}\right]^{-1}$$
$$(A_{T-5} + B_{T-5}\mu_{T-6}$$
$$+ \int_{\Theta} C_{T-5} \frac{A_{T-4} + \int_{\Theta} C_{T-4} \frac{A_{T-3} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + \int_{\Theta} C_{T-2} A_{T-1} d\theta_{T-1}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}} d\theta_{T-2}}{1 - \int_{\Theta} \frac{C_{T-3} B_{T-2}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}}}}{1 - \int_{\Theta} \frac{C_{T-4} B_{T-3}}{1 - \int_{\Theta} \frac{C_{T-3} B_{T-2}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}}}} \Bigg). \tag{36}$$

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 μ_{T-5} as

$$\mu_{T-5} = \frac{A_{T-5} + B_{T-5}\mu_{T-6} + \int_{\Theta} C_{T-5} \frac{A_{T-4} + \int_{\theta_{T-3}} C_{T-4} \frac{A_{T-3} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + \int_{\Theta} C_{T-2} A_{T-1}}{D_{T-2}}}{D_{T-3}}}{D_{T-4}}}{D_{T-5}}. \tag{37}$$

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 μ_{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

$$\mu_2 = \frac{A_2 + B_2\mu_1 + E_2}{D_2}. \tag{38}$$

and get

$$\mu_1 = \frac{A_1 + E_1}{D_1}. \tag{39}$$

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

In equation (39) one can see that the $\mu_1(\theta_1) = 0$ if savings taxes are zero. Recursive calculation reveals that all μ_t are equal to zero.

A.2 Labor Income Taxes

A.2.1 T Periods

Dividing (25) by $\Psi' \frac{1}{\theta_s}$ and adding (24) yields

$$\frac{\mathcal{T}_s'(y_s(\theta_s))}{1 - \mathcal{T}_s'(y_s(\theta_s))} = \left(1 + \frac{1}{\varepsilon(\theta_s)}\right) \frac{\eta_s(\theta^s)}{\lambda \frac{1}{(1+r)^{s-1}} \theta_s \int_{\Theta^{s-1}} f_s(\theta_s|\theta_{s-1}) h_{s-1}(\theta^{s-1}) d\theta^{s-1}}. \tag{40}$$

Inserting (28) in (41) yields (20).

A.2.2 Two Periods

Dividing (25) by $\Psi' \frac{1}{\theta_s}$ and adding (24) yields

$$\frac{\mathcal{T}_s'(y_s(\theta_s))}{1 - \mathcal{T}_s'(y_s(\theta_s))} = \left(1 + \frac{1}{\varepsilon(\theta_s)}\right) \frac{\eta_s(\theta^s)}{\lambda \frac{1}{(1+r)^{s-1}} \theta_s \int_{\Theta^{s-1}} f_s(\theta_s|\theta_{s-1}) h_{s-1}(\theta^{s-1}) d\theta^{s-1}}. \tag{41}$$

Inserting (28) in (41) and taking into account the fact that $\mu_2(\theta^2) = 0$ as period 2 is the terminal period yields Proposition 7.

A.2.3 Two Periods - Relation to Savings Responses

For the multiplier function $\mu_1(\theta_1)$ we have (see Appendix A.1.3):

$$\mu_1(\theta_1) = \frac{\lambda\tau}{SOC(\theta_1)}$$

where $SOC(\theta_1)$ are the second-order conditions of the savings decision for an individual of type θ_1. Thus we obtain:

$$\eta_1(\theta_1) = \lambda(1 - F_1(\theta_1)) - \int_{\theta_1}^{\overline{\theta}} U'(\tilde{\theta}_1) d\tilde{F}_1(\tilde{\theta}_1) - \lambda\tau \int_{\theta_1}^{\overline{\theta}} \frac{U''(\tilde{\theta}_1)}{SOC(\theta_1)} dF_1(\tilde{\theta}_1) \tag{42}$$

and

$$\eta_2(\theta_1, \theta_2) = \frac{1}{1+r} \int_{\Theta} \int_{\theta_2}^{\overline{\theta}} dF_2(\tilde{\theta}_2|\theta_1) dF_1(\theta_1) - \beta \int_{\Theta} \int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1) d\tilde{F}_1(\tilde{\theta}_1) \tag{43}$$

$$+ \lambda\tau \int_{\Theta} \frac{\int_{\theta_2}^{\overline{\theta}} U''(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{SOC(\theta_1)} dF_1(\theta_1). \tag{44}$$

By implicitly differentiating the Euler equation of an individual of type θ_1, one can show that the last terms in (42) and (44) capture the impact of savings responses on the government budget and therefore one arrives at the results in Proposition 8.

