CESifo WORKING PAPERS

10684
2023

Original Version: October 2023
This Version: March 2024

Redistribution with Unequal Life Expectancy

Sebastian Koehne

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
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: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10684*

Redistribution with Unequal Life Expectancy

Abstract

This paper introduces life expectancy inequality into a tractable Mirrleesian life-cycle model and characterizes the optimal income tax policy using theory and calibration. A positive association between life expectancy and income counteracts the well-known static pattern of declining marginal utility. As a result, the mechanical value of redistribution is reduced at all income levels. Moreover, the pension wedge becomes a novel determinant of optimal taxation, motivating relatively lower optimal tax rates for low earners and relatively higher optimal tax rates for high earners. Quantitatively, the effects of the mechanical value of redistribution dominate, and the optimal marginal tax rates fall by up to 10 percentage points when life expectancy is heterogeneous.

JEL-Codes: D820, H210.

Keywords: optimal taxation, redistribution, life expectancy, inequality.

Sebastian Koehne
Kiel University / Germany
koehne@economics.uni-kiel.de

March 22, 2024
I gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through grant KO 6330/1-1.

1 Introduction

Income and life expectancy are strongly correlated (Goldman, 2001; Cutler et al., 2011). Recent empirical evidence documents that individuals in the top decile of the US income distribution can expect to live about 10 years longer than those in the bottom decile (Chetty et al., 2016). Similar socioeconomic health gaps exist around the world, even in countries with lower levels of economic inequality and more universal health care systems (Mackenbach et al., 2008; European Commission, 2013).[1] Since income and survival are fundamental determinants of well-being, redistributive policies should take into account inequalities in both dimensions. However, existing studies of optimal redistribution typically focus on economic inequality and, with very few exceptions, neglect differences in life expectancy.[2] The goal of this paper is to assess the implications of unequal lifespans for optimal redistribution. To this end, I introduce life expectancy inequality into a tractable Mirrleesian life-cycle model and characterize the optimal income tax policy using economic theory and calibration.

In general, life expectancy inequality can both affect and be affected by redistribution policy. In the main specification of the model, I consider life expectancy inequality to be exogenous.[3] The reason for this choice is that the current state of empirical knowledge about the income-health gradient in rich countries seems far too limited to be able to model a "production function" for life expectancy. The surveys by Cutler et al. (2006) and Cutler et al. (2011) suggest that a causal role of income is not evident and that the gradient is primarily the result of reverse causality and a number of other factors (including education, access to health care, health behavior, parental resources, and genetic differences) whose relative importance remains poorly understood. Several recent studies using panel data or quasi-experimental variations in income or wealth also suggest that the impact of economic resources on health is small to negligible in developed countries (Adams et al., 2003; Meer et al., 2003; Frijters et al., 2005; Banerjee et al., 2010; Cesarini et al., 2016; Gerdtham et al., 2016; Erixson, 2017). Although there is no final consensus yet, the exclusion of a causal pathway from income to health is a

[1] In Germany, for example, the gap in life expectancy between the top and bottom deciles of the income distribution is about 7 years (Lampert et al., 2019).

[2] Cremer et al. (2004) and Shourideh and Troshkin (2017) model heterogeneous life expectancy in Mirrleesian frameworks to explore optimal retirement incentives. Hosseini and Shourideh (2019) use a similar approach to study asset taxation.

[3] I also extend the theory to a setting with endogenous life expectancy. However, the calibration of such settings lacks a reliable empirical counterpart.

viable working assumption.

I introduce life expectancy inequality into a two-period extension of the seminal Mirrlees (1971) model. The first period represents the working life of the individuals and the second period describes the retirement phase. The individuals differ from each other on the basis of a one-dimensional, unobservable characteristic that determines both their labor productivity and their probability of reaching the retirement phase. Thus, the model is able to replicate the positive association between between income and life expectancy observed in the data. The government optimizes a nonlinear income tax function and employs an exogenous pension system that transfers resources to the retirement phase. Individuals can supplement their retirement benefits by annuitizing the disposable income they do not spent during their working years. With a uniform life expectancy, the model condenses to the classic static Mirrlees setup, allowing me to connect to the well-known characterizations of the optimal policy by Diamond (1998), Saez (2001) and many others.

The main findings are as follows. First, the pension wedge (which measures the implicit tax imposed by the pension system) becomes a novel determinant of optimal income taxes. Intuitively, the efficiency costs of income taxation are large when there is a pre-existing distortion to labor supply. Therefore, the optimal statutory tax rate decreases with the size of the pension wedge.[4] Provided that the pension system is progressive, this factor depresses optimal tax rates particularly for high earners (except for top earners whose income exceeds the limits of the pension system). However, as life expectancy inequality increases, the pension wedges of high and low earners move closer together. Through this channel, life expectancy inequality leads to relatively lower optimal tax rates for low earners and relatively higher optimal tax rates for high earners.

Second, the mechanical value of redistribution differs from its static counterpart due to a "lifespan factor". While in static models an individual's disposable income can be converted directly into her marginal utility, the conversion process is affected by the individual's life expectancy in the present environment. Individuals who expect to live longer will place a higher weight on the future and thus reduce their current consumption. This increases their marginal utility from each tax dollar. In addition, individuals with long life expectancies may face higher

[4] Fiscal externalities provide an alternative explanation for this result. Note that a positive wedge creates fiscal externalities for the government. In this case, the government receives an extra benefit from the generation of income and should not discourage income choices too much.

prices for annuitizing their wealth. Provided that the income effect of the annuity price dominates the substitution effect, this channel amplifies the positive impact of life expectancy on marginal utility. Overall, the positive association between life expectancy and income counteracts the well-known static pattern of declining marginal utility. As a result, the mechanical value of redistribution decreases at all income levels.

Third, I show that the main optimality principle generalizes to a wide range of alternative scenarios. In particular, I explore a setting without annuity markets, where individuals rely on storage to transfer wealth. In addition, I study a model with health investment. I also explore the combined choice of tax and pension policies and I extend the analysis to multiple periods. In all of these settings, optimal redistribution is shaped by the same forces as in the main model, with only minor modifications.

Fourth, I evaluate the quantitative importance of life expectancy inequality for optimal redistribution. I calibrate the model to the US economy and simulate optimal income taxes with and without life expectancy inequality. Quantitatively, the "lifespan factor" strongly dominates the "pension wedge factor". Thus, optimal marginal tax rates drop at all income levels when life expectancy inequality is introduced. Averaged across taxpayers, the optimal marginal tax rates decrease by 6 percentage points in the baseline calibration. The impact is particularly pronounced at low incomes, where optimal marginal tax rates fall by more than 10 percentage points. For high earners, the impact of life expectancy inequality on optimal marginal tax rates is rather small. This is because life expectancy rises steeply with income at low levels and flattens out at higher levels.[5] Finally, I show that the results are qualitatively and quantitatively robust to alternative model parameterizations.

Related literature. Building on the seminal Mirrlees (1971) model, this paper relates to the extensive literature on optimal nonlinear income taxation surveyed, for example, by Mankiw et al. (2009), Diamond and Saez (2011), and Piketty and Saez (2013). To my knowledge, this is the first paper that links the characterization of nonlinear income taxes to heterogeneous life expectancy.

With its life-cycle perspective, the paper also contributes to the literature on dynamic optimal taxation. Most works in this literature employ models with homogeneous life expectancy.[6]

[5] The relationship between life expectancy and income percentiles is nearly linear (Chetty et al., 2016). However, the relationship becomes concave when percentiles are replaced by income levels.

[6] See for example Grochulski and Kocherlakota (2010), Farhi and Werning (2013), Koehne and Kuhn (2015),

The paper is particularly related to studies by Michau (2014), Shourideh and Troshkin (2017) and Hosseini and Shourideh (2019), which also use Mirrleesian life-cycle frameworks with deterministic productivity paths. Michau (2014) assumes *homogeneous* lifespans and focuses on the retirement margin. He shows that optimal allocations involve distortions to the retirement decision and that their implementation requires history-dependent instruments such as a social security system. Hosseini and Shourideh (2019) consider heterogeneous lifespans and focus on intertemporal distortions. Orthogonal to the distributional aspects highlighted in the current paper, they show that asset subsidies late in life play an important role in Pareto optimal reforms. The current paper takes asset taxation (and the resulting net annuity prices) as given and focuses on income taxation as the main policy instrument. Shourideh and Troshkin (2017) study optimal retirement incentives using a Mirrleesian approach. In particular, they show that labor wedges and retirement wedges are closely connected in any constrained efficient allocation. In their quantitative analysis, they explore the implications of heterogeneous lifespans for efficient retirement ages and the pension system. The current paper complements this work by characterizing the implications of heterogeneous life expectancy for the income tax system.

The paper also relates to several earlier works on the implications of heterogeneous life expectancy for alternative policy questions. In particular, Sheshinski (2008) studies the optimal pricing of annuities when lifespans differ across individuals. Bommier et al. (2011a), Bommier et al. (2011b) and Leroux and Ponthiere (2013) explore redistribution with only heterogeneity in longevity. I diverge from those works by adding skill heterogeneity and a labor-leisure margin. Cremer et al. (2004) study optimal life-cycle policies in a model with two or three types of unobserved productivity and health status. In the absence of first-best instruments, they show that income tax rates should be positive and the retirement decision should be distorted. Abstracting from the retirement margin, the current paper extends their setup to a continuous type space in which skill and life expectancy are positively related, and derives a complete characterization of the optimal nonlinear income tax.

Heterogeneous mortality also plays a key role in the quantitative literature on optimal pension design (e.g., Bagchi, 2019; Abraham et al., 2023; Jones and Li, 2023; Pashchenko et al., 2023). Most papers in this literature examine calibrated models with rich heterogeneity and with parameterized functions for the policy instruments. Based on a parsimonious Mirrleesian

Golosov et al. (2016), Koehne (2018) and the survey by Stantcheva (2020).

framework, I add to this literature by providing a theoretical characterization of optimal nonlinear redistribution policies. Following Bagchi (2019), Abraham et al. (2023) and Jones and Li (2023), I apply a utilitarian welfare criterion. Pashchenko et al. (2023) propose an alternative welfare criterion with aversion to inequality in lifetime utilities.

Finally, the paper relates to the analysis of capital taxation with heterogeneous discount rates by Diamond and Spinnewijn (2011). To keep the analysis of two-dimensional heterogeneity tractable, their approach replaces the continuous Mirrlees (1971) framework with a four-types, two-jobs model. They show that savings taxes increase welfare when skills and discount rates are sufficiently correlated. The current paper works with explicit discount rates that are uniform across agents, but generates implicit discount rates that increase with skills due to higher life expectancy. Adding a positive relationship between explicit discount rates and skills would strengthen the results by amplifying the lifespan factor in the mechanical value of redistribution.

The rest of the paper is organized as follows. Section 2 extends the Mirrlees (1971) model to include heterogeneous life expectancy. Section 3 performs a theoretical analysis and derives a novel "$\mathcal{ABCD}$ formula" for optimal redistribution. Section 4 presents several theoretical extensions of the optimality condition. Section 5 calibrates the model and simulates optimal tax systems with homogeneous and heterogeneous life expectancy. Section 6 concludes. All proofs are relegated to the appendix.

2 Model

I consider a two-period version of the Mirrlees (1971) model of optimal nonlinear taxation. The first period represents the individual's working life. In this period, individuals work, consume and save. The second period represents the individual's retirement phase. In this period, individuals have left the labor market and consume retirement benefits and their annuitized wealth. Individuals differ with respect to a one-dimensional, unobservable characteristic that determines both their labor productivity and their probability of reaching the retirement phase.

2.1 Preferences and skills

There is a continuum of individuals with identical, time-separable von Neumann–Morgenstern preferences. The individuals live for up to two periods and discount the future with the factor

$\beta \in (0,1)$.

Individuals differ with respect to their skill $\theta \in \Theta := \left[\underline{\theta}, \overline{\theta}\right] \subset \mathbb{R}_{++}$. An individual with skill θ and labor hours l earns a gross income of $y = \theta l$. Skill and labor hours are private information, whereas gross income y is publicly observable. The distribution of skill types in the economy is defined by a smooth probability density $f : \Theta \to \mathbb{R}_{++}$ with full support. The cumulative distribution function of this distribution is denoted by $F : \Theta \to [0,1]$.

In the first period, individual utility depends on consumption $c_1 \in \mathbb{R}_+$ and labor hours $l \in \mathbb{R}_+$ and is given by $u\left(c_1 - v\left(l\right)\right)$. In this specification, the utility function $u : \mathbb{R} \to \mathbb{R}$ is strictly increasing, strictly concave and twice continuously differentiable. The labor disutility function $v : \mathbb{R}_+ \to \mathbb{R}_+$ is strictly increasing, strictly convex, twice continuously differentiable and satisfies $v\left(0\right) = 0$. Note that income effects on labor supply are zero for this specification.

With probability $p\left(\theta\right) \in (0,1)$, an individual with skill θ reaches the second period. In this period, the individual does not work and her utility $u(c_2) = u\left(c_2 - v\left(0\right)\right)$ is determined exclusively by her consumption c_2.

2.2 Individual problem

At the beginning of the first period, individuals observe their skill realization θ. Then, they choose their labor supply $l\left(\theta\right)$ and, as a consequence, their gross income $y\left(\theta\right) = \theta \cdot l\left(\theta\right)$. Individuals pay income taxes $T\left(y\left(\theta\right)\right)$ and social security contributions $S\left(y\left(\theta\right)\right)$ and spend their disposable income on contemporaneous consumption $c_1\left(\theta\right)$ and an annuity $a\left(\theta\right)$ with unit price $q\left(\theta\right)$ that is paid out during the retirement phase. In the second period (retirement), which is reached with probability $p\left(\theta\right)$, individuals do not work and consume the sum $c_2\left(\theta\right) = a\left(\theta\right) + B\left(y\left(\theta\right)\right)$ of their annuity payments $a\left(\theta\right)$ and social security benefits $B\left(y\left(\theta\right)\right)$.

