



Working Papers

www.cesifo.org/wp

Redistribution and Insurance with Simple Tax Instruments

Sebastian Findeisen
Dominik Sachs

CESIFO WORKING PAPER NO. 5400
CATEGORY 1: PUBLIC FINANCE
JUNE 2015

Presented at CESifo Area Conference on Public Sector Economics, April 2015

An electronic version of the paper may be downloaded

- *from the SSRN website:* www.SSRN.com
- *from the RePEc website:* www.RePEc.org
- *from the CESifo website:* www.CESifo-group.org/wp

ISSN 2364-1428

Redistribution and Insurance with Simple Tax Instruments

Abstract

We analyze optimal taxation of labor and capital income in a life-cycle framework with idiosyncratic income risk. We provide a novel 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 tax rates and provides an informative lower bound on optimal taxes: even for welfare functions that do not value redistribution, marginal tax rates are positive for all income levels. Optimal capital taxes are significant and yield sizable welfare gains; in particular if labor income taxes are suboptimal.

JEL-Code: H210, H230.

Keywords: optimal taxation, capital taxation, redistribution, insurance.

Sebastian Findeisen
University of Mannheim
Mannheim / Germany
findeisen@uni-mannheim.de

Dominik Sachs
University of Cologne
Cologne / Germany
sachs@wiso.uni-koeln.de

June 8, 2015

An older version of the paper was previously circulated under 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 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

A comprehensive literature in public economics has characterized optimal labor income taxation. This body of work has shown that labor supply elasticities and the distribution of abilities are the key forces determining optimal nonlinear income tax schedules.¹ A relatively recent literature, often called the *New Dynamic Public Finance (NDPF)*, has expanded the Mirrlees model to take into account dynamics and idiosyncratic uncertainty.² The *NDPF* studies properties of second-best efficient allocations that are only constrained by informational asymmetries. The implementation of such allocations therefore requires complex history-dependent tax functions.³

In this paper, we bring together elements from the public economics and the *NDPF*-literature. In the tradition of the public economics literature, we study the optimal design of relatively simple tax instruments that are within the realm of current tax practices:⁴ nonlinear labor income taxes of current income – i.e. history-independent taxes – and linear capital income taxes. In line with the *NDPF*, we study a life-cycle framework with idiosyncratic uncertainty. These life-cycle considerations allow us to complement the public economics literature in the following two ways: First, it allows us to study the effect of income risk on labor income taxation and distinguish the role of taxes for redistribution and insurance against idiosyncratic wage risk. Second, the dynamics allow us to study the question of optimal capital taxation. Importantly, our approach is flexible enough to also study the optimal design of capital taxes for some given suboptimal labor income tax code and vice versa.

Main Results and Contribution. First, we derive a novel formula for optimal marginal labor income tax rates in a dynamic setting. Additionally, we show how the mechanical effect (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 difference in the welfare weights on different income levels at a 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). Interestingly we find numerically that for our benchmark calibration the redistribution and insurance component

¹See, e.g., Piketty (1997), Diamond (1998) and Saez (2001) or Piketty and Saez (2013a) for a recent survey article.

²Golosov, Kocherlakota, and Tsyvinski (2003) is often seen as the starting point of this literature. More recently Farhi and Werning (2013) and Golosov, Troshkin, and Tsyvinski (2015) study the properties of dynamic optimal allocations and especially labor wedges in more detail.

³ See Kocherlakota (2004, 2005) and Werning (2011).

⁴Social insurance and tax codes in most countries typically also contain some *history-dependent* elements, in particular retirement and unemployment benefits. Personal income taxes – although they contain some form of history dependence in the form of deductions – are arguably better described as being history-independent.

are roughly equally important for the optimal Utilitarian tax rates. Based on the above considerations, we also provide a social insurance lower bound on labor 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. We find that tax rates are strictly positive, starting at about 25%, then fall before they slightly increase again and converge to a level of around 10%.

Next, we derive a novel formula for the optimal linear capital tax rate⁵ that 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 model with risk optimal capital taxes are, in general, non-zero. In our model savings taxes are not redundant as individuals are heterogenous with respect to labor income and capital income over their life cycle. It is, hence, beneficial for the government to employ multiple instruments. This logic is related to the inheritance tax model by Piketty and Saez (2013b). In our simulations, the government strongly relies on capital income taxation and the optimal tax rate is around 15%.

Our framework also allows to investigate the interaction between labor and capital taxation, which is not possible in the static Mirrlees model. This interaction is, in general, also not studied in the *NDPF*-approach, since both labor and capital wedges are part of the dynamic mechanism design solution. We examine optimal labor income taxes for a given capital tax and vice versa. Strikingly, we find that for given labor income taxes, optimal capital tax rates differ substantially, depending on how labor income taxes are set. This depends on two mechanisms. First, lower labor income taxes lead to a more concentrated distribution of wealth 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. Conducting the opposing exercise, the simulation reveals that optimal marginal labor income tax rates are decreasing in the level of capital taxation, the quantitative effect is very small, however.

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 tractable. If labor income taxes are only a function of current income y_t , the income that individuals optimally choose in a decentralized economy only depends on their current productivity θ_t and not on accumulated wealth. For the allocation, this implies that income is solely a function of θ_t and not of the history of shocks $\theta^t = (\theta_1, \theta_2, \dots, \theta_t)$. This guarantees that the individuals can easily be ordered among the y_t dimension. A second advantage of this specification is that the

⁵The practice of taxing capital linearly in most countries, is often linked to arbitrage opportunities in the case of capital. If individual A face a higher marginal tax rate on savings than individual B, there would be a deal where both could be better off by making individual B save individual A's money. The assumption of linear capital taxes can therefore be grounded on the idea that the government cannot observe consumption on the individual level and is often made in the public finance literature. See Hammond (1987) for a more general theoretical discussion of that issue.

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 researchers who would like to study dynamic life-cycle problems like, for example, models with retirement, and would like to restrict labor taxes to be history-independent but fully nonlinear.

Closely Related Literature. As already discussed in the first paragraph, our paper is closely related to the public economics taxation literature and work in the *NDPF* (Piketty 1997, Diamond 1998, Piketty and Saez 2013a, Golosov, Troshkin, and Tsyvinski 2015, Farhi and Werning 2013).

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

Stantcheva (2015a) and Piketty and Saez (2013b) 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. There is a small but growing literature on age-dependent income taxation. Most recent and related to our work, Weinzierl (2011) and Bastani, Blomquist, and Micheletto (2011) study the welfare gains of age-dependent labor income taxation. Blomquist and Micheletto (2008) is an important earlier paper in this literature.

Also studying age dependency as well as standard income taxes, Best and Kleven (2013) augment the canonical optimal tax framework by incorporating career effects into a 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. Related dynamic tax models with human capital are, for example, developed in Da Costa and Maestri (2007), Kapicka and Neira (2013), Findeisen and Sachs (2015b) and Stantcheva (2015b). 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. Conesa, Kitao, and Krueger (2009), in tradition with the Ramsey approach to optimal taxation, study optimal labor and capital income taxes in a computational life-cycle framework. While our approach shares some features from a Ramsey type of exercise, we allow labor income taxes to be an arbitrarily nonlinear function in the Mirrlees tradition and theoretically highlight the forces driving labor and capital taxation.