A.2.4 Two periods and no Capital Tax

Dividing (25) by $\Psi' \frac{1}{\theta_s}$ and adding (24) yields

$$\frac{\mathcal{T}_s'(y_s(\theta_s))}{1-\mathcal{T}_s'(y_s(\theta_s))} = \left(1+\frac{1}{\varepsilon(\theta_s)}\right) \frac{\eta_s(\theta_s)}{\lambda \frac{1}{(1+r)^{s-1}} \theta_s \int_{\Theta^{s-1}} f_s(\theta_s|\theta_{s-1}) h_{s-1}(\theta^{s-1}) d\theta^{s-1}}. \tag{45}$$

with $s = 1, 2$. Now insert $\eta_s(\theta_s)$ (for $s = 1, 2$) as defined in (28). Then first use the fact that $\mu_2(\theta^2) = 0$ as period 2 is the terminal period and the fact that $\mu_1(\theta_1) = 0$ if capital taxes are zero; for the latter see the arguments in Appendix A.1.3.

A.2.5 Derivation of Insurance vs. Redistribution Decomposition in the Presence of Wealth Taxes

In this appendix we show how the decomposition of the mechanical effect can intuitively derived via a tax reform as described in Section 3.1.2. We will directly look at the case of T periods. The two period result is then a special case. We look at a reform, where we increase the marginal tax rate for individuals with income $y_t(\theta_t)$ such that all individuals with income $y_t > y_t(\theta_t)$ will pay exactly one more dollar of income taxes. This will cause a labor supply response which will have a first-order impact on public funds and savings effects that will also have first-order effects on public funds. In addition, it will have a mechanical effect on public funds given by

$$\begin{aligned} M_t(\theta_t) =& \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\theta_t}^{\overline{\theta}} dF_t(\tilde{\theta}_t|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \\ & - \beta^{t-1} \int_{\Theta^{t-1}} \int_{\theta_t}^{\overline{\theta}} U'(\theta^{t-1}, \theta_t) dF_t(\tilde{\theta}_t|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1}. \end{aligned}$$

We now show how this mechanical effect can be decomposed into a redistribution and an insurance effect as in Section 3.1.2. We therefore (as in Section 3.1.2) assume that the raised tax revenue will then be lump sum redistributed to all individuals in period t. As in Section 3.1.2, we will define a constant utility term:

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

Similarly, we can now define an insurance term that gives the welfare gain from the increase in resource due to the additional insurance provided by the tax reform:

$$\mathcal{M}_t^I(\theta_t) = \lambda \int_{\Theta^{t-1}} \left(1 - F_t(\theta_t|\theta_{t-1}) - \mathcal{CU}_t(\theta^{t-1}, \theta_t)\right) h_{t-1}(\theta^{t-1}) d\theta^{t-1}.$$

Not every type θ^{t-1} will receive the additional transfer that would leave his utility constant, however. Instead everyone receives $\Delta L(\theta_t) = \int_{\Theta^{t-1}} \int_{\theta_t}^{\overline{\theta}} dF_t(\theta|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}$. Thus the reform implies the following monetary gain for each θ^{t-1}: $\Delta L(\theta_t) - \mathcal{CU}_t(\theta^{t-1}, \theta_t)$. For some θ^{t-1} this term will be positive, for some it will be negative. The impact on welfare of these utility changes is given by

$$M_t^R(\theta_t) = \int_{\Theta^{t-1}} \left(\frac{\lambda}{(1+r)^{t-1}} - \beta^{t-1} \int_{\theta_t} U'(\theta^{t-1}, \theta_t) dF_t(\theta_t|\theta_{t-1}) \right) \left(CU(\theta^{t-1}, \theta_t) - \Delta L(\theta_t) \right) h_{t-1}(\theta^{t-1}) d\theta^{t-1}.$$

Now, however, the lump sum redistribution will cause an additional savings response. The reinterpretation of the reform will now additionally affect the Euler equations of all agents in period t. This has the following welfare implications:

$$M_t^{S1}(\theta_t) = \Delta L(\theta_t) \int_{\Theta^{t-1}} \int_{\Theta} \mu_t(\theta^{t-1}, \theta_t) U''(\theta^{t-1}, \theta_t) d\theta_t d\theta^{t-1}.$$

The reinterpretation of the reform will now additionally affect the Euler equations of agents in period $t-1$. This has the following effect on welfare:

$$M_t^{S2}(\theta_t) = -\Delta L(\theta_t)(1-\tau_t)\beta(1+r) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \Delta L(\theta_t) \int_{\Theta} U''(\theta^{t-1}, \theta_t) dF_t(\theta_t|\theta_{t-1}) d\theta^{t-1}.$$

Now we use (28) and the transversality condition, i.e. $\eta_t(\underline{\theta}) = 0$ to obtain that

$$M_t^R(\theta_t) + M_t^{S1}(\theta_t) + M_t^{S2}(\theta_t) = \int_{\Theta^{t-1}} \left(\frac{\lambda}{(1+r)^{t-1}} - \beta^{t-1} \int_{\theta_t} U'(\theta^{t-1}, \theta_t) dF_t(\theta_t|\theta_{t-1}) \right) CU(\theta^{t-1}, \theta_t) - h_{t-1}(\theta^{t-1}) d\theta^{t-1},$$

which is the redistribution term (21).

