Redistribution and Insurance with Simple Tax Instruments*

Sebastian Findeisen
University of Mannheim

Dominik Sachs
European University Institute

December 14, 2016

Abstract

We analyze optimal taxation of labor and capital income in a life cycle framework with idiosyncratic income risk and ex-ante heterogeneity. Tax instruments are simple in that they can only condition on current income. We provide a decomposition of labor income tax formulas into a *redistribution* and an *insurance* component. The latter is independent of the social welfare function and determined by the degree of income risk and risk aversion. The optimal linear capital tax is non-zero and trades off redistribution and insurance against savings distortions. Our quantitative results reveal that the insurance component contributes significantly to optimal labor income tax rates and provides a lower bound on optimal taxes. Optimal capital taxes are significant.

JEL-**classification:** H21, H23

Keywords: Optimal Taxation, Capital Taxation, Redistribution, Insurance

*Contact: findeisen@uni-mannheim.de, dominik.sachs@eui.eu. An older version of the paper had the title "Efficient Labor and Capital Income Taxation over the Life Cycle". We thank Alan Auerbach, Spencer Bastani, Felix Bierbrauer, Pierre Boyer, Friedrich Breyer, Carlos da Costa, Bas Jacobs, Emanuel Hansen, Leo Kaas, Etienne Lehmann, Emmanuel Saez, Eytan Sheshinski, Stefanie Stantcheva, Fabrizio Zilibotti, Josef Zweimueller and seminar participants in Berkeley, Mannheim, Konstanz, Zurich, the SAET 2013 (Paris), SEEK 2014 (Mannheim), CESifo Public Economics 2015 and the IIPF 2013 (Taormina) for helpful suggestions and discussions. We gratefully acknowledge 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 was partly 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 in a life cycle framework. Individuals face idiosyncratic labor income risk but are already heterogeneous when they enter the labor market, consistent with a large empirical literature.[1] Our key innovation is to focus on simple, history-independent tax instruments: nonlinear taxes on *current* labor income and linear taxes on *current* capital income taxes. This approach contrasts with and complements the recent New Dynamic Public Finance (NDPF) literature that considers history-dependent tax instruments.

In particular, our approach allows us to address a set of policy relevant questions theoretically and quantitatively: What are the most important factors shaping optimal tax schedules? What are the roles of income risk and ex-ante heterogeneity for optimal tax rates? Do labor and capital taxes interact and how does that influence their optimal design?

Our model has a simple life cycle structure. Agents enter the labor market with heterogeneous productivity levels. They live and work for T periods. An agent's productivity is stochastic and evolves as a Markov process. Each period, *after* the shock is realized, an agent makes a labor-leisure and consumption-savings decision. The return on savings is deterministic. Thus, in period t, there is uncertainty about productivity in period $t+1$ but not about the level of assets in period $t+1$.

We first derive a novel formula for optimal marginal labor income tax rates. We show how a version of the mechanical effect that is well-known from static Mirrlees models (Diamond 1998, Saez 2001) can be cleanly decomposed into an insurance and a redistribution component. Intuitively, taxes serve a social insurance role which depends on the degree of risk aversion and wage risk in the economy. The redistribution component reflects how much redistribution of resources between individuals who are ex-ante different is valued. Concretely in the life cycle context, young individuals already start out with very different income levels. The redistribution component is then mostly pinned down by differences in welfare weights on different income levels at young age. We calibrate the model based 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). We also provide a social insurance lower bound on taxes. The experiment we consider shuts down the redistributive benefits of labor taxes by adjusting the welfare weights in such a way that labor taxes would be zero in a static setting. Any positive level of taxes for these Pareto weights purely captures the insurance motive. In our benchmark calibration, we find that tax rates are strictly positive, starting at about 31%, then fall before they slightly increase again and converge to a level of around 20%.

Next, we derive a novel formula for the optimal linear capital tax rate. The optimal capital tax follows a very simple and intuitive equity-efficiency relationship: the gains from redistributing wealth are traded off against the negative incentive effects on the savings margin. In contrast to the famous Atkinson-Stiglitz result (Atkinson and Stiglitz 1976), in a dynamic

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

model with risk optimal capital taxes are, in general, non-zero. In our model savings taxes are not redundant as individuals are heterogeneous with respect to both labor income and capital income over their life cycle. It is, hence, beneficial for the government to employ two instruments with two-dimensional heterogeneity. This logic is related to the inheritance tax model by Piketty and Saez (2013). In our simulations, the government strongly relies on capital income taxation and the optimal tax rate is around 19% – even though the only savings motive in the model is the precautionary one.

In addition to those quantitative baseline results, we conduct several experiments to investigate how the social welfare function and idiosyncratic risk influence optimal policies. We also study the welfare losses from simplicity by comparing our policies to the dynamic mechanism design solution (NDPF) and optimal age-dependent taxes, which can be considered as an intermediate case.

Our framework also allows to investigate the interaction between labor and capital taxation, which is not possible in the static Mirrlees model. We examine optimal capital income taxes for a given labor income tax.[2] Strikingly, we find that for given labor income taxes, optimal capital tax rates differ substantially, depending on how labor income taxes are set. This depends on two mechanisms. First, lower labor income taxes lead to a more concentrated distribution of wealth which increases the redistributive power of capital income taxes. Second, the lower the degree of social insurance through labor income taxation, the stronger the desire to self insure in the form of precautionary savings and the lower the elasticity of savings with respect to capital income taxes.

Finally, our contribution is also of technical nature. We show that assuming preferences without income effects on labor supply is the key simplification to make the problem of choosing optimal history independent but fully nonlinear labor income taxes tractable.[3] 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 the history of shocks $\theta^t = (\theta_1, \theta_2, ..., \theta_t)$. This guarantees that the individuals can easily be ordered among the y_t dimension. 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. We believe our approach is also attractive for other life cycle settings, where the focus should be on history independent but fully nonlinear labor taxes, for example, settings with retirement.

[2]In an earlier version of this paper, we also studied the optimal labor income tax for a given level of the capital tax. The level of the capital tax did not matter a lot for optimal labor income taxes, see Findeisen and Sachs (2014).

[3]This also allows to look at welfare losses from simplicity as compared to the full mechanism-design optimum. We provide such an analysis in Section 4.7.

Related Literature. Our paper builds on the work by (Piketty 1997, Diamond 1998, Saez 2001). They were among the first to write down the optimal marginal tax rate formula as a function of elasticities and the skill distribution. Our analysis has the same goal in a dynamic framework and we show how the formula from Diamond (1998) is augmented in this context. The novel force we find here is that the interaction with savings taxes matters. This happens because the savings decision is endogenous with respect to labor taxes, which in turn affects the government's budget by the presence of savings taxes. Another difference is that in our dynamic setting, the so-called 'mechanical effect' captures two things: redistribution between ex-ante heterogeneous agents and social insurance against idiosyncratic wage risk.

Two related papers in the so-called *New Dynamic Public Finance* (NDPF) are Golosov, Troshkin, and Tsyvinski (2016) and Farhi and Werning (2013). They characterize the solution to the dynamic mechanism design problem when the planner's constraint on policy instruments only comes from the asymmetric information problem. Our approach is complementary: we restrict the instruments to only condition on current income. Arguably, this brings the policies closer to reality, as complex history-dependent taxation is required in the NDPF. Because of this restriction to not use all available information from the past but to tax only based on current income, the optimal marginal tax rate formulas we obtain are simpler to interpret.[4]

A recent related paper, that also studies optimal nonlinear labor income taxation in the presence of risk and uncertainty is Boadway and Sato (2015). Our approach differs in two respects: (i) their timing structure is different in that individuals choose their labor supply before wage risk is realized – we assume that individuals choose their labor supply after the realization of uncertainty. We share the view of Boadway and Sato (2015, p.12) that "In reality, there are elements of both approaches present" and therefore consider our approach as complementary. We relate our formulas for optimal labor taxes to theirs. (ii) They study a static setting and therefore do not study the question of capital taxation.

Two recent papers also study simpler policies in dynamic stochastic environments. Weinzierl (2011) and Bastani, Blomquist, and Micheletto (2013) study age-independent and age-dependent income taxation to quantify the welfare gains from age-dependent taxation.[5] These papers 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 and the optimal tax results can be connected to the literature above (Diamond 1998, Saez 2001, Golosov, Troshkin, and Tsyvinski 2016). Whereas we focus on age-independent taxation in this

[4] 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.

[5] Blomquist and Micheletto (2008) is an important earlier theoretical contribution in this literature, where it is shown that age-dependent taxes can Pareto improve on age-independent taxes and that capital should be taxed. Da Costa and Santos (2015) study optimal age-dependent taxation in an OLG economy and find that parts of the welfare gains from age-dependent taxes are lost in the transition due to the endogeneity of human capital.

paper, our method also allows us to study optimal age-dependent taxation, see Findeisen and Sachs (2014), which is an earlier version of this paper.

The paper is also related to Golosov, Tsyvinski, and Werquin (2014), who study general dynamic tax reforms and elaborate the welfare gains from the sophistication of the tax code such as age dependence, history dependence or joint taxation of labor and capital income. Like 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. In contrast to them, our focus is on the properties of optimal policies and the elaboration of the distinction between the insurance and redistribution motive of taxation.

In a series of papers, Cremer and Gahvari (1995a, 1995b, 1999) study the differential tax treatment of goods that are precommitted before the realization of uncertainty and those that are not. In our framework, consumption today can be considered as the former whereas consumption of tomorrow can be considered as the latter. Positive capital income taxation is, therefore, in line with their finding that precommitted goods should be taxed at a lower rate.

Stantcheva (2015) and 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 precautionary life cycle savings instead of intergenerational considerations for capital taxes. Saez and Stantcheva (2016) show that the nonlinear capital tax problem is isomorphic to the nonlinear labor income tax problem. In their quantitative exercise they show that the optimal capital tax schedule is close to linear because wealth is very concentrated and the constant top tax rate kicks in very early.

Best and Kleven (2013) augment the canonical optimal tax framework by incorporating career effects into a deterministic model. 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.[6]

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.[7] While our approach shares some features with a Ramsey type of exercise because we restrict tax instruments, 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.

Structure. We start with a description of the formal framework in Section 2. We also describe our technical contribution in this section. Section 3 contains our theoretical results on labor and capital taxes. Optimal policy simulations are presented in Section 4. Finally, in Section 5 we conclude.

[6] Related dynamic tax models with human capital are, for example, developed in Da Costa and Maestri (2007), Kapicka and Neira (2015), Findeisen and Sachs (2016a) and Stantcheva (2016). 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.

[7] Erosa and Gervais (2002) study optimal age-dependent and age-independent linear taxation of capital and labor income in a life cycle growth model.

2 The Model

2.1 Environment

We consider a life cycle framework with T periods. Individuals already differ in their first period skill θ_1 which is distributed according to the cumulative distribution function (cdf) $F_1(\theta_1)$ with density $f_1(\theta_1)$. At the beginning of each period $t = 2, ..., T$, individuals draw a new productivity shock θ_t according to cdf $F_t(\theta_t|\theta_{t-1})$ with respective density $f_t(\theta_t|\theta_{t-1})$. We assume for all t that $\theta_t \in \Theta = [\underline{\theta}, \overline{\theta}]$.

In every period t individuals first observe their shock θ_t and then make a labor-leisure and a consumption-savings 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 we assume $U' > 0, U'' < 0$, and $\Psi', \Psi'' > 0$. c_t is consumption in period t, y_t is gross income in period t and $\frac{y_t}{\theta_t}$ captures labor effort. The discount factor is β. The interest on savings r is fixed; thus we either consider a small open economy or a linear production technology. We assume incomplete markets – individuals only have access to risk-free one period bonds.[8] Finally, we assume $\beta(1 + r) = 1$, which simplifies the exposition.

Importantly, our functional form assumption about preferences eliminates income effects on labor supply, we discuss this choice below in Section 2.1.2. This assumption is crucial for the tractability of the dynamic optimal tax problem with simple instruments as we describe in detail in this section.

We denote the history of shocks in period t by $\theta^t = (\theta_1, \theta_2, \ldots, \theta_t)$ with $h_t(\theta^t)$ being the probability of history t, i.e. $h_t(\theta^t) = f_t(\theta_t|\theta_{t-1})f_{t-1}(\theta_{t-1}|\theta_{t-2}).....f_1(\theta_1)$. Denote by Θ^t the set of possible histories in t. 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 $U''(\theta^t)$.

Finally, we denote by $f_t(\theta_t)$ the cross-sectional skill distribution at time t, i.e. $f_t(\theta_t) = \int_{\Theta^{t-1}} f_t(\theta_t|\theta_{t-1})h_t(\theta^{t-1})d\theta^{t-1}$.