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. Section 3.1 and 3.3 derive the new formulas, and in Section 3.2 we present the insurance-redistribution decomposition. Optimal policy simulations are presented in Section 4. Section 4.2 contains the baseline simulation. In Section 4.3 we study the interaction of labor and capital taxes and Section 4.4 presents the social insurance lower bound on optimal taxes. Finally, in Section 5 we conclude.

2 The Model

2.1 Environment

We consider a life-cycle framework with T periods where individuals at any point in time t are characterized by their productivity $\theta_t \in \Theta = [\underline{\theta}, \bar{\theta}]$.⁶ We denote the history of shocks by $\theta^t = (\theta_1, \theta_2, \dots, \theta_t)$. Flow utility is given by

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

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.

Abusing notation, we sometimes write the utility function or its derivatives as a function of the history of shocks only, i.e. $U(\theta^t)$, $U'(\theta^t)$ and $U''(\theta^t)$. When we present optimal tax formulas, we denote by ε the labor supply elasticity and sometimes suppress the dependence of ε on the current skill type, consistent with the utility functions we use for the simulations, where the elasticity is constant. As our appendix reveals, the theoretical results are more general and apply also to utility functions where the elasticity changes with the skill type and taxes. Importantly, the functional form of U eliminates income effects on labor supply, while allowing for risk aversion.⁷ 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 assume that agents already differ in the first period. The conditional density function (*cdf*) of the initial distribution of productivities is denoted by $F_1(\theta_1)$ and captures the ex-ante heterogeneity of agents. The reader should think about this heterogeneity as the level of heterogeneity of individuals at age of roughly 25. In the following periods, productivities evolve stochastically over time according to a Markov process. The respective *cdf* is $F_t(\theta_t|\theta_{t-1})$. Denote by $h_t(\theta^t)$ the probability of history θ^t , i.e. $h_t(\theta^t) = f_t(\theta_t|\theta_{t-1})f_{t-1}(\theta_{t-1}|\theta_{t-2})\dots f_1(\theta_1)$. We assume that all conditional density functions are continuously differentiable. Denote by Θ^t

⁶For notational convenience, we assume that the support of θ_t does not change over time. Allowing for a moving support would be a straightforward extension.

⁷The 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 (2012) using the universe of Danish tax records).

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

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. Further, we assume incomplete markets in a sense that individuals only have access to risk-free one period bonds.⁸ Finally, we assume $\beta(1+r) = 1$, which does not come at the expense of generality but simplifies the exposition.

2.2 Policy Instruments and Planner's Objective

We are interested in the Pareto efficient set of nonlinear labor income tax schedule and linear capital income tax rates that only condition on current income. Thus, we are not solving for a second-best Pareto problem, where the government could condition policy instruments on all public information (typically the history of income and savings), but rather restrict the set of policy instruments in a Ramsey manner. However, our approach shares the feature of the Mirrlees approach that labor income taxes can be an arbitrarily nonlinear function of current income, see also the discussion at the end of this section. Finally, note that our approach also allows to study age-dependent taxation, which we explored in an earlier version of this paper, see Findeisen and Sachs (2014).

The government chooses a labor income tax schedules $\mathcal{T}(\cdot)$ and a linear capital tax rate τ_s . The preferences of the social planner are described by the set of Pareto weights $\{\tilde{f}_1(\theta_1)\}_{\theta_1 \in [\underline{\theta}, \bar{\theta}]}$. 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 $\tilde{F}_1(\bar{\theta}) = 1$. Different sets of Pareto weights refer to different points on the Pareto frontier. The set of weights where $\tilde{f}_1(\theta_1) = f_1(\theta_1) \forall \theta_1$, e.g., refers to the Utilitarian planner. Similar as $h_t(\theta^t)$, define $\tilde{h}_t(\theta^t) = f_t(\theta_t|\theta_{t-1})f_{t-1}(\theta_{t-1}|\theta_{t-2})\dots\tilde{f}_1(\theta_1)$ to express the Pareto weights for individuals with certain histories.

In the remainder of this paper, we use the notions wealth, savings or capital for a_t interchangeably. Also note that the way we define τ_s , it is a stock tax 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 defined τ_s . In the following, we use the notions capital taxes, wealth taxes and capital income taxes interchangeably. Further note that the way we define capital taxes implies that borrowing is subsidized at the same rate as saving is taxed.⁹

Relation to Ramsey and Mirrlees Approaches. The tax problem we look at lies at the intersection of previously considered instruments. As in the public economics taxation

⁸We 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 $\sum_{t=s}^T \left(y_t(\underline{\theta}) - \Psi\left(\frac{y_t(\underline{\theta})}{\underline{\theta}}\right) \right)$.

⁹One 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.

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 the set of allocations that we can attain with our policy instruments 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:

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

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

Lemma 1. *The optimal gross income y_t that solves (1) is independent of assets, the savings tax rate τ_s and the labor income tax schedule \mathcal{T} . 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 tax problem of the social planner is then:

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

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:¹⁰

¹⁰If we integrate over different histories θ^t , then within the integral we use θ_t to denote the last element of θ^t .

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

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.

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

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

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

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

These conditions are only necessary and not sufficient for the agents' choices to be optimal. Due to the assumption about preferences, however, the second-order conditions are of particularly simple form. The derivative of the first-order condition of labor supply with respect to consumption, i.e. the cross derivative of the value function, is zero. By symmetry of the Hessian, the same holds for the derivative of the Euler equation with respect to labor supply. Thus, the minor diagonal of the Hessian matrix contains only zeros. For (4) and (5) to represent a maximum, only the second derivatives of the value function with respect to labor supply and

consumption have to be ≤ 0 . For labor supply, a familiar argument from the standard Mirrlees model implies that this holds if and only if¹¹

$$y'(\theta) \geq 0 \quad \forall \theta \in \Theta. \quad (6)$$

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

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

Inserting (4) into this derivative yields:

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

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 $\{M(\theta), y(\theta)\}_{\theta \in \Theta}$, $\{a_t(\theta^{t-1})\}_{\theta^{t-1} \in \Theta^{t-1}}_{t=1, \dots, T}$ 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. 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 ex-post check whether the monotonicity condition is fulfilled or not.

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.

¹¹See, e.g., Salanié (2003, p.87 ff).

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:*

$$\frac{\mathcal{T}'(y(\theta))}{1 - \mathcal{T}'(y(\theta))} = \left(1 + \frac{1}{\varepsilon}\right) \frac{1}{\theta \sum_{t=1}^T \frac{1}{(1+r)^{t-1}} f_t(\theta)} \times \sum_{t=1}^T [\mathcal{M}_t(\theta) + \mathcal{S}_t(\theta)], \quad (8)$$

where

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

and

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

where $\mu_t(\theta^t)$ is the Lagrangian multiplier on the Euler equation of individuals with history θ^t and λ is the Lagrangian multiplier on the government budget constraint, i.e. the marginal value of public funds. In Appendix A.2, we derive expressions for these multipliers. Further, we have $\mathcal{T}'(y(\underline{\theta})) = 0$, and $\mathcal{T}'(y(\bar{\theta})) = 0$ – i.e. no distortion at the bottom and top – if $\bar{\theta} < \infty$.