A.2.6 Proof of Proposition 5

Consider the following small perturbations of the labor income tax schedules

- Increase the marginal tax rate in period 2 in some small income interval around $y_2(\theta_2)$ with length $\Delta y_2(\theta_2)$ by $\Delta T'(y_2(\theta_2)$.
- Then change the tax function in period 1 such that the expected lifetime utilities of all θ_1-types stay unaffected.
- By definition this has no direct welfare consequences
- However its effects on the resource constraint that should all cancel out for the tax schedule to be optimal. Those three effects are:

1. **Insurance effect:** Note that the marginal tax rate increase in period 2 affected the expected utility for each θ_1-type as follows:

$$-\Delta y_2(\theta_2) \Delta T'(y_2(\theta_2)) \beta \int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)$$

   Hence, the change in period 1 taxes that leaves expected lifetime utility unaffected must obey:

$$\Delta y_2(\theta_2) \Delta T'(y_2(\theta_2)) \frac{\beta \int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{U'(\theta_1)}$$

   for each θ_1. These resources must be less than the additional mechanical tax revenue in period 2 because of risk aversion. Formally this overall insurance effect reads as

$$dW_I(\theta_2) = \Delta y_2(\theta_2) \Delta T'(y_2(\theta_2)) \lambda \int_{\Theta} \int_{\theta_2}^{\overline{\theta}} \left(1 - \frac{\beta \int_{\theta_2}^{\overline{\theta}} U'(R(\theta_1, \tilde{\theta}_2))}{U'(R(\theta_1))} \right) dF_2(\tilde{\theta}_2|\theta_1) dF_1(\theta_1).$$

2. **Behavioural Effect in Period 2:** This is rather standard. We know that all individuals with skill θ_2 will change their labor supply by

$$\frac{\partial y_2(\theta_2)}{\partial T'(y_2(\theta_2))} \cdot \Delta T'(y_2(\theta_2)) = \varepsilon \frac{y_2(\theta_2)}{1 - T'(y_2(\theta_2))} \Delta T'(y_2(\theta_2)).$$

The mass of those individuals is

$$\Delta\theta_2 \int_{\theta_1} f_2(\theta_2|\theta_1) dF_1(\theta_1) = f_2(\theta_2|\theta_1) \Delta y_2(\theta_2) \frac{\theta_2}{y_2(\theta_2)} \frac{1}{\varepsilon_{y_2,\theta_2}}.$$

Hence the impact on welfare of this behavioral effect reads as

$$dW_{B2}(\theta_2) = \Delta y_2(\theta_2) \Delta T'(y_2(\theta_2)) \lambda \frac{T'}{1-T'} \theta_2 \frac{\varepsilon}{\varepsilon+1} \int_{\theta_1} f_2(\theta_2|\theta_1) dF_1(\theta_1)$$

because $\varepsilon_{y_2,\theta_2}(\theta_2) = 1 + \varepsilon(\theta_2)$.

3. **Behavioural Effect in Period 1:** This effect is rather complex. We know that the tax function in period 1 is changed by this reform such that the lifetime utilities are unchanged. We therefore have to calculate the changes in marginal tax rates that result from this perturbation. Then we can easily calculate the effects on public funds due to the resulting change in labor supply.

   We know that the change in the tax payment for each agent in period 1 is

$$\frac{\beta \int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{U'(\theta_1)}$$

   The question is how marginal tax rates have to change such that these changes in tax payments result. The following formula tells us how marginal tax rates for each θ_1 have to change

$$\frac{\partial \frac{\int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{U'(\theta_1)}}{\partial \theta_1} = \Delta T'(\theta_1) * \frac{\partial y_1(\theta_1)}{\partial \theta_1}$$

   Hence,

$$\Delta T_1'(\theta_1) = \frac{\partial \frac{\int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{U'(\theta_1)}}{\partial \theta_1} \frac{\partial \theta_1}{\partial y_1(\theta_1)}$$

   Hence, individuals change labor supply as follows:

$$\frac{\partial y_1(\theta_1)}{\partial T'} \times \frac{\partial \frac{\int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{U'(\theta_1)}}{\partial \theta_1} \frac{\partial \theta_1}{\partial y_1(\theta_1)} = -\varepsilon \frac{y}{1-T'} \frac{\partial \theta_1}{\partial y_1(\theta_1)} \frac{\partial \frac{\int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{U'(\theta_1)}}{\partial \theta_1}$$

   Hence, the effect on the government budget for each type θ_1 reads as

$$-f_1(\theta_1) \frac{T'}{1-T'} \theta_1 \frac{\varepsilon}{\varepsilon+1} \frac{\partial \frac{\int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{U'(\theta_1)}}{\partial \theta_1}.$$

   The overall welfare effect is thus:

$$dW_{B1}(\theta_2) = \lambda \int_{\Theta} -\frac{T'}{1-T'} \theta_1 \frac{\varepsilon}{\varepsilon+1} \frac{\partial \frac{\int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{U'(\theta_1)}}{\partial \theta_1} dF_1(\theta_1).$$

For the sequence of tax function $\mathcal{T}_1, \mathcal{T}_2$ to be efficient, these effects have to add up to zero. $dW_{B1}(\theta_2) + dW_{B2}(\theta_2) + dW_I(\theta_2) = 0$ the yields the condition in Proposition 5.