Formally, individuals solve the following optimization problem:

$$\begin{aligned} V\left(\theta\right) := \max_{c_1, c_2, a, y} \; & u\left(c_1 - v\left(\frac{y}{\theta}\right)\right) + p\left(\theta\right)\beta u\left(c_2\right) \\ \text{subject to} \quad & c_1 + q\left(\theta\right) a = y - T\left(y\right) - S\left(y\right) \quad \text{and} \quad c_2 = a + B\left(y\right). \end{aligned} \tag{1}$$

The two budget constraints can be easily combined to obtain the following intertemporal budget constraint:

$$c_1 + q\left(\theta\right) c_2 = y - T\left(y\right) - S\left(y\right) + q\left(\theta\right) B\left(y\right). \tag{2}$$

Using the intertemporal budget constraint, the individual problem can be written as follows:

$$V(\theta) = \max_{c_2, y} u\left(y - T(y) - S(y) + q(\theta) B(y) - v\left(\frac{y}{\theta}\right) - q(\theta) c_2\right) + p(\theta) \beta u(c_2) \tag{3}$$

This decision problem can be decomposed into two steps. First, individuals maximize their disposable income minus the disutility of labor:

$$\max_{y} y - T(y) - S(y) + q(\theta) B(y) - v\left(\frac{y}{\theta}\right). \tag{4}$$

Given the solution of this problem, individuals optimize intertemporally:

$$\max_{c_2} u\left(y - T(y) - S(y) + q(\theta) B(y) - v\left(\frac{y}{\theta}\right) - q(\theta) c_2\right) + p(\theta) \beta u(c_2). \tag{5}$$

From the point of view of the individual, the labor supply problem (4) is equivalent to that of the typical static Mirrlees problem, except that the statutory income tax is replaced by the *effective* income tax,

$$\hat{T}(y; \theta) := T(y) + S(y) - q(\theta) B(y), \tag{6}$$

to account for the contributions and benefits of the pension system. Note that the effective marginal tax rate,

$$\hat{T}'(y; \theta) = T'(y) + S'(y) - q(\theta) B'(y), \tag{7}$$

is the sum of the statutory marginal tax rate T' and the *pension wedge* $S' - qB'$ (which captures the marginal labor distortion imposed by the pension system).

2.3 Government problem

The government does not observe skills or labor hours and is setting taxes T as a (nonlinear) function of income y. For the main analysis, I focus on the statutory tax function T as the government's design object and consider the pension system (S, B) and the annuity price q to be exogenous. In an extension, I discuss the implications of an endogenous pension system.[7]

The government chooses income taxes to maximize a weighted utilitarian welfare function

[7] For a model with asset taxation that yields endogenous intertemporal prices, see Hosseini and Shourideh (2019).

with nonnegative weights defined by the function $\chi : \Theta \to \mathbb{R}_+$. The tax function T is *optimal* if it solves the following problem:

$$\max_T \int_{\underline{\theta}}^{\overline{\theta}} \chi(\theta) V(\theta) dF(\theta), \tag{8}$$

where $V(\theta)$ is the indirect utility defined by Equation (1).

2.4 Optimal control approach

Following standard procedures in the optimal taxation literature, I reformulate the government problem (8) as an optimal control problem.

First, I map the labor supply problem to a reporting problem in which the individuals report their skill to a fictitious social planner, who allocates consumption resources and income targets based on the reports. An individual with skill θ who reports skill $\hat{\theta}$ to the planner is required to work $y\left(\hat{\theta}\right)/\theta$ hours and pay taxes $T\left(y\left(\hat{\theta}\right)\right) + S\left(y\left(\hat{\theta}\right)\right)$ in the first period and obtains a social security benefit $B\left(y\left(\hat{\theta}\right)\right)$ in the second period. Hence, if the individual reports a skill of $\hat{\theta}$, her disposable income minus labor disutility is given by:

$$\hat{x}\left(\theta, \hat{\theta}\right) := y\left(\hat{\theta}\right) - T\left(y\left(\hat{\theta}\right)\right) - S\left(y\left(\hat{\theta}\right)\right) + q(\theta) B\left(y\left(\hat{\theta}\right)\right) - v\left(\frac{y\left(\hat{\theta}\right)}{\theta}\right). \tag{9}$$

The tax system induces truthful reporting if it satisfies

$$x(\theta) := \hat{x}(\theta, \theta) = \max_{\hat{\theta}} \hat{x}\left(\theta, \hat{\theta}\right). \tag{10}$$

Following common practice (Mirrlees, 1971, 1976; Diamond, 1998; Saez, 2001), I replace the incentive compatibility constraint (10) by the associated envelope condition:

$$\frac{dx(\theta)}{d\theta} = \frac{\partial \hat{x}(\theta, \theta)}{\partial \theta} = \frac{l(\theta)}{\theta} v'(l(\theta)) + q'(\theta) B(\theta l(\theta)). \tag{11}$$

Next, I specify a feasibility constraint. The tax system is budget balanced if

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[T(y(\theta)) + S(y(\theta)) - q(\theta) B(y(\theta))\right] dF(\theta) - E \geq 0, \tag{12}$$

where E represents other government expenditures. Equivalently, using the variables x and l, this condition can be expressed as

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[\theta l\left(\theta\right) - x\left(\theta\right) - v\left(l\left(\theta\right)\right)\right] dF\left(\theta\right) - E \geq 0. \tag{13}$$

Finally, I express the objective function in terms of variables that are convenient for the control problem. As explained above, for a given level x of disposable income minus labor disutility, individual welfare is obtained by the following intertemporal value function:

$$V\left(\theta\right) = \tilde{V}\left(x\left(\theta\right);\theta\right) := \max_{c_2} u\left(x\left(\theta\right) - q\left(\theta\right)c_2\right) + p\left(\theta\right)\beta u\left(c_2\right). \tag{14}$$

By the envelope theorem, I obtain

$$\frac{\partial \tilde{V}\left(x\left(\theta\right);\theta\right)}{\partial x\left(\theta\right)} = u'\left(x\left(\theta\right) - q\left(\theta\right)c_2\left(\theta\right)\right), \tag{15}$$

where $c_2\left(\theta\right)$ is the maximizer of the intertemporal problem (14).

With these preparations, the optimal control problem is formulated as follows:

$$\max_{l} \int_{\underline{\theta}}^{\overline{\theta}} \chi\left(\theta\right)\tilde{V}\left(x\left(\theta\right);\theta\right) dF\left(\theta\right) \tag{16}$$

$$\text{subject to} \quad \int_{\theta_0}^{\theta_1} \left[\theta l\left(\theta\right) - x\left(\theta\right) - v\left(l\left(\theta\right)\right)\right] dF\left(\theta\right) - E \geq 0,$$

$$\text{and} \quad \frac{dx}{d\theta} = \frac{l\left(\theta\right)}{\theta}v'\left(l\left(\theta\right)\right) + q'\left(\theta\right)B\left(\theta l\left(\theta\right)\right).$$

3 The $\mathcal{ABCD}$ formula for optimal redistribution

In this section, I characterize the optimal tax system based on the first-order conditions of the control problem (16).[8] There are two main insights. First, heterogeneous life expectancies change the mechanical value of redistribution. Second, the pension wedge becomes a novel determinant of the optimal policy. Through this channel, heterogeneous life expectancies have an additional effect on optimal redistribution.

The following elasticities will help to obtain simple and intuitive expressions of the optimality

[8] In Appendix A.2, I derive the same result using a tax perturbation approach based on Saez (2001).

condition.

Lemma 1 (Elasticities) *Assuming a linearized tax and pension system, i.e.,* $T'' = S'' = B'' = 0$*, the elasticities of taxable income are as follows:*

$$\varepsilon_{y,\theta} = 1 + \frac{v'(l)}{lv''(l)} + \frac{\theta^2 q'(\theta) B'}{lv''(l)} \quad , \tag{17}$$

$$\varepsilon_{y,(1-T')} = \frac{v'(l)}{lv''(l)} + \frac{\theta (S' - q(\theta) B')}{lv''(l)} \quad , \tag{18}$$

$$\varepsilon_{y,(1-S')} = \frac{v'(l)}{lv''(l)} + \frac{\theta (T' - q(\theta) B')}{lv''(l)} \quad , \tag{19}$$

$$\varepsilon_{y,B'} = \frac{v'(l)}{lv''(l)} - \frac{\theta (1 - T'(y) - S'(y))}{lv''(l)} \quad , \tag{20}$$

$$\varepsilon_{y,(1-\hat{T}')} = \frac{v'(l)}{lv''(l)} \quad . \tag{21}$$

In particular, assuming a linearized tax and pension system and a zero pension wedge, the elasticity of taxable income with respect to the net-of tax rate is given by:

$$\varepsilon_{y,(1-T')}\Big|_{S'-qB'=0} = \frac{v'(l)}{lv''(l)}. \tag{22}$$

Next, I state the main optimality condition for the tax system. I provide two different versions of this condition. Equation (23) is stated in terms of the exogenous skill distribution. This version is helpful for the simulating optimal tax policy and forms the basis for the numerical analysis in Section 5. Equation (24), expressed in terms of the endogenous income distribution, is easily connected to well-known sufficient statistics results (Saez, 2001; Chetty, 2009) and can be used to test the optimality of a given tax system.

Proposition 1 ($\mathcal{ABCD}$ formula) *If the tax function T is optimal, then it satisfies*

$$\frac{T'}{1-T'} = \frac{\varepsilon_{y,\theta}}{\varepsilon_{y,(1-T')}} \left(1 - \frac{\int_\theta^{\overline{\theta}} \chi u' dF}{\lambda (1-F)}\right) \frac{1-F}{\theta f} \frac{T'}{T' + S' - qB'}, \tag{23}$$

where λ is the multiplier of the government budget constraint (13)*. Equivalently, the optimality condition stated in terms of the cumulative income distribution H and its density function*

$h = H'$ *is given by*

$$\frac{T'}{1-T'} = \frac{1}{\varepsilon_{y,(1-T')}} \left(1 - \frac{\int_y^{\overline{y}} \chi u' dH}{\lambda \left(1 - H\left(y\right)\right)}\right) \frac{\left(1 - H\left(y\right)\right)}{y \cdot h\left(y\right)} \frac{T'}{T' + S' - qB'}. \tag{24}$$

The optimality condition differs from the standard "$\mathcal{ABC}$ formula" by Diamond (1998) in one obvious and one more subtle way. First, there is a novel "$\mathcal{D}$ term" given by $\mathcal{D} := T'/\hat{T}'$, where $\hat{T}' = T' + S' - qB'$ is the effective marginal tax rate and $S' - qB'$ is the pension wedge. Second, the mechanical value of redistribution, captured by the term $\mathcal{B} := 1 - \int_\theta^{\overline{\theta}} \chi u' dF / \left(\lambda \left(1 - F\right)\right)$, also changes relative to the static case.

Below, I examine the consequences of these changes and briefly discuss the remaining terms in the optimality condition.

3.1 Elasticity of taxable income

The elasticity term, given by $\mathcal{A} := \varepsilon_{y,\theta}/\varepsilon_{y,(1-T')}$ in the first version of the optimality condition and by $1/\varepsilon_{y,(1-T')}$ in the second version, captures the efficiency losses induced by the local marginal tax rate. When the elasticity of taxable income is high, taxation generates a large deadweight loss and thus the marginal tax rate should be small.

Without a pension system, the elasticity term in Equation (23) can be simplified to the form $\mathcal{A} = 1 + 1/\varepsilon_{y,(1-T')}$ known from Diamond (1998), as the next result shows.

Lemma 2 (Simplification of $\mathcal{A}$) *If $q'B' = 0$, the elasticity ratio $\varepsilon_{y,\theta}/\varepsilon_{y,(1-T')}$ is given by:*

$$\frac{\varepsilon_{y,\theta}}{\varepsilon_{y,(1-T')}} = \frac{1 + \varepsilon_{y,(1-T')}\big|_{S'-qB'=0}}{\varepsilon_{y,(1-T')}}. \tag{25}$$

If $q'B' \neq 0$, then the ratio $\varepsilon_{y,\theta}/\varepsilon_{y,(1-T')}$ cannot be expressed in terms of the taxable income elasticities $\varepsilon_{y,(1-T')}$, $\varepsilon_{y,(1-S')}$, $\varepsilon_{y,B'}$.

To see how the pension system prevents the simplification of the "$\mathcal{A}$ term, note that the first-order condition for labor supply is given by the condition

$$\theta \left(1 - T' - S' + q\left(\theta\right) B'\right) = v', \tag{26}$$

which states that the marginal return to an hour of work (measured in terms of disposable lifetime income) equals the marginal disutility of an hour of work. The comparative statics of labor supply with respect to skill are therefore driven by two effects. First, a higher skill raises the gross return to work (as indicated by the factor θ that multiplies the left-hand side). Second, a higher skill has an additional influence on the net return to work if the discounted value of marginal pension entitlements $q(\theta) B'$ depends on skill, i.e., if $q'B'$ is nonzero. This novel term does not exist in the static case and is not captured by conventional elasticity concepts with respect to marginal tax or transfer rates. Therefore, the elasticity $\varepsilon_{y,\theta}$ does not vanish in the optimality condition (23) if $q'B'$ is nonzero. However, when the condition is stated in terms of the income distribution as in Equation (24), the transformation between income and skills drops out and the elasticity term takes the form that is familiar from static sufficient statistics results. In this sense, the trade-offs of the tax designer (as viewed with the tax perturbation approach) do not change in any deeper way, even though the "$\mathcal{A}$ term" is somewhat more elaborate than in the static case.

3.2 Mechanical value of redistribution

As usual, the "$\mathcal{B}$ term" in the optimal tax formula is given by

$$\mathcal{B} := 1 - \frac{\int_{\theta}^{\overline{\theta}} \chi u' dF}{\lambda (1 - F)} \tag{27}$$

and captures the mechanical value of redistribution. The marginal tax rate (MTR henceforth) at a given skill raises tax revenue from individuals with higher skills without affecting their labor supply. The redistributive value of the MTR is thus obtained by comparing the average social marginal welfare weight in the population (which equals 1) to the average social marginal welfare weight of individuals with skills above θ, which is obtained by integrating $\chi u' / \lambda$ in the expression above.

Although the definition of this term is standard, its value will differ substantially from the static case if heterogeneous life expectancies affect the cross-sectional profile of marginal utility. Importantly, the marginal utility in this term measures the marginal utility of *current* consumption and thus depends not only on disposable income, as in a static model, but also on the individual's life expectancy. Next, I characterize this effect using some manipulations of the intertemporal decision problem. Further below, I quantify the effect in a calibrated model.