The mechanical effect $\mathcal{M}_t(\theta)$ is similar as in the static model. In the life-cycle model, it is defined for every period t . An important difference is, however, that it now also captures the insurance motive of taxation on top of redistributive concerns. We disentangle the welfare gains from redistribution and insurance in Section 3.2.

$\mathcal{S}_t(\theta)$ 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 (Piketty (1997), Golosov, Tsyvinski, and Werquin (2014)). 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 $\Delta\mathcal{T}'$ at an infinitesimal income interval with length $\Delta y(\theta)$ 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 $\mathcal{M}_t(\theta) \times \Delta \mathcal{T}' \Delta y(\theta)$ 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 $> y(\theta)$ 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 . It is well known from the tax perturbation literature that this can be captured by $\mathcal{L}\mathcal{S}_t(\theta) \times \Delta \mathcal{T}' \Delta y(\theta)$, where $\mathcal{L}\mathcal{S}_t(\theta) \equiv -\lambda \frac{\mathcal{T}'(y(\theta))}{1-\mathcal{T}'(y(\theta))} \theta^{-\varepsilon} \frac{1}{(1+r)^{t-1}} f_t(\theta)$ for each period t , see e.g. Piketty (1997). The change in labor supply has no direct impact on welfare by the envelope condition.

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 $\Delta \mathcal{T}' \Delta y(\theta) \times \mathcal{S}^{pert}(\theta)$, where:¹²

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

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 reduced the demand for savings as a self-insurance device. In our simulations we come back to the issue and check how much $\mathcal{S}^{pert}(\theta)$ contributes quantitatively to optimal labor taxes.

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

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

Solving (12) for $\frac{\mathcal{T}'(y(\theta))}{1-\mathcal{T}'(y(\theta))}$ reveals that we have to have $\mathcal{S}^{pert}(\theta) = \sum_{t=1}^T \mathcal{S}_t(\theta)$ given that (12) has to imply 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 $\mathcal{S}_t(\theta)$ instead provides a way to analytically capture these effects that we also exploit in our numerical simulations.¹³ For intuition and interpretation on the other hand, (11) is more suitable.

¹²This effect on savings behavior also has no direct first-order effect on welfare because of the envelope theorem.

¹³In 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.

3.2 Insurance – Redistribution Decomposition

We now show how the role of income taxes in a dynamic environment can be cleanly decomposed into an insurance and a redistribution component.¹⁴ Specifically, we decompose the mechanical effect from the last section. Recall its definition for an increase of the marginal tax rate at income level $y(\theta)$:

$$\mathcal{M}_t(\theta) = \int_{\Theta^{t-1}} \int_{\theta}^{\bar{\theta}} \left(\frac{\lambda h_t(\theta^{t-1})}{(1+r)^{t-1}} - \frac{U'(\theta^{t-1}, \theta_t) \tilde{h}_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} \right) dF_t(\theta_t | \theta_{t-1}) d\theta^{t-1}. \quad (13)$$

$\mathcal{M}_1(\theta)$ 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 3. *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

$$\mathcal{M}_t^I(\theta) = \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} [(1 - F_t(\theta | \theta_{t-1})) - \mathcal{CU}(\theta^{t-1}; \theta)] h_{t-1}(\theta^{t-1}) d\theta^{t-1} \quad (14)$$

and

$$\mathcal{M}_t^R(\theta) = \int_{\Theta^{t-1}} \left(\frac{\lambda h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} - \frac{\tilde{h}_{t-1}(\theta^{t-1}) \int_{\theta_t} U'(\theta^{t-1}, \theta_t) dF_t(\theta_t | \theta_{t-1})}{(1+r)^{t-1}} \right) \times \mathcal{CU}(\theta^{t-1}; \theta) d\theta^{t-1} \quad (15)$$

with

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

It can easily be shown that adding (14) and (15) yields (13). Whereas (15) depends on the set of Pareto weights, (14) 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.

¹⁴Relatedly, Boadway and Sato (2012) derive a formula for the optimal marginal tax rate in a static setting with heterogeneity and uncertainty. They also show how their formula addresses the desire to redistribute *and* to provide insurance. Their timing is different, however. In their setup, individuals do not perfectly know the gross income they will earn when making their labor supply decision because gross income will be a function of labor supply and a stochastic term.

For this purpose we slightly reinterpret the classical tax perturbation method. For ease of exposition, we consider the case without capital taxation.¹⁵ 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 net present value (NPV) mechanical tax revenue increase of this reform is 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:¹⁶ The uniform lump-sum tax increase is given by

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

Individuals with $\theta_1 < \theta$ will enjoy higher period one consumption of $\Delta(\theta)$, whereas period one consumption for individuals with $\theta_1 \geq \theta$ will decrease by $1 - \Delta(\theta)$. Whether period t consumption will be increased or decreased depends on the realization of the shock. From a period $t - 1$ perspective, the reform of the tax function in period t will provide an insurance value. To obtain an expression for this insurance value, define for each type θ^{t-1} a ‘constant utility term’:

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

The numerator captures the (expected) utility loss in period t due to the tax increase (absent the lump-sum tax adjustment). Dividing it by the expected marginal utility in period t says by how much consumption had to be increased in period t , for every possible realization of the t -period shock, in order to make the individual of type 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

¹⁵The presence of a non-zero capital tax τ_s has no impact on the decomposition in Proposition 3, however, the perturbation argument is a bit more involved. In Appendix A.4 we show how the perturbation argument differs in that case.

¹⁶This lump sum increase has no first-order impact on welfare via the implied responses of savings behavior. The reason is that the change in behavior has no first-order impact on individual utilities by the envelope theorem and that there is no effect on the government budget because wealth taxes are zero. An equivalent option would be to redistribute the lump sum in the first period. But also in the presence of a savings tax a slightly related tax reform can be constructed to obtain the decomposition of the mechanical effect. See Appendix A.4 for the proof.

is simple to show that it is always positive. Adding up and integrating over all histories yields (14). $\mathcal{M}_t^I(\theta)$ 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 for individuals of type θ^{t-1} increases by

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

in monetary terms, 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:

$$\begin{aligned} \mathcal{M}_t^{R*}(\theta) &= \int_{\Theta^{t-1}} \left(\frac{\lambda h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} - \frac{\tilde{h}_{t-1}(\theta^{t-1}) \int_{\Theta_t} U'(\theta^{t-1}, \theta_t) dF_t(\theta_t | \theta_{t-1})}{(1+r)^{t-1}} \right) \\ &\quad \times (\mathcal{CU}(\theta^{t-1}; \theta) - \Delta(\theta)) d\theta^{t-1}. \end{aligned} \quad (16)$$

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

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

This follows from the expression for λ which says that average period one social marginal utility is equal to λ (see Appendix A.2) and the Euler equations. Combining this with (16) yields $\mathcal{M}_t^R(\theta)$ as defined in (15).

These arguments reveal that (15) indeed measures the welfare effect through redistribution between ex-ante heterogeneous agents. 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.