A.2.7 Relation Between μ-Formula and Savings Tax Effect formula in a Three-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 θ_1, i.e. himself one period ago. We will then, in a second step, ask how type θ_1 will react to the savings change of type θ^2. But of course this reaction of type θ_1 will cause a reaction of type θ^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:

$$a_3'(\theta^2) \equiv -\frac{U''(\theta^2)}{SOC_2(\theta^2)}, \tag{46}$$

where $SOC_t(\theta^t)$ is the second-order condition for savings of an agent with history θ^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:

$$A_1(\theta_1) \equiv \frac{\beta(1+r)(1-\tau_2)\int_{\theta_2}^{\overline{\theta}} U''(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{SOC_1(\theta_1)}. \tag{47}$$

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 (46) and (47) will trigger second-round effects.

Second Round Effects: The savings adjustment of period 2, (46), will make individuals in period 1 adjust savings:

$$A_2(\theta_1) \equiv \frac{\beta(1+r)(1-\tau_2)\int_{\theta_2}^{\overline{\theta}} U''(\theta_1, \tilde{\theta}_2) a_3'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{SOC_1(\theta_1)}. \tag{48}$$

The first round savings adjustment in period 1, captured by (47), will trigger savings adjustments in period 2 for all θ_2:

$$\frac{U''(\theta_1, \theta_2)(1+r)(1-\tau_2)A_1(\theta_1)}{SOC_2(\theta^2)} = a_3'(\theta_1, \theta_2)(1+r)(1-\tau_2)A_1(\theta_1). \tag{49}$$

Third Round Effects: The savings adjustment of period 2 (49) will now trigger savings adjustments in period 1

$$\frac{\beta(1+r)^2(1-\tau_2)^2 \int_{\Theta} U''(\theta_1, \theta_2) a_3' A_1(\theta_1) dF_2(\theta_2|\theta_1)}{SOC_1(\theta_1)} = A_1(\theta_1)A_3(\theta_1), \tag{50}$$

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

The savings adjustment of period 1 (48) will now trigger savings adjustments in period 2

$$a_3'(\theta_1,\theta_2)(1+r)(1-\tau_2)A_2(\theta_1). \tag{51}$$

Fourth Round Effects: The savings adjustment of period 2 (51) will now trigger savings adjustments in period 1

$$\frac{\beta(1+r)^2(1-\tau_2)^2\int_\Theta U''(\theta_1,\theta_2)a_3'A_2(\theta_1)dF_2(\theta_2|\theta_1)}{SOC_1(\theta_1)} = A_2(\theta_1)A_3(\theta_1) \tag{52}$$

The savings adjustment of period 1 (52) will now trigger savings adjustments in period 2

$$a_3'(\theta_1,\theta_2)(1+r)(1-\tau_2)A_1(\theta_1)A_3(\theta_1) \tag{53}$$

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

$$\frac{\partial a_3(\theta^2)}{\partial \mathcal{T}_2'(y_2(\theta_2))} = a_3'(1+r)(1-\tau_2)(A_1+A_2)\sum_{i=0}^{\infty} A_3^i. \tag{54}$$

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

$$\frac{\partial a_3(\theta^2)}{\partial \mathcal{T}_2'(y_2(\theta_2))} = a_3' + a_3'(1+r)(1-\tau_2)(A_1+A_2)\sum_{i=0}^{\infty} A_3^i. \tag{55}$$

Further, period 1 savings adjust according to:

$$\frac{\partial a_2(\theta_1)}{\partial \mathcal{T}_2'(y_2(\theta_2))} = (A_1+A_2)\sum_{i=0}^{\infty} A_3^i. \tag{56}$$

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

$$\begin{aligned}&-\int_\Theta\int_{\theta_2}^{\bar{\theta}} \mu_2(\theta_1,\tilde{\theta}_2)U''(\theta_1,\tilde{\theta}_2)d\tilde{\theta}_2 d\theta_1\\ &+\beta(1+r)(1-\tau_2)\int_\Theta \mu_1(\theta_1)\int_{\theta_2}^{\bar{\theta}} U''(\theta_1,\tilde{\theta}_2))dF_2(\tilde{\theta}_2|\theta_1)d\theta_1\end{aligned} \tag{57}$$

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

$$\frac{\lambda\tau_2}{1+r}\int_\Theta\left((A_1(\theta_1)+A_2(\theta_1))\sum_{i=0}^{\infty}A_3(\theta_1)^i\right)dF_1(\theta_1)+\tau_3\frac{\lambda}{(1+r)^2}\left(a_3'+a_3'(1+r)(1-\tau_2)(A_1+A_2)\sum_{i=0}^{\infty}A_3^i\right).$$