Proposition 2 (Marginal utility) *(i) For CRRA utility with* $u(c) = \frac{1}{1-\sigma} c^{1-\sigma}$ *for some parameter* $\sigma > 0$*, the social marginal welfare weight in Equation* (27) *equals*

$$\frac{\chi(\theta)}{\lambda} u'(x(\theta) - q(\theta) c_2(\theta)) = \frac{\chi(\theta)}{\lambda} \left(1 + q(\theta)^{1-\frac{1}{\sigma}} (\beta p(\theta))^{\frac{1}{\sigma}}\right)^{\sigma} u'(x(\theta)). \tag{28}$$

(ii) For CARA utility with $u(c) = -\frac{1}{\alpha} e^{-\alpha c}$ *for some parameter* $\alpha > 0$*, the social marginal welfare weight in Equation* (27) *equals*

$$\frac{\chi(\theta)}{\lambda} u'(x(\theta) - q(\theta) c_2) = \frac{\chi(\theta)}{\lambda} \left(\frac{\beta p(\theta)}{q(\theta)}\right)^{\frac{q(\theta)}{1+q(\theta)}} (u'(x(\theta)))^{\frac{1}{1+q(\theta)}}. \tag{29}$$

By relating the marginal utility of current consumption to the marginal utility of disposable income, Proposition 2 uncovers the effect of the life expectancy on the social marginal welfare weight that is otherwise hidden in the intertemporal decision problem. Since the marginal utility of disposable income is the basis of the welfare weights in static models, this representation allows the value of redistribution to be interpreted as resulting from a static model with transformed welfare weights.

For CRRA utilities, Equation (28) shows that, ceteris paribus, individuals with higher survival rates $p(\theta)$ receive more social weight. Intuitively, holding the level of disposable income constant, individuals who expect to live longer will shift more income into the future and will thus have a higher marginal utility of current consumption. This effect increases their social weight.

Furthermore, Equation (28) implies that the effect of the annuity price $q(\theta)$ on the social weight depends on the intertemporal elasticity of substitution $1/\sigma$. If $\sigma \geq 1$, income effects dominate and the annuity price has a negative impact on current consumption.[9] The marginal utility and the social weight are then increasing in $q(\theta)$. If $\sigma < 1$, substitution effects dominate and the annuity price has a positive impact on current consumption. The marginal utility and the social weight are then decreasing in the annuity price. However, assuming that annuities are fairly priced, i.e., $(1+r) q(\theta) = p(\theta)$ for some interest rate r, the combined effect of the survival rate and the annuity price on the social weight remains positive even in the case with $\sigma < 1$.[10]

[9] See Appendix A.3 for details.

[10] Note that $1 + q(\theta)^{1-\frac{1}{\sigma}} (\beta (1+r) q(\theta))^{\frac{1}{\sigma}} = 1 + (\beta(1+r))^{\frac{1}{\sigma}} q(\theta)$ increases in $q(\theta)$ for any $\sigma > 0$.

Exponential utility is another common case where the marginal utility of current consumption can be related to the marginal utility of disposable income in closed form. In analogy to CRRA utility, Equation (29) shows that the social weight increases with the survival rate $p(\theta)$. Moreover, assuming fairly priced annuities and $\beta(1+r)=1$, the right-hand side of the equation simplifies to

$$\frac{\chi(\theta)}{\lambda} e^{-\frac{\alpha x(\theta)}{1+q(\theta)}} \tag{30}$$

and has a positive derivative with respect to $q(\theta)$. Thus, similar to the CRRA case, the combined effect of the survival rate and the annuity price on the social weight is positive under mild restrictions.

In summary, provided that life expectancies increase with skill, Proposition 2 suggests that the mechanical value of redistribution is smaller than in models with uniform life expectancy because long-lived individuals economize on current consumption. All else equal, this "lifespan factor" calls for lower optimal tax rates across the entire income distribution.

3.3 Inverse Pareto coefficient

The term $\mathcal{C} := (1-F)/(\theta f)$ has its usual form and captures the inverse Pareto coefficient of the skill distribution. Intuitively, this term relates equity gains and efficiency losses induced by the local MTR. The term increases with the proportion of taxpayers with skills above θ (whose tax level increases with the local MTR) and decreases with the productivity and the proportion of taxpayers with skill θ (whose choices are distorted by the local MTR).

3.4 Pension wedge factor

The novel term $\mathcal{D} = T'/\hat{T}'$ reflects the fiscal externalities of the pension system. Note that perturbations of T' cause income changes. To convert these income changes into public funds, the effective MTR $\hat{T}'$ rather than the statutory MTR T' is appropriate. When the local fiscal externality of the pension system is positive, i.e., when $T' < \hat{T}'$ or $S' - qB' > 0$, the "$\mathcal{D}$ term" pushes down the optimal tax rate, which leads to higher individual income and a bigger revenue from the fiscal externality.

A complementary interpretation of this factor can be based on efficiency costs rather than fiscal externalities. Intuitively, when the pension wedge is positive, income choices are already distorted before any income tax is levied. Ceteris paribus, the efficiency costs of income taxation are then higher, which motivates lower statutory tax rates.

In the cross section, the "$\mathcal{D}$ term" generates relatively higher statutory MTRs for low earners and relatively lower MTRs for high earners, provided that the pension wedge, $S' - qB'$, increases with income. Life expectancy inequality moderates this effect via the annuity price q. If life expectancies (and annuity prices q) rise with skill, the gradient of the pension wedge with respect to skill will be smaller and the effect of the pension wedge on optimal statutory MTRs will be dampened.

To explore the consequences of life expectancy inequality via the "$\mathcal{D}$ term" more formally, I consider a mean-preserving spread of an arbitrary annuity price function $q_0(\theta)$ with mean $\bar{q}$. Let

$$q(\theta) = q_0(\theta) + \eta(q_0(\theta) - \bar{q}),$$

where $\eta > 0$ is the spread parameter. The consequences of the mean-preserving spread are summarized below.

Proposition 3 (Comparative statics of $\mathcal{D}$) *Let* $q(\theta) = q_0(\theta) + \eta(q_0(\theta) - \bar{q})$. *Let* $\mathcal{D} = T'/(T' + S' - qB')$. *Assuming* $T' > 0$ *and* $B' > 0$, *the comparative statics of* $\mathcal{D}$ *with respect to the spread parameter* η *are as follows:*

$$\frac{\partial \mathcal{D}}{\partial \eta} \gtreqqless 0 \quad \iff \quad q_0(\theta) \gtreqqless \bar{q}. \tag{31}$$

Moreover, $\frac{\partial \mathcal{D}}{\partial \eta} = 0$ *if* $B' = 0$.

Assuming a monotonic relationship between individual life expectancies and annuity prices, the spread parameter η is a measure of life expectancy inequality. Following this interpretation, Proposition 3 states that life expectancy inequality affects optimal tax rates differently across the skill distribution. At low skills (annuity prices below the mean), higher life expectancy inequality tends to reduce the optimal tax rates via the "$\mathcal{D}$ term". At high skills (annuity prices above the mean), the opposite is true. Intuitively, as life expectancy inequality increases, the pension wedges created by a progressive pension system differ less between low and high earners. As a consequence, the efficiency costs of taxation increase for low earners (via the higher pre-existing distortion created by the pension wedge), pushing towards lower optimal statutory tax rates. For high earners, this effect is reversed. Finally, for top earners with a marginal pension replacement rate of zero, the pension wedges are constant regardless of the annuity price. Hence, the "$\mathcal{D}$ term" does not depend on life expectancy inequality at the top of the income distribution.

4 Extensions

This section shows that the condition for optimal redistribution extends to a number of alternative models with only minor modifications. Motivated by the annuity puzzle (e.g., Benartzi et al., 2011), one extension assumes that individuals cannot annuitize their wealth and rely on a simple storage technology to transfer wealth. Another extension creates endogenous variations in life expectancy by modeling a health investment. The third extension considers the combined choice of tax and pension policies. Finally, the results are extended two models with more than two periods.

4.1 Missing annuity market

Next, I extend the $\mathcal{ABCD}$ formula to a setting without an annuity market. For individuals, the only way to transfer resources across time is through storage, i.e., they face a uniform intertemporal price of $q(\theta) = \bar{q}$. The government accounts for future payments at the true fiscal costs, i.e., its intertemporal price is equal to $q^G(\theta) = p(\theta)/(1+r)$, where r represents the interest rate on government bonds.

The associated optimal control problem is then formulated as follows:

$$\max_{l} \int_{\underline{\theta}}^{\overline{\theta}} \chi(\theta)\, \tilde{V}(x(\theta);\theta)\, dF(\theta) \tag{32}$$

$$\text{subject to} \quad \int_{\theta_0}^{\theta_1} \left[\theta l(\theta) - x(\theta) - v(l(\theta)) - \left[q^G(\theta) - \bar{q}\right] B(\theta l(\theta))\right] dF(\theta) - E \geq 0,$$

$$\text{and} \quad \frac{dx}{d\theta} = \frac{l(\theta)}{\theta} v'(l(\theta)).$$

Proceeding exactly as in the proof of Proposition 1, I find that the $\mathcal{ABCD}$ formula in Equation (23) remains valid, except that the government's intertemporal price q^G replaces the individual intertemporal price q in the pension wedge factor. This change leaves the optimality condition qualitatively unaffected. Quantitatively, the storage technology is particularly punitive for individuals with low survival probabilities, who would face low annuity prices in an actuarially fair market. Depending on the intertemporal elasticity of substitution, this channel can dampen or amplify the "lifespan factor" in the mechanical value of redistribution, as noted above.

4.2 Endogenous life expectancy

Next, I replace the exogenous survival probability $p(\theta)$ with a model of health investment. In the first period, individuals make a health investment $h(\theta)$ that determines their survival probability $p(h(\theta))$. The probability function p is positive, strictly increasing and strictly concave in investment. To ensure a positive return on health investment, the period utility function u is assumed to be positive.[11]

In this setting, the intertemporal value function of the individuals is given by:

$$V(\theta) = \tilde{V}(x(\theta);\theta) := \max_{c_2,h} u(x(\theta) - q(\theta)c_2 - h) + p(h)\beta u(c_2). \tag{33}$$

Apart from this modification, the statement of the optimal control problem (16) remains unchanged. Thus, the approach of Proposition 1 establishes the same $\mathcal{ABCD}$ formula as in the baseline model with exogenous survival. Again, the survival probabilities affect the optimality condition only indirectly through the marginal utility of consumption.

As the next result shows, the survival probabilities remain monotonic in skill under mild restrictions. Thus, the implications of the health investment model for optimal taxation closely resemble those of the baseline model.

Proposition 4 (Monotonic survival) *Consider the model extension with endogenous life expectancy. Suppose that the period utility function u and the probability function p are positive. Moreover, suppose that p is twice continuously differentiable and satisfies $p' > 0 > p''$. If $c_2(\theta)$ and $q(\theta)$ are weakly (strictly) increasing in θ, then $p(h(\theta))$ is weakly (strictly) increasing in θ when solving* (33).

[11]Standard utility functions such as CRRA utilities satisfy this condition after the addition of a sufficiently large constant.

4.3 Optimal pension system

If the policy maker jointly chooses the tax system T and the pension system (S, B), there are two degrees of freedom and the optimization yields insights for the effective marginal tax rates $\hat{T}'$, but not for the rates T', S', B' separately. Following the proof of Proposition 1, the pension wedge factor drops out and optimal effective marginal tax rates $\hat{T}'$ are characterized by an $\mathcal{ABC}$ formula. In this way, life expectancy inequality still affects optimal policy choices through the mechanical value of redistribution. The implications for effective marginal tax rates thus follow from the arguments discussed in Section 3.2.

4.4 Many periods

The two-period model facilitates the exposition, but is not essential for the results. In this section, I extend the $\mathcal{ABCD}$ formula to a multi-period setting with age-dependent wages, preferences and survival rates.

In this setting, individuals work in periods $t = 1, \ldots, R-1$ and are retired in periods $t = R, \ldots, T$. Individuals differ on the basis of the one-dimensional type $\theta \in \Theta := \left[\underline{\theta}, \overline{\theta}\right] \subset \mathbb{R}_{++}$ which determines the probability $p_t(\theta)$ of being alive in period $t = 1, \ldots, T$ and the wage rate $w_t(\theta)$ for periods $t = 1, \ldots, R-1$. Individuals have access to a linear technology for intertemporal transfers, allowing them to convert $q_{t+1}(\theta)\, a_{t+1}$ units of resources in period t into a_{t+1} units of resources in period $t+1$. During the working life, individuals pay income taxes $T_t(y_t(\theta))$ and social security contributions $S_t(y_t(\theta))$ for $t = 1, \ldots, T$. Upon retirement, the pension system pays a benefit $B(y_1(\theta), \ldots y_{R-1}(\theta)) = \sum_{t=1}^{R-1} B_t(y_t(\theta))$ for as long as the individual is alive. Individuals choose consumption $c_t(\theta)$ for $t = 1, \ldots, T$, intertemporal transfers $a_{t+1}(\theta)$ for $t = 1, \ldots, T-1$ and income $y_t(\theta) = w_t(\theta)\, l_t(\theta)$ for $t = 1, \ldots, R-1$ to maximize their expected lifetime utility:

$$V(\theta) := \max_{c_t, a_t, y_t} \sum_{t=1}^{R-1} \beta^{t-1} p_t(\theta)\, u_t\left(c_t - v_t\left(\frac{y_t}{w_t(\theta)}\right)\right) + \sum_{t=R}^{T} \beta^{t-1} p_t(\theta)\, u(c_t) \tag{34}$$

subject to the period budget constraints

$$\begin{aligned} c_t + q_{t+1}(\theta)\, a_{t+1} &= y_t + a_t - T_t(y_t) - S_t(y_t) \quad \text{for } t < R, \\ c_t + q_{t+1}(\theta)\, a_{t+1} &= a_t + B(y_1, \ldots, y_{R-1}) \quad \text{for } R \leq t < T, \\ c_t &= a_t + B(y_1, \ldots, y_{R-1}) \quad \text{for } t = T. \end{aligned} \tag{35}$$