3.3 Capital Income Taxes

We now turn to the question of 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 4. *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

$$\sum_{t=2}^T \frac{1}{(1+r)^{t-2}} \int_{\Theta^{t-1}} a_t(\theta^{t-1}) h_{t-1}(\theta^{t-1}) d\theta^{t-1}. \quad (17)$$

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

$$\sum_{t=2}^T \beta^{t-1} (1+r) \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \int_{\underline{\theta}}^{\bar{\theta}} \frac{U'(\theta^{t-1}, \tilde{\theta})}{\lambda} dF_t(\tilde{\theta} | \theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1}. \quad (18)$$

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:

$$\begin{aligned} & \sum_{t=2}^T \frac{\tau_s}{(1+r)^{t-2}} \int_{\Theta^{t-1}} \frac{\partial a_t(\theta^{t-1})}{\partial \tau_s} h(\theta^{t-1}) d\theta^{t-1} \\ &= \sum_{t=2}^T -\frac{1}{(1+r)^{t-2}} \frac{\tau_s}{1 - \tau_s} \int_{\Theta^{t-1}} \zeta_{a_t, 1-\tau_s}(\theta^{t-1}) a_t(\theta^{t-1}) h(\theta^{t-1}) d\theta^{t-1}. \end{aligned} \quad (19)$$

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 $(17)|_{\tau=0} + (18)|_{\tau=0} \neq 0$.

For $\tau \neq 0$, however, it holds that $(19) \neq 0$. Optimality of τ_s then requires $(17) + (18) + (19) = 0$, which yields (4) (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 (4). 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 (23).

Whereas formula (4) 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 (4). 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 2015). 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 homogenous agents, which can end up as a high or low skilled agent in period 2. Savings are taxed to relax the incentive constraint in period 2 due to an income effect on labor supply. Savings are not taxed for redistributive issues because ex-ante homogeneous all save the same amount.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 (2013b) derive a formula for the optimal linear inheritance tax in an overlapping generations framework. As for the formulas presented in this paper, equity-efficiency considerations are key to understand optimal bequest taxation.

¹⁷Bastani, Blomquist, and Micheletto (2011) 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.

4 Simulation

4.1 Calibration

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

We base our parameterization on their results. Given the estimates of Karahan and Ozkan (2013) for the evolution of income over the life cycle¹⁸, 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.

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 5 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).¹⁹

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 $\rho = 1.5$ (Chetty 2006).

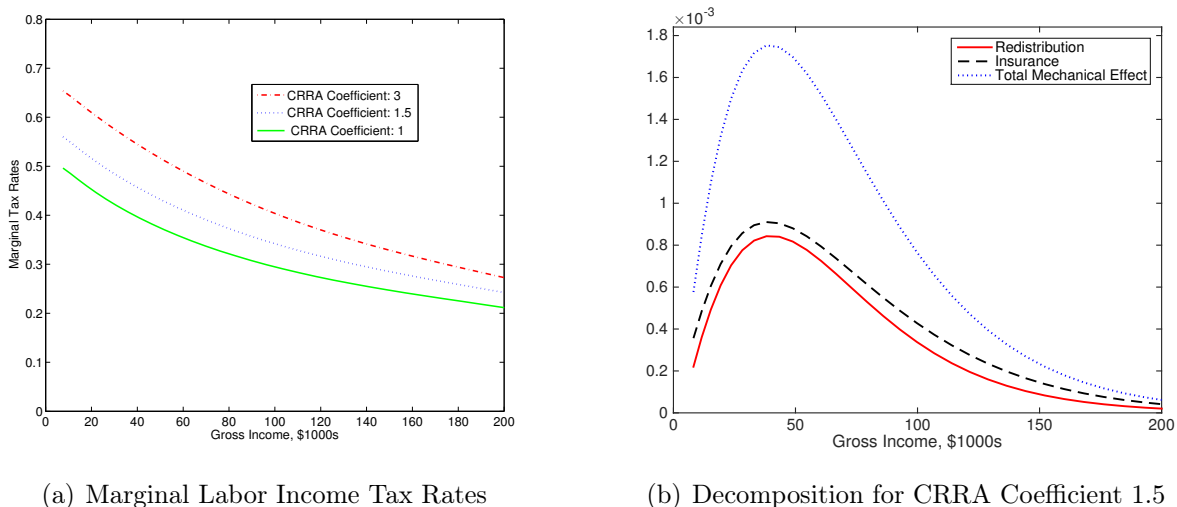


Figure 1: ‘Laissez-Faire’ Pareto Weights

4.2 Optimal Policies

In Figure 1(a) we plot optimal marginal tax rates on labor income for different values of risk aversion. 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.

Moreover, taxes are increasing in risk aversion. If risk aversion increases, so does the insurance value of taxation. At the same time also the redistributive desire increases because marginal utility is decreasing faster in income.

Figure 1(b) illustrates the insurance and redistribution components of the mechanical effect at the Utilitarian optimum as discussed in Section 3.2 for $\rho = 1.5$. We plot each component added up over all periods: $\sum_{t=1}^3 \mathcal{M}_t(\theta)$, $\sum_{t=1}^3 \mathcal{M}_t^I(\theta)$ and $\sum_{t=1}^3 \mathcal{M}_t^R(\theta)$. By definition, $\mathcal{M}_1^I(\theta) = 0$, as in the first period taxes purely redistribute but do not provide any insurance value. All three functions are hump-shaped. Moreover, the insurance and redistribution effects co-move closely across the income distribution. Interestingly, for this Utilitarian case, the insurance component is slightly bigger than the redistribution component for all income levels, however, the magnitude of effects are close together.

The optimal tax rate on capital income is 14.82%. 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 (2011).²⁰ In the next section, we show that the level of the

¹⁸We gratefully acknowledge that they shared some estimates with us that are not directly available from their paper.

¹⁹We 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.

²⁰As Farhi, Sleet, Werning, and Yeltekin (2011), we calculate the tax on annual capital income as a relative geometric average $\tau_{CI,annual}$ according to

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

optimal capital tax rate can change significantly if the labor income tax code is not set at the optimal but some exogenous level.

4.3 The Interaction of Labor and Capital 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.

4.3.1 Optimal Capital Taxes for Given Labor Taxes

The optimal capital income tax formula (4) 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 schedule²¹, (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 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 3% and 51%. In line with the argumentation from above, the less progressive the labor income tax schedule the higher the optimal capital tax rate and also the implied welfare gains from capital taxation.²² Wealth inequality increases if the labor tax code is less redistributive: going from the case with zero labor taxes to the high tax scenario (iv), the wealth share held by the top 10% of the wealth distribution in the last period increases from 64% to over 80%.²³

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 (4) 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,

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

²¹Here, we use the Gouveia-Strauss specification of Guner, Kaygusuz, and Ventura (2013), who provide different parametric approximations of the US tax system.

²²As 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.

²³For brevity, we do not provide comparative statics w.r.t. the CRRA coefficient. A higher CRRA coefficient increases both, the size and the welfare gain of capital taxation, see also an earlier version of this paper (Findeisen and Sachs 2014).

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

Table 1: Optimal Capital Tax Rates For Given Labor Tax Systems (Utilitarian Planner, CRRA Coefficient of 1.5)

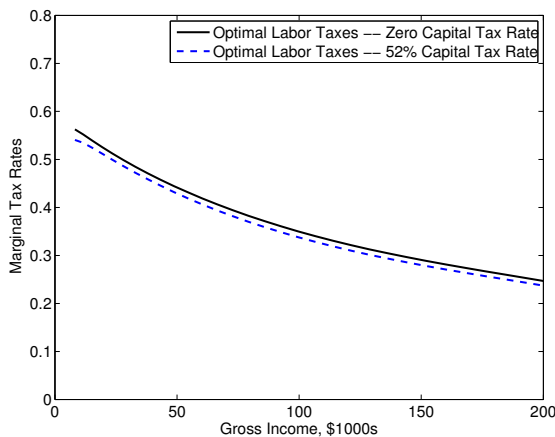


Figure 2: Optimal Labor Income Taxes for given Capital Income Taxes

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.