Using (39) and evaluating for $T=3$, yields

$$\mu_1(\theta_1) = \frac{\lambda\frac{(1-\tau_2)\int_\Theta \tau_3 a_3' dF_2(\theta_2|\theta_1)f_1(\theta_1)}{SOC_1}+\lambda\frac{f_1(\theta_1)\tau_2}{SOC_1}}{1-\frac{(1-\tau_2)^2(1+r)^2\int_\Theta a_3'U''(\theta_1,\theta_2)dF_2(\theta_2|\theta_1)}{\mathrm{SOC}_1}}$$

and then inserting into (38), yields

$$\mu_2(\theta_1,\theta_2)=\mu_1(\theta_1)\beta(1-\tau_2)(1+r)\frac{U''(\theta_1,\theta_2)\,f_2(\theta_2|\theta_1)}{\mathrm{SOC}_2}+\frac{\lambda}{1+r}\frac{h_2(\theta^2)\tau_3}{\mathrm{SOC}_2} \tag{58}$$

Inserting these two terms into (57) yields

$$\begin{aligned}
&-\frac{\lambda\tau_3}{1+r}\int_{\Theta}\int_{\theta_2}^{\overline{\theta}}\frac{U''(\theta_1,\tilde{\theta}_2)}{\mathrm{SOC}_2}dF_2(\tilde{\theta}_2|\theta_1)dF_1(\theta_1)\\
&+\beta(1+r)(1-\tau_2)\int_{\Theta}\mu_1(\theta_1)\int_{\theta_2}^{\overline{\theta}}U''(\theta_1,\tilde{\theta}_2)\left(1-\frac{U''(\theta_1,\tilde{\theta}_2)}{SOC_2}\right)dF_2(\tilde{\theta}_2|\theta_1)d\theta_1
\end{aligned} \tag{59}$$

$$\begin{aligned}
&\frac{\lambda\tau_3}{1+r}\int_{\Theta}\int_{\theta_2}^{\overline{\theta}}a_3'(\theta_1,\tilde{\theta}_2)dF_2(\tilde{\theta}_2|\theta_1)dF_1(\theta_1)\\
&+\lambda\tau_2\beta(1+r)(1-\tau_2)\int_{\Theta}\left(\sum_{i=0}^{\infty}A_3^i\right)\int_{\theta_2}^{\overline{\theta}}\frac{U''(\theta_1,\tilde{\theta}_2)}{SOC_1}\left(1+a_3'\right)dF_2(\tilde{\theta}_2|\theta_1)dF_1(\theta_1)\\
&+\lambda\tau_3\beta(1+r)(1-\tau_2)\\
&\int_{\Theta}\left((1-\tau_2)\sum_{i=0}^{\infty}A_3^i\right)\left(\int_{\theta_2}a_3'dF_2(\theta_2|\theta_1)\right)\int_{\theta_2}^{\overline{\theta}}\frac{U''(\theta_1,\tilde{\theta}_2)}{SOC_1}\left(1+a_3'\right)dF_2(\tilde{\theta}_2|\theta_1)dF_1(\theta_1)
\end{aligned} \tag{60}$$

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

$$\begin{aligned}
&\frac{\lambda\tau_3}{1+r}\int_{\Theta}\int_{\theta_2}^{\overline{\theta}}a_3'(\theta_1,\tilde{\theta}_2)dF_2(\tilde{\theta}_2|\theta_1)dF_1(\theta_1)\\
&+\lambda\tau_2\int_{\Theta}\left(\sum_{i=0}^{\infty}A_3(\theta_1)^i\right)(A_1(\theta_1)+A_2(\theta_1))dF_1(\theta_1)\\
&+\lambda\tau_3\int_{\Theta}\left(\sum_{i=0}^{\infty}A_3(\theta_1)^i\right)(A_1(\theta_1)+A_2(\theta_1))\int_{\Theta}(1-\tau_2)a_3'(\theta_1,\theta_2)dF_2(\theta_2|\theta_1)dF_1(\theta_1)
\end{aligned} \tag{61}$$

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

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

and in period 3 it reads as

$$\begin{aligned}
\frac{\mathcal{T}_2'(y_3(\theta_3))}{1-\mathcal{T}_3'(y_3(\theta_3))}&=\left(\frac{1}{\varepsilon(\theta_3)}+1\right)\frac{1}{\theta_3\lambda\int_{\theta^2}f_3(\theta_3|\theta_2)h_2(\theta^2)d\theta^2}\\
&\times\left[\frac{\lambda}{(1+r)^2}\int_{\theta^2}\int_{\theta_3}^{\overline{\theta}}dF_3(\tilde{\theta}_3|\theta_2)h_2(\theta^2)d\theta^2-\beta^2\int_{\theta^2}\int_{\theta_3}^{\overline{\theta}}U'(\theta^2,\tilde{\theta}_3)dF_3(\tilde{\theta}_3|\theta_2)\tilde{h}_2(\theta^2)d\theta^2\right.\\
&\left.+\beta(1+r)(1-\tau_3)\int_{\theta^2}\mu_2(\theta^2)\int_{\theta_3}^{\overline{\theta}}U''(\theta^2,\tilde{\theta}_3)dF_3(\tilde{\theta}_3|\theta_2)d\theta^2\right]
\end{aligned}$$

Using the definition of μ_1 and μ_2 from above, one can show that the 'μ-terms' in these optimal tax formulas reflect the fiscal externalities from the savings adjustment.

A.3 Capital Taxes

A.3.1 Capital Income Taxes: T Periods

Simply rearranging (27) yields formula (23).