As usual, I transform the decentralized individual problem into a reporting problem in which the individuals report their type to a social planner. If an individual of type θ reports a type of $\hat{\theta}$, her disposable income minus labor disutility in period $t = 1, \ldots, R-1$ is given by:

$$\hat{x}_t\left(\theta, \hat{\theta}\right) := y_t\left(\hat{\theta}\right) - T_t\left(y_t\left(\hat{\theta}\right)\right) - S_t\left(y_t\left(\hat{\theta}\right)\right) + \sum_{\tau=R}^{T} q^{t,\tau}(\theta)\, B_t\left(y_t\left(\hat{\theta}\right)\right) - v_t\left(\frac{y_t\left(\hat{\theta}\right)}{w_t(\theta)}\right), \tag{36}$$

where $q^{t,\tau}(\theta) = q_{t+1}(\theta)\, q_{t+2}(\theta) \cdots q_\tau(\theta)$ is a factor that converts income in period τ into income in period t. Setting $x_t(\theta) := \hat{x}_t(\theta, \theta)$ and $Q_t(\theta) := \sum_{\tau=R}^{T} q^{t,\tau}(\theta)$, the envelope condition for period $t = 1, \ldots, R-1$ is given by:

$$\frac{dx_t}{d\theta} = \frac{\partial \hat{x}_t(\theta, \theta)}{\partial \theta} = l_t(\theta)\, \frac{w_t'(\theta)}{w_t(\theta)} v_t'(l_t(\theta)) + Q_t'(\theta)\, B_t(w_t(\theta)\, l_t(\theta)). \tag{37}$$

The feasibility constraint is

$$\int_{\underline{\theta}}^{\overline{\theta}} \sum_{t=1}^{R-1} q^{1,t}(\theta)\left[w_t(\theta)\, l_t(\theta) - x_t(\theta) - v_t(l_t(\theta))\right] dF(\theta) - E \geq 0 \tag{38}$$

and implies intertemporal budget balance for the tax system:

$$\int_{\underline{\theta}}^{\overline{\theta}} \sum_{t=1}^{R-1} q^{1,t}(\theta)\left[T_t(y_t(\theta)) + S_t(y_t(\theta)) - Q_t(\theta)\, B_t(y_t(\theta))\right] dF(\theta) - E \geq 0. \tag{39}$$

For a given profile $(x_1, \ldots, x_{R-1})$ of disposable income minus labor disutility, the intertemporal optimization problem of an individual of type θ is

$$\begin{aligned} \tilde{V}(x_1, \ldots, x_{R-1}; \theta) := \max_{a_t} & \sum_{t=1}^{R-1} \beta^{t-1} p_t(\theta)\, u_t(x_t + a_t - q_{t+1}(\theta)\, a_{t+1}) \\ & + \sum_{t=R}^{T} \beta^{t-1} p_t(\theta)\, u_t(a_t - q_{t+1}(\theta)\, a_{t+1}) \end{aligned} \tag{40}$$

where the maximization is with respect to asset holdings $a_2, \ldots, a_T$ and assuming $a_1 = a_{T+1} = 0$. The intertemporal value function of an individual of type θ is defined as

$$V(\theta) := \tilde{V}(x_1(\theta), \ldots, x_{R-1}(\theta); \theta). \tag{41}$$

With these preparations, the government problem can be stated as the following optimal control problem with state vector $(x_1, \ldots, x_{R-1})$ and control vector $(l_1, \ldots, l_{R-1})$:

$$\max_{l} \int_{\underline{\theta}}^{\overline{\theta}} \chi(\theta) \tilde{V}(x_1(\theta), \ldots, x_{R-1}(\theta); \theta) dF(\theta) \tag{42}$$

$$\text{subject to} \quad \int_{\theta_0}^{\theta_1} \sum_{t=1}^{R-1} q^{1,t}(\theta)\left[w_t(\theta) l_t(\theta) - x_t(\theta) - v_t(l_t(\theta))\right] dF(\theta) - E \geq 0,$$

$$\text{and} \quad \frac{dx_t}{d\theta} = l_t(\theta) \frac{w_t'(\theta)}{w_t(\theta)} v_t'(l_t(\theta)) + Q_t'(\theta) B_t(w_t(\theta) l_t(\theta)).$$

In this multi-period setting, the optimality conditions for the tax system are straightforward extensions of the $\mathcal{ABCD}$ formula given in Proposition (1), as the next result shows.

Proposition 5 ($\mathcal{ABCD}$ formula for many periods) *Consider the model extension with many periods. If the tax functions $T_1, \ldots, T_{R-1}$ are optimal, then for each $t = 1, \ldots, R-1$ they satisfy*

$$\frac{T_t'}{1 - T_t'} = \frac{\varepsilon_{y_t, w_t}}{\varepsilon_{y_t, (1-T_t')}} \left(1 - \frac{\int_{\theta}^{\overline{\theta}} \chi \frac{\beta^{t-1} p_t}{q^{1,t}} u_t' dF}{\lambda (1 - F)}\right) \frac{w_t'(1 - F)}{w_t f} \frac{T_t'}{T_t' + S_t' - Q_t B_t'}, \tag{43}$$

where λ is the multiplier of the intertemporal government budget constraint (38).

A comparison of Equations (23) and (43) shows that the marginal welfare weights are rescaled by the intertemporal discount factor β^{t-1}, the survival probability p_t and the reciprocal of the resource discount factor $q^{1,t}$. Moreover, the inverse Pareto coefficient of the type distribution F is rescaled by $\theta w_t'/w_t$ to map it to the wage distribution in period t. The economic interpretation of the formula is unaffected by these minimal adjustments.

5 Numerical simulation

In this section, I assess the quantitative relevance of heterogeneous life expectancies for optimal redistribution policy. I calibrate the model to the US economy and simulate optimal tax systems with and without life expectancy inequality.

5.1 Parameterization

The utility function is $u(c) = \frac{1}{1-\sigma}c^{1-\sigma}$ with a coefficient of relative risk-aversion given by $\sigma = 2$. Considering that individuals retire approximately 40 years after entering the labor force, I assume that each model period has a duration of 40 years and set the discount factor to $\beta = 0.98^{40}$. Labor disutility is of the iso-elastic form $v(l) = \left(1+\frac{1}{\phi}\right)^{-1} l^{1+\frac{1}{\phi}}$ with parameter $\phi = 0.5$. With this specification, the labor supply problem has a closed-form solution given by $l = [\theta(1 - T' - S' + qB')]^{\phi}$. For given marginal tax and benefit rates, income and skills are thus related as follows:

$$\theta = y^{\frac{1}{1+\phi}} \left[1 - T' - S' + qB'\right]^{-\frac{\phi}{1+\phi}}. \tag{44}$$

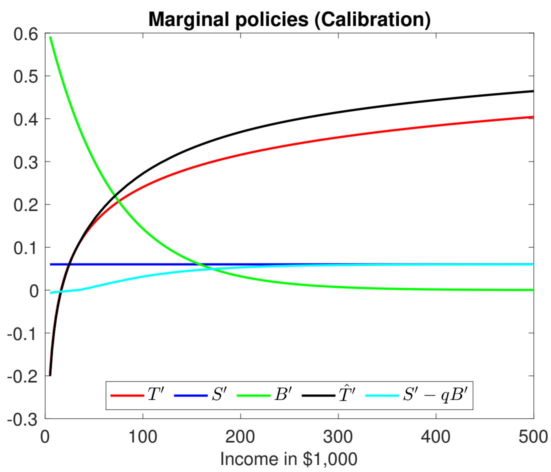
I exploit this relationship to infer the skill distribution from the income distribution.

In line with Sachs et al. (2020) and Koehne and Sachs (2022), the baseline income distribution is lognormal with an appended Pareto tail. Following Diamond and Saez (2011), the Pareto tail begins at \$150,000 with a Pareto coefficient that decreases from $\Pi = 2.2$ at income \$150,000 to a constant coefficient $\Pi = 1.5$ for incomes above \$350,000. All income thresholds are converted from the 2005 levels used by Diamond and Saez (2011) into current 2022 dollars. Incomes below the Pareto tail follow a lognormal distribution with parameters $\mu = 10.66$ and $\sigma = 0.968$. The parameter μ is chosen to match the 2022 level of US income per capita of \$73,360. The parameter σ is set to obtain a Pareto coefficient of 2.2 at the boundary of the lognormal part.

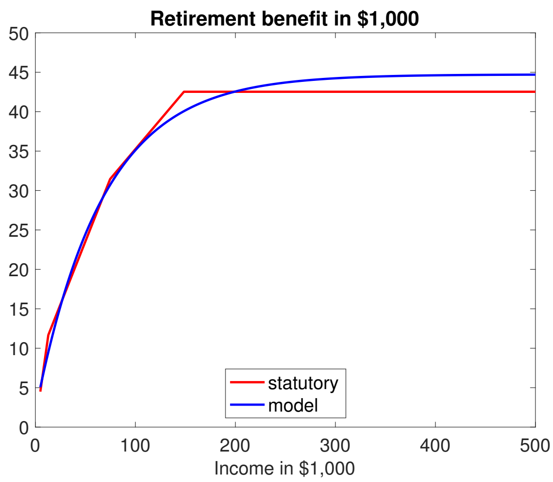
The baseline tax system has the well-known parametric form $T(y) = y - \nu y^{1-\tau}$used, for example, by Benabou (2002) and Heathcote et al. (2017). Because the social security system is modeled separately in my model, I choose a parameter of $\tau = 0.151$ that represents the progressivity of the US tax and transfer system abstracting from intergenerational redistribution (Heathcote et al., 2014).[12] The parameter ν controls the level of taxation and is chosen to match the average rate of the US income tax. Recent statistics published by the IRS (2022) report an adjusted gross income of \$12.592 trillion and a total income tax of \$1.711 trillion, generating an average tax rate of 13.6 percent. To match this rate, I set $\nu = 5.089$.

[12]Due to the progressivity of the social security system, the estimated progressivity parameter increases to $\tau = 0.181$ if the intergenerational dimension of redistribution is included (Heathcote et al., 2017).

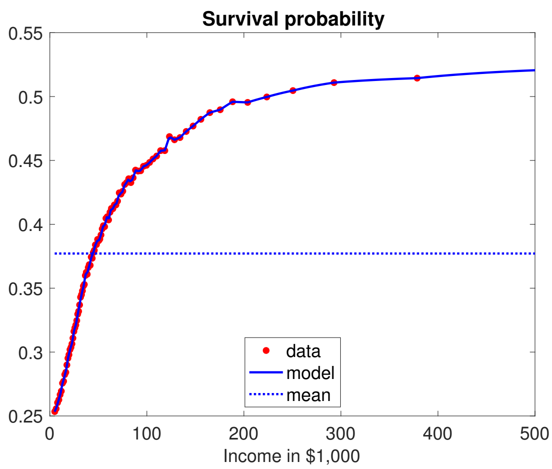
The social security system in the U.S. has a progressive, piece-wise linear schedule for retirement benefits. The marginal replacement rate for newly eligible individuals in 2022 equals 90% for incomes below \$12,288, 32% for incomes between \$12,288 and \$74,064, and 15% for incomes above \$74,064 up to the maximum creditable income of \$147,000 (SSA, 2022). I approximate the retirement benefits by fitting an exponential function, $B(y) = \delta e^{\varepsilon \cdot y} + \gamma$, with parameters $\delta = -42738.126$, $\varepsilon = 1.492 \cdot 10^{-5}$ and $\gamma = 44719.133$ obtained by least-squares estimation. This function provides an excellent fit of the statutory benefits (Figure 1b), yielding an R^2 of 0.994. The benefits are financed by a budget-balancing proportional income tax of 6.02 percent. Given the approximated tax and pension system (Figure 1a) and the lognormal-Pareto distribution of income, I use Equation (44) to calibrate the skill distribution following the approach of Saez (2001).



(a) Baseline marginal policy rates



(b) Retirement benefit B



(c) Survival probability p by baseline income

Figure 1: Calibration

The probability $p(\theta)$ of reaching the retirement phase is calibrated based on estimates of male life expectancy by income percentile constructed by Chetty et al. (2016). To approximate the remaining life expectancy of individuals reaching the full retirement age in 2022, I subtract 66 years and 4 months (the full retirement age for the cohort born in 1956) from the estimated life expectancy. Then, I convert life expectancy years into survival probabilities assuming that the retirement period has a duration of 40 years, and interpolate the probabilities across the income range. Following these steps, I obtain survival probabilities that range from 0.253 for bottom earners to 0.525 for top earners (Figure 1c), corresponding to expected retirement durations between 10 and 21 years.

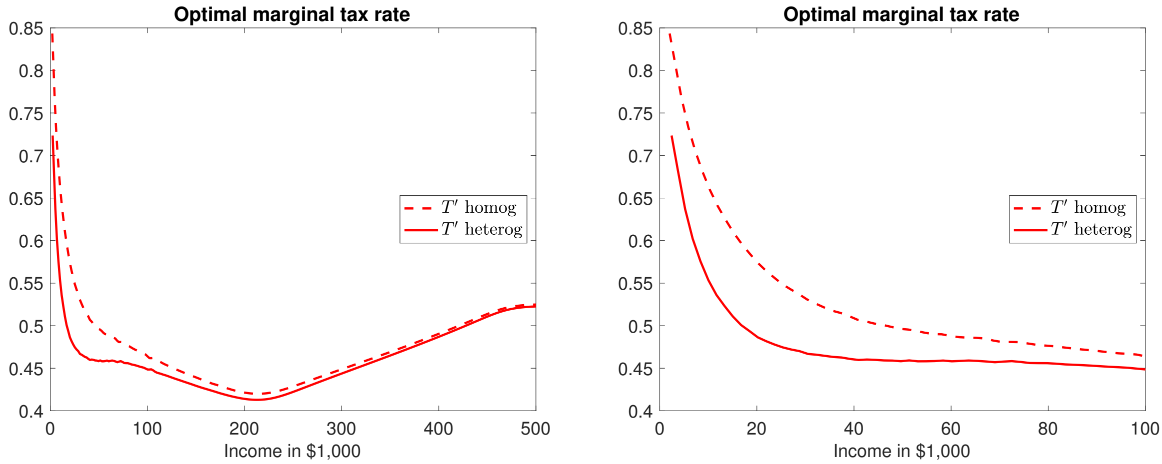
Finally, I assume that annuities are priced actuarially fairly: $q = \beta p$. Given the annuity prices and the baseline tax and pension system, I infer government expenditures E from the government budget constraint (12).

5.2 Optimal tax policy

Given the calibrated parameters, I solve for the optimal tax policy.[13] My computational approach follows Mankiw et al. (2009) and iterates the right-hand side of the optimality condition (23) until a fixed point is found. To confirm the optimality of the fixed point tax policy, I verify that the resulting income profile $y(\theta)$ is increasing in skill θ.