2.1.1 Timing and the Realization of Uncertainty

Before turning to the social planner's problem in the next section, we briefly elaborate on the exact timing in the model, the role of uncertainty and how it connects to other recent approaches in the literature. We assume that individuals in period t decide about labor supply and savings *after* the observation of their shock θ_t. In period t, labor income y_t and assets a_{t+1} are deterministic, whereas y_j with $j = t + 1, ..., T$ and assets a_k with $k = t + 2, ..., T$ are

[8] 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 zero consumption and a negative argument in $U(\cdot)$ because of the disutility of labor. Therefore, in the absence of taxes, the maximal amount of debt is $\sum_{t=s}^{T}\left(y_t(\underline{\theta}) - \Psi\left(\frac{y_t(\underline{\theta})}{\underline{\theta}}\right)\right)$.

uncertain. These assumptions about the timing choices and the realization of uncertainty are standard in the macroeconomic literature (Ljungqvist and Sargent (2012)) and the more recent NPDF literature following Golosov, Kocherlakota, and Tsyvinski (2003).

Certainly, other assumptions are not less plausible. E.g., one could assume that the shock θ_t realizes after labor effort is chosen, as assumed by Jacobs and Schindler (2012) and Boadway and Sato (2015). In that case, labor income taxes in period t would have an insurance role also within period t, whereas in our setting it only provides insurance from the point of view of the period before t. In reality, the timing of the realization of uncertainty and choices is probably a mix and we, therefore, consider our work as complementary.

2.1.2 The Role of Income Effects

The assumption that there are no income effects on labor supply is crucial for tractability of the optimal tax problem, see the remainder of Section 2. In particular, our preference specification implies that there is no intertemporal substitution or – in other words – that individuals adjust income in period t in the same way for a permanent or a temporary wage shock. The Hicksian and the Frisch elasticity of labor supply are equivalent. Further, the assumption implies that there is no precautionary labor supply. Although we do not want to claim that the absence of income effects describes reality in the best possible way, there is some empirical evidence that supports our simplifying assumption.[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 (2014) using the universe of Danish tax records). Chetty, Guren, Manoli, and Weber (2011) provide a meta study to reconcile different findings on labor supply elasticities. Whereas they argue that the Frisch elasticity is slightly larger than the Hicksian one, the difference is rather small. Finally, 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).

2.2 Policy Instruments

We are interested in the Pareto efficient set of nonlinear labor income tax schedules and linear capital income tax rates that only condition on current income. The government chooses a labor income tax schedule $\mathcal{T}(\cdot)$ and a linear capital tax rate τ_s. In the remainder of this paper, we use the notions wealth, savings or capital for a_t interchangeably. Also note that we define the savings tax τ_s as a stock tax and not a flow tax. However, there is always a one-to-one mapping between such a stock tax on a_t and a tax on capital income ra_t. Thus, there is no loss of generality in the way we define τ_s. In the following, we use the notions capital taxes, wealth

[9]In Section 3.4 we provide a discussion how the results might plausibly change with income effects.

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.[10]

One reason why our focus is on linear savings taxes is tractability. Additionally, the majority of OECD uses flat taxes on personal capital income (Harding 2013).[11] On the theoretical side, taxing capital linearly is often linked to arbitrage opportunities. If individual A faces a higher marginal tax rate on savings than individual B, there exists an arbitrage opportunity, if the government cannot perfectly monitor transactions between all agents. The assumption of linear capital taxes can, hence, be grounded on the idea that the government cannot observe consumption on the individual level and is often commonly in the public finance literature. See Hammond (1987) for a more general theoretical discussion of that issue. Finally, in a very recent paper, Saez and Stantcheva (2016) show that the optimal capital tax schedule is close to linear. The reason is that capital is very concentrated (empirically) and therefore a constant top tax result, similar as for labor income taxes (Saez 2001), applies throughout a large part of the wealth distribution. Thus restricting capital income taxes to be linear seems to be a minor issue against this background.

Relation to Ramsey and Mirrlees Approaches. The tax problem we look at lies at the intersection of previously considered instruments. As in the Mirrlees taxation approach, the labor tax is fully nonlinear with no parametric restrictions. However, as we consider a life cycle we also characterize linear savings taxes. In the Ramsey approach one typically considers linear labor and savings taxes. 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 set of allocations that can be attained is a subset of the allocations that can be reached via the *NDPF* approach.

2.3 Individual Problem Given Taxes

Each period, individuals make a work and savings decision. Formally, the recursive problem of individuals given taxes reads as:

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

[10] 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.

[11] See Table 5 in Harding (2013), which shows that 19 of the listed countries use flat taxes and 13 use the relevant personal income tax rate.

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 and the savings tax rate τ_s. It is thus only a function of the current shock and of the labor income tax schedule: $y(\theta_t, \mathcal{T})$.*

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: $a_{t+1}(\theta_t, a_t, \mathcal{T}, \tau_s)$. Recursively inserting, one can also write $a_{t+1}(\theta^t, \mathcal{T}, \tau_s)$. 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 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 to $h_t(\theta^t)$, 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.

The tax problem of the social planner is then:

$$\max_{\mathcal{T}, \tau_s} \int_{\Theta} V_1(\theta_1, 0, \mathcal{T}, \tau_s) d\tilde{F}_1(\theta_1) \tag{2}$$

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:[12]

$$\begin{aligned} &\sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^t} \mathcal{T}(y(\theta_t)) h_t(\theta^t) d\theta^t \\ &+ \sum_{t=2}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \tau_s (1+r) a_t(\theta^{t-1}, \mathcal{T}, \tau_s) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \geq \mathcal{R} \end{aligned} \tag{3}$$

where $\mathcal{R}$ is some exogenous revenue requirement of the government.

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.

[12] If we integrate over different histories θ^t, then within the integral we use θ_t to denote the last element of θ^t.

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 $y(\theta_t)$ instead of $y(\theta_t, \mathcal{T})$ and $a_t(\theta^{t-1})$ instead of $a_t(\theta^{t-1}, \mathcal{T}, \tau_s)$.

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, they are particularly simple because the problem does not depend on time t but only on the realized shock. The reason is that labor income taxes do not depend on time. Thus, we have $\forall\, \theta \in \Theta$:

$$1 - \mathcal{T}'(y(\theta)) = \Psi' \left(\frac{y(\theta)}{\theta} \right) \frac{1}{\theta}. \tag{4}$$

For the savings decision, the first order condition $\forall t = 1, ..., T-1$ and $\forall \theta^t \in \Theta^t$ is given by:

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

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 smaller or equal to zero. For labor supply, a familiar argument from the standard Mirrlees model implies that this holds if and only if[13]

$$y'(\theta) \geq 0 \;\forall\, \theta \in \Theta. \tag{6}$$

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

In a final step, we make use of a change of variables and define $M(\theta_t) = y(\theta_t) - \mathcal{T}(y(\theta_t))$. Applying this for (4) is still problematic as it contains $\mathcal{T}'$, however. To tackle this problem, we make use of the following derivative

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

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

Inserting (4) into this derivative yields:

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

Thus, (7) is equivalent to (4). Next, we apply the change of variables $M(\theta) = y(\theta) - \mathcal{T}(y(\theta))$ for the LHS of (7), which leads to a problem that can be solved with calculus of variation or optimal control:

Proposition 1. *Instead of choosing $\mathcal{T}$ and τ_s to maximize (2) subject to (1) and (3), the planner can also choose $\{M(\theta), y(\theta)\}_{\theta\in\Theta}$, $\{\{a_t(\theta^{t-1})\}_{\theta^{t-1}\in\Theta^{t-1}}\}_{t=1,\ldots,T}$ and τ_s subject to (3), (5), (6) and (7), where $y(\theta) - \mathcal{T}(y(\theta)) = M(\theta)$.*

The approach can also be interpreted as a restricted direct mechanism that is augmented by a savings choice. Agents report their type θ_t and the planner assigns bundles $(M(\theta_t), y(\theta_t))$ – it is a restricted mechanism because the planner cannot make M and y conditional on the history of shocks. We provide the Lagrangian and the first-order conditions in Appendix A.1. $\eta(\theta)$ denotes the Lagrangian multiplier function on (7) and $\mu_t(\theta^t)$ the one on (5). Further denote by λ the Lagrangian multiplier on the resource constraint, i.e. the marginal value of public funds. 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 check ex-post whether the monotonicity condition is fulfilled.

3 Optimal Taxation

We now characterize optimal labor and capital income taxes. Importantly, the formulas for labor and savings taxes are also valid if the other instrument is not chosen optimally.

3.1 Labor Income Taxes

We directly start with the characterization of optimal marginal tax rates.

Proposition 2. *Optimal marginal tax rates on labor income $y(\theta)$ satisfy:*

$$\frac{\mathcal{T}'(y(\theta))}{1 - \mathcal{T}'(y(\theta))} = \left(1 + \frac{1}{\varepsilon(\theta)}\right)\frac{1}{\theta \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} f_t(\theta)} \times \eta(\theta), \tag{8}$$

where $\eta(\theta) = \sum_{t=1}^{T}\left[\mathcal{M}_t(\theta) + \mathcal{S}_t(\theta)\right]$ with

$$
\begin{aligned}
\mathcal{M}_t(\theta) =& \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\underline{\theta}}^{\overline{\theta}} dF_t(\tilde{\theta}|\theta_{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1} \\
&- \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\underline{\theta}}^{\overline{\theta}} U'(\theta^{t-1},\tilde{\theta}) dF_t(\tilde{\theta}|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1}
\end{aligned} \tag{9}
$$

and

$$
\begin{aligned}
\mathcal{S}_t(\theta) =& - \int_{\Theta^{t-1}} \int_{\underline{\theta}}^{\overline{\theta}} \mu_t(\theta^{t-1},\tilde{\theta}) U''(\theta^{t-1},\tilde{\theta}) d\tilde{\theta} d\theta^{t-1} \\
&+ (1-\tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\underline{\theta}}^{\overline{\theta}} U''(\theta^{t-1},\tilde{\theta}) dF_t(\tilde{\theta}|\theta_{t-1}) d\theta^{t-1}.
\end{aligned} \tag{10}
$$

In Appendix A.2, we derive an expression for the multiplier function $\mu_t(\theta^t)$. Further, we have $\eta(\underline{\theta}) = \eta(\overline{\theta}) = 0$ and therefore $\mathcal{T}'(y(\underline{\theta})) = 0$, and $\mathcal{T}(y(\overline{\theta})) = 0$ – i.e. no distortion at the bottom and top – if $\overline{\theta} < \infty$.

Compared to the static case with quasi-linear preferences (Diamond 1998), there are three differences: (i) naturally in our life cycle setting, there is an aggregation over different points in time and periods later in life are discounted. (ii) The so-called mechanical effect $\mathcal{M}_t(\theta)$ captures both, insurance and redistribution motive of taxation that we decompose in Section 3.3. (iii) The term $\mathcal{S}_t(\theta)$ captures the impact of slightly increasing the marginal tax rate at income $y(\theta_t)$ on the Euler equations constraints of the problem described in Proposition 1. As we show below, one can intuitively connect it to perturbation arguments.

In the life cycle model, $\mathcal{M}_t(\theta)$ is defined for every period t. $\mathcal{M}_t(\theta)$ measures the impact on welfare of taking one (marginal) dollar from individuals in period t with income above $y(\theta)$. One further small difference are the terms $\frac{1}{(1+r)^{t-1}}$ that solely capture discounting to the first period. Also integration over types is slightly more involved as there are individuals with one and the same productivity in period t but different histories of productivities. A more important difference that goes beyond notation issues is that it now also captures the insurance motive of taxation on top of redistributive concerns. This is not directly obvious, but we devote Section 3.3 to this issue and formally disentangle the welfare gains from redistribution and insurance.

Before we turn to this decomposition, we discuss the second term that is novel in comparison to the static Mirrlees model. $\mathcal{S}_t(\theta)$ captures the impact of labor income taxes on the Euler equation constraints (5). For $\tau_s = 0$ this additional term is equal to zero because the values of the Lagrangian multipliers on the Euler equations are equal to zero. Relaxing or tightening the Euler equations has no first-order effect on welfare because it does not affect incentives to supply labor. Only for $\tau_s \neq 0$, relaxing or tightening the Euler equations has a first-order impact on welfare through the implied change in capital tax revenue.

Alternatively, one can derive (8) with a tax perturbation. In that case, the term $\mathcal{S}_t(\theta)$ gets a straightforward economic interpretation. We briefly sketch this perturbation argument.

Consider – starting from the optimal tax system – an infinitesimal increase of the marginal tax rate $d\mathcal{T}'$ at an infinitesimal income interval with length $dy(\theta)$ around income level $y(\theta)$. Since the tax system was optimal, this should have no first-order effect on welfare. The impact on welfare can be decomposed into three effects.