4.3.2 Optimal Labor Income Taxes for Given Capital Taxes

The optimal labor income tax schedule is still described by (8) – the formula holds for any level of the capital tax rate τ_s . The optimal labor income tax schedule is nevertheless influenced by the degree of capital taxation because the level of the capital tax rate impacts the terms $\mathcal{M}_t(\theta_t)$ and $\mathcal{S}^I(\theta)$. The higher the capital tax rate, the weaker the desire for redistribution through labor income taxes as captured by $\mathcal{M}_t(\theta_t)$. The impact on $\mathcal{S}^I(\theta)$ is ambiguous.

Our quantitative exploration reveals that the level of the capital tax rate is relatively unimportant for the design of optimal labor income taxes. For example, if we compare the two extreme cases (i) zero capital taxes and (ii) very high capital taxes of 52% (which was the highest level in Table 1), the difference in the optimal labor income tax schedule is relatively small, see Figure 2. The schedules differ by at most two percentage points.

This result strongly contrasts with the finding that the optimal capital tax rate is strongly influenced by the labor income tax schedule. What is driving this asymmetry? The reason is that capital taxes hardly influence the redistributive power of labor income taxes. In particular, the distribution of labor income which is of key importance for optimal marginal tax rates, is not affected by the level of capital income taxation. The level of the capital tax rate only influences the mechanical effects (9) and the savings effect. However, these forces are of second order numerically. By contrast, the redistributive power of capital taxes heavily depends on inequality in savings, which in turn is strongly affected by the labor income tax code.

A Naive Planner: Using a Static Labor Tax Formula Given that the level of optimal labor income taxes hardly depends on the level of the savings tax, a natural question is how bad a planner who ignores the direct impact of labor taxes on savings tax revenue performs in terms of welfare. The planner in this scenario uses the optimal tax formula (8) with (10) set to zero. Thus, a government is considered that does not take into account the impact on public funds through the implied savings responses of individuals. Our simulation reveals that both, the effects on the optimal tax schedule and on welfare of using this static formula is minuscule. The welfare loss is only 0.01% in consumption equivalents.

4.4 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 3(a); richer individuals obtain a higher weight as compared to their population share.

Figure 3(b) contains the results for optimal marginal tax rates in this case. Tax rates start at about 25%, then they fall, slightly increase again and finally converge to a level around 10%. 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 3.2 provides the analytical tools to illustrate the insurance value. In Figure 4, we illustrate the 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, which is illustrated by the red dashed curve. However, the planner positively values the amount of insurance against income risk that is provided by taxation because also initially high-skilled

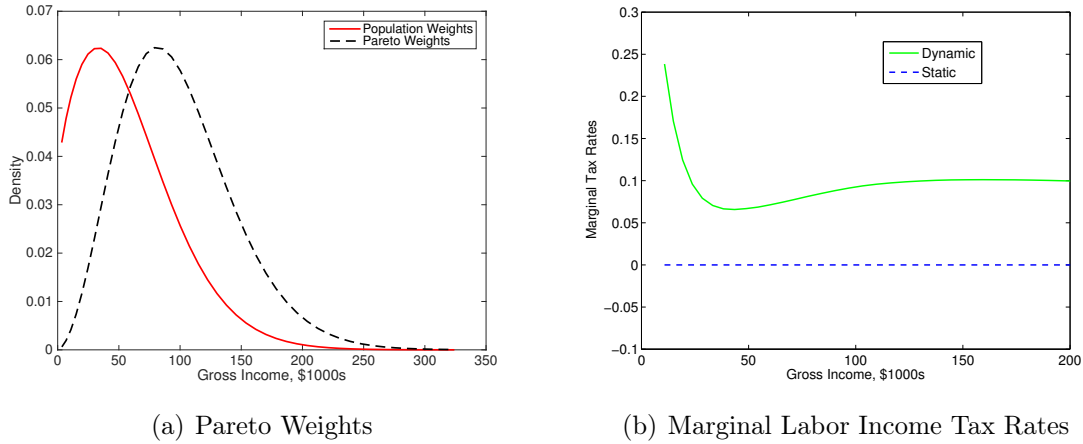


Figure 3: ‘Laissez-Faire’ Pareto Weights

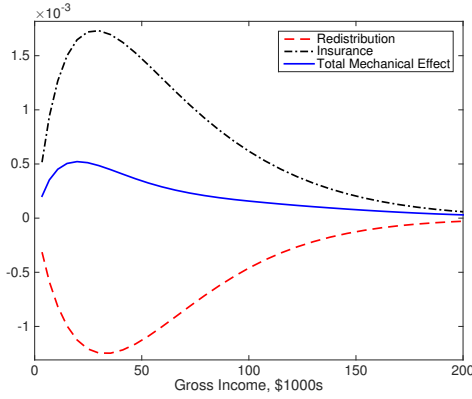


Figure 4: Decomposition of Mechanical Effect for ‘Laissez-Faire’ Pareto Weights

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 (compare Figure 1(b)). 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.

5 Conclusion

We contribute to the optimal taxation literature in public economics by studying labor and capital income taxation in life-cycle model with risk. This allows to decompose optimal taxes into an insurance and redistribution component. For the optimal capital income tax, simple equity-efficiency considerations are first-order. Optimal capital taxes are, hence in general, non-zero. In the model savings taxes are not redundant as individuals are heterogenous 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.²⁴ Examples are models of retirement or a life-cycle model with also the extensive margin of labor supply.

²⁴We apply this approach in the context of education in a model with multidimensional heterogeneity in Findeisen and Sachs (2015a).

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.
- BASTANI, S., S. BLOMQUIST, AND L. MICHELETTO (2011): “The welfare gains of age related optimal income taxation,” *International Economic Review* (forthcoming).
- BEST, M., AND H. KLEVEN (2013): “Optimal Income Taxation with Career Effects of Work Effort,” *Working Paper, London School of Economics*.
- BLOMQUIST, S., AND L. MICHELETTO (2008): “Age-related Optimal Income Taxation*,” *The Scandinavian Journal of Economics*, 110(1), 45–71.
- BOADWAY, R., AND M. SATO (2012): “Optimal Income Taxation with Uncertain Earnings: A Synthesis,” *Working Paper*.
- CHETTY, R. (2006): “A new method of estimating risk aversion,” *The American Economic Review*, 96(5), 1821–1834.
- (2012): “Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply,” *Econometrica*, 80(3), 969–1018.
- CONESA, J. C., S. KITAO, AND D. KRUEGER (2009): “Taxing Capital? Not a Bad Idea after All,” *The American Economic Review*, pp. 25–48.
- 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.
- 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.
- FARHI, E., C. SLEET, I. WERNING, AND S. YELTEKIN (2011): “Nonlinear Capital Taxation Without Commitment,” *The Review of Economic Studies* (forthcoming).
- 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 (2015a): “Designing efficient college and tax policies,” Discussion paper, Working Paper.