The solid line in Figure 2a displays the optimal tax policy. The optimal marginal tax rates are approximately U-shaped, ranging from 70 percent for bottom earners to 41 percent for middle earners. For top earners in the Pareto tail, the optimal tax rates converge to 52 percent. To explore the role of life expectancy inequality, I contrast these results with the optimal tax rates obtained in a model with homogeneous life expectancies $p = \bar{p}$. In this model, I set the individual survival rates to the mean of the heterogeneous life expectancy model. The annuity price in this model is given by $\bar{q} = \beta\bar{p}$. I maintain the skill distribution, the pension system and the level of government expenditure from the baseline model. The optimal marginal tax rates in the homogeneous life expectancy model are similarly U-shaped (dashed line in Figure 2a), but higher than in the heterogeneous life expectancy model. The difference is particularly pronounced for low-to-middle incomes and exceeds 10 percentage points for the lowest earners (Figure 2b).

[13] The pension system is held constant in the analysis.

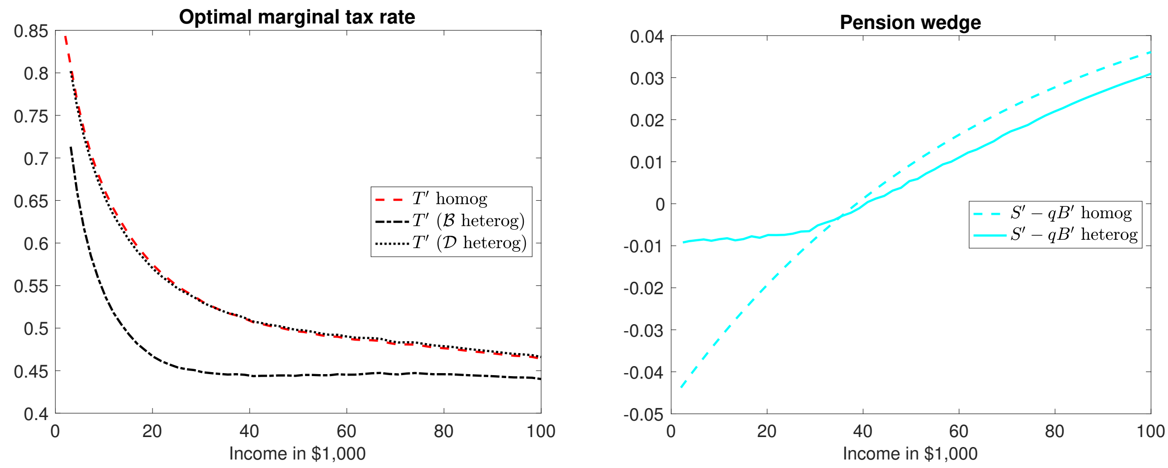


(a) Optimal marginal tax rates (b) Optimal marginal tax rates for low incomes

Figure 2: Optimal policy

To uncover the mechanism behind this finding, I isolate the roles of the mechanical value of redistribution (represented by the "$\mathcal{B}$ term" in the optimal tax formula) and the pension wedge (captured by the "$\mathcal{D}$ term"). The dash-dotted line in Figure 3a shows how the optimal tax rate varies with the mechanical value of redistribution. When the "$\mathcal{B}$ term" is set to the level of the heterogeneous life expectancy model, the optimal tax rates drop significantly below those of the homogeneous life expectancy model. This result mirrors the theoretical prediction of Proposition 2 and underscores the quantitative importance of accounting for lifecycle effects in the value of redistribution. The dotted line in Figure 3a shows how the optimal tax rate depends on the cross-sectional profile of the pension wedge. In this case, however, the effect is much more modest. Consistent with Proposition 3, setting the "$\mathcal{D}$ term" to the level of the heterogeneous life expectancy model leads to a reduction in optimal tax rates for low earners and to an increase in optimal tax rates further up in the income distribution. The non-monotonic shift of the pension wedge in response to life expectancy inequality is responsible for this differential effect. Progressive pension systems are generally associated with pension wedges that rise with income. However, as the life expectancy of low earners falls and that of higher earners increases, the profile of pension wedges becomes flatter (Figure 3b). As a result, the efficiency cost of taxation increases for low earners, which rationalizes lower tax rates. For higher earners, the effect is reversed.[14]

[14]Top earners are omitted from Figure 3a. For them, the marginal replacement rate drops to zero and life expectancy inequality obtains a null effect on the "$\mathcal{D}$ term", consistent with the last part of Proposition 3.



(a) Optimal marginal tax rates (b) Pension wedges

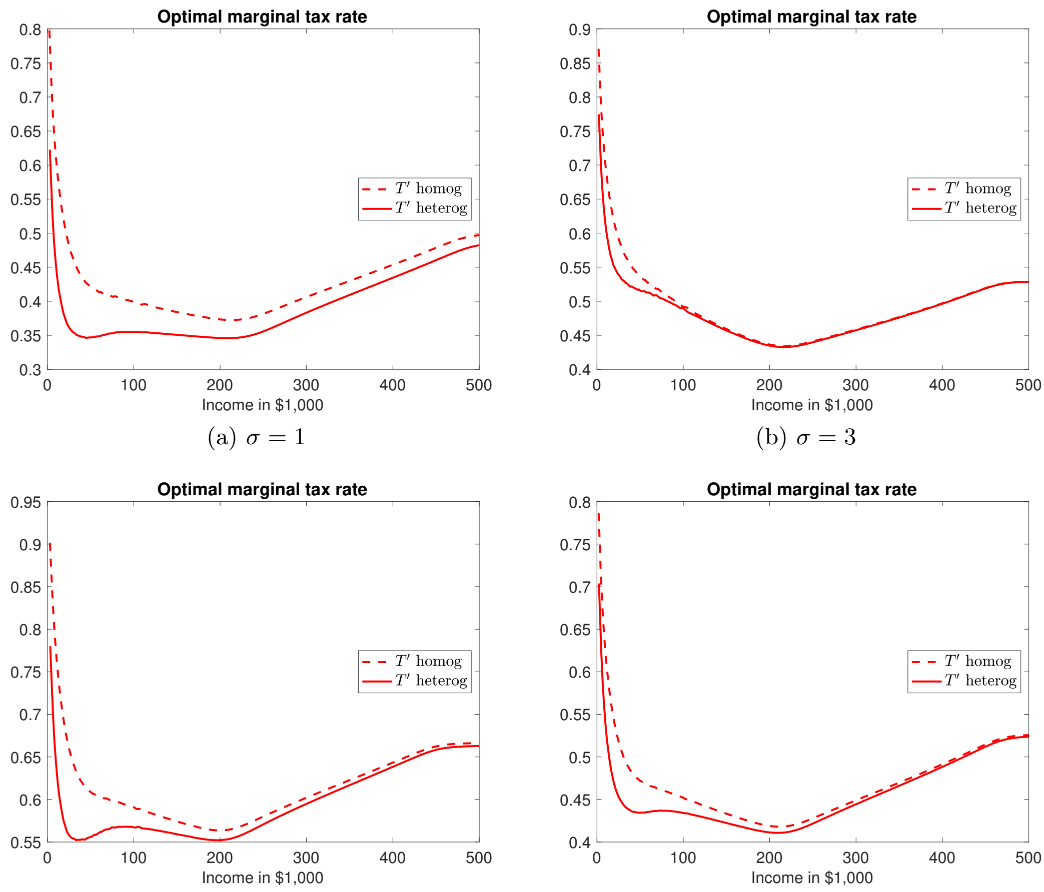
Figure 3: Impact of mechanical value of redistribution and pension wedge factor. Note: In the left panel, the dashed red line displays the optimal marginal tax rate in the homogeneous life expectancy model. The black lines display how the optimal marginal tax rate changes when the "$\mathcal{B}$ term" or "$\mathcal{D}$ term" are set to the level of the heterogeneous life expectancy model.

Overall, the policy experiment demonstrates that life expectancy inequality justifies lower income tax rates, especially at the bottom of the income distribution. In other words, the tax system becomes less redistributive as life expectancy inequality increases. The result is mainly due to the reduced mechanical value of redistribution when higher earners spread their resources over a longer lifetime and lower earners spread them over a shorter lifetime.

	Baseline	$\sigma = 1$	$\sigma = 3$	$\phi = 0.25$	$\beta = 0.95^{40}$
Mean MTR (homog)	0.563	0.496	0.600	0.666	0.534
Mean MTR (heterog)	0.500	0.396	0.554	0.593	0.482
Mean income (homog)	62,716	66,160	61,028	62,991	62,951
Mean income (heterog)	64,361	69,087	62,133	64,426	64,527

Table 1: Comparison of optimal policies between homogeneous and heterogeneous life expectancy models. The second column reports mean marginal tax rates and mean incomes in the baseline parameterization of the model. The remaining columns report the results for alternative levels of risk aversion, elasticity of labor supply and intertemporal discount factor.

Table 1 quantifies the difference between the optimal policies in the two models. In the heterogeneous life expectancy model, the mean marginal tax rate is 6 pp lower and mean income is 2.6 percent higher. Table 1 also documents the robustness of the results for alternative levels of risk aversion, elasticity of labor supply and intertemporal discount factor. The impact of heterogeneous life expectancies on optimal tax rates remains sizable, exceeding 4 pp in all parameterizations. Moreover, in all cases life expectancy inequality tends to lower the optimal marginal tax rates particularly at the bottom of the income distribution (Figure 4).



(a) $\sigma = 1$ (b) $\sigma = 3$

(c) $\phi = 0.25$ (d) $\beta = 0.95^{40}$

Figure 4: Optimal marginal tax rates for alternative parameterizations

6 Concluding remarks

This paper has introduced heterogeneous life expectancy into a Mirrleesian life-cycle model. Heterogeneous life expectancy strongly affects the mechanical value of redistribution. Moreover, it affects the optimal redistribution policy indirectly through changes in the pension wedge. Optimal marginal tax rates decrease significantly relative to the homogeneous life expectancy benchmark.

The baseline results are derived in a tractable two-period model. They generalize easily to longer time horizons when individual skills are fixed over time. Moreover, the implications for the mechanical value of redistribution extend to alternative models of the asset market as long as the pricing of future payoffs is weakly monotonic in individual life expectancy. Finally, the key role of the mechanical value of the redistribution remains valid even if the government also optimizes the pension system. However, this setting includes a degree of freedom between pension progressivity and tax progressivity.

A Appendix

A.1 Proofs

Proof of Lemma 1. The individual labor supply problem (4) generates the first-order condition

$$\theta \left(1 - T'(y) - S'(y) + q(\theta) B'(y)\right) = v'\left(\frac{y}{\theta}\right). \tag{45}$$

Assuming a linearized tax and pension system, differentiation of (45) with respect to θ yields:

$$\left(1 - T' - S' + q(\theta) B'\right) + \theta q'(\theta) B' = v''\left(\frac{y}{\theta}\right) \frac{\frac{dy}{d\theta}\theta - y}{\theta^2}.$$

Equivalently,

$$\frac{dy}{d\theta} = \frac{\theta\left(1 - T' - S' - q(\theta) B'\right) + \theta^2 q'(\theta) B'}{v''\left(\frac{y}{\theta}\right)} + \frac{y}{\theta}.$$

Equivalently,

$$\frac{dy}{d\theta} = \frac{v'\left(\frac{y}{\theta}\right) + \theta^2 q'(\theta) B'}{v''\left(\frac{y}{\theta}\right)} + \frac{y}{\theta}.$$

Hence, the elasticity of income with respect to skill is

$$\varepsilon_{y,\theta} = \frac{dy}{d\theta}\frac{\theta}{y} = 1 + \frac{v'\left(\frac{y}{\theta}\right)}{\frac{y}{\theta} v''\left(\frac{y}{\theta}\right)} + \frac{\theta^2 q'(\theta) B'}{\frac{y}{\theta} v''\left(\frac{y}{\theta}\right)} = 1 + \frac{v'(l)}{l v''(l)} + \frac{\theta^2 q'(\theta) B'}{l v''(l)}.$$

Assuming once more a linearized tax and pension system, differentiation of (45) with respect to the net-of-tax rate $1 - T'$ yields:

$$\theta = v''\left(\frac{y}{\theta}\right) \frac{1}{\theta} \frac{dy}{d\left(1 - T'\right)}.$$

Equivalently,

$$\frac{dy}{d\left(1 - T'\right)} = \frac{\theta^2}{v''\left(\frac{y}{\theta}\right)} = \frac{\theta^2}{v''(l)}.$$

Hence, the elasticity of income with respect to the net-of-tax rate is

$$\varepsilon_{y,(1-T')} = \frac{dy}{d(1-T')}\frac{(1-T')}{y} = \frac{\theta(1-T')}{lv''(l)}.$$

Using (45), I obtain

$$\theta\left(1-T'\right) = v'(l) + \theta\left(S' - q(\theta)B'\right).$$

Hence,

$$\varepsilon_{y,(1-T')} = \frac{v'(l)}{lv''(l)} + \frac{\theta(S' - q(\theta)B')}{lv''(l)}$$

Similarly, we can derive the elasticity with respect to the effective net-of-tax rate $1-\hat{T}'$. Assuming a linearized tax and pension system, differentiation of the first-order condition

$$\theta\left(1-\hat{T}'\right) = v'\left(\frac{y}{\theta}\right)$$

yields

$$\theta = v''\left(\frac{y}{\theta}\right)\frac{1}{\theta}\frac{dy}{d\left(1-\hat{T}'\right)}.$$

Hence,

$$\varepsilon_{y,(1-\hat{T}')} = \frac{dy}{d\left(1-\hat{T}'\right)}\frac{\left(1-\hat{T}'\right)}{y} = \frac{\theta^2}{v''\left(\frac{y}{\theta}\right)}\frac{\frac{v'\left(\frac{y}{\theta}\right)}{\theta}}{y} = \frac{v'(l)}{lv''(l)}.$$

Finally, I derive the elasticities $\varepsilon_{y,(1-S')}$ and $\varepsilon_{y,B'}$. Differentiation of the first-order condition (45) with respect to $1-S'$, assuming a linearized tax and pension system, yields:

$$\theta = v''\left(\frac{y}{\theta}\right)\frac{1}{\theta}\frac{dy}{d(1-S')}.$$

Hence,

$$\varepsilon_{y,(1-S')} = \frac{dy}{d(1-S')}\frac{(1-S')}{y} = \frac{v'(l)}{lv''(l)} + \frac{\theta(T' - q(\theta)B')}{lv''(l)}.$$