First, there is a mechanical welfare effect from taking money from individuals with income $> y(\theta)$. It is given by $\mathcal{M}_t(\theta) \times d\mathcal{T}'dy(\theta)$ for each period t. It depends on 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 larger than y$y(\theta)$ in period t.

Second, an increase in the marginal tax rate triggers a loss in tax revenue which is induced by lower labor supply of individuals of type θ. From the tax perturbation literature we know that this can be captured by $\mathcal{LS}_t(\theta) \times d\mathcal{T}'dy(\theta)$, where $\mathcal{LS}_t(\theta) \equiv -\lambda \frac{\mathcal{T}'(y(\theta))}{1-\mathcal{T}'(y(\theta))}\theta \frac{\varepsilon}{\varepsilon+1}\frac{1}{(1+r)^{t-1}} f_t(\theta)$ for each period t, see e.g. Piketty (1997). The change in labor supply has no direct impact on welfare by the envelope condition.[14]

Third, the tax perturbation also impacts the savings decision in each period and therefore changes welfare through the implied change in revenue from the savings tax by $d\mathcal{T}'dy(\theta) \times \mathcal{S}^{pert}(\theta)$, where:[15]

$$\mathcal{S}^{pert}(\theta) = \lambda\tau_s \sum_{j=2}^{T}(1+r)^{j-2} \int_{\Theta^{j-1}} \int_{\theta}^{\overline{\theta}} \frac{\partial a_j(\theta^{j-1})}{\partial \mathcal{T}(y(\tilde{\theta}_t))} d\tilde{\theta}_t \; h_{j-1}(\theta^{j-1}) d\theta^{j-1}. \tag{11}$$

The sign of $\mathcal{S}^{pert}(\theta)$ is ambiguous. On the one hand, higher labor taxes reduce savings by an income effect as it reduces resources available to save for all individuals. On the other hand, higher labor taxes also mean that tax burdens in the future are higher, which drives up the need for savings. Finally, higher labor taxes also imply a higher degree of social insurance which reduces the demand for savings as a self-insurance device.

Given that the initial tax function was optimal, the overall impact on welfare must add up to zero:

$$\Delta\mathcal{T}'\Delta y(\theta) \times \left(\mathcal{S}^{pert}(\theta) + \sum_{t=1}^{T} \left(\mathcal{LS}_t(\theta) + \mathcal{M}_t(\theta)\right) \right) = 0. \tag{12}$$

Solving (12) for $\frac{T'(y(\theta))}{1-T'(y(\theta))}$ reveals that $\mathcal{S}^{pert}(\theta) = \sum_{t=1}^{T} S_t(\theta)$, with $S_t(\theta)$ defined in (10), must hold because (12) implies the same optimality condition as Proposition 2.

Obtaining analytical expressions for $\mathcal{S}^{pert}(\theta)$, is quite complex in a stochastic environment because expressions for $\frac{\partial a_j(\theta^{j-1})}{\partial \mathcal{T}(y(\tilde{\theta}_t))}$ are needed. Using $S_t(\theta)$ instead provides a way to analytically

[14] Less heuristic versions of the perturbation are found in Jacquet and Lehmann (2016) and Golosov, Tsyvinski, and Werquin (2014).

[15] This effect on savings behavior also has no direct first-order effect on welfare because of the envelope theorem.

capture these effects that we also exploit in our numerical simulations.[16] For intuition and interpretation on the other hand, (11) is more suitable.

3.2 Capital Income Taxes

We now turn to the optimal capital tax. As for the marginal labor income tax formula, our optimal capital tax formula does not only hold for optimal labor income taxes but also for suboptimal labor income taxes.

Proposition 3. *The optimal linear capital tax rate τ_s satisfies:*

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

where $\zeta_{a_t,1-\tau_s}(\theta^{t-1})$ is the elasticity of savings with respect to (one minus) the tax rate on capital for individuals of history θ^{t-1}.

We now provide a brief intuitive derivation. Assume that starting from the optimal capital tax rate, the government slightly increases it. This small change will mechanically increase tax revenue in present value terms by

$$\sum_{t=2}^{T} \frac{1}{(1+r)^{t-2}} \int_{\Theta^{t-1}} a_t(\theta^{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}. \tag{13}$$

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

$$\sum_{t=2}^{T} \beta^{t-1}(1+r) \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \int_{\underline{\theta}}^{\bar{\theta}} \frac{U'(\theta^{t-1},\tilde{\theta})}{\lambda} dF_t(\tilde{\theta}|\theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1}. \tag{14}$$

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

$$\begin{aligned} &\sum_{t=2}^{T} \frac{\tau_s}{(1+r)^{t-2}} \int_{\Theta^{t-1}} \frac{\partial a_t(\theta^{t-1})}{\partial \tau_s} h(\theta^{t-1}) d\theta^{t-1} \\ &= \sum_{t=2}^{T} -\frac{1}{(1+r)^{t-2}} \frac{\tau_s}{1-\tau_s} \int_{\Theta^{t-1}} \zeta_{a_t,1-\tau_s}(\theta^{t-1}) a_t(\theta^{t-1}) h(\theta^{t-1}) d\theta^{t-1}. \end{aligned} \tag{15}$$

[16] In an earlier version of this paper, we formally show the equivalence between the two approaches for a three-period economy (Findeisen and Sachs 2014). Deriving savings responses for the three-period case is already quite involved.

Note that for $\tau_s = 0$ 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 $(13)|_{\tau=0} + (14)|_{\tau=0} \neq 0$.

For $\tau \neq 0$, however, it holds that $(15) \neq 0$. Optimality of τ_s then requires $(13)+(14)+(15) = 0$, which yields (3) (assuming $\beta(1+r) = 1$). Developing a novel dynamic tax reform approach, Golosov, Tsyvinski, and Werquin (2014) look at the welfare effects of an increase of a linear capital tax rate starting from any given tax system and obtain a formula similar to (3). In Appendix A.1 we also provide a first-order condition for the optimal linear capital tax rate in terms of the Lagrangian multiplier functions on the Euler equations, see equation (25).

Whereas the capital tax formula in Proposition (3) applies for optimal and suboptimal labor income taxes, its quantitative implications are sensitive with respect to the labor income tax schedule. On the one hand, the more progressive the labor income tax function, the less concentrated is wealth, which lowers the power of capital taxes for redistribution and insurance as captured by the numerator of (3). On the other hand, the lower the level of labor income taxation, the less insured individuals are against labor income risk and the stronger the need for self-insurance through savings. A strong need for self-insurance implies a lower responsiveness of savings with respect to capital taxes. Both effects call for higher capital taxation if labor income taxes are lower. We study this in detail in the numerical section.

Relation To Previous Public Finance Literature. In the *NDPF*-literature, savings wedges are usually interpreted to arise because of income effects on labor supply and complementarities between consumption and labor (Golosov, Troshkin, and Tsyvinski 2016). The first channel of income effects is shut down in our framework by assumption. Moreover, our analysis show that complementarities do not show up to have a direct effect on optimal linear capital tax rates when the tax instruments are history-independent and there are no income effects. Only the trade-off between redistribution of capital income and incentives for savings matter for the optimal level of the capital tax.

In Blomquist and Micheletto (2008), who consider age-dependent nonlinear taxes in a two period model with ex-ante identical 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 agents all save the same amount.[17] 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. Finally, Piketty and Saez (2013) derive a formula for the optimal linear inheritance tax in an overlapping generations framework. As

[17]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.

for the formulas presented in this paper, equity-efficiency considerations are key to understand optimal bequest taxation.

3.3 Insurance – Redistribution Decomposition

We now show how the role of income taxes in a dynamic environment can be decomposed into an insurance and a redistribution component. Specifically, we decompose the mechanical effect as defined in Proposition 2. Recall its definition for an increase of the marginal tax rate at income level $y(\theta)$:

$$\mathcal{M}_t(\theta) = \int_{\Theta^{t-1}} \int_{\theta}^{\overline{\theta}} \left(\frac{\lambda h_t(\theta^{t-1})}{(1+r)^{t-1}} - \frac{U'(\theta^{t-1}, \theta_t)\tilde{h}_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} \right) dF_t(\theta_t|\theta_{t-1}) d\theta^{t-1}. \tag{16}$$

$\mathcal{M}_1(\theta)$ measures the redistributive gain from income taxation in period one. $\mathcal{M}_t(\theta)$ with $t > 1$ captures both, welfare gains of taxation from redistribution between ex-ante heterogeneous individuals and insurance against idiosyncratic uncertainty in the t-th period. Whereas the gains from redistribution depend on the particular set of Pareto weights, gains from insurance are independent of the welfare criterion – this is formalized in the next proposition:

Proposition 4. *The mechanical effect $\mathcal{M}_t(\theta)$ can be decomposed into two parts:*

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

where

$$\mathcal{M}_t^I(\theta) = \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \left[(1 - F_t(\theta|\theta_{t-1})) - \mathcal{CU}(\theta^{t-1}; \theta) \right] h_{t-1}(\theta^{t-1}) d\theta^{t-1} \tag{17}$$

and

$$\mathcal{M}_t^R(\theta) = \int_{\Theta^{t-1}} \left(\frac{\lambda h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} - \frac{\tilde{h}_{t-1}(\theta^{t-1}) \int_{\theta_t} U'(\theta^{t-1}, \theta_t)) dF_t(\theta_t|\theta_{t-1})}{(1+r)^{t-1}} \right) \times \mathcal{CU}(\theta^{t-1}; \theta) d\theta^{t-1} \tag{18}$$

with

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

It can easily be shown that adding (17) and (18) yields (16). Whereas (18) depends on the set of Pareto weights, (17) is independent of the redistributive preferences. In the following we use an intuitive perturbation argument to show that $\mathcal{M}_t^I(\theta)$ indeed captures the mechanical insurance value of taxation and $\mathcal{M}_t^R(\theta)$ captures the mechanical impact of taxation on welfare through redistribution between ex-ante heterogeneous agents.

For this purpose we slightly 'reinterpret' the classical tax perturbation by assuming that the revenue raised is directly redistributed lump sum. Further, w.l.o.g. we normalize the tax perturbation such that $\Delta\mathcal{T}'\Delta y(\theta) = 1$. Thus, for incomes above $y(\theta)$ tax payment $\mathcal{T}$ is increased by one dollar. The overall mechanical tax revenue increase of this reform in net present value (NPV) is therefore given by: $\sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} (1 - F_t(\theta))$. Then, assume that the additional tax revenue generated by this increase is redistributed in a lump sum fashion. The uniform lump-sum transfer increase is given by

$$\Delta(\theta) = \frac{\sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} (1 - F_t(\theta))}{\sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}}}.$$

How are the considerations from the standard perturbation as discussed in Section 3.1 changed? In the absence of income effects, labor supply incentives are not altered additionally through the lump sum rebate. Incentives for savings are slightly changed, however. This has additional implications for the government budget. We address this below.

Who gains and who loses from this tax perturbation where the lump sum benefit is adjusted by $\Delta(\theta)$? Individuals with $\theta_1 < \theta$ enjoy higher period one consumption of $\Delta(\theta)$, whereas period one consumption for individuals with $\theta_1 \geq \theta$ decreases by $1 - \Delta(\theta)$. Whether period t consumption is increased or decreased depends on the realization of the shock. From a period $t-1$ perspective, the reform of the tax function in period t provides an insurance value. To obtain an expression for this insurance value, define for each type θ^{t-1} a 'constant utility term':

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

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

To measure the welfare gain through this insurance role of income taxation, we ask the following question: If the government could increase the lump-sum transfer in period t by a different amount for each θ^{t-1}-history such that expected period t utilities are unchanged for all θ^{t-1}-types (i.e. by $\mathcal{CU}(\theta^{t-1};\theta)$ respectively), how much resources could the government save due to this insurance against income risk? From each individual of type θ^{t-1}, the government obtains tax revenue of $\frac{1-F_t(\theta|\theta_{t-1})}{(1+r)^{t-1}}$ in NPV. To hold utility constant for that individual from a period $t-1$ perspective, only $\frac{\mathcal{CU}(\theta^{t-1};\theta)}{(1+r)^{t-1}}$ of resources have to be spent (in NPV). The difference $\frac{1-F_t(\theta|\theta_{t-1})}{(1+r)^{t-1}} - \frac{\mathcal{CU}(\theta^{t-1};\theta)}{(1+r)^{t-1}}$ then captures the NPV resource gain from this hypothetical reform. It is simple to show that it is always positive. Adding up and integrating over all histories yields (17). $\mathcal{M}_t^I(\theta)$ reflects the gains from insurance for individuals in period $t-1$ against their period t shocks. This insurance gain is measured in first period resources. The more pronounced labor income risk, conditional on θ^{t-1}, and the stronger risk aversion, the larger is this insurance effect.