- (2015b): “Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans,” *Working Paper*.
- GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): “Optimal Indirect and Capital Taxation,” *Review of Economic Studies*, 70(3), 569–587.
- GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2015): “Redistribution and Social Insurance,” *Working Paper*.
- GOLOSOV, M., A. TSYVINSKI, AND N. WERQUIN (2014): “A Variational Approach to the Analysis of Tax Systems,” *Working Paper*.
- GRUBER, J., AND E. SAEZ (2002): “The elasticity of taxable income: evidence and implications,” *Journal of Public Economics*, 84(1), 1–32.
- GUNER, N., R. KAYGUSUZ, AND G. VENTURA (2013): “Income Taxation of U.S. Households: Facts and Parametric Estimates,” *Review of Economic Dynamics (forthcoming)*.
- HAMMOND, P. J. (1987): “Markets as constraints: multilateral incentive compatibility in continuum economies,” *The Review of Economic Studies*, 54(3), 399–412.
- JACOBS, B., AND D. SCHINDLER (2012): “On the desirability of taxing capital income in optimal social insurance,” *Journal of Public Economics*, 96(9-10), 853–868.
- 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 (2013): “Optimal taxation in a life-cycle economy with endogenous human capital formation,” *Working paper, UC Santa Barbara*.
- KARAHAN, F., AND S. OZKAN (2013): “On the persistence of income shocks over the life cycle: Evidence, theory, and implications,” *Review of Economic Dynamics*, 26(3), 452–476.
- KLEVEN, H. J., AND E. A. SCHULTZ (2012): “Estimating Taxable Income Responses using Danish Tax Reforms,” .
- KOCHERLAKOTA, N. R. (2004): “Wedges and taxes,” *American Economic Review*, 94(2), 109–113.
- KOCHERLAKOTA, N. R. (2005): “Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation,” *Econometrica*, 73(5), 1587–1621.
- MEGHIR, C., AND L. PISTAFERRI (2011): “Earnings, consumption and life cycle choices,” *Handbook of Labor Economics*, 4, 773–854.

- PIKETTY, T. (1997): “La redistribution fiscale face au chômage,” *Revue française d’économie*, 12(1), 157–201.
- PIKETTY, T., AND E. SAEZ (2013a): “Optimal labor income taxation,” *Handbook of Public Economics*, Vol 5.
- (2013b): “A Theory of Optimal Inheritance Taxation,” *Econometrica*, forthcoming.
- SAEZ, E. (2001): “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68(1), 205–229.
- SALANIÉ, B. (2003): *The Economics of Taxation*. The MIT Press.
- STANTCHEVA, S. (2015a): “Optimal Income, Education and Bequest Taxes in an Intergenerational Model,” .
- STANTCHEVA, S. (2015b): “Optimal Taxation and Human Capital Policies Over the Lifecycle,” *Working Paper, Harvard University*.
- WEINZIERL, M. (2011): “The Surprising Power of Age-Dependent Taxes,” *The Review of Economic Studies*, 78(4), 1490–1518.
- WERNING, I. (2011): “Nonlinear Capital Taxation,” *Working Paper, MIT*.

A Appendix

A.1 Lagrangian and First-Order Conditions

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

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

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

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

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial M(\theta)} = & - \sum_{t=1}^T \frac{\lambda}{(1 + r)^{t-1}} \int_{\Theta^{t-1}} f(\theta | \theta_{t-1}) h(\theta^{t-1}) d\theta^{t-1} \\
& + \sum_{t=1}^T \beta^{t-1} \int_{\Theta^{t-1}} U'(\theta^{t-1}, \theta) f(\theta | \theta_{t-1}) \tilde{h}(\theta^{t-1}) d\theta^{t-1} \\
& + \sum_{t=1}^{T-1} \int_{\Theta^{t-1}} \mu_t(\theta^{t-1}, \theta) U''(\theta^{t-1}, \theta) d\theta^{t-1} \\
& - \sum_{t=2}^T \beta(1 + r)(1 - \tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) U''(\theta^{t-1}, \theta) f(\theta | \theta_{t-1}) d\theta^{t-1} \\
& - \eta'(\theta) = 0
\end{aligned} \tag{20}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial y(\theta)} &= \sum_{t=1}^T \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta} f(\theta|\theta_{t-1})h(\theta^{t-1})d\theta^{t-1} \\
&\quad - \sum_{t=1}^T \beta^{t-1} \int_{\Theta} U'(\theta^{t-1}, \theta)\Psi' \left(\frac{y(\theta)}{\theta} \right) \frac{1}{\theta} f(\theta|\theta_{t-1})\tilde{h}(\theta^{t-1})d\theta^{t-1} \\
&\quad + \sum_{t=1}^{T-1} \int_{\Theta^{t-1}} \mu_t(\theta^{t-1}, \theta)U''(\theta^{t-1}, \theta)\Psi' \left(\frac{y(\theta)}{\theta} \right) \frac{1}{\theta} d\theta^{t-1} \\
&\quad - \sum_{t=2}^T \beta(1+r)(1-\tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1})U''(\theta^{t-1}, \theta)\Psi' \left(\frac{y(\theta)}{\theta} \right) \frac{1}{\theta} f(\theta|\theta_{t-1})d\theta^{t-1} \\
&\quad - \eta'(\theta)\Psi' \left(\frac{y(\theta)}{\theta} \right) \frac{1}{\theta} - \eta(\theta) \left(\Psi' \left(\frac{y(\theta)}{\theta} \right) \frac{1}{\theta^2} + \Psi'' \left(\frac{y(\theta)}{\theta} \right) \frac{y(\theta)}{\theta^2} \right) = 0 \tag{21}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_{t+1}(\theta^t)} &= \frac{\lambda}{(1+r)^{t-1}} \tau_s h_t(\theta^t) - \mu_t(\theta^t)U''(\theta^t) \\
&\quad - (1-\tau_s)^2 \beta(1+r)^2 \mu_t(\theta^t) \int_{\Theta} U''(\theta^t, \theta_{t+1})dF(\theta_{t+1}|\theta_t) \\
&\quad + (1-\tau_s)\beta(1+r)\mu_{t-1}(\theta^{t-1})U''(\theta^t) f(\theta_t|\theta_{t-1}) \\
&\quad + (1-\tau_s)(1+r) \int_{\Theta} \mu_{t+1}(\theta^t, \theta_{t+1})U''(\theta^t, \theta_{t+1})d\theta_{t+1} = 0 \tag{22}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \tau_s} &= \sum_{t=2}^T \frac{\lambda}{(1+r)^{t-2}} \int_{\Theta^{t-1}} a_t(\theta^{t-1})h(\theta^{t-1})d\theta^{t-1} \\
&\quad - \sum_{t=2}^T \beta^{t-1}(1+r) \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \int_{\Theta} U'(\theta^{t-1}, \theta_t)dF(\theta_t|\theta_{t-1})\tilde{h}(\theta^{t-1})d\theta^{t-1} \\
&\quad - \sum_{t=1}^{T-1} (1+r) \int_{\Theta^{t-1}} a_t(\theta^{t-1}) \int_{\Theta} \mu_t(\theta^{t-1}, \theta_t)U''(\theta^{t-1}, \theta_t)d\theta_t d\theta^{t-1} \\
&\quad + \sum_{t=2}^T \beta(1+r)^2(1-\tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\Theta} U''(\theta^{t-1}, \theta_t)a_t(\theta^{t-1})dF(\theta_t|\theta_{t-1})d\theta^{t-1} \\
&\quad + \sum_{t=2}^T \beta(1+r) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\Theta} U'(\theta^{t-1}, \theta_t)dF(\theta_t|\theta_{t-1})d\theta^{t-1} = 0. \tag{23}
\end{aligned}$$