Differentiation of (45) with respect to B', assuming a linearized tax and pension system, yields

$$\theta q(\theta) = v''\left(\frac{y}{\theta}\right)\frac{1}{\theta}\frac{dy}{dB'}.$$

Hence,

$$\varepsilon_{y,B'} = \frac{dy}{dB'}\frac{B'}{y} = \frac{\theta^2 q(\theta)}{v''\left(\frac{y}{\theta}\right)}\frac{B'}{y} = \frac{v'(l)}{lv''(l)} - \frac{\theta\left(1 - T'(y) - S'(y)\right)}{lv''(l)}.$$

■

Proof of Proposition 1. Problem (16) is a constrained optimal control problem with state variable x and control variable l. Denote the co-state variable by $\mu(\theta)$ and the multiplier of the feasibility constraint by λ. The Hamiltonian of problem (16) is:

$$\begin{aligned} H =& \chi(\theta)\tilde{V}(x(\theta);\theta) f(\theta) + \lambda f(\theta)\left[\theta l(\theta) - x(\theta) - v(l(\theta))\right] \\ & - \lambda E + \mu(\theta)\left[\frac{l(\theta)}{\theta}v'(l(\theta)) + q'(\theta) B(\theta l(\theta))\right]. \end{aligned}$$

The maximum principle generates the following necessary conditions:

$$\begin{aligned} 0 &= \frac{\partial H}{\partial l} = \mu(\theta)\left[\frac{v'(l(\theta))}{\theta} + \frac{l(\theta)}{\theta}v''(l(\theta)) + \theta q'(\theta) B'(\theta l(\theta))\right] + \lambda f(\theta)\left[\theta - v'(l(\theta))\right], \\ \dot{\mu} &= -\frac{\partial H}{\partial x} = -\left[\chi(\theta) u'(x(\theta) - q(\theta) c_2(\theta)) - \lambda\right] f(\theta). \end{aligned}$$

where $c_2(\theta)$ is the solution of the individual's intertemporal problem (14).

Using the transversality condition, $\mu\left(\overline{\theta}\right) = 0$, I obtain:

$$\mu(\theta) = \mu(\theta) - \mu\left(\overline{\theta}\right) = -\int_{\theta}^{\overline{\theta}} \dot{\mu}(\theta)\, d\theta = \int_{\theta}^{\overline{\theta}} \left[\chi(\theta) u'(x(\theta) - q(\theta) c_2(\theta)) - \lambda\right] f(\theta)\, d\theta. \qquad (46)$$

Combining this result with the first-order condition for l yields:

$$-\lambda\left[\theta - v'\right] f = \frac{v' + lv'' + \theta^2 q' B'}{\theta}\int_{\theta}^{\overline{\theta}} \left[\chi u' - \lambda\right] f d\theta,$$

where I have dropped the arguments of all functions for brevity. Equivalently,

$$\frac{\theta - v'}{v'} = \frac{v' + lv'' + \theta^2 q' B'}{v'}\frac{1}{\theta f}\int_{\theta}^{\overline{\theta}} \left[1 - \frac{\chi u'}{\lambda}\right] f d\theta.$$

Equivalently,

$$\frac{\theta - v'}{v'} = \frac{v' + lv'' + \theta^2 q' B'}{v'}\frac{1 - F}{\theta f}\frac{\int_{\theta}^{\overline{\theta}} \left[1 - \frac{\chi u'}{\lambda}\right] f d\theta}{1 - F}.$$

Equivalently,

$$\frac{\theta - v'}{v'} = \frac{v' + lv'' + \theta^2 q' B'}{v'} \frac{1-F}{\theta f} \left(1 - \frac{\int_\theta^{\overline{\theta}} \chi u' f d\theta}{\lambda (1-F)}\right).$$

Next, I connect this optimality condition to the marginal tax rates using individual first-order conditions. The first-order condition of problem (4) is

$$1 - \hat{T}' = \frac{v'}{\theta}.$$

Hence,

$$\frac{\theta - v'}{v'} = \frac{1 - \frac{v'}{\theta}}{\frac{v'}{\theta}} = \frac{\hat{T}'}{1 - \hat{T}'}.$$

Hence,

$$\frac{\hat{T}'}{1 - \hat{T}'} = \frac{v' + lv'' + \theta^2 q' B'}{v'} \frac{1-F}{\theta f} \left(1 - \frac{\int_\theta^{\overline{\theta}} \chi u' f d\theta}{\lambda (1-F)}\right).$$

Equivalently,

$$\frac{T'}{1 - T'} = \frac{v' + lv'' + \theta^2 q' B'}{v'} \frac{1-F}{\theta f} \left(1 - \frac{\int_\theta^{\overline{\theta}} \chi u' f d\theta}{\lambda (1-F)}\right) \frac{T'}{\hat{T}'} \frac{1 - \hat{T}'}{1 - T'}.$$

The effective marginal tax rate and the statutory marginal tax rate are related as follows:

$$\hat{T}' = T' + \left(S' - qB'\right).$$

Using this relationship and the first-order condition

$$1 - \hat{T}' = \frac{v'}{\theta},$$

I obtain

$$\frac{T'}{1 - T'} = \frac{v' + lv'' + \theta^2 q' B'}{v'} \frac{1-F}{\theta f} \left(1 - \frac{\int_\theta^{\overline{\theta}} \chi u' f d\theta}{\lambda (1-F)}\right) \frac{T'}{T' + S' - qB'} \frac{\frac{v'}{\theta}}{\frac{v'}{\theta} + S' - qB'}.$$

Equivalently,

$$\frac{T'}{1-T'}=\frac{v'+lv''+\theta^2q'B'}{v'+\theta\left(S'-qB'\right)}\frac{1-F}{\theta f}\left(1-\frac{\int_{\theta}^{\overline{\theta}}\chi u'fd\theta}{\lambda\left(1-F\right)}\right)\frac{T'}{T'+S'-qB'}.$$

Next, I rewrite the optimality condition with the help of the elasticities obtained in Lemma 1. Using Equations (17) and (18), I obtain

$$\frac{v'+lv''+\theta^2q'B'}{v'+\theta\left(S'-qB'\right)}=\frac{1+\frac{v'+\theta^2q'B'}{lv''}}{\frac{v'+\theta(S'-qB')}{lv''}}=\frac{\varepsilon_{y,\theta}}{\varepsilon_{y,(1-T')}}.$$

The optimality condition for the tax system is hence

$$\frac{T'}{1-T'}=\frac{\varepsilon_{y,\theta}}{\varepsilon_{y,(1-T')}}\left(1-\frac{\int_{\theta}^{\overline{\theta}}\chi u'fd\theta}{\lambda\left(1-F\right)}\right)\frac{1-F}{\theta f}\frac{T'}{T'+S'-qB'},$$

which establishes Equation (23).

Alternatively, using the result

$$\frac{\hat{T}'}{1-\hat{T}'}=\frac{v'+lv''+\theta^2q'B'}{v'}\frac{1-F}{\theta f}\left(1-\frac{\int_{\theta}^{\overline{\theta}}\chi u'fd\theta}{\lambda\left(1-F\right)}\right)$$

and Equation (21), the optimality condition can be stated in terms of effective marginal tax rates as follows:

$$\frac{\hat{T}'}{1-\hat{T}'}=\frac{\varepsilon_{y,\theta}}{\varepsilon_{y,(1-\hat{T}')}}\left(1-\frac{\int_{\theta}^{\overline{\theta}}\chi u'fd\theta}{\lambda\left(1-F\right)}\right)\frac{1-F}{\theta f}. \tag{47}$$

Next, I state the optimality condition in terms of the income distribution. Let H be the cumulative distribution function of income y and $h=H'$ be its density function. Assuming a linearized tax and pension system, the individual first-order condition for labor supply (26) implies

$$dy=\left(\frac{v'\left(\frac{y}{\theta}\right)+\theta^2q'\left(\theta\right)B'}{v''\left(\frac{y}{\theta}\right)}+\frac{y}{\theta}\right)d\theta.$$

Using $h\left(y\right)dy=f\left(\theta\right)d\theta$, I obtain

$$\frac{1}{y\cdot h\left(y\right)}=\frac{dy}{y\cdot f\left(\theta\right)d\theta}=\frac{1}{\theta\cdot f\left(\theta\right)}\left(1+\frac{v'\left(l\right)}{lv''\left(l\right)}+\frac{\theta^2q'\left(\theta\right)B'}{lv''\left(l\right)}\right)=\frac{\varepsilon_{y,\theta}}{\theta\cdot f\left(\theta\right)}.$$

Hence, using $H(y) = F(\theta)$, I obtain

$$\frac{1 - H(y)}{y \cdot h(y)} = \frac{1 - F(\theta)}{\theta \cdot f(\theta)} \varepsilon_{y,\theta}$$

and can transform Equation (23) into Equation (24). ■

Proof of Lemma 2. By Equations (17) and (22), I obtain

$$\begin{aligned} \varepsilon_{y,\theta} &= 1 + \frac{v'(l)}{lv''(l)} + \frac{\theta^2 q'(\theta) B'}{lv''(l)}, \\ 1 + \varepsilon_{y,(1-T')}\big|_{S'-qB'=0} &= 1 + \frac{v'(l)}{lv''(l)}, \end{aligned}$$

where $\varepsilon_{y,(1-T')}\big|_{S'-qB'=0}$ represents the elasticity of taxable income with respect to the net-of-tax rate, assuming a pension wedge of zero. If $q'(\theta) B' = 0$, I obtain

$$\varepsilon_{y,\theta} = 1 + \frac{v'(l)}{lv''(l)} = 1 + \varepsilon_{y,(1-T')}\big|_{S'-qB'=0}.$$

If $q'(\theta) B' \neq 0$, the elasticity $\varepsilon_{y,\theta}$ involves the term $q'(\theta)$ which is not contained in any of the elasticities $\varepsilon_{y,(1-T')}$, $\varepsilon_{y,(1-S')}$, $\varepsilon_{y,B'}$ stated in Equations (18), (19) and (20). ■

Proof of Proposition 2. (i) The first-order condition of the intertemporal problem (14) is given by

$$-q(\theta) u'(x(\theta) - q(\theta) c_2) + \beta p(\theta) u'(c_2) = 0.$$

Using $u'(c) = c^{-\sigma}$, I obtain

$$c_2 = \frac{x(\theta)}{q(\theta) + q(\theta)^{\frac{1}{\sigma}} (\beta p(\theta))^{-\frac{1}{\sigma}}}.$$

Hence,

$$V(\theta) = u\left(x(\theta) - \frac{q(\theta) x(\theta)}{q(\theta) + q(\theta)^{\frac{1}{\sigma}} (\beta p(\theta))^{-\frac{1}{\sigma}}}\right) + \beta p(\theta) u\left(\frac{x(\theta)}{q(\theta) + q(\theta)^{\frac{1}{\sigma}} (\beta p(\theta))^{-\frac{1}{\sigma}}}\right).$$

Equivalently,

$$V(\theta) = \left[\left(\frac{q(\theta)^{\frac{1}{\sigma}} (\beta p(\theta))^{-\frac{1}{\sigma}}}{q(\theta) + q(\theta)^{\frac{1}{\sigma}} (\beta p(\theta))^{-\frac{1}{\sigma}}}\right)^{1-\sigma} + \beta p(\theta) \left(\frac{1}{q(\theta) + q(\theta)^{\frac{1}{\sigma}} (\beta p(\theta))^{-\frac{1}{\sigma}}}\right)^{1-\sigma}\right] u(x(\theta)).$$

Equivalently,

$$V(\theta) = \frac{q(\theta)^{\frac{1-\sigma}{\sigma}} (\beta p(\theta))^{-\frac{1-\sigma}{\sigma}} + \beta p(\theta)}{\left(q(\theta) + q(\theta)^{\frac{1}{\sigma}} (\beta p(\theta))^{-\frac{1}{\sigma}}\right)^{1-\sigma}} u(x(\theta)).$$

Equivalently,

$$V(\theta) = \frac{1 + \beta p(\theta) q(\theta)^{-\frac{1-\sigma}{\sigma}} (\beta p(\theta))^{\frac{1-\sigma}{\sigma}}}{\left(q(\theta)^{1-\frac{1}{\sigma}} (\beta p(\theta))^{\frac{1}{\sigma}} + 1\right)^{1-\sigma}} u(x(\theta)).$$

Equivalently,

$$V(\theta) = \frac{1 + q(\theta)^{1-\frac{1}{\sigma}} (\beta p(\theta))^{\frac{1}{\sigma}}}{\left(q(\theta)^{1-\frac{1}{\sigma}} (\beta p(\theta))^{\frac{1}{\sigma}} + 1\right)^{1-\sigma}} u(x(\theta)).$$

Equivalently,

$$V(\theta) = \left(1 + q(\theta)^{1-\frac{1}{\sigma}} (\beta p(\theta))^{\frac{1}{\sigma}}\right)^{\sigma} u(x(\theta)).$$

Equation (28) follows immediately.