A.3.2 Capital Income Taxes: Two Periods

In case of two periods, (27) reads as (after inserting $\mu_1(\theta_1)$)

$$\begin{aligned}
0 =& \lambda \int_\Theta a_2(\theta_1) dF_1(\theta_1) \\
&- \beta(1+r) \int_\Theta a_2(\theta_1) \int_\Theta U'(\theta_1,\theta_2) dF_2(\theta_2|\theta_1) d\tilde{F}_1(\theta_1) \\
&+ \lambda\tau\beta(1+r)^2(1-\tau) \int_\Theta \frac{\int_\Theta U''(\theta_1,\theta_2) a_2(\theta_1) dF_2(\theta_2|\theta_1)}{SOC(\theta_1)} d\theta_1 \\
&+ \lambda\tau\beta(1+r) \int_\Theta \frac{\int_\Theta U'(\theta_1,\theta_2) dF_2(\theta_2|\theta_1)}{SOC(\theta_1)} d\theta_1 = 0. \qquad (62)
\end{aligned}$$

By implicitly differentiating the Euler equation, one can show that line three and four of (62) capture the impact of savings responses on the government budget, where line three captures the income effect on savings and line four captures the price effect. Simple rearranging then yields the formula as in Proposition 6.

A.4 Indeterminacy of Transfers

Assume that a tax system $\mathcal{T}_1, \mathcal{T}_2$ implements a certain allocation. Then there always exists the following tax system that also implements the desired allocation:

- $\mathcal{T}_1^*(0) = \mathcal{T}_1(0) - X$
- $\mathcal{T}_2^*(0) = \mathcal{T}_2(0) + (1-\tau)X$

If the Euler equation of individuals was fulfilled at the original tax system with savings $a_2(\theta_1)$, then the euler equation is also fulfilled under the new tax system with savings $a_2^*(\theta_1) = a_2^*(\theta_1) + X$. In addition, the impact on the government budget of this reform is zero. The impact of transfer changes is $(1-\tau)X - X$. However the government also obtains higher revenue via savings, which is given by τX.

B Age-Independent Taxes

B.1 Lagrangian, First-Order Conditions and Multipliers

B.1.1 The Lagrangian

Here we have $y_t(\theta_t) = y(\theta_t)$ and $M_t(\theta_t) = M(\theta_t)$. The Lagrangian then reads as

$$
\begin{aligned}
\mathcal{L} = \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^t} U \Bigg(& M(\theta_t) - a_{t+1}(\theta^t) \\
& + (1-\tau)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y(\theta_t)}{\theta_t}\right)\Bigg) \tilde{h}_t(\theta^t) d\theta^t \\
& + \lambda \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\Theta} y(\theta_t) - M(\theta_t) + \tau(1+r)a_t(\theta^{t-1}) dF_t(\theta_t|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \\
& + \sum_{t=1}^{T-1} \int_{\Theta^t} \mu_t(\theta^t) \Bigg[U'\left(M(\theta_t) - a_{t+1}(\theta^t) + (1-\tau)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y(\theta_t)}{\theta_t}\right)\right) \\
& - \beta(1+r)(1-\tau_{t+1}) \int_{\Theta} U' \Bigg(M(\theta_{t+1}) - a_{t+2}(\theta^t, \theta_{t+1}) \\
& + (1-\tau)(1+r)a_{t+1}(\theta^t) - \Psi\left(\frac{y(\theta_{t+1})}{\theta_{t+1}}\right)\Bigg) dF_{t+1}(\theta_{t+1}|\theta_t) \Bigg] d\theta^t \\
& + \int_{\Theta} \eta(\theta) \frac{\partial\left(M(\theta) - \Psi\left(\frac{y(\theta)}{\theta}\right)\right)}{\partial\theta} d\theta - \int_{\Theta} \eta(\theta) \Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{y(\theta)}{\theta^2} d\theta.
\end{aligned}
$$

Partially integrating $\int_\Theta \eta(\theta) \frac{\partial\left(M(\theta)-\Psi\left(\frac{y(\theta)}{\theta}\right)\right)}{\partial\theta} d\theta$ yields

$$
\eta(\overline{\theta})\left(M(\overline{\theta}) - \Psi\left(\frac{y(\overline{\theta})}{\overline{\theta}}\right)\right) - \eta(\underline{\theta})\left(M(\underline{\theta}) - \Psi\left(\frac{y(\underline{\theta})}{\underline{\theta}}\right)\right) - \int_{\Theta} \eta'(\theta)\left(M(\theta) - \Psi\left(\frac{y(\theta)}{\theta}\right)\right) d\theta.
$$