A.2 Multiplier Functions

First, we integrate (20) over all realizations and use the transversality conditions $\eta(\underline{\theta}) = \eta(\bar{\theta}) = 0$ to obtain:

$$\begin{aligned} \lambda &= \left(\sum_{t=1}^T \frac{1}{(1+r)^{t-1}} \right)^{-1} \times \\ &\sum_{t=1}^T \left\{ \beta^{t-1} \int_{\Theta^{t-1}} U'(\theta^{t-1}, \theta_t) f_t(\theta_t | \theta_{t-1}) \tilde{h}_{t-1}(\theta^{t-1}) d\theta^{t-1} \right. \\ &+ \int_{\Theta^{t-1}} \mu_t(\theta^{t-1}, \theta_t) U''(\theta^{t-1}, \theta_t) d\theta^{t-1} \\ &\left. - (1 - \tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) U''(\theta^{t-1}, \theta_t) f_t(\theta_t | \theta_{t-1}) d\theta^{t-1} \right\}. \end{aligned} \quad (24)$$

From (20) we obtain through integrating:

$$\begin{aligned} \eta(\theta) &= \sum_{t=1}^T \frac{\lambda}{(1+r)^{t-1}} \int_{\Theta^{t-1}} \int_{\theta}^{\bar{\theta}} dF(\tilde{\theta} | \theta^{t-1}) dH(\theta^{t-1}) \\ &- \sum_{t=1}^T \beta^{t-1} \int_{\Theta^{t-1}} \int_{\theta}^{\bar{\theta}} U'(R_t(\theta^{t-1}, \theta_t)) dF(\tilde{\theta} | \theta^{t-1}) d\tilde{H}(\theta^{t-1}) \\ &- \sum_{t=1}^T \int_{\Theta^{t-1}} \int_{\theta}^{\bar{\theta}} \mu_t(\theta^{t-1}, \tilde{\theta}) U''(R_t(\theta^{t-1}, \tilde{\theta})) d\tilde{\theta} d\theta^{t-1} \\ &+ \sum_{t=1}^T (1 - \tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta}^{\bar{\theta}} U''(R_t(\theta^{t-1}, \tilde{\theta})) dF(\tilde{\theta} | \theta^{t-1}) d\theta^{t-1}. \end{aligned} \quad (25)$$

Next, we derive μ_t . Use (22) to obtain, with $SOC_t(\theta^t)$ being the second-order condition for savings from the individuals problem:

$$\begin{aligned} \mu_t(\theta^t) &= \frac{\frac{\lambda}{(1+r)^{t-1}} \tau_s h_{t-1}(\theta^{t-1}) + (1 - \tau_s) \beta (1+r) \mu_{t-1}(\theta^{t-1}) U''(\theta^t) f_t(\theta_t | \theta_{t-1})}{SOC_t(\theta^t)} \\ &+ \frac{(1 - \tau_s)(1+r) \int_{\Theta} \mu_{t+1}(\theta^t, \theta_{t+1}) U''(\theta^t, \theta_{t+1}) d\theta_{t+1}}{SOC_t(\theta^t)}. \end{aligned} \quad (26)$$

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

$$\begin{aligned} A_t(\theta^t) &= \frac{\frac{\lambda}{(1+r)^{t-1}} \tau_s h_{t-1}(\theta^{t-1})}{SOC_t} \\ B_t(\theta^t) &= \frac{(1 - \tau_s) \beta (1+r) U''(\theta^t) f_t(\theta_t | \theta_{t-1})}{SOC_t} \end{aligned}$$

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

then, we can rewrite (26) 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 $t = T - 2$:

$$\begin{aligned} \mu_{T-2}(\theta^{T-2}) &= A_{T-2}(\theta^{T-2}) + B_{T-2}(\theta^{T-2})\mu_{T-3}(\theta^{T-3}) \\ &\quad + \int_{\Theta} C_{T-2}(\theta^{T-2}, \theta_{T-1})\mu_{T-1}(\theta^{T-2}, \theta_{T-1})d\theta_{T-1}. \end{aligned} \quad (27)$$

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

$$\mu_{T-1}(\theta^{T-1}) = A_{T-1}(\theta^{T-1}) + B_{T-1}(\theta^{T-1})\mu_{T-2}(\theta^{T-2}). \quad (28)$$

Now insert (28) into (27). 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}

$$\mu_{T-3} = A_{T-3} + B_{T-3}\mu_{T-4} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + B_{T-2}\mu_{T-3} + \int_{\Theta} C_{T-2}A_{T-1}d\theta_{T-1}}{1 - \int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}} d\theta_{T-2}, \quad (29)$$

yielding

$$\mu_{T-3} = \frac{A_{T-3} + B_{T-3}\mu_{T-4} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + \int_{\Theta} C_{T-2}A_{T-1}d\theta_{T-1}}{1 - \int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}} d\theta_{T-2}}{1 - \int_{\Theta} \frac{C_{T-3}B_{T-2}}{1 - \int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}}}. \quad (30)$$

Now insert this into μ_{T-4}

$$\begin{aligned} \mu_{T-4} &= A_{T-4} + B_{T-4}\mu_{T-5} \\ &\quad + \int_{\Theta} C_{T-4} \frac{A_{T-3} + B_{T-3}\mu_{T-4} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + \int_{\Theta} C_{T-2}A_{T-1}d\theta_{T-1}}{1 - \int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}} d\theta_{T-2}}{1 - \int_{\Theta} \frac{C_{T-3}B_{T-2}}{1 - \int_{\Theta} C_{T-2}B_{T-1}d\theta_{T-1}}} d\theta_{T-3}. \end{aligned} \quad (31)$$

Rewrite to obtain

$$\begin{aligned} \mu_{T-4} = & \left[1 - \int_{\Theta} C_{T-4} B_{T-3} \left[1 - \int_{\theta_{T-2}} C_{T-3} B_{T-2} \left[1 - \int_{\Theta} C_2 B_{T-1} d\theta_{T-1} \right]^{-1} \right]^{-1} \right]^{-1} \\ & \left(A_{T-4} + B_{T-4} \mu_{T-5} \right. \\ & \left. + \int_{\Theta} C_{T-4} \frac{A_{T-3} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + \int_{\Theta} C_{T-2} A_{T-1} d\theta_{T-1}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}} d\theta_{T-2}}{1 - \int_{\Theta} \frac{C_{T-3} B_{T-2}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}}} d\theta_{T-3} \right). \end{aligned} \quad (32)$$