This consideration, however, was hypothetical because the lump-sum transfer in period t is not increased by $\mathcal{CU}(\theta^{t-1};\theta)$ for type θ^{t-1} but by $\Delta(\theta)$ instead, expected period t utility therefore does not stay constant. Instead expected utility (in monetary terms) for individuals of type θ^{t-1} increases by

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

which can be either positive or negative. The larger θ_{t-1}, the larger $\mathcal{CU}(\theta^{t-1};\theta)$ and the lower $\mathcal{R}(\theta^{t-1};\theta)$. $\mathcal{R}(\theta^{t-1};\theta)$ captures the redistributive element of this tax reform for period t. To derive the welfare consequences of this implied redistribution, recall that $\mathcal{R}(\theta^{t-1};\theta)$ measures the expected utility increase of type θ^{t-1} in monetary terms. A marginal increase in consumption in period t for individuals of type θ^{t-1} (and for each realization of the shock θ_t) is valued

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

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

$$\begin{aligned}\mathcal{M}_t^{R*}(\theta) = \int_{\Theta^{t-1}} & \left(\frac{\lambda h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} - \frac{\tilde{h}_{t-1}(\theta^{t-1})\int_{\theta_t} U'(\theta^{t-1},\theta_t)dF_t(\theta_t|\theta_{t-1})}{(1+r)^{t-1}}\right) \\ & \times \left(\mathcal{CU}(\theta^{t-1};\theta) - \Delta(\theta)\right)d\theta^{t-1}, \qquad (19)\end{aligned}$$

which is not equivalent to (18) because of the '$\Delta(\theta)$-term'. To arrive at (18), we also have to take into account that the increase of the lump-sum transfer by $\Delta(\theta)$ affects all Euler equations. The implied effect on welfare is given by:

$$\begin{aligned}S^*(\theta) = & \Delta(\theta)\Bigg\{\sum_{t=1}^{T-1}\int_{\Theta^{t-1}}\int_{\theta}\mu_t(\theta^{t-1},\theta_t)U''(\theta^{t-1},\theta_t)d\theta_t d\theta^{t-1} \\ & - \sum_{t=2}^{T}(1-\tau_s)\int_{\Theta^{t-1}}\mu_{t-1}(\theta^{t-1})\int_{\Theta}U''(\theta^{t-1},\theta_t)dF_t(\theta_t|\theta_{t-1})d\theta^{t-1}\Bigg\}. \qquad (20)\end{aligned}$$

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

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

This follows from the transversality condition $\eta(\underline{\theta}) = 0$ and implies:

$$\sum_{t=1}^{T}\mathcal{M}_t^{R*}(\theta_t) + S^*(\theta) = \sum_{t=1}^{T}\mathcal{M}_t^{R}(\theta_t).$$

and therefore this alternative perturbation also yields Proposition 4.

Interpretation and implications. The $\mathcal{CU}(\theta^{t-1};\theta)$-term should be higher for high θ_{t-1}-types because they are likely to draw a better shock in Period t. For redistributive Pareto-weights (i.e. $F(\theta_1) \leq \tilde{F}(\theta_1)$ for each θ_1) the term $\mathcal{M}_t^R(\theta)$ should therefore be positive at every θ. If Pareto weights are sufficiently strong in favor of high innate types, the welfare effect from redistribution can, intuitively, become negative. We investigate and illustrate the decomposition in our numerical simulations.

Finally, note that the decomposition should be best interpreted as a lower bound on insurance, since any redistribution component for period t could also be interpreted as providing insurance from the perspective of periods $t-2$ and further away.

Relation to the literature. Boadway and Sato (2015) 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. In particular, Boadway and Sato (2015) show that the ex ante identical agents case and the no-risk case with heterogenous agents come out as special cases. Further, they do not only derive an expression for the optimal marginal income tax rate for a given ex post income but also an expression for the expected marginal tax rate faced by workers of different skills.

We also complement the analysis of Golosov, Troshkin, and Tsyvinski (2016) who decompose their optimal history-dependent labor wedge into an intra- and intertemporal component, where they interpret the former also as an insurance component. The comparison of their and our results is interesting. They show that in the unrestricted optimum the insurance component is given by the optimal static tax formula for a Utilitarian planner, conditional on the history of skill shocks. That is, the planner provides efficient insurance for each history of shocks separately. In our case with history-independent instruments this is not possible by definition; the planner only has one tax schedule. Nevertheless, we can provide an intuitive decomposition, where the insurance component is the average over the insurance component of all histories. This aggregate insurance component then captures how much resources the planner saves in net present value due to this provision of insurance. Related is the contribution by Stantcheva (2016), who derives a recursive representation of the optimal human capital wedge, where she distinguishes an insurance effect and a redistribution effect (recursive element).

Capital Income Taxation. Similar as for labor income taxes, one can also decompose the mechanical effect of the capital income tax into a redistribution and insurance element. The formulas are summarized in the following corollary.

Corollary 1. *For $t > 2$, the mechanical effect of the capital income tax reads as*

$$\frac{1}{(1+r)^{t-2}} \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \left[h_{t-1}(\theta^{t-1}) - \int_{\Theta} \frac{U'(\theta^{t-1}, \tilde{\theta}_t) dF_t(\tilde{\theta}_t|\theta_{t-1})}{\lambda} \tilde{h}_{t-1}(\theta^{t-1}) \right] d\theta^{t-1} \tag{21}$$

and can be decomposed into a redistribution part

$$\frac{1}{(1+r)^{t-2}} \int_{\Theta^{t-1}} \left(h_{t-1}(\theta^{t-1}) - \int_{\Theta} \int_{\Theta} U'(\theta^{t-1}, \theta_t, \theta_{t+1}) dF_{t+1}(\theta_{t+1}|\theta_t) dF_t(\theta_t|\theta_{t-1}) \right) CU(\theta^{t-1}) d\theta^{t-1}.$$

and an insurance part

$$\frac{1}{(1+r)^{t-2}} \int_{\Theta^{t-2}} \left[\int_{\Theta} a_t(\theta^{t-2}, \theta) dF_{t-1}(\theta_{t-1}|\theta_t) - CU(\theta^{t-2}) \right] h_{t-2}\left(\theta^{t-2}\right) d\theta^{t-2},$$

where

$$CU(\theta^{t-1}) = \frac{\int_{\Theta} a_{t+1}(\theta^{t-1}, \theta_t) \int_{\Theta} U'(\theta^{t-1}, \theta_t, \theta_{t+1}) dF_{t+1}(\theta_{t+1}|\theta_t) dF_t(\theta_t|\theta_{t-1})}{\int_{\Theta} \int_{\Theta} U'(\theta^{t-1}, \theta_t, \theta_{t+1}) dF_{t+1}(\theta_{t+1}|\theta_t) dF_t(\theta_t|\theta_{t-1})}.$$

The intuition behind the terms is a bit different than for the labor income tax because of the timing. A tax on capital income in period t provides insurance against skill risk that is realized in period $t-1$ and therefore provides insurance from a period $t-2$ perspective. Otherwise the same logic applies than in Proposition 4. In Section 4, we apply this formal decomposition to our calibrated economy.

3.4 The Role of Income Effects

As explained in Section 2, the first-order approach relies on the absence of income effects. What would be the expected changes if, hypothetically, the assumption was relaxed? For the capital income tax, the presence of income effects would matter in the following way. An increase in capital taxes would boost labor supply (as in Jacobs and Schindler (2012)) tomorrow. This would show up as it increases labor tax revenue.[18] However, it increases consumption today, which creates a counteracting negative income effect on labor supply today, resulting in a negative fiscal externality (see also Jacobs and Schindler (2012)). Whether savings taxes are, ceteris paribus, higher or lower in an economy with income effects depends on the relative strength of those effects.[19] For optimal labor income taxes, we expect that the presence of

[18] A similar logic is also underlying the inverse Euler equation result in the NDPF literature, where one intuition is that a small positive capital tax has only a second-order effect on individual utility but implies a first-order relaxation of incentive constraints (which is equivalent to boosting labor supply) in the presence of income effects: "Intuitively, the planner should be able to lower costs by offering a contract that pays the second-order cost of reducing smoothing to get the first-order benefit of improving insurance." (Kocherlakota 2010, p. 56)

[19] In Jacobs and Schindler (2012), a savings tax also provides insurance against consumption risk for the period where the savings decision is made. This is because skill risk realizes after the leisure-consumption decision. This effect does not show up in our framework because the leisure-consumption decision is made after the skill shock is observed, see also the discussion of the timing structure in Section 2.1.1.

income effects should (quantitatively) lead to higher marginal rates like in the static model by Saez (2001). In addition, the following dynamic considerations should play a role, at the very least theoretically. A labor income tax reform will change savings behavior, as also argued in Section 3.2. This changed savings behavior will trigger income effects in the presence and future. Again, these changes in labor supply will show up as fiscal externalities, weighted by the relevant marginal tax rates.

4 Simulation

4.1 Calibration

There is a large literature on the estimation of earnings dynamics over the life cycle – see Jappelli and Pistaferri (2010) and Meghir and Pistaferri (2011) for recent surveys. For the parameterization of our model, we use the recent empirical approach taken by Karahan and Ozkan (2013). In their analysis, they estimate the persistence of permanent shocks as well as the variance of permanent and transitory income shocks for US workers. Innovatively, and in contrast to most previous work in this strand of the literature, they allow these parameters to be age-dependent and to change over the life cycle. They find two structural breaks in how the key parameters change over the life cycle. This results in 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[20], we simulate a large number of labor income histories. We describe the process for earnings risk over the life cycle in detail in Appendix B. In order to study income taxation of top earners in the next section we append Pareto tails – the exact procedure is also described in Appendix B. Without high-quality administrative data on the dynamics of top incomes, how to append the tails involves some judgement calls. For this reason we present simulations for both cases: with appended tails and without.

We then partition individuals into three age-groups (corresponding to three periods in the model), namely 24-36, 37-49 and 50-62. Last, we calibrate the cross sectional income distributions for each age group and the respective transition probabilities. Figure 8 in Appendix B shows the three cross-sectional income distributions for each age group. To complete the parametrization of the model, we calibrate all conditional skill distributions from their income counterparts, as pioneered by Saez (2001).[21]

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

[21]We back out the skill from the first-order condition of individual labor supply given an approximation of the current US-tax system, a linear tax rate of 30%. Our results are robust to using parametric nonlinear tax functions fitted to the US system.

We assume that the utility function is of the form

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

For the benchmark, we set $\varepsilon = 0.33$ (Chetty 2012). Our conclusions from the simulations are not sensitive to the choice of the labor supply elasticity. For the coefficient of relative risk aversion we look at different values; in the benchmark scenario we chose $\rho = 2$, a conventional number from the literature. Finally, the model is calibrated to a yearly interest rate of 3% and β such that $\beta(1 + r) = 1$.

The simulations contain comparative statics w.r.t the taste for redistribution. As in Saez (2002), we vary the Pareto weights such that:

$$\tilde{f}_1(\theta_1) = \frac{1}{\theta_1^v} \times f_1(\theta_1).$$

This is a transparent and neat way to change the preference for distribution with one parameter. For $v = 0$, we arrive at the Utilitarian case, and as $v \to \infty$ we approach the Rawlsian criterion (Saez (2002)). Numerically, we have found that for values $v \geq 4$ the Rawlsian case is well approximated.

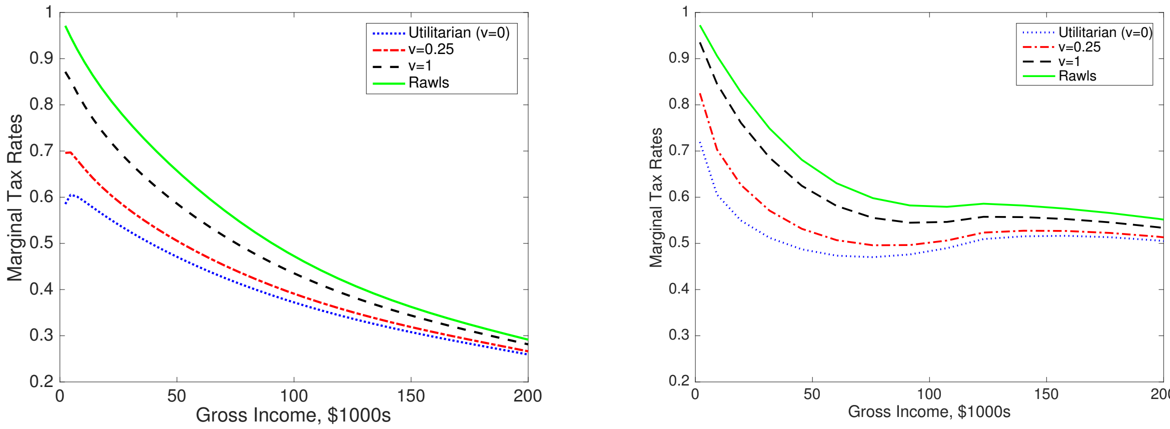
4.2 Main Results

4.2.1 Labor Taxes

Figure 1(a) illustrates optimal marginal tax rates for four welfare criteria. We let v vary such that it goes from the Utilitarian to the Rawlsian case and becomes more redistributive along the way. In Figure 1(a), marginal tax rates are decreasing in income which is similar to the pattern found in the static Mirrlees model when productivities are log-normal distributed. We can learn from the graph that the influence of the social welfare function can be quite big on tax rates at the bottom. However, this effect vanishes for higher incomes. This is intuitive since the social marginal utility of income also goes to zero for high incomes also for the Utilitarian case.