(ii) The first-order condition of the intertemporal problem (14) implies

$$u'(x(\theta) - q(\theta) c_2) = \frac{\beta p(\theta)}{q(\theta)} u'(c_2).$$

Using $u'(c) = e^{-\alpha c}$, I obtain

$$e^{-\alpha(x(\theta) - q(\theta) c_2)} = \frac{\beta p(\theta)}{q(\theta)} e^{-\alpha c_2}.$$

Equivalently,

$$e^{-\alpha x(\theta) + \alpha(1+q(\theta)) c_2} = \frac{\beta p(\theta)}{q(\theta)}.$$

After taking logs and rearranging, I obtain

$$c_2 = \frac{x(\theta)}{1 + q(\theta)} + \frac{1}{\alpha(1 + q(\theta))} \ln \frac{\beta p(\theta)}{q(\theta)}.$$

Hence,

$$u'(x(\theta) - q(\theta) c_2) = \frac{\beta p(\theta)}{q(\theta)} e^{-\frac{1}{1+q(\theta)}\left(\alpha x(\theta) + \ln \frac{\beta p(\theta)}{q(\theta)}\right)}.$$

Equivalently,

$$u'(x(\theta) - q(\theta) c_2) = \frac{\beta p(\theta)}{q(\theta)} \left(\frac{q(\theta)}{\beta p(\theta)} e^{-\alpha x(\theta)}\right)^{\frac{1}{1+q(\theta)}}.$$

Equivalently,

$$u'\left(x\left(\theta\right)-q\left(\theta\right)c_{2}\right)=\left(\frac{\beta p\left(\theta\right)}{q\left(\theta\right)}\right)^{\frac{q(\theta)}{1+q(\theta)}}\left(u'\left(x\left(\theta\right)\right)\right)^{\frac{1}{1+q(\theta)}},$$

which establishes Equation (29). ■

Proof of Proposition 3. Differentiation of $\mathcal{D}$ yields

$$\frac{\partial\mathcal{D}}{\partial\eta}=\frac{T'B'\left(q_{0}\left(\theta\right)-\bar{q}\right)}{\left(T'+S'-qB'\right)^{2}}.$$

The result follows immediately. ■

Proof of Proposition 4. The first-order conditions of (33) are given by

$$\begin{aligned}q\left(\theta\right)u'\left(x\left(\theta\right)-q\left(\theta\right)c_{2}\left(\theta\right)-h\left(\theta\right)\right) &= \beta p\left(h\left(\theta\right)\right)u'\left(c_{2}\left(\theta\right)\right),\\ u'\left(x\left(\theta\right)-q\left(\theta\right)c_{2}\left(\theta\right)-h\left(\theta\right)\right) &= \beta p'\left(h\left(\theta\right)\right)u\left(c_{2}\left(\theta\right)\right).\end{aligned}$$

Division yields

$$q\left(\theta\right)\frac{u\left(c_{2}\left(\theta\right)\right)}{u'\left(c_{2}\left(\theta\right)\right)}=\frac{p\left(h\left(\theta\right)\right)}{p'\left(h\left(\theta\right)\right)}.$$

Let $c_2(\theta)$ and $q(\theta)$ be weakly (strictly) increasing in θ. Then, the first factor on the left-hand side is weakly (strictly) increasing in θ. Moreover, since u is strictly increasing in c_2 and u' is strictly decreasing in c_2, the second factor on the left-hand side is also weakly (strictly) increasing in θ. Hence, the right-hand side must also be weakly (strictly) increasing in θ. Since p/p' is assumed to be monotonic, inverting the equation shows that $h(\theta)$ is weakly (strictly) increasing in θ. Therefore, $p(h(\theta))$ is weakly (strictly) increasing in θ. ■

Proof of Proposition 5. Denote the co-state variables of Problem (42) by $\mu_1, \ldots, \mu_{R-1}$ and the multiplier of the feasibility constraint by λ. The Hamiltonian of problem (42) is:

$$\begin{aligned}H=&\chi\left(\theta\right)\tilde{V}\left(x_{1}\left(\theta\right),\ldots,x_{R-1}\left(\theta\right);\theta\right)f\left(\theta\right)\\ &+\lambda f\left(\theta\right)\sum_{t=1}^{R-1}q^{1,t}\left(\theta\right)\left[w_{t}\left(\theta\right)l_{t}\left(\theta\right)-x_{t}\left(\theta\right)-v_{t}\left(l_{t}\left(\theta\right)\right)\right]-\lambda E\\ &+\sum_{t=1}^{R-1}\mu_{t}\left(\theta\right)\left[l_{t}\left(\theta\right)\frac{w_{t}'\left(\theta\right)}{w_{t}\left(\theta\right)}v_{t}'\left(l_{t}\left(\theta\right)\right)+Q_{t}'\left(\theta\right)B_{t}\left(w_{t}\left(\theta\right)l_{t}\left(\theta\right)\right)\right].\end{aligned}$$

For each $t = 1, \ldots, R-1$, the maximum principle implies:

$$0 = \frac{\partial H}{\partial l_t} = \mu_t \left[\frac{w_t'}{w_t} v_t' + l_t \frac{w_t'}{w_t} v_t'' + w_t Q_t' B_t' \right] + \lambda f q^{1,t} \left[w_t - v_t' \right],$$
$$\dot{\mu_t} = -\frac{\partial H}{\partial x_t} = -\left[\chi \beta^{t-1} p_t u_t' - \lambda q^{1,t} \right] f,$$

where I have dropped the arguments of all functions for brevity. Using the transversality condition $\mu_t \left(\overline{\theta} \right) = 0$, I obtain:

$$\mu_t (\theta) = \mu_t (\theta) - \mu_t \left(\overline{\theta} \right) = -\int_{\theta}^{\overline{\theta}} \dot{\mu_t} (\theta) \, d\theta = \int_{\theta}^{\overline{\theta}} \left[\chi (\theta) \beta^{t-1} p_t (\theta) u_t' (\theta) - \lambda q^{1,t} (\theta) \right] f (\theta) \, d\theta.$$

Combining this result with the first-order condition for l_t yields

$$-\lambda f q^{1,t} \left[w_t - v_t' \right] = \frac{w_t' v_t' + l_t w_t' v_t'' + (w_t)^2 Q_t' B_t'}{w_t} \int_{\theta}^{\overline{\theta}} \left[\chi \beta^{t-1} p_t u_t' - \lambda q^{1,t} \right] f d\theta.$$

Equivalently,

$$w_t - v_t' = \left(v_t' + l_t v_t'' + \frac{(w_t)^2}{w_t'} Q_t' B_t' \right) \int_{\theta}^{\overline{\theta}} \left[1 - \frac{\chi \beta^{t-1} p_t u_t'}{\lambda q^{1,t}} \right] f d\theta \frac{w_t'}{w_t f}.$$

Equivalently,

$$w_t - v_t' = \left(v_t' + l_t v_t'' + \frac{(w_t)^2}{w_t'} Q_t' B_t' \right) \left(1 - \frac{\int_{\theta}^{\overline{\theta}} \chi \frac{\beta^{t-1} p_t}{q^{1,t}} u_t' f d\theta}{\lambda (1 - F)} \right) \frac{w_t' (1 - F)}{w_t f}.$$

Equivalently,

$$\frac{1 - \frac{v_t'}{w_t}}{\frac{1}{w_t}} = \left(v_t' + l_t v_t'' + \frac{(w_t)^2}{w_t'} Q_t' B_t' \right) \left(1 - \frac{\int_{\theta}^{\overline{\theta}} \chi \frac{\beta^{t-1} p_t}{q^{1,t}} u_t' f d\theta}{\lambda (1 - F)} \right) \frac{w_t' (1 - F)}{w_t f}. \tag{48}$$

Next, I rewrite this condition with the help of elasticities. The first-order condition of the individual labor supply problem in period t is

$$w_t (\theta) \left(1 - T_t' - S_t' + Q_t (\theta) B_t' \right) = v_t' \left(\frac{y_t (\theta)}{w_t (\theta)} \right). \tag{49}$$

Assuming a linearized tax and pension system, differentiation of (49) with respect to $w_t (\theta)$

yields:

$$\left(1 - T_t' - S_t' + Q_t B_t'\right) + w_t Q_t' \frac{1}{w_t'} B_t' = v_t'' \left(\frac{y_t}{w_t}\right) \frac{\frac{dy_t}{dw_t} w_t - y_t}{(w_t)^2}.$$

Equivalently,

$$\frac{dy_t}{dw_t} = \frac{w_t \left(1 - T_t' - S_t' + Q_t B_t'\right) + \frac{(w_t)^2}{w_t'} Q_t' B_t'}{v_t''\left(\frac{y_t}{w_t}\right)} + \frac{y_t}{w_t}.$$

Equivalently,

$$\frac{dy_t}{dw_t} = \frac{v_t' + \frac{(w_t)^2}{w_t'} Q_t' B_t'}{v_t''} + \frac{y_t}{w_t}.$$

Hence, the elasticity of income with respect to the wage is

$$\varepsilon_{y_t, w_t} = \frac{dy_t}{dw_t} \frac{w_t}{y_t} = 1 + \frac{v_t'}{l_t v_t''} + \frac{(w_t)^2 Q_t' B_t'}{w_t' l_t v_t''}.$$

Assuming once more a linearized tax and pension system, differentiation of (49) with respect to the net-of-tax rate $1 - T_t'$ yields:

$$w_t = v_t'' \frac{1}{w_t} \frac{dy_t}{d\left(1 - T_t'\right)}.$$

Equivalently,

$$\frac{dy_t}{d\left(1 - T_t'\right)} = \frac{(w_t)^2}{v_t''}.$$

Hence, the elasticity of income with respect to the net-of-tax rate is

$$\varepsilon_{y_t, (1 - T_t')} = \frac{dy_t}{d\left(1 - T_t'\right)} \frac{\left(1 - T_t'\right)}{y_t} = \frac{w_t \left(1 - T_t'\right)}{l_t v_t''}.$$

Using (49), I obtain

$$w_t \left(1 - T_t'\right) = v_t' + w_t \left(S_t' - Q_t B_t'\right).$$

Hence,

$$\varepsilon_{y_t, (1 - T_t')} = \frac{v_t'}{l_t v_t''} + \frac{w_t \left(S_t' - Q_t B_t'\right)}{l_t v_t''}.$$

Hence, the elasticity ratio $\varepsilon_{y_t,w_t}/\varepsilon_{y_t,(1-T_t')}$ equals

$$\frac{\varepsilon_{y_t,w_t}}{\varepsilon_{y_t,(1-T_t')}} = \frac{1+\frac{v_t'}{l_t v_t''}+\frac{(w_t)^2 Q_t' B_t'}{w_t' l_t v_t''}}{\frac{v_t'}{l_t v_t''}+\frac{w_t(S_t'-Q_t B_t')}{l_t v_t''}} = \frac{v_t'+l_t v_t''+\frac{(w_t)^2}{w_t'}Q_t' B_t'}{v_t'+w_t\left(S_t'-Q_t B_t'\right)}.$$

Using this result, the optimality condition (48) for the tax system can be written as

$$\frac{1-\frac{v_t'}{w_t}}{\frac{v_t'}{w_t}+S_t'-Q_t B_t'} = \frac{\varepsilon_{y_t,w_t}}{\varepsilon_{y_t,(1-T_t')}}\left(1-\frac{\int_\theta^{\overline{\theta}}\chi\frac{\beta^{t-1}p_t}{q^{1,t}}u_t' f d\theta}{\lambda\left(1-F\right)}\right)\frac{w_t'\left(1-F\right)}{w_t f}.$$

Using (49), I obtain

$$1-\frac{v_t'}{w_t} = T_t'+S_t'-Q_t B_t'$$

and

$$1-T_t' = \frac{v_t'}{w_t}+S_t'-Q_t B_t'.$$

Hence, the optimality condition can be written as

$$\frac{T_t'+S_t'-Q_t B_t'}{1-T_t'} = \frac{\varepsilon_{y_t,w_t}}{\varepsilon_{y_t,(1-T_t')}}\left(1-\frac{\int_\theta^{\overline{\theta}}\chi\frac{\beta^{t-1}p_t}{q^{1,t}}u_t' f d\theta}{\lambda\left(1-F\right)}\right)\frac{w_t'\left(1-F\right)}{w_t f}.$$

Equivalently,

$$\frac{T_t'}{1-T_t'} = \frac{\varepsilon_{y_t,w_t}}{\varepsilon_{y_t,(1-T_t')}}\left(1-\frac{\int_\theta^{\overline{\theta}}\chi\frac{\beta^{t-1}p_t}{q^{1,t}}u_t' f d\theta}{\lambda\left(1-F\right)}\right)\frac{w_t'\left(1-F\right)}{w_t f}\frac{T_t'}{T_t'+S_t'-Q_t B_t'}.$$

Finally, note that the wage in period t is distributed according to the cdf $\tilde{F}_t\left(w_t\left(\theta\right)\right)=F\left(\theta\right)$. Differentiation of this condition yields

$$\tilde{f}_t\left(w_t\left(\theta\right)\right)w_t'\left(\theta\right)=f\left(\theta\right).$$

Hence, the inverse Pareto coefficient of the wage distribution in period t equals

$$\frac{1-\tilde{F}_t\left(w_t\left(\theta\right)\right)}{w_t\left(\theta\right)\tilde{f}_t\left(w_t\left(\theta\right)\right)} = \frac{w_t'\left(\theta\right)\left(1-F\left(\theta\right)\right)}{w_t\left(\theta\right)f\left(\theta\right)}.$$

■

A.2 Tax perturbation approach

Using the tax perturbation approach of Saez (2001), I derive an optimal tax condition in terms of the endogenous income distribution rather than the exogenous skill distribution.

Denote the cumulative distribution function of income y by H and its density function by $h = H'$. Consider an increase of T' by $d\tau$ in a small band $(y, y + dy)$. Mechanically, this reform raises an extra tax revenue of

$$dM = dy\, d\tau\, (1 - H(y)).$$

Since the extra revenue is financed by all individuals with incomes above y, there is a mechanical welfare cost (in terms of public funds) of

$$dW = -dy\, d\tau\, (1 - H(y)) \frac{\int_y^{\overline{y}} \chi u' dH}{\lambda (1 - H(y))}.$$

Moreover, individuals in the band $(y, y + dy)$ change their income in response to the increased marginal tax rate. There are $h(y)\, dy$ individuals in the band and they change their income by $\varepsilon_{y,(1-T')} y\, d\tau / (1 - T')$. To convert the income change into a change in public funds, it is multiplied by the marginal *effective* tax rate $\hat{T}'$. Summing up, the behavioral effect changes public funds by

$$dB = dy\, d\tau\, h(y)\, \varepsilon_{y,(1-T')} y \frac{\hat{T}'}{(1 - T')}.$$

If the tax system is optimal, $dM + dW + dB = 0$ and, hence,

$$\frac{\hat{T}'}{(1 - T')} = \frac{1}{\varepsilon_{y,(1-T')}} \left(1 - \frac{\int_y^{\overline{y}} \chi u' dH}{\lambda (1 - H(y))} \right) \frac{(1 - H(y))}{y \cdot h(y)}.$$

Equivalently,

$$\frac{T'}{(1 - T')} = \frac{1}{\varepsilon_{y,(1-T')}} \left(1 - \frac{\int_y^{\overline{y}} \chi u' dH}{\lambda (1 - H(y))} \right) \frac{(1 - H(y))}{y \cdot h(y)} \frac{T'}{\hat{T}'}.$$

A.3 Shadow value of a marginal individual dollar

Consider an individual who maximizes $u(c_1) + \beta \cdot p \cdot u(c_2)$ subject to the budget constraint $c_1 + qc_2 = x$. The Lagrangian

$$L = u(c_1) + \beta p u(c_2) + \mu [x - c_1 - qc_2]$$

has the first-order conditions

$$u'(c_1) = \mu,$$

$$\beta p u'(c_2) = \mu q.$$

Hence, the shadow value of a marginal dollar is given by $\mu = u'(c_1)$.