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

$$\begin{aligned} \mu_{T-5} = & \left[1 - \int_{\Theta} C_{T-5} B_{T-4} \left[\dots \left[1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1} \right]^{-1} \dots \right]^{-1} \right]^{-1} \\ & \left(A_{T-5} + B_{T-5} \mu_{T-6} \right. \\ & \left. + \int_{\Theta} C_{T-5} \frac{A_{T-4} + \int_{\Theta} C_{T-4} \frac{A_{T-3} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + \int_{\Theta} C_{T-2} A_{T-1} d\theta_{T-1}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}} d\theta_{T-2}}{1 - \int_{\Theta} \frac{C_{T-3} B_{T-2}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}}} d\theta_{T-3}}{1 - \int_{\Theta} \frac{C_{T-4} B_{T-3}}{1 - \int_{\Theta} \frac{C_{T-3} B_{T-2}}{1 - \int_{\Theta} C_{T-2} B_{T-1} d\theta_{T-1}}} d\theta_{T-4}} \right). \end{aligned} \quad (33)$$

Now define

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

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

$$\mu_{T-5} = \frac{A_{T-5} + B_{T-5} \mu_{T-6} + \int_{\Theta} C_{T-5} \frac{A_{T-4} + \int_{\theta_{T-3}} C_{T-4} \frac{A_{T-3} + \int_{\Theta} C_{T-3} \frac{A_{T-2} + \int_{\Theta} C_{T-2} A_{T-1}}{D_{T-2}} d\theta_{T-2}}{D_{T-3}} d\theta_{T-3}}{D_{T-4}}}{D_{T-5}}. \quad (34)$$

It now turns out helpful to make another definition:

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

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

$$\mu_2 = \frac{A_2 + B_2\mu_1 + E_2}{D_2}. \quad (35)$$

and get

$$\mu_1 = \frac{A_1 + E_1}{D_1}. \quad (36)$$

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

In equation (36) 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 (21) by $\Psi'^{\frac{1}{\theta}}$ and adding (20) yields

$$\frac{\mathcal{T}'(y(\theta))}{1 - \mathcal{T}'(y(\theta))} = \left(1 + \frac{1}{\varepsilon_{y_t, 1 - \mathcal{T}'(\theta)}}\right) \frac{\eta(\theta)}{\lambda\theta \sum_{t=1}^T \frac{1}{(1+r)^{t-1}} \int_{\theta_{t-1}} \int_{\theta_t} f(\theta_t|\theta_{t-1})h(\theta^{t-1})d\theta^{t-1}}. \quad (37)$$

Inserting (25) into (37) yields the formula for optimal labor tax rates.

A.4 Decomposition of the Mechanical Effect in the Presence of Capital Taxes

In this appendix we show that the decomposition of the mechanical effect as derived in Section 3.2 also holds in the presence of a non-zero capital tax rate τ_s . Again, we consider a reform, where the marginal tax rate for individuals with income $y(\theta)$ is increased such that all individuals with income $y > y(\theta)$ will pay exactly one more dollar of income taxes. Besides the labor supply effect on welfare, this affects welfare by $\sum_{t=1}^T (\mathcal{M}_t(\theta) + \mathcal{S}_t(\theta))$. We now slightly reinterpret the tax perturbation to intuitively derive that the effect on welfare can also be written as $\sum_{t=1}^T \mathcal{M}_t^R(\theta) + \mathcal{M}_t^I(\theta) + \mathcal{S}^I(\theta)$, where $M^I(\theta)$ is as defined in (14) and $\mathcal{M}_t^R(\theta)$ is as defined in (15).

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

In the presence of a non-zero capital tax τ_s , one also has to take into account that the increase in the lump sum element by $\Delta(\theta)$ (as defined in (3.2)) causes savings responses that affect welfare. Formally this effect on welfare through the change in the Euler equations is given by:

$$S^*(\theta) = \Delta(\theta) \left\{ \sum_{t=1}^{T-1} \int_{\Theta^{t-1}} \int_{\Theta} \mu_t(\theta^{t-1}, \theta_t) U''(\theta^{t-1}, \theta_t) d\theta_t d\theta^{t-1} - \sum_{t=2}^T (1 - \tau_s) \int_{\Theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\Theta} U''(\theta^{t-1}, \theta_t) dF_t(\theta_t | \theta_{t-1}) d\theta^{t-1} \right\}. \quad (38)$$

Equations (14) and (16) can be derived in the same way as in the main text. However, now we do not have $\mathcal{M}_t^{R^*}(\theta) = \mathcal{M}_t^R(\theta)$ but instead $\sum_{t=1}^T \mathcal{M}_t^{R^*}(\theta) + S^*(\theta) = \sum_{t=1}^T \mathcal{M}_t^R(\theta)$:

$$\begin{aligned} \sum_{t=1}^T \mathcal{M}_t^{R^*}(\theta_t) + S^*(\theta) &= \int_{\theta}^{\bar{\theta}} \lambda f_1(\theta_1) - \tilde{f}_1(\theta_1) U'(\theta_1) d\theta_1 \\ &+ \sum_{t=2}^T \int_{\Theta^{t-1}} \left(\frac{\lambda h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} - \frac{\tilde{h}_{t-1}(\theta^{t-1}) \int_{\Theta} U'(\theta^{t-1}, \theta_t) dF_t(\theta_t | \theta_{t-1})}{(1+r)^{t-1}} \right) \times \mathcal{CU}(\theta^{t-1}; \theta) d\theta^{t-1} \\ &- \Delta(\theta) \sum_{t=1}^T \int_{\Theta^{t-1}} \left(\frac{\lambda h_{t-1}(\theta^{t-1})}{(1+r)^{t-1}} - \frac{\tilde{h}_{t-1}(\theta^{t-1}) \int_{\Theta} U'(\theta^{t-1}, \theta_t) dF_t(\theta_t | \theta_{t-1})}{(1+r)^{t-1}} \right) d\theta^{t-1} + S^*(\theta). \end{aligned} \quad (39)$$

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

B Details on Calibration

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

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

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

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

where the AR(1) process is given by

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

and where the error term 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: $\sigma_{\eta,Y}^2 = 0.027$, $\sigma_{\eta,M}^2 = 0.013$ and $\sigma_{\eta,O}^2 = 0.026$
- Variances of the transitory shock: $\sigma_{\epsilon,Y}^2 = 0.056$, $\sigma_{\epsilon,M}^2 = 0.059$ and $\sigma_{\epsilon,O}^2 = 0.068$

Age-independent parameters:

- Variance of individual fixed effect: $\sigma_\alpha^2 = 0.0707$
- Variance of z_1 (i.e. the starting value of the persistence term): $\sigma_z^2 = 0.0767$

Time-dependent parameters:

- As we consider only one cohort, we assume the time dependent loading factors 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 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),

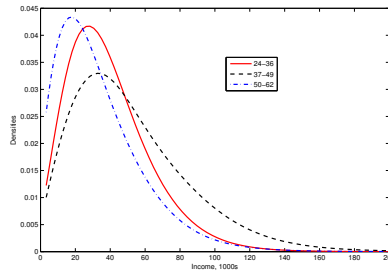


Figure 5: Income Distribution for Each Age Group

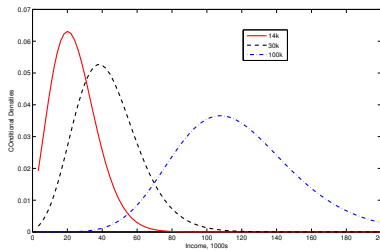


Figure 6: Conditional Income Distributions, Middle Aged Workers

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 6 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.