Pareto Calibration. How do these results change when we calibrate the model to include Pareto tails for top incomes? Theoretically, we show in Appendix C.1, that the optimal top tax formula in this case is an extension of the famous Saez (2001) formula with two differences: i) a savings effect shows up as for the general T' formula ii) the relevant Pareto parameter for the formula is a weighted average of each Pareto parameter of every conditional distribution, where thicker tails are overweighted and contribute more.

Numerically, as expected, the top tax rate converges. Figure 1(b) shows this. As before, the influence of redistributive preferences pushes towards higher tax rates on average, and more so for marginal tax rates on low incomes.



(a) Baseline

(b) Pareto Calibration

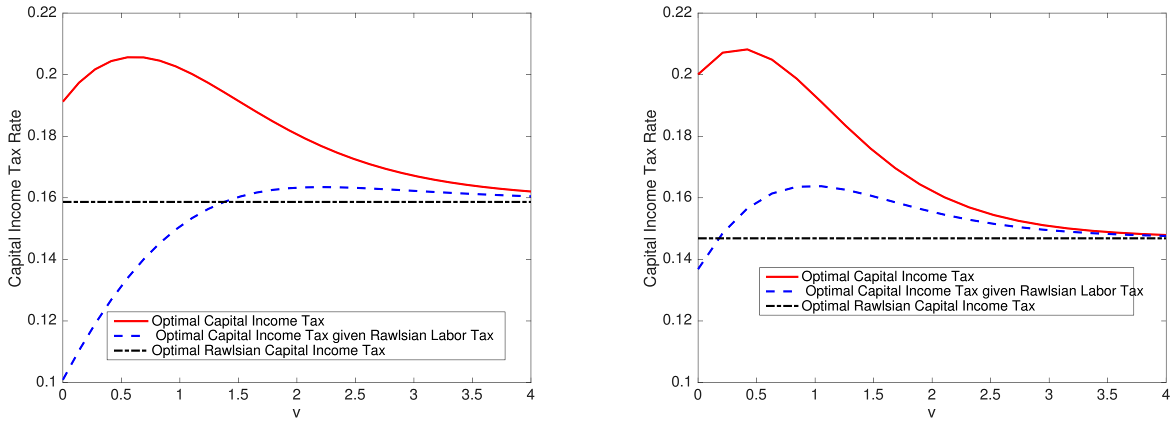
Figure 1: Optimal Labor Tax Rates

4.2.2 Capital Taxes

We recalculate the tax on the stock of capital τ_s such that it can be interpreted as an annual tax rate on capital income, following Farhi, Sleet, Werning, and Yeltekin (2012).[22] The optimal Utilitarian tax rate on capital income is 19.11%. Figure 2(a) shows the comparative statics w.r.t. to v, which increases along the x-axis. First, as expected, a higher taste for redistribution leads to higher capital taxes. As v approaches one, however, the optimal tax starts to decrease. Where does this surprising effect come from? It comes from the fact that as v increases, the level of the labor income tax rates also increases and therefore the concentration of wealth decreases. To see this effect, Figure 2(a) plots the optimal capital tax for fixed-labor taxes, set at the optimal Rawlsian level. We see that the dashed line is monotone in v up to $v = 2$. Thus, again the question arises why the level of the optimal capital income tax rate decreases in v (even if the labor income tax rate is fixed at the Rawlsian level). The answer is as follows: savings are not monotone. Thus, for very low income levels, savings are decreasing in income and therefore increasing v does not necessarily imply that the weight on those with the lowest level of wealth is increased. Finally, Figure 2(b) shows the same qualitative result holds for the Pareto calibration, and the dashed line starts to decrease already for v around a value of 1.

[22] As Farhi, Sleet, Werning, and Yeltekin (2012), we calculate the tax on annual capital income as a relative geometric average $\tau_{CI,annual}$ according to

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



(a) Baseline (b) Pareto Calibration

Figure 2: Optimal Capital Taxation For Different Redistributive Preferences

4.3 Decomposition

Figure 3 illustrates the insurance and redistribution components of the mechanical effect at the Utilitarian optimum as discussed in Section 3.3 for $\rho = 2$. We plot each component added up over all periods: $\sum_{t=1}^{3} \mathcal{M}_t(\theta)$, $\sum_{t=1}^{3} \mathcal{M}_t^I(\theta)$ and $\sum_{t=1}^{3} \mathcal{M}_t^R(\theta)$. By definition, $\mathcal{M}_1^I(\theta) = 0$, as in the first period taxes purely redistribute but do not provide any insurance value. All three functions are hump-shaped. There is no direct interpretation for the absolute value of the mechanical effect, what the graph rather illustrates is the relative importance of the redistribution and insurance components. Moreover, the insurance and redistribution effects co-move closely across the income distribution. The insurance component is slightly smaller than the redistribution component for low and intermediate income levels but this reverses for high income levels. Finally, for the optimal capital tax, our computations reveal that 29.10% of the welfare gains from redistributing capital can be attributed to the insurance motive.

4.4 High Risk Scenario

What is the effect of an increase in the amount of idiosyncratic risk on optimal policies? This is an interesting question from a policy perspective since risk seems to go up in recessions, see Storesletten, Telmer, and Yaron (2004) and Guvenen, Ozkan, and Song (2014). To study this question, we consider a scenario where we increase the standard deviation of persistent and transitory shocks (as stated in Appendix B) by the factor 2. Figure 4 shows the resulting optimal marginal tax rates. Clearly, marginal tax rates increase. This leads to more progression (in terms of average rates) to insure agents. It supports the conclusion that labor taxes should be countercyclical but more definite research on this questions seems desirable. Relatedly, for the optimal capital income tax, the rate increases from 19.10% to 22.90%.

---

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

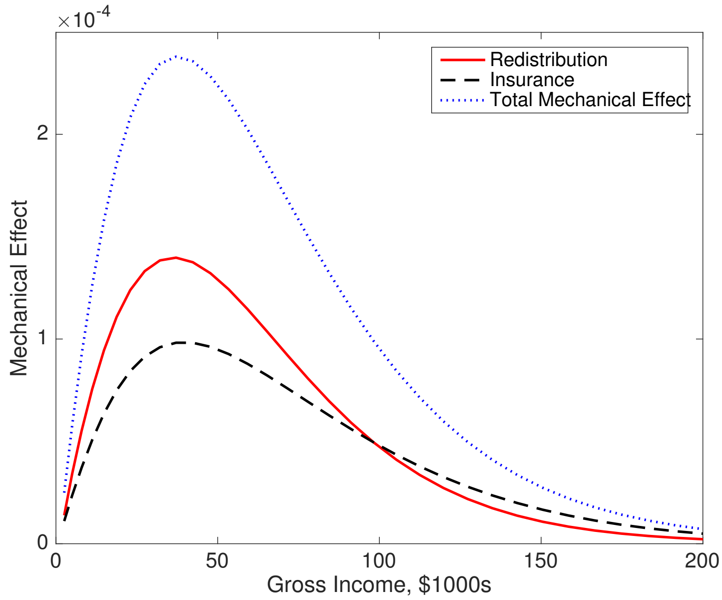


Figure 3: Decomposition of Marginal Labor Tax Rate into Redistribution and Insurance Component

4.5 A Social Insurance Lower Bound on Taxes

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. that point on the Pareto frontier with zero redistribution in a static economy. The respective Pareto weights are illustrated in Figure 5(a); richer individuals obtain a higher weight as compared to their population share.[23]

Figure 5(b) contains the results for optimal marginal tax rates in this case. Tax rates start at about 33%, then they fall, slightly increase again and finally converge to a level around 20%. Given the choice of the welfare objective, this result must be based on the insurance value of taxation. Our decomposition of the mechanical effect into an insurance and a redistribution effect that we derived in Section 3.3 provides the analytical tools to illustrate the insurance value. In Figure 6, we illustrate this decomposition. Given that the planner does not value redistribution, the redistribution effect is negative. In fact the planner dislikes the fact that positive marginal tax rates imply redistribution from initially high-skilled to initially low-skilled

[23] An alternative would have been to look at an economy without ex-ante heterogeneity. Such an economy, however, would be a very different counterfactual economy compared to the baseline case as it would ignore the important dimension of ex-ante heterogeneity. Further, it would not be clear how to set θ_1 for the ex-ante homogenous agents.

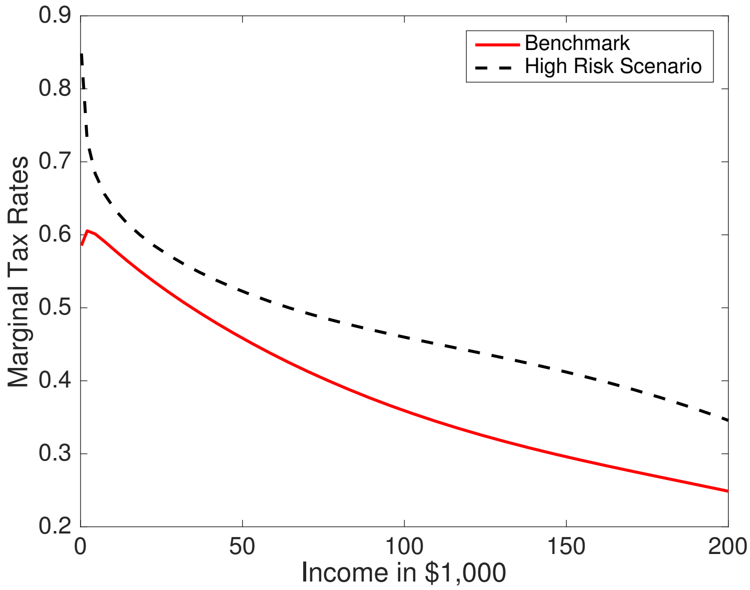


Figure 4: Comparison of Optimal Labor Income Taxes: Benchmark versus High Risk Scenario

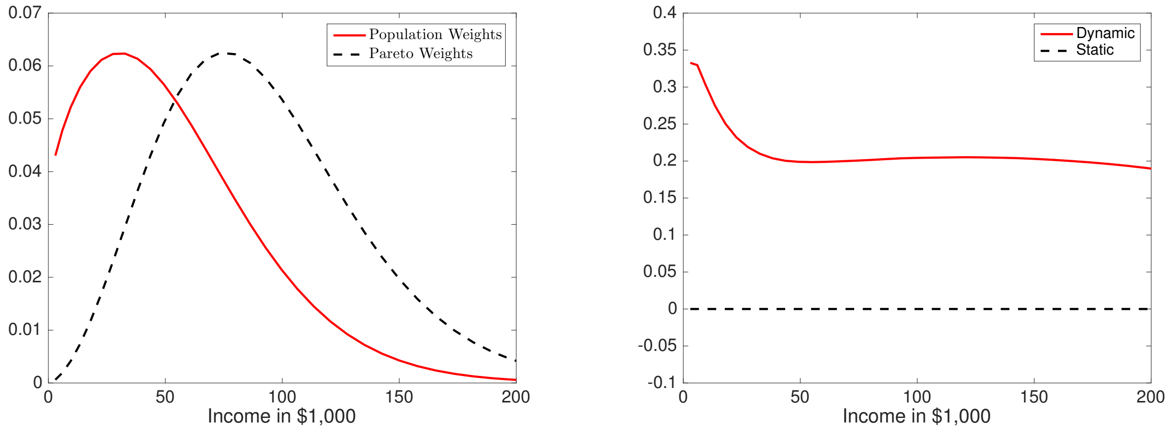
agents, which is illustrated by the red bold curve. However, the planner positively values the amount of insurance against income risk that is provided by taxation because also initially high-skilled benefit from social insurance; this is illustrated by the black dashed-dotted curve. Interestingly, the value of insurance is even higher than in the Utilitarian case. The reason is that tax rates are larger in the Utilitarian case and therefore the level of insurance provided is larger, which in turn implies that the marginal gain from additional insurance is smaller. Finally, note that the optimal capital income tax rate is -4.7% in this case, in line with the intuition that the welfare weights imply a higher weight for savers.