Inserting then yields:

$$
\begin{aligned}
\mathcal{L} = \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^t} U \Bigg(& M(\theta_t) - a_{t+1}(\theta^t) \\
& + (1-\tau)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y(\theta_t)}{\theta_t}\right)\Bigg) \tilde{h}_t(\theta^t) d\theta^t \\
& + \lambda \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\Theta} y(\theta_t) - M(\theta_t) + \tau(1+r)a_t(\theta^{t-1}) dF_t(\theta_t|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \\
& + \sum_{t=1}^{T-1} \int_{\Theta^t} \mu_t(\theta^t) \Bigg[U'\left(M(\theta_t) - a_{t+1}(\theta^t) + (1-\tau)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y(\theta_t)}{\theta_t}\right)\right) \\
& - \beta(1+r)(1-\tau_{t+1}) \int_{\Theta} U' \Bigg(M(\theta_{t+1}) - a_{t+2}(\theta^t, \theta_{t+1}) \\
& + (1-\tau)(1+r)a_{t+1}(\theta^t) - \Psi\left(\frac{y(\theta_{t+1})}{\theta_{t+1}}\right)\Bigg) dF_{t+1}(\theta_{t+1}|\theta_t) \Bigg] d\theta^t \\
& - \int_{\Theta} \eta'(\theta)\left(M(\theta) - \Psi\left(\frac{y(\theta)}{\theta}\right)\right) d\theta - \int_{\Theta} \eta(\theta) \Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{y(\theta)}{\theta^2} d\theta \\
& + \eta(\overline{\theta})\left(M(\overline{\theta}) - \Psi\left(\frac{y(\overline{\theta})}{\overline{\theta}}\right)\right) - \eta(\underline{\theta})\left(M(\underline{\theta}) - \Psi\left(\frac{y(\underline{\theta})}{\underline{\theta}}\right)\right).
\end{aligned}
$$

B.1.2 First-Order Conditions

The derivatives with respect to the endpoint conditions yield $\forall t: \ \eta_t(\overline{\theta}) = \eta_t(\underline{\theta}) = 0$. The first-order conditions read as

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial M(\theta)} = & -\sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} f_t(\theta|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \\
& + \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^{t-1}} U'(\theta^{t-1},\theta) f_t(\theta|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1} \\
& + \sum_{t=1}^{T-1} \int_{\Theta^{t-1}} \mu_t(\theta^{t-1},\theta) U''(\theta^{t-1},\theta) d\theta^{t-1} \\
& - \sum_{t=2}^{T} \beta(1+r)(1-\tau) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) U''(\theta^{t-1},\theta) f_t(\theta|\theta_{t-1}) d\theta^{t-1} \\
& - \eta'(\theta) = 0 \qquad (63)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial y(\theta)} = & \sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta} f_t(\theta|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \\
& - \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta} U'(\theta^{t-1},\theta) \Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} f_t(\theta|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1} \\
& + \sum_{t=1}^{T-1} \int_{\Theta^{t-1}} \mu_t(\theta^{t-1},\theta) U''(\theta^{t-1},\theta) \Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} d\theta^{t-1} \\
& - \sum_{t=2}^{T} \beta(1+r)(1-\tau) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) U''(\theta^{t-1},\theta) \Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} f_t(\theta|\theta_{t-1}) d\theta^{t-1} \\
& - \eta'(\theta) \Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} - \eta(\theta) \left(\Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta^2} + \Psi''\left(\frac{y(\theta)}{\theta}\right) \frac{y(\theta)}{\theta^2} \right) = 0 \qquad (64)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_{s+1}(\theta^s)} = & \frac{\lambda}{(1+r)^{s-1}} \tau h_s(\theta^s) - \mu_s(\theta^s) U''(\theta^s) \\
& - (1-\tau)^2 \beta (1+r)^2 \mu_s(\theta^s) \int_{\Theta} U''(\theta^s, \theta_{s+1})) dF_{s+1}(\theta_{s+1}|\theta_s) \\
& + (1-\tau)\beta(1+r) \mu_{s-1}(\theta^{s-1}) U''(\theta^s) f_s(\theta_s|\theta_{s-1}) \\
& + (1-\tau)(1+r) \int_{\Theta} \mu_{s+1}(\theta^s, \theta_{s+1}) U''(\theta^s, \theta_{s+1}) d\theta_{s+1} = 0 \qquad (65)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \tau} =& \sum_{t=2}^{T} \frac{\lambda}{(1+r)^{t-2}} \int_{\Theta^{t-1}} a_t(\theta^{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \\
&- \sum_{t=2}^{T} \beta^{t-1}(1+r) \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \int_{\Theta} U'(\theta^{t-1},\theta_t) dF_t(\theta_t|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1} \\
&- \sum_{t=1}^{T-1} (1+r) \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \int_{\Theta} \mu_t(\theta^{t-1},\theta_t) U''(\theta^{t-1},\theta_t) d\theta_t d\theta^{t-1} \\
&+ \sum_{t=2}^{T} \beta(1+r)^2(1-\tau_t) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\Theta} U''(\theta^{t-1},\theta_t) a_t(\theta^{t-1}) dF_t(\theta_t|\theta_{t-1}) d\theta^{t-1} \\
&+ \sum_{t=2}^{T} \beta(1+r) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\Theta} U'(\theta^{t-1},\theta_t) dF_t(\theta_t|\theta_{t-1}) d\theta^{t-1} = 0.
\end{aligned} \tag{66}
$$