Using the budget constraint, c_2 is given by

$$c_2 = \frac{x - c_1}{q}.$$

For CRRA utility, $u'(c) = c^{-\sigma}$, the first-order conditions imply

$$c_1^{-\sigma} = \mu = \frac{\beta p}{q} c_2^{-\sigma}.$$

Equivalently,

$$\left(\frac{\beta p}{q}\right)^{\frac{1}{\sigma}} c_1 = c_2.$$

Combined with the budget constraint, I obtain

$$\left(\frac{\beta p}{q}\right)^{\frac{1}{\sigma}} c_1 = c_2 = \frac{x - c_1}{q}.$$

Hence,

$$c_1 = \frac{x}{1 + q^{1-\frac{1}{\sigma}} (\beta p)^{\frac{1}{\sigma}}}.$$

If $\sigma \geq 1$, consumption in the first period is decreasing in the annuity price q. In that case, the shadow value of a marginal dollar is increasing in q. The opposite is true if $\sigma < 1$.

A.4 Solution of parameterized model

A.4.1 Individual problem

Using $v(l) = \gamma \left(1 + \frac{1}{\phi}\right)^{-1} l^{1+\frac{1}{\phi}}$, the first-order condition for labor supply implies

$$l = \left[\frac{\theta}{\gamma}\left(1 - T' - S' + qB'\right)\right]^{\phi}.$$

Income is given by

$$y = \theta l = \frac{\theta^{1+\phi}}{\gamma^{\phi}} \left[1 - T' - S' + qB'\right]^{\phi}.$$

Inverting this relationship, the skill can be expressed as

$$\theta = y^{\frac{1}{1+\phi}} \gamma^{\frac{\phi}{1+\phi}} \left[1 - T' - S' + qB'\right]^{-\frac{\phi}{1+\phi}}.$$

To find the marginal utility of consumption, I define the net lifetime income

$$y^n := y - T(y) - S(y) + q(\theta) B(y)$$

and solve the intertemporal problem:

$$\max_{c_2} u\left(y^n - v(l) - q(\theta) c_2\right) + \beta p(\theta) u(c_2).$$

The first-order condition

$$q(\theta)\left(y^n - v(l) - q(\theta) c_2\right)^{-\sigma} = \beta p(\theta) c_2^{-\sigma}$$

implies

$$c_2 = \frac{y^n - v(l)}{\left(\frac{q(\theta)}{\beta p(\theta)}\right)^{\frac{1}{\sigma}} + q(\theta)}.$$

Hence, the marginal utility of consumption (generated by a dollar received in the first period) is

$$u' = \left(y^n - v(l) - q(\theta) \frac{y^n - v(l)}{\left(\frac{q(\theta)}{\beta p(\theta)}\right)^{\frac{1}{\sigma}} + q(\theta)}\right)^{-\sigma}.$$

Equivalently,

$$u' = \left(1 + q(\theta)^{1-\frac{1}{\sigma}} \left(\beta p(\theta)\right)^{\frac{1}{\sigma}}\right)^{\sigma} \left[y^n - v(l)\right]^{-\sigma}.$$

A.4.2 Optimal tax

Equation (47) states the optimality condition for the effective marginal tax rate as:

$$\frac{\hat{T}'}{1-\hat{T}'} = \frac{\varepsilon_{y,\theta}}{\varepsilon_{y,(1-\hat{T}')}} \left(1 - \frac{\int_{\theta}^{\overline{\theta}} \chi u' f d\theta}{\lambda (1-F)}\right) \frac{1-F}{\theta f}.$$

The condition can be decomposed into 3 terms:

$$\begin{aligned}
\hat{\mathcal{A}}(\theta) &= \frac{\varepsilon_{y,\theta}}{\varepsilon_{y,1-\hat{T}'}} = 1 + \frac{1}{\phi} + \frac{\theta^2 q' B'}{v'}, \\
\mathcal{B}(\theta) &= 1 - \frac{\int_{\theta}^{\overline{\theta}} \chi(\theta) u'(\theta) f(\theta) d\theta}{\lambda (1 - F(\theta))}, \\
\mathcal{C}(\theta) &= \frac{1 - F(\theta)}{\theta f(\theta)}.
\end{aligned}$$

Using the transversality condition (46), I obtain:

$$\mu(\theta) = \int_{\theta}^{\overline{\theta}} \left[\chi(\theta) u'(x(\theta) - q(\theta) c_2(\theta)) - \lambda\right] f(\theta) d\theta.$$

Due to the free initial state, I obtain the transversality condition $\mu(\underline{\theta}) = 0$ and hence

$$\lambda = \int_{\underline{\theta}}^{\overline{\theta}} \chi(\theta) u'(\theta) f(\theta) d\theta.$$

Alternatively, using Equation (23) the optimality condition can be expressed in terms of the statutory marginal tax rates as

$$\frac{T'}{1-T'} = \frac{\varepsilon_{y,\theta}}{\varepsilon_{y,(1-T')}} \left(1 - \frac{\int_{\theta}^{\overline{\theta}} \chi u' dF}{\lambda (1-F)}\right) \frac{1-F}{\theta f} \frac{T'}{T' + S' - qB'}.$$

The right-hand side can be decomposed into 4 factors $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$ defined as follows:

$$\begin{aligned}
\mathcal{A}(\theta) &= \frac{\varepsilon_{y,\theta}}{\varepsilon_{y,1-T'}} = \frac{1 + \phi + \frac{\theta^2 q' B'}{l v''}}{\phi + \frac{\theta(S' - qB')}{l v''}}, \\
\mathcal{D}(\theta) &= \frac{T'}{T' + S' - qB'} = \frac{T'}{\hat{T}'},
\end{aligned}$$

with $\mathcal{B}(\theta)$ and $\mathcal{C}(\theta)$ defined as above.

References

Abraham, A., P. Brendler, and E. Carceles (2023): "Optimal Income Redistribution," Mimeo. 1

Adams, P., M. D. Hurd, D. McFadden, A. Merrill, and T. Ribeiro (2003): "Healthy, wealthy, and wise? Tests for direct causal paths between health and socioeconomic status," *Journal of Econometrics*, 112, 3–56, analysis of data on health: 2. 1

Bagchi, S. (2019): "Differential mortality and the progressivity of social security," *Journal of Public Economics*, 177, 104044. 1

Banerjee, A., E. Duflo, G. Postel-Vinay, and T. Watts (2010): "Long-Run Health Impacts of Income Shocks: Wine and Phylloxera in Nineteenth-Century France," *Review of Economics and Statistics*, 92, 714–728. 1

Benabou, R. (2002): "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica*, 70, 481–517. 5.1

Benartzi, S., A. Previtero, and R. H. Thaler (2011): "Annuitization puzzles," *Journal of Economic Perspectives*, 25, 143–164. 4

Bommier, A., M.-L. Leroux, and J.-M. Lozachmeur (2011a): "Differential mortality and social security," *Canadian Journal of Economics/Revue canadienne d'économique*, 44, 273–289. 1

——— (2011b): "On the public economics of annuities with differential mortality," *Journal of Public Economics*, 95, 612–623. 1

Cesarini, D., E. Lindqvist, R. Östling, and B. Wallace (2016): "Wealth, Health, and Child Development: Evidence from Administrative Data on Swedish Lottery Players," *Quarterly Journal of Economics*, 131, 687–738. 1

Chetty, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1, 451–488. 3

Chetty, R., M. Stepner, S. Abraham, S. Lin, B. Scuderi, N. Turner, A. Bergeron, and D. Cutler (2016): "The Association Between Income and Life Expectancy in the

United States, 2001-2014," *Journal of the American Medical Association*, 315, 1750–1766. 1, 5, 5.1

Cremer, H., J.-M. Lozachmeur, and P. Pestieau (2004): "Social security, retirement age and optimal income taxation," *Journal of Public Economics*, 88, 2259–2281, trans-Atlantic Public Economics Seminar. 2, 1

Cutler, D., A. Deaton, and A. Lleras-Muney (2006): "The determinants of mortality," *Journal of Economic Perspectives*, 20, 97–120. 1

Cutler, D. M., A. Lleras-Muney, and T. Vogl (2011): "Socioeconomic status and health: dimensions and mechanisms," in *The Oxford Handbook of Health Economics*, ed. by S. Glied and P. C. Smith, Oxford University Press. 1

Diamond, P. and E. Saez (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25, 165–90. 1, 5.1

Diamond, P. and J. Spinnewijn (2011): "Capital Income Taxes with Heterogeneous Discount Rates," *American Economic Journal: Economic Policy*, 3, 52–76. 1

Diamond, P. A. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 83–95. 1, 2.4, 3, 3.1

Erixson, O. (2017): "Health responses to a wealth shock: evidence from a Swedish tax reform," *Journal of Population Economics*, 30, 1281–1336. 1

European Commission (2013): *Health inequalities in the EU – Final report of a consortium - Consortium lead: Sir Michael Marmot*, European Commission Directorate-General for Health and Consumers. 1

Farhi, E. and I. Werning (2013): "Insurance and taxation over the life cycle," *Review of Economic Studies*, 80, 596–635. 6

Frijters, P., J. P. Haisken-DeNew, and M. A. Shields (2005): "The causal effect of income on health: Evidence from German reunification," *Journal of Health Economics*, 24, 997–1017. 1

Gerdtham, U.-G., P. Lundborg, C. H. Lyttkens, and P. Nystedt (2016): "Do Education and Income Really Explain Inequalities in Health? Applying a Twin Design," *The Scandinavian Journal of Economics*, 118, 25–48. 1

Goldman, N. (2001): "Social inequalities in health: disentangling the underlying mechanisms," *Annals of the New York Academy of Sciences*, 954, 118–139. 1

Golosov, M., M. Troshkin, and A. Tsyvinski (2016): "Redistribution and social insurance," *American Economic Review*, 106, 359–386. 6

Grochulski, B. and N. Kocherlakota (2010): "Nonseparable preferences and optimal social security systems," *Journal of Economic Theory*, 145, 2055 – 2077. 6

Heathcote, J., K. Storesletten, and G. L. Violante (2014): "Optimal Tax Progressivity: An Analytical Framework," IFS Working Papers W14/27, Institute for Fiscal Studies. 5.1

——— (2017): "Optimal Tax Progressivity: An Analytical Framework," *Quarterly Journal of Economics*, 132, 1693–1754. 5.1, 12

Hosseini, R. and A. Shourideh (2019): "Retirement financing: An optimal reform approach," *Econometrica*, 87, 1205–1265. 2, 1, 7

IRS (2022): *Statistics of Income—2020 Individual Income Tax Returns*, 1304 (Rev. 11-2022), Internal Revenue Service. 5.1

Jones, J. B. and Y. Li (2023): "Social Security reform with heterogeneous mortality," *Review of Economic Dynamics*, 48, 320–344. 1

Koehne, S. (2018): "On the Taxation of Durable Goods," *International Economic Review*, 59, 825–857. 6

Koehne, S. and M. Kuhn (2015): "Optimal taxation in a habit formation economy," *Journal of Public Economics*, 122, 31–39. 6

Koehne, S. and D. Sachs (2022): "Pareto-improving reforms of tax deductions," *European Economic Review*, 148. 5.1

Lampert, T., J. Hoebel, and L. E. Kroll (2019): “Soziale Unterschiede in der Mortalität und Lebenserwartung in Deutschland - Aktuelle Situation und Trends,” *Journal of Health Monitoring*, 4, 3–15. 1

Leroux, M.-L. and G. Ponthiere (2013): “Utilitarianism and unequal longevities: A remedy?” *Economic Modelling*, 30, 888–899. 1

Mackenbach, J. P., I. Stirbu, A.-J. R. Roskam, M. M. Schaap, G. Menvielle, M. Leinsalu, and A. E. Kunst (2008): “Socioeconomic inequalities in health in 22 European countries,” *New England Journal of Medicine*, 358, 2468–2481. 1

Mankiw, N. G., M. Weinzierl, and D. Yagan (2009): “Optimal Taxation in Theory and Practice,” *Journal of Economic Perspectives*, 23, 147–74. 1, 5.2

Meer, J., D. L. Miller, and H. S. Rosen (2003): “Exploring the health–wealth nexus,” *Journal of Health Economics*, 22, 713–730. 1

Michau, J.-B. (2014): “Optimal redistribution: A life-cycle perspective,” *Journal of Public Economics*, 111, 1–16. 1

Mirrlees, J. A. (1971): “An Exploration in the Theory of Optimum Income Taxation,” *Review of Economic Studies*, 38, 175–208. 1, 2, 2.4

——— (1976): “Optimal tax theory: A synthesis,” *Journal of Public Economics*, 6, 327–358. 2.4

Pashchenko, S., P. Porapakkarm, and Y. Jang (2023): “Mortality Regressivity and Pension Design,” Available at SSRN 4509697. 1

Piketty, T. and E. Saez (2013): “Chapter 7 - Optimal Labor Income Taxation,” in *Handbook of Public Economics*, ed. by A. J. Auerbach, R. Chetty, M. Feldstein, and E. Saez, Elsevier, vol. 5 of *Handbook of Public Economics*, 391–474. 1

Sachs, D., A. Tsyvinski, and N. Werquin (2020): “Nonlinear tax incidence and optimal taxation in general equilibrium,” *Econometrica*, 88, 469–493. 5.1

Saez, E. (2001): “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68, 205–29. 1, 2.4, 8, 3, 5.1, A.2

Sheshinski, E. (2008): *The Economic Theory of Annuities*, Princeton University Press. 1

Shourideh, A. and M. Troshkin (2017): "Incentives and efficiency of pension systems," Mimeo. 2, 1

SSA (2022): *Social Security Bulletin: Annual statistical supplement*, 13-11700, Social Security Administration. 5.1

Stantcheva, S. (2020): "Dynamic Taxation," *Annual Review of Economics*, 12, 801–831. 6