4.6 Interaction Between Capital and Labor Taxation

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

The optimal capital income tax formula (3) applies for optimal and suboptimal labor income taxes. We quantitatively explore savings taxes under different scenarios and consider the optimal capital tax given (i) the optimal labor income tax (as in Section 4.2), (ii) an approximation of the current US labor income tax schedule[24], (iii) zero labor income taxes and (iv) a very progressive labor income tax code which is as in (ii) but marginal tax rates are increased by 50 percentage points at each point. In terms of the amount of redistribution carried through the

[24]Here, we use the Gouveia-Strauss specification of Guner, Kaygusuz, and Ventura (2014), who provide different parametric approximations of the US tax system.



(a) Pareto Weights

(b) Marginal Labor Income Tax Rates

Figure 5: ‘Laissez-Faire’ Pareto Weights

labor tax system, scenario (iv) features the most progressive schedule, followed by the optimal schedule (i), and then the current schedule (iii) and finally the zero tax case.

Table 1 summarizes the results. There are very large differences across the four scenarios. The optimal tax rate on annual capital income varies between 5% and 84% in Panel A for the baseline calibration. In line with the argumentation from above, the less progressive the labor income tax schedule, the higher the optimal capital tax rate and also the implied welfare gains from capital taxation.[25] Wealth inequality increases if the labor tax code is less redistributive: going from the high tax scenario (iv) to the case with zero labor taxes, the savings of the 99 income percentile of the first period increases by 429% .[26] Panel B shows the results are very similar in the calibration with Pareto tails.

As argued above, a more progressive labor income tax schedule calls for lower capital income taxation not only because of lower wealth concentration but also because individuals need to provide less self-insurance through precautionary savings and should therefore be more elastic w.r.t. capital taxes in their savings decision. To disentangle these effects, we decompose (3) to see what part of the variation in optimal capital tax rates across the four scenarios is driven by the the desire to redistribute wealth, captured by the numerator, and behavioral responses, captured by the denominator. Going from the zero labor tax case to the high tax scenario, the numerator decreases by a factor around six and the denominator increases by a factor around three. Wealth inequality and the strength of behavioral responses, hence, both matter in explaining the variation in optimal capital tax rates and the decomposition results imply a larger role for the wealth inequality component.

[25] As is standard, we measure the welfare gains in consumption equivalents. Concretely, we ask by how much percent consumption in the zero-capital tax scenario would have to be increased for each individual in each period such that welfare is as high as in the optimal capital tax scenario.

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

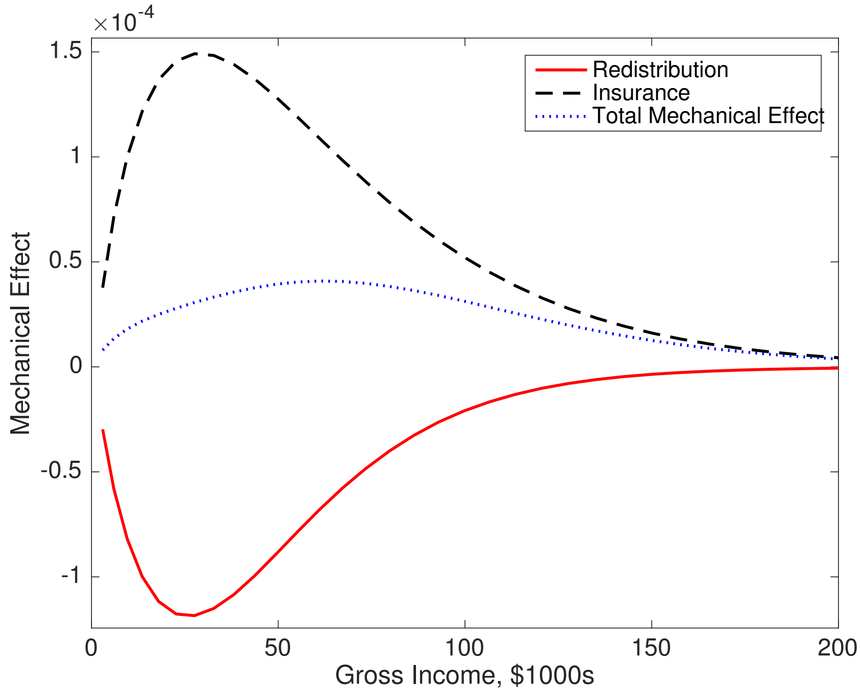


Figure 6: Decomposition of Mechanical Effect for 'Laissez-Faire' Pareto Weights

	Panel A: Baseline Calibration			
	Optimal Nonlinear	Current System	Zero Labor Taxes	High and Progressive
Optimal Tax Rate on Annual Capital Income	19.21%	47.37%	83.97%	5.25%
	Panel B: Pareto Calibration			
Optimal Tax Rate on Annual Capital Income	20.01%	45.03%	87.70%	4.64%

Table 1: Optimal Capital Tax Rates For Given Labor Tax Systems (Utilitarian Planner)

To put these numbers for the optimal capital income tax into perspective, we compare our results to two important papers on capital taxation. Conesa, Kitao, and Krueger (2009) find a value of 36%. In their paper, the main motive to tax savings is to mimic age-dependent taxes, as the relevant elasticity changes over the life cycle. In our model in contrast, the motive to tax is not coming from changing elasticities at all. In a recent contribution, Saez and Stantcheva (2016) identify optimal rates from the observed distribution of capital, which leads to optimal rates between 45% and 75% when the Pareto tail kicks in, depending on the elasticity. In our model taxes on capital are pure taxes on life cycle wealth, which should be expected to be much more equally distributed. Still, our number of 47% for the current labor income tax, interestingly, comes very close to their exercises.

4.7 Welfare Comparison To History-Dependent Taxation

How well do history-independent policies perform in relation to more complex ones? We now compare our policies to the optimal solution of the dynamic mechanism design problem which allows the planner to condition wedges on the history of shocks. The papers by Farhi and Werning (2013) and Golosov, Troshkin, and Tsyvinski (2016) treat this case. As is standard, we summarize the welfare losses from simpler instruments as follows: we ask what constant percentage increase in consumption is necessary, at all dates and histories, to achieve the same welfare as in the full optimum.

We also add age-dependent taxation as an informative intermediate case. Note that the study of optimal age-dependent labor income taxes generally involves the same problems as optimal age-independent taxation. This problem also becomes tractable through our first-order approach derived in Section 2.5.[27] For this case, we let the planner choose nonlinear age-dependent labor wedges and age-independent linear capital taxes in equivalence to the history-independent case. We chose to focus on an age-independent capital tax here because arbitrage opportunities would probably make it difficult to tax savings in an age-dependent manner.[28]

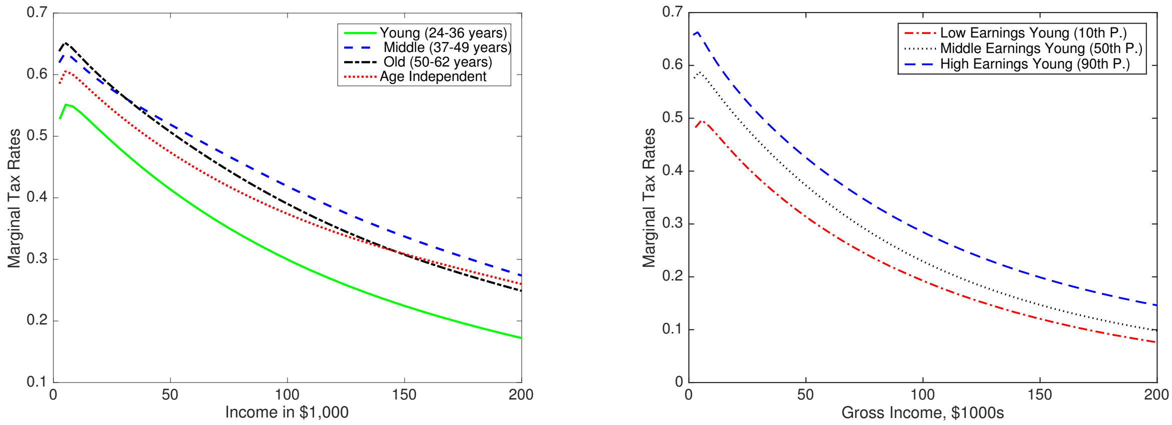
	$\rho = 1$	$\rho = 2$	$\rho = 3$
Welfare Gains from Full Second Best	0.22%	0.46%	0.71%
Welfare Gains from Age-Dependent Labor Income Taxes	0.10%	0.18%	0.26%

Table 2: Utilitarian Welfare Gains

Table 2 reports the welfare gains for three levels of risk-aversion; going from log utility ($\rho = 1$) to $\rho = 3$. As expected, the welfare gains are increasing in risk-aversion. In the first column, one can see that age-dependent taxes deliver around half of the total welfare gains from full history-dependence ($\frac{0.10}{0.22} = 0.45$). From the second column, with $\rho = 2$, we get that age-dependency reaps around 40% of the welfare gains ($\frac{0.18}{0.46} = 0.39$). Finally, for $\rho = 3$ age-dependent taxes deliver a similar value with 37% of the full the welfare gains ($\frac{0.26}{0.71} = 0.37$). Our findings complement Farhi and Werning (2013) and Weinzierl (2011). The former compare linear age-dependent and age-independent policies with the full optimum. In the log utility case, their estimates imply that 76% of the gains are captured by the age-dependent tax. The availability of nonlinear taxation seems, hence, to reduce the welfare advantage of age-dependent taxes. Weinzierl (2011) is closer to our setting as he considers nonlinear taxation. He considers a small type space (3 skill types) as he is not applying a first-order approach. Whereas the overall numbers for the welfare gains are significantly higher in his setting, the relative gains from age dependence are also around 60%.

[27] In an earlier version of this paper (Findeisen and Sachs 2014), we worked out the age-dependent case analytically in more detail.

[28] When we allow the planner to also set the capital tax rate in an age-dependent way the results are very similar.



(a) Age-(In)dependent Labor Taxes

(b) History Dependent Labor Taxes

Figure 7: Illustration of Difference in Policy Implications

To understand these welfare losses, it is of course important to understand how the policies in fact differ. In Figure 7(a) we display the optimal age-dependent labor wedges for $\rho = 2$. Taxes are highest for the middle-aged for the most part; a hazard-rate effect as an inspection of the conditional income distributions reveals, see Figure 8 in Appendix B. We also compare them with the optimal age-independent tax (red dotted line). Age-independent taxes are roughly an average of age-dependent taxes except for high incomes, where age-independent tax rates are close to the age-dependent taxes for the middle and old-age group. The tailoring of marginal tax rates to different age groups allows the planner to set significantly better incentives, which explains the welfare gains. Finally, the optimal capital income tax in the age-dependent labor tax scenario is 22.48%, which is slightly higher than in the age-independent case where it was 19.11%. The reason is that the wealth concentration is slightly higher due to less progressive taxes for the young. For example, the 99 income percentile of the first period save 130% more in the age-dependent tax scenario as compared to the age-independent tax scenario.

Finally, Figure 7(b) illustrates the labor wedges in the second period for three different income histories (10th, 50th and 90th percentile in the earnings distribution when young). In line with Golosov, Troshkin, and Tsyvinski (2013), wedges are higher for higher earnings histories. In the full optimum, marginal labor distortions are not only tailored to different age groups but to different histories of incomes. One has to be cautious when interpreting these wedges really as policy implications because the implementation would be very complicated.[29]

[29] Finally, in the full optimum, the average history-dependent capital income wedge are 17.24% and 11.06% for periods 2 and 3 respectively if weighted by the probability that a history occurs and 21.57% and 25.25% if we take an unweighted average.

5 Conclusion

We contribute to the optimal taxation literature in public economics by studying labor and capital income taxation in a life cycle model with risk. This allows to decompose optimal taxes into an insurance and redistribution component. Concerning the optimal capital income tax, simple equity-efficiency considerations are first-order. For this reason, optimal capital taxes are, in general, non-zero. In the model, savings taxes are not redundant as individuals are heterogeneous with respect to labor income and capital income over their life cycle. Finally, our contribution is also technical. We believe our approach laid out in Section 2 is also attractive for future applications where dynamics and nonlinear taxation are important.[30] Examples are models of retirement or a life cycle model with also the extensive margin of labor supply.

[30]We apply this approach in the context of college education in a model with multidimensional heterogeneity in Findeisen and Sachs (2016b).

References

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

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

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

BASTANI, S., S. BLOMQUIST, AND L. MICHELETTO (2013): "The welfare gains of age related optimal income taxation," *International Economic Review*, 54(4), 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 (2015): "Optimal Income Taxation with Uncertain Earnings: A Synthesis," *Journal of Public Economic Theory*, 17, 773–801.

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

CHETTY, R., A. GUREN, D. MANOLI, AND A. WEBER (2011): "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins," *American Economic Review Papers and Proceedings*, 101, 471–475.

