Rising Earnings Inequality and Optimal Income Tax And Social Security Policies[⋆,⋆⋆]

Pavel Brendler

University of Bonn

March 16, 2022

Abstract

Cross-sectional earnings inequality has risen sharply since the late 1970s in the United States. I ask: How should the government optimally respond to this development? I set up a rich quantitative model in which a Ramsey government optimally decides on income taxation and Social Security and is able to discriminate agents by their age and education. I find that the optimal income tax and Social Security system induces a welfare gain to U.S. households equal to 1.2% in consumption equivalent terms compared to the status quo. Quantitatively, three factors exert the most pronounced impact on the optimal solution: 1) a larger dispersion in initial skills and innate abilities between college graduates and non-college graduates, 2) a higher variance in the education-specific fixed effects, and 3) a widening gap in education-specific mortality rates.

Keywords: Optimal tax, Income taxation, Public pension program, Earnings inequality, Idiosyncratic risk

JEL: D3, E6, H2, H3

[⋆]Institute for macroeconomics, Adenaueralle 24-42, 53113 Bonn, Germany.

[⋆⋆]I benefited a lot from insightful discussions with Pawel Doligalski, Emanuel Hansen, Thomas Hintermaier, Moritz Kuhn, and Benjamin Larin. I also received very valuable feedback from Arpad Abraham, Marco Bassetto, Ctirad Slavik, Fatih Guvenen, Jonathan Heathcote, Urban Jermann, Joachim Jungherr, Sagiri Kitao, Winfried Koeniger, Tim Krieger, Dirk Krueger, Patrick Macnamara, Alisdair McKay, Christian Moser, Roberto Pancrazi, Facundo Piguillem, Dominik Sachs, Almuth Scholl, Kjetil Storesletten, Arnau Valladares-Esteban as well as the seminar participants at several seminars and conferences. I thank Oliver Freyermuth at the Physics Department of the University of Bonn for his excellent IT support. As always, all mistakes are mine.

Email address: `pavel.brendler@uni-bonn.de` (Pavel Brendler)

1. Introduction

Cross-sectional earnings inequality has risen sharply since the late 1970s in the United States.[1] Recent literature has analyzed how this development has affected the optimal progressivity of the *overall* tax-and-transfer system.[2] However, our understanding of how inequality affects *individual* government programs and, more importantly, how these programs may complement each other to mitigate the adverse effects of inequality on households is still very limited.

This paper focuses on income taxation and Social Security. The analysis of these programs raises interest for two reasons. First, these are large programs that play vital roles for many U.S. households.[3] Second, both policies target different population groups and have distinct institutional designs. In particular, income taxation redistributes incomes based on the individual's *current* economic conditions, while pension benefits depend on the worker's earnings *history*. These two features lead to non-trivial distributional conflicts between and within generations over insurance provision and redistribution, which the policymaker must trade off.

To analyze the impact of rising earnings inequality on the optimal composition of both programs, I extend the general equilibrium overlapping generations model in the style of Huggett (1996). The model features a Ramsey government that optimally chooses income taxation and Social Security. The government has access to three instruments. First, it can set a linear income tax rate denoted by τ_I to finance lump-sum transfers paid to workers and retirees. Although the income tax rate is linear, the implied *effective* income tax rates are progressive in the model.[4] I follow the literature and allow the government to discriminate between the sources of income such that capital income is taxed separately at an exogenous rate.

The other two policy instruments control the replacement rate schedule in the public pension system. The schedule determines the individual's pension benefit as a function of their average lifetime earnings. I approximate the statutory schedule in the data using a

[1] Heathcote et al. (2010a) and Heathcote et al. (2010b) provide extensive empirical evidence on the trends in income inequality in the United States.

[2] Among others, Lockwood and Weinzierl (2016), Heathcote et al. (2020), and Wu (2021) have made significant contributions in this field.

[3] While income taxation constitutes the largest source of tax revenues, the public pension system also Social Security's payouts amount to 30% of total government outlays. According to Hosseini and Shourideh (2019), Social Security benefits constitute as much as 40% of older people's total income.

[4] Numerous studies in the optimal income taxation literature also assume a linear income tax, see Corbae et al. (2009), Chang et al. (2018), among others.

flexible non-linear function with two variables (α_1, α_2). The policy variable α_1 is equal to the replacement rate of an agent whose average lifetime earnings at retirement are exactly equal to the economy-wide average taxable earnings. The policy variable α_2 controls the slope of the replacement rate schedule, i.e., pension system progressivity. Hence, the government is able to control the progressivity of income taxes and Social Security. Since the programs are assumed to run separate budgets, a given choice of (α_1, α_2) pins down the payroll tax rate in equilibrium. Both income and Social Security taxation have efficiency costs. At the intensive margin, workers reduce their labor supply and spend less time acquiring new skills. At the extensive margin, they may drop out of the labor force altogether by choosing to retire early.