B.1.3 Multiplier Functions

Use (63) to obtain

$$
\begin{aligned}
\eta(\theta) =& \sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\theta}^{\overline{\theta}} dF_t(\tilde{\theta}|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \\
&- \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^{t-1}} \int_{\theta}^{\overline{\theta}} U'(\theta^{t-1},\theta_t) dF_t(\tilde{\theta}|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1} \\
&- \sum_{t=1}^{T} \int_{\Theta^{t-1}} \int_{\theta}^{\overline{\theta}} \mu_t(\theta^{t-1},\tilde{\theta}) U''(\theta^{t-1},\tilde{\theta}) d\tilde{\theta} d\theta^{t-1} \\
&+ \sum_{t=1}^{T} \beta(1+r)(1-\tau) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta}^{\overline{\theta}} U''(\theta^{t-1},\tilde{\theta}) dF_t(\tilde{\theta}|\theta_{t-1}) d\theta^{t-1}.
\end{aligned} \tag{67}
$$

Obtaining μ_t is equivalent to the age-dependent case in Appendix A.1.3.

B.2 Labor Income Taxes

B.2.1 Labor Income Taxes: T Periods

Dividing (64) by $\Psi' \frac{1}{\theta}$ and adding (63) yields

$$
\frac{\mathcal{T}'(y(\theta))}{1-\mathcal{T}'(y(\theta))} = \left(1 + \frac{1}{\varepsilon(\theta)}\right) \frac{\eta(\theta)}{\lambda\theta \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\Theta} f_t(\theta_t|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}}. \tag{68}
$$

Inserting (67) into (68) yields the formula for optimal labor tax rates.

B.2.2 Labor Income Taxes: Two Periods

As in Appendix A.2.3, for the multiplier function $\mu_1(\theta_1)$ we now have:

$$
\mu_1(\theta_1) = \frac{\lambda\tau}{SOC(\theta_1)}.
$$

For $\eta(\theta)$ we then have

$$\eta(\theta) = \lambda(1 - F_1(\theta)) - \int_\theta^{\overline{\theta}} U'(\tilde{\theta}_1) d\tilde{F}_1(\tilde{\theta}_1) - \lambda\tau \int_\theta^{\overline{\theta}} \frac{U''(\tilde{\theta}_1)}{SOC(\theta_1)} dF_1(\tilde{\theta}_1)$$

$$\frac{1}{1+r} \int_\Theta \int_\theta^{\overline{\theta}} dF_2(\tilde{\theta}_2|\theta_1) dF_1(\theta_1) - \beta \int_\Theta \int_\theta^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1) d\tilde{F}_1(\tilde{\theta}_1) \quad (69)$$

$$+ \lambda\tau \int_\Theta \frac{\int_\theta^{\overline{\theta}} U''(\theta_1, \tilde{\theta}_2) dF_2(\tilde{\theta}_2|\theta_1)}{SOC(\theta_1)} dF_1(\theta_1). \quad (70)$$

By similar reasoning as in Proposition A.2.3, we arrive at Proposition 9.

B.2.3 Two Periods without Capital Taxes

Dividing (64) by $\Psi' \frac{1}{\theta}$ and adding (63) yields

$$\frac{\mathcal{T}'(y(\theta))}{1 - \mathcal{T}'(y(\theta))} = \left(1 + \frac{1}{\varepsilon(\theta)}\right) \frac{\eta(\theta)}{\lambda\theta \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} f_t(\theta_t|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}}. \quad (71)$$

with $s = 1, 2$. Now insert $\eta(\theta)$ (for $s = 1, 2$) as defined in (67). Then first use the fact that $\mu_2(\theta^2) = 0$ as period 2 is the terminal period and the fact that $\mu_1(\theta_1) = 0$ if capital taxes are zero; for the latter see the arguments in Appendix A.1.3.

B.3 Capital Income Taxes

Rearranging (66) would yield the age-independent equivalent to the age-dependent formula A.3.1.

C Details on Numerical Simulations

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

$$y^i_{h,t} = f(X^i_a; \theta_t) + \tilde{y}^i_{h,t},$$

where $f(X^i_a)$ is a function of the observable characteristics. Residual income is then decomposed into a fixed effect (α^i), an AR(1) component ($z^i_{h,t}$) and a transitory component ($\phi_t \epsilon^i_h$):

$$\tilde{y}^i_{h,t} = \alpha^i + z^i_{h,t} + \phi_t \epsilon^i_{h,t},$$

where the AR(1) process is given by

$$z^i_{h,t} = \rho_{h-1} z^i_{h-1,t-1} + \pi_t \eta^i_h,$$

and where the error term η^i_h captures persistent shocks, π_t is a time dependent loading factor and ρ_{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: $\sigma^2_{\eta,Y} = 0.027$, $\sigma^2_{\eta,M} = 0.013$ and $\sigma^2_{\eta,O} = 0.026$

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

Age-independent parameters:

- Variance of individual fixed effect: $\sigma^2_{\alpha} = 0.0707$

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

Time-dependent parameters:

- As we consider only one cohort, we assume the time dependent loading factors π_t and ϕ_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 $f(X^i_a; \theta_t)$:

- 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, 9.916471and 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 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).