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

CREMER, H., AND F. GAHVARI (1995a): "Uncertainty and optimal taxation: In defense of commodity taxes," *Journal of Public Economics*, 56(2), 291–310.

——— (1995b): "Uncertainty, optimal taxation and the direct versus indirect tax controversy," *The Economic Journal*, pp. 1165–1179.

——— (1999): "Uncertainty, commitment, and optimal taxation," *Journal of Public Economic Theory*, 1(1), 51–70.

DA COSTA, C. E., AND L. J. MAESTRI (2007): "The risk Properties of Human Capital and the Design of Government Policies," *European Economic Review*, 51(3), 695–713.

DA COSTA, C. E., AND M. R. SANTOS (2015): "Age-dependent taxes with endogenous human capital," *Mimeo, Escola de Pós-Graduação em Economia da FGV*.

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

Erosa, A., and M. Gervais (2002): "Optimal taxation in life-cycle economies," *Journal of Economic Theory*, 105(2), 338–369.

Farhi, E., C. Sleet, I. Werning, and S. Yeltekin (2012): "Non-linear Capital Taxation Without Commitment," *The Review of Economic Studies*, 79(4), 1469–1493.

Farhi, E., and I. Werning (2013): "Insurance and taxation over the life cycle," *The Review of Economic Studies*, 80(2), 596–635.

Findeisen, S., and D. Sachs (2014): "Efficient Labor and Capital Income Taxation over the Life Cycle," *Working Paper 14-17, University of Mannheim/Department of Economics Working Paper Series.*

Findeisen, S., and D. Sachs (2016a): "Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans," *Journal of Public Economics*, 138, 1–21.

——— (2016b): "Optimal Need-Based Financial Aid," *Mimeo, University of Mannheim/European University Institute.*

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 (2016): "Redistribution and Social Insurance," *American Economic Review*, 106(2), 359–86.

Golosov, M., A. Tsyvinski, and N. Werquin (2014): "A Variational Approach to the Analysis of Tax Systems," *NBER Working Paper 20780.*

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.

Guner, N., R. Kaygusuz, and G. Ventura (2014): "Income Taxation of U.S. Households: Facts and Parametric Estimates," *Review of Economic Dynamics*, 17(4), 559–581.

Guvenen, F., S. Ozkan, and J. Song (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*, 122(3), 621–660.

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

Harding, M. (2013): "Taxation of dividend, interest, and capital gain income," *OECD Publishing.*

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., and E. Lehmann (2016): "Optimal Taxation with Heterogeneous Skills and Elasticities: Structural and Sufficient Statistics Approaches," *Working Paper.*

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.

Kapicka, M., and J. Neira (2015): "Optimal taxation with risky human capital," *Mimeo, CERGE-EI/University of Exeter.*

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 (2014): "Estimating Taxable Income Responses using Danish Tax Reforms," *American Economic Journal: Economic Policy*, 6(4), 271–301.

Kocherlakota, N. (2010): *The new dynamic public finance.* Princeton Univ Press.

Ljungqvist, L., and T. J. Sargent (2012): *Recursive macroeconomic theory.* MIT press.

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.

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*, 71, 1851 – 1886.

Ramsay, C. M. (2006): "The distribution of sums of certain iid Pareto variates," *Communications in Statistics- Theory and Methods*, 35(3), 395–405.

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

Saez, E. (2002): "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses," *Quarterly Journal of Economics*, 117(3), 1039–1073.

Saez, E., and S. Stantcheva (2016): "A Simpler Theory of Optimal Capital Taxation," *NBER Working Paper 22664*.

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

Stantcheva, S. (2015): "Optimal Income, Education, and Bequest Taxes in an Intergenerational Model," *NBER Working Paper 21177*.

Stantcheva, S. (2016): "Optimal Taxation and Human Capital Policies Over the Lifecycle," *Journal of Political Economy (forthcoming)*.

Storesletten, K., C. I. Telmer, and A. Yaron (2004): "Cyclical dynamics in idiosyncratic labor market risk," *Journal of Political Economy*, 112(3), 695–717.

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

A Appendix

A.1 Lagrangian and First-Order Conditions

$$\mathcal{L} = \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^t} U\left(M(\theta_t) - a_{t+1}(\theta^t) + (1-\tau_s)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y(\theta_t)}{\theta_t}\right)\right)\tilde{h}(\theta^t)d\theta^t$$
$$+ \lambda \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\Theta} y(\theta_t) - M(\theta_t) + \tau_s(1+r)a_t(\theta^{t-1})dF(\theta_t|\theta_{t-1})h(\theta^{t-1})d\theta^{t-1}$$
$$+ \sum_{t=1}^{T-1} \int_{\Theta^t} \mu_t(\theta^t) \Bigg[U'\left(M(\theta_t) - a_{t+1}(\theta^t) + (1-\tau_s)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y(\theta_t)}{\theta_t}\right)\right)$$
$$- \beta(1+r)(1-\tau_s) \int_{\Theta} U'\Bigg(M(\theta_{t+1}) - a_{t+2}(\theta^t, \theta_{t+1})$$
$$+ (1-\tau_s)(1+r)a_{t+1}(\theta^t) - \Psi\left(\frac{y(\theta_{t+1})}{\theta_{t+1}}\right)\Bigg) dF(\theta_{t+1}|\theta_t)\Bigg] d\theta^t$$
$$+ \int_{\Theta} \eta(\theta) \frac{\partial \left(M(\theta) - \Psi\left(\frac{y(\theta)}{\theta}\right)\right)}{\partial \theta} d\theta - \int_{\Theta} \eta(\theta)\Psi'\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta^2}d\theta.$$

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

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

The derivatives with respect to the endpoint conditions yield $\eta(\overline{\theta}) = \eta(\underline{\theta}) = 0$. Substituting this into the Lagrangian, yields the following first-order conditions:

$$\frac{\partial \mathcal{L}}{\partial M(\theta)} = -\sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} f(\theta|\theta_{t-1})h(\theta^{t-1})d\theta^{t-1}$$
$$+ \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^{t-1}} U'(\theta^{t-1},\theta)f(\theta|\theta_{t-1})\tilde{h}(\theta^{t-1})d\theta^{t-1}$$
$$+ \sum_{t=1}^{T-1} \int_{\Theta^{t-1}} \mu_t(\theta^{t-1},\theta)U''(\theta^{t-1},\theta)d\theta^{t-1}$$
$$- \sum_{t=2}^{T} \beta(1+r)(1-\tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1})U''(\theta^{t-1},\theta)f(\theta|\theta_{t-1})d\theta^{t-1}$$
$$- \eta'(\theta) = 0 \tag{22}$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial y(\theta)} =& \sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta} f(\theta|\theta_{t-1}) h(\theta^{t-1}) d\theta^{t-1} \\
&- \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta} U'(\theta^{t-1},\theta) \Psi' \left(\frac{y(\theta)}{\theta} \right) \frac{1}{\theta} f(\theta|\theta_{t-1}) \tilde{h}(\theta^{t-1}) d\theta^{t-1} \\
&+ \sum_{t=1}^{T-1} \int_{\Theta^{t-1}} \mu_t(\theta^{t-1},\theta) U''(\theta^{t-1},\theta) \Psi' \left(\frac{y(\theta)}{\theta} \right) \frac{1}{\theta} d\theta^{t-1} \\
&- \sum_{t=2}^{T} \beta(1+r)(1-\tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) U''(\theta^{t-1},\theta) \Psi' \left(\frac{y(\theta)}{\theta} \right) \frac{1}{\theta} f(\theta|\theta_{t-1}) d\theta^{t-1} \\
&- \eta'(\theta) \Psi' \left(\frac{y(\theta)}{\theta} \right) \frac{1}{\theta} - \eta(\theta) \left(\Psi' \left(\frac{y(\theta)}{\theta} \right) \frac{1}{\theta^2} + \Psi'' \left(\frac{y(\theta)}{\theta} \right) \frac{y(\theta)}{\theta^2} \right) = 0
\end{aligned} \tag{23}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_{t+1}(\theta^t)} =& \frac{\lambda}{(1+r)^{t-1}} \tau_s h_t(\theta^t) - \mu_t(\theta^t) U'' \left(\theta^t\right) \\
&- (1-\tau_s)^2 \beta (1+r)^2 \mu_t(\theta^t) \int_{\Theta} U''(\theta^t,\theta_{t+1}) dF(\theta_{t+1}|\theta_t) \\
&+ (1-\tau_s)\beta(1+r)\mu_{t-1}(\theta^{t-1}) U'' \left(\theta^t\right) f(\theta_t|\theta_{t-1}) \\
&+ (1-\tau_s)(1+r) \int_{\Theta} \mu_{t+1}(\theta^t,\theta_{t+1}) U''(\theta^t,\theta_{t+1}) d\theta_{t+1} = 0
\end{aligned} \tag{24}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \tau_s} =& \sum_{t=2}^{T} \frac{\lambda}{(1+r)^{t-2}} \int_{\Theta^{t-1}} a_t(\theta^{t-1}) h(\theta^{t-1}) d\theta^{t-1} \\
&- \sum_{t=2}^{T} \beta^{t-1}(1+r) \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \int_{\Theta} U'(\theta^{t-1},\theta_t) dF(\theta_t|\theta_{t-1}) \tilde{h}(\theta^{t-1}) d\theta^{t-1} \\
&- \sum_{t=1}^{T-1} (1+r) \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \int_{\Theta} \mu_t(\theta^{t-1},\theta_t) U''(\theta^{t-1},\theta_t) d\theta_t d\theta^{t-1} \\
&+ \sum_{t=2}^{T} \beta(1+r)^2(1-\tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\Theta} U''(\theta^{t-1},\theta_t) a_t(\theta^{t-1}) dF(\theta_t|\theta_{t-1}) d\theta^{t-1} \\
&+ \sum_{t=2}^{T} \beta(1+r) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\Theta} U'(\theta^{t-1},\theta_t) dF(\theta_t|\theta_{t-1}) d\theta^{t-1} = 0.
\end{aligned} \tag{25}
$$

A.2 Multiplier Functions

First, we integrate (22) over all realizations and use the transversality conditions $\eta(\underline{\theta}) = \eta(\overline{\theta}) = 0$ to obtain:

$$\begin{aligned}
&\lambda = \left(\sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}}\right)^{-1} \times \\
&\sum_{t=1}^{T}\left\{\beta^{t-1}\int_{\Theta^{t-1}} U'(\theta^{t-1},\theta_t) f_t(\theta_t|\theta_{t-1})\tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1}\right. \\
&+ \int_{\Theta^{t-1}} \mu_t(\theta^{t-1},\theta_t) U''(\theta^{t-1},\theta_t) d\theta^{t-1} \\
&\left. - (1-\tau_s)\int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) U''(\theta^{t-1},\theta_t) f_t(\theta_t|\theta_{t-1}) d\theta^{t-1}\right\}. \qquad (26)
\end{aligned}$$

From (22) we obtain through integrating:

$$\begin{aligned}
\eta(\theta) =& \sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\theta}^{\overline{\theta}} dF(\tilde{\theta}|\theta^{t-1}) dH(\theta^{t-1}) \\
&- \sum_{t=1}^{T} \beta^{t-1} \int_{\Theta^{t-1}} \int_{\theta}^{\overline{\theta}} U'(R_t(\theta^{t-1},\theta_t)) dF(\tilde{\theta}|\theta^{t-1}) d\tilde{H}(\theta^{t-1}) \\
&- \sum_{t=1}^{T} \int_{\Theta^{t-1}} \int_{\theta}^{\overline{\theta}} \mu_t(\theta^{t-1},\tilde{\theta}) U''(R_t(\theta^{t-1},\tilde{\theta})) d\tilde{\theta} d\theta^{t-1} \\
&+ \sum_{t=1}^{T} (1-\tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta}^{\overline{\theta}} U''(R_t(\theta^{t-1},\tilde{\theta})) dF(\tilde{\theta}|\theta^{t-1}) d\theta^{t-1}. \qquad (27)
\end{aligned}$$

Next, we derive μ_t. Use (24) to obtain, with $SOC_t(\theta^t)$ being the second-order condition for savings from the individuals problem:

$$\begin{aligned}
\mu_t(\theta^t) =& \frac{\frac{\lambda}{(1+r)^{t-1}} \tau_s h_{t-1}(\theta^{t-1}) + (1-\tau_s)\beta(1+r)\mu_{t-1}(\theta^{t-1}) U''(\theta^t) f_t(\theta_t|\theta_{t-1})}{SOC_t(\theta^t)} \\
&+ \frac{(1-\tau_s)(1+r)\int_{\Theta} \mu_{t+1}(\theta^t,\theta_{t+1}) U''(\theta^t,\theta_{t+1}) d\theta_{t+1}}{SOC_t(\theta^t)}. \qquad (28)
\end{aligned}$$

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

$$A_t(\theta^t) = \frac{\frac{\lambda}{(1+r)^{t-1}} \tau_s h_{t-1}(\theta^{t-1})}{SOC_t}$$

$$B_t(\theta^t) = \frac{(1-\tau_s)\beta(1+r) U''(\theta^t) f_t(\theta_t|\theta_{t-1})}{SOC_t}$$

$$C_t(\theta^t, \theta_{t+1}) = \frac{(1-\tau_s)(1+r)U''(\theta^t, \theta_{t+1})}{SOC_t},$$

then, we can rewrite (28) as

$$\mu_t(\theta^t) = A_t(\theta^t) + B_t(\theta^t)\mu_{t-1}(\theta^{t-1}) + \int_\Theta C_t(\theta^t, \theta_{t+1})\mu_{t+1}(\theta^t, \theta_{t+1})d\theta_{t+1}.$$

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

$$\begin{aligned}\mu_{T-2}(\theta^{T-2}) =& A_{T-2}(\theta^{T-2}) + B_{T-2}(\theta^{T-2})\mu_{T-3}(\theta^{T-3}) \\ &+ \int_\Theta C_{T-2}(\theta^{T-2}, \theta_{T-1})\mu_{T-1}(\theta^{T-2}, \theta_{T-1})d\theta_{T-1}.\end{aligned} \tag{29}$$

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

$$\mu_{T-1}(\theta^{T-1}) = A_{T-1}(\theta^{T-1}) + B_{T-1}(\theta^{T-1})\mu_{T-2}(\theta^{T-2}). \tag{30}$$

Now insert (30) into (29). 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}

$$\mu_{T-3} = A_{T-3} + B_{T-3}\mu_{T-4} + \int_\Theta C_{T-3}\frac{A_{T-2} + B_{T-2}\mu_{T-3} + \int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1 - \int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}, \tag{31}$$

yielding

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

Now insert this into μ_{T-4}

$$\begin{aligned}\mu_{T-4} =& A_{T-4} + B_{T-4}\mu_{T-5} \\ &+ \int_\Theta C_{T-4}\frac{A_{T-3} + B_{T-3}\mu_{T-4} + \int_\Theta C_{T-3}\frac{A_{T-2} + \int_\Theta C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}}{1 - \int_\Theta \frac{C_{T-3}B_{T-2}}{1-\int_\Theta C_{T-2}B_{T-1}d\theta_{T-1}}}d\theta_{T-3}.\end{aligned} \tag{33}$$

Rewrite to obtain

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

$$\Bigg(A_{T-4} + B_{T-4}\mu_{T-5}$$

$$+ \int_{\Theta} C_{T-4} \frac{A_{T-3} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + \int_{\Theta} C_{T-2} A_{T-1} d\theta_{T-1}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}} d\theta_{T-2}}{1 - \int_{\Theta} \frac{C_{T-3} B_{T-2}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}}} d\theta_{T-3} \Bigg). \tag{34}$$

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

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

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

$$+ \int_{\Theta} C_{T-5} \frac{A_{T-4} + \int_{\Theta} C_{T-4} \frac{A_{T-3} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + \int_{\Theta} C_{T-2} A_{T-1} d\theta_{T-1}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}} d\theta_{T-2}}{1 - \int_{\Theta} \frac{C_{T-3} B_{T-2}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}}}}{1 - \int_{\Theta} \frac{C_{T-4} B_{T-3}}{1 - \int_{\Theta} \frac{C_{T-3} B_{T-2}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}}}} \Bigg). \tag{35}$$

Now define

$$D_t = \left[1 - \int_{\Theta} C_t B_{t+1} \left[1 - \int_{\Theta} C_{t+1} B_{t+2} \left[... \left[1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}\right]^{-1} .... \right]^{-1} d\theta_{t+2}\right]^{-1} d\theta_{t+1}\right]^{-1}.$$

Using this definition, we can write μ_{T-5} as

$$\mu_{T-5} = \frac{A_{T-5} + B_{T-5}\mu_{T-6} + \int_{\Theta} C_{T-5} \frac{A_{T-4} + \int_{\theta_{T-3}} C_{T-4} \frac{A_{T-3} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + \int_{\Theta} C_{T-2} A_{T-1}}{D_{T-2}}}{D_{T-3}}}{D_{T-4}}}{D_{T-5}}. \tag{36}$$

It is helpful to make another definition:

$$E_t = \int_{\Theta} C_t \frac{A_{t+1} \int_{\Theta} C_{t+1} \frac{A_{t+2} + \int_{\Theta} C_{t+2} \frac{A_{t+3} + \int_{\Theta} C_{t+3} \frac{A_{t+4} \cdots\cdots}{D_{t+4}}}{D_{t+3}}}{D_{t+3}}}{D_{t+1}}.$$

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 general, we thus obtain:

$$\mu_t = \frac{A_t + B_t\mu_{s-1} + E_t}{D_t}.$$

For the second period, we obtain

$$\mu_2 = \frac{A_2 + B_2\mu_1 + E_2}{D_2}. \tag{37}$$

and get

$$\mu_1 = \frac{A_1 + E_1}{D_1}. \tag{38}$$

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

In equation (38) 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.3 Labor Income Taxes

Dividing (23) by $\Psi' \frac{1}{\theta}$ and adding (22) yields

$$\frac{\mathcal{T}'(y(\theta))}{1 - \mathcal{T}'(y(\theta))} = \left(1 + \frac{1}{\varepsilon_{y_t, 1-\mathcal{T}'}(\theta)}\right) \frac{\eta(\theta)}{\lambda\theta \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\theta_{t-1}} \int_{\theta_t} f(\theta_t|\theta_{t-1}) h(\theta^{t-1}) d\theta^{t-1}}. \tag{39}$$

Inserting (27) into (39) yields the formula for optimal labor tax rates.

B Details on Calibration

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

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

and where the error term η_h^i 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 $f(X_a^i; \theta_t)$:

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

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

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

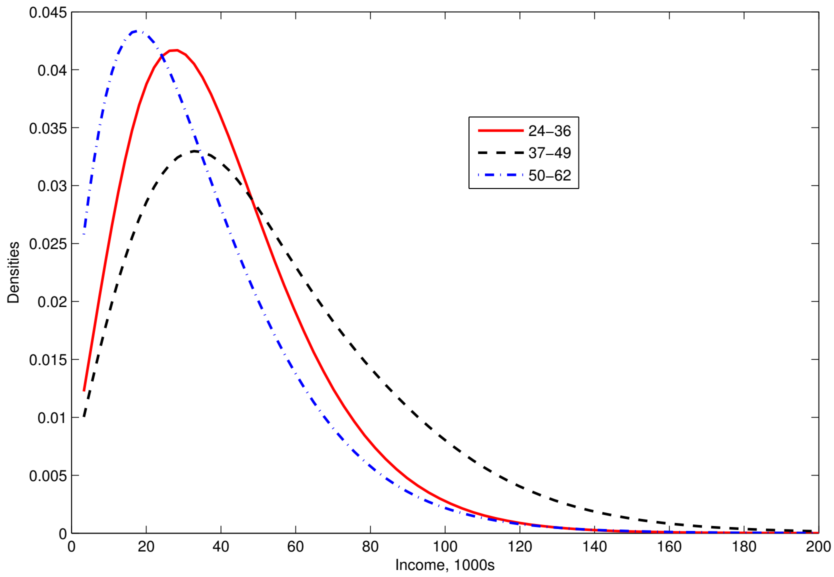


Figure 8: Income Distribution for Each Age Group

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

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

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

In the calibration with Pareto tails, we append tails at income of >\$140,000 for each conditional distribution. These tails are age-dependent and informed by empirical estimates (Badel and Hugget 2014). For the young the parameter is 3.5. For the middle 2 and the old 1.75.

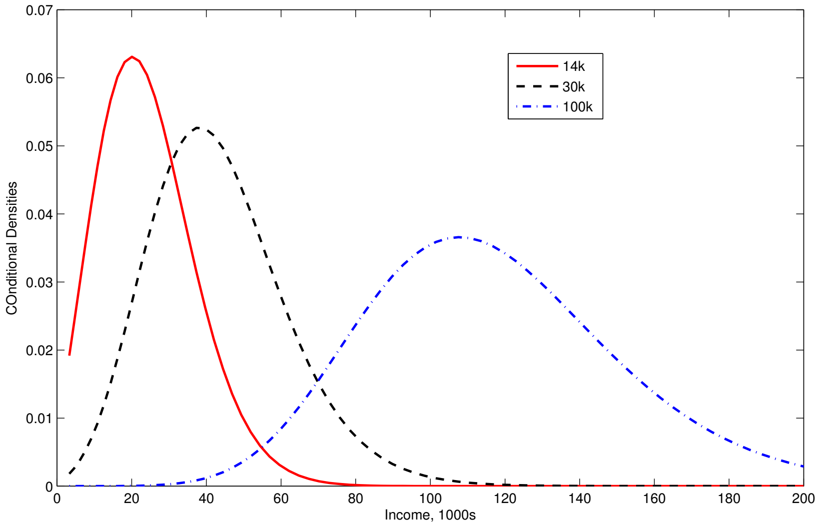


Figure 9: Conditional Income Distributions, Middle Aged Workers

C Robustness Pareto Tails

C.1 Top Tax Rate Result – Pareto Distribution

Saez (2001) documents the importance of modeling top incomes for optimal taxation in an empirical realistic way. We now explore the implications of Pareto tails in the dynamic life cycle framework. We first provide a limit result for the optimal labor tax rate at the top of the income distribution. In our stochastic multi-period environment, this forms the counterpart to the famous Saez (2001) formula for the static model. We follow Saez and express the optimal tax formula in terms of income instead of skill distributions. Our formula generalizes and nests that of Saez (2001)

Formally, let $e_{t|t-1}(y_t|y_{t-1})$ denote the conditional distribution in period t. The standard formula assumes, in line with the data, that the right tail of the cross-sectional income distribution is Pareto distributed. In the dynamic model more assumptions are needed. One needs an assumption for the right tail for every conditional distribution $e_{t|t-1}(y_t|y_{t-1})$ in every time period t and for any previous income level y_{t-1}. Given assumptions on the value of the Pareto parameter for each conditional distribution, it would then be possible to derive a top tax rate formula, depending on all Pareto parameters across the conditional distributions.

In what follows, instead of placing assumptions on every single conditional distribution $e_{t|t-1}(y_t|y_{t-1})$, we consider a scenario where Pareto parameters differ by age groups only. We do this for two reasons. First, there is no empirical guidance on how Pareto parameters differ across different conditional distributions, however, some recent papers have shown how these parameters differ by age groups, see, e.g., Badel and Huggett (2014). Second, it simplifies the exposition while delivering the same intuition. Formally, we assume that at every age above a high-income threshold y^H, earnings are Pareto distributed with an age-dependent thickness parameter a_t. We obtain the following top tax result for the dynamic economy:

Proposition 5. *Assume that welfare weights for top incomes converge to a constant $\overline{\mathcal{W}}$ across age groups and elasticities at the top converge to a constant ϵ. Then the optimal top marginal tax rate above income level y^H satisfies:*

$$\frac{\overline{\mathcal{T}'}}{1-\overline{\mathcal{T}'}} = \frac{\left(1-\overline{\mathcal{W}}\right)\sum_{t=1}^{T}\frac{1}{(1+r)^{t-1}}\frac{1}{a_t-1}+\overline{S}}{\epsilon\cdot\sum_{t=1}^{T}\frac{1}{(1+r)^{t-1}}\frac{a_t}{a_t-1}}, \tag{40}$$

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

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

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

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

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

Moreover, simple numerical exercises on formula (40) reveal that with heterogeneous Pareto tails across different distributions, the relatively thicker tails dominate optimal tax rates. Suppose, for example, we take two age groups and that the Pareto parameter of the young a_1 is around 3.5 and that of the old a_2 is 1.75 – this corresponds to the numbers found for the US for the age groups around 25 years (the young) and 50 years (the old) based on administrative data (Badel and Huggett 2014). Using formula (40), setting $r = \overline{S} = \overline{\mathcal{W}} = 0$ for simplicity and an elasticity parameter of 0.3, would yield an optimal top tax around 61%. In comparison, the optimal tax rate would be around 65.5% if both tails were thick ($a_1 = a_2 = 1.75$) but only 48.8% if both tail were thin ($a_1 = a_2 = 3.5$). In this sense, the optimal tax formula allowing for heterogeneous Pareto coefficients, strongly overweighs those distributions with a thicker tail.[31] This result is robust when we consider more than just two age groups.

[31] This is in line with Ramsay (2006) who shows that sum of two Pareto distributions tends to behave like a Pareto distribution, where the heavier tail distribution seems to dominate.