I follow the related literature and assume that policies are chosen once-and-for-all. At the same time, I relax the assumption that the government maximizes the welfare of *newborn* agents. Instead, I allow the government to care about all agents who are alive at the time when the policy is implemented. The reasoning behind this departure is because newborn agents in the calibrated model prefer a counterfactually small average replacement rate α_1, regardless of their initial characteristics.[5] Furthermore, I allow the government to discriminate between alive agents by their age and education. I calibrate the model to the 1970s and identify a joint distribution of age- and education-specific Pareto weights (under parametric restrictions), such that the optimal and the calibrated policies $(\tau_I, \alpha_1, \alpha_2)$ coincide. In this regard, I merge two approaches in the literature that have introduced education-specific weights to explain the observed income tax policy, on the one hand, and the age-specific Pareto weights to account for the actual average replacement rate, on the other hand.[6]

I find that the Pareto weight distribution is skewed toward younger, less educated households in the 1970s. In particular, a newborn agent receives a 13.5 times larger weight than an agent at the normal retirement age of the same education. This finding is in stark contrast to the optimal income taxation literature, in which the age distribution of Pareto weights locates its entire mass on newborn agents. At the same time, non-college graduates receive a 2.7 times larger weight than college graduates of the same age. After identifying the Pareto weight distribution, I re-calibrate the model to the 2010s, augmenting rising earnings inequality with other significant developments in the U.S. economy since the 1970s. In particular, I adjust education-specific survival probability rates such that the model generates a widening gap in longevity between college graduates and non-college graduates, consistent with the

[5] This issue does not arise in the literature on optimal income taxation because it either takes Social Security as given or treats it implicitly as part of a broad tax-and-transfer system.

[6] See the next section for a literature review.

data.

My thought experiment then proceeds as follows. Suppose that the U.S. economy is in a steady-state equilibrium in the 2010s, with the income tax and Social Security policies fixed at the 1970s levels. By construction of the Pareto weights, these policies were optimal under the 1970s calibration, but they will become suboptimal in the 2010s because the fundamentals have changed. So I ask: What choice of $(\tau_I, \alpha_1, \alpha_2)$ should the current U.S. government make to address the economic and demographic challenges?

I find that the optimal policy response induces a welfare gain equal to 1.2% in consumption equivalent terms compared to the counterfactual scenario when the government remains committed to the 1970s policy despite the changed fundamentals. To give intuition behind the optimal policy response, consider first an exogenous increase in the degree of complementarity between educated and uneducated workers, which is one of the major driving forces behind rising earnings inequality in the model. As skills become more complementary in production, the wage premium increases. This leads to a more dispersed distribution of cross-sectional earnings, which results in a less equal distribution of average lifetime earnings at retirement. Since the vast majority of households at the bottom of the average lifetime earnings distribution are non-college graduates, who the government cares relatively more about, pension system progressivity α_2 becomes an effective instrument to target those agents. Since the government can target pensions precisely at poor households, it optimally chooses to lower the average replacement rate α_1, thus reducing the total Social Security tax burden on workers. Finally, the government complements its choice by raising the income tax level τ_I. This measure allows the government to redistribute incomes towards non-college graduates already during the early stages of their lives.

In contrast to the rising degree of complementarity between college and non-college graduates, an exogenous increase in the supply of educated workers has an opposite effect on the optimal outcome. A higher supply of college graduates reduces the wage premium in general equilibrium, compared to the 1970s. Since the wage distribution becomes more compressed, the share of educated agents at the bottom of the lifetime earnings distribution increases. So the government optimally chooses to maintain the average replacement rate level to direct resources toward non-college graduates. Since this measure exerts upward pressure on the Social Security tax, the government optimally reduces the income tax.

Finally, I exploit the model to gain some understanding for why the actual policies deviate from the model-based optimal outcome. Based on my calibration, both Social Security and income taxation changed insignificantly during 1970–2010. To account for the conservative response of the U.S. government, I compute a new distribution of Pareto weights such that the existing income tax and Social Security policies solve the government's optimization problem. I find that Pareto weights must shift towards younger and more educated households during 1970–2010. The shift in government preferences has a welfare cost to U.S. households. Compared to the optimal solution under the 1970s weights, the existing policy causes a welfare loss of 0.6% in consumption equivalent terms.

Below I proceed as follows. Section 2 highlights the paper's contribution to the closely related literature. Section 3 lays out the quantitative life-cycle model that I will employ in the optimal policy analysis. The main quantitative experiment is explained in Section 4. Section 5 describes model calibration. The paper's findings are presented in Section 6. Section 7 concludes the paper.

2. Relation to the literature

This paper relates to three strands in the literature. The first strand applies the inverse optimum approach to recover social preferences for income redistribution. Several studies have inverted government preferences by looking at the progressivity of the tax-and-transfer policy in the United States. This literature's appealing feature is that the notion of the tax-and-transfer system comprises a very broad range of redistributive programs, such as income taxation, social security, Medicare, child support, etc. Tsujiyama and Heathcote (2015) and Chang et al. (2018) study the progressivity of the tax-and-transfer system in the cross-sectional U.S. data and identify a relatively high Pareto weight attached to more productive agents. Wu (2021) and Heathcote et al. (2020) analyze the *time trends* in social preferences for redistribution in a framework, where a Ramsey government chooses income tax progressivity once-and-for-all to maximize the welfare of newborn agents. Wu (2021) estimates that the progressivity of the overall tax-and-transfer system declined during 1978–2016.[7] He finds that a larger Pareto weight on high-ability households rationalizes a significant portion of the drop in income tax progressivity.[8] Heathcote et al. (2020) challenge this finding. Their empirical investigation concludes that income tax progressivity has remained constant between 1980 and 2016. In their quantitative model, a *utilitarian* solution is consistent with the data. The authors show that a dynamic distortion to skill investment is crucial for this outcome.

I depart from these studies in two respects. First, I show explicitly how income tax and Social Security policies optimally interact to mitigate the adverse effects of inequality on

[7]The progressivity measure in his analysis excludes Social Security benefits.

[8]Lockwood and Weinzierl (2016) apply a sufficient statistics approach and obtain a similar finding that the average marginal social welfare weights for high-income households increased during 1980–1990.

households. Second, I relax the assumption that the government maximizes the welfare of newborn agents and allow the government to discriminate agents by their age and education. This departure allows the model to account for the observed income tax *and* Social Security policies.

The second strand in the literature asks a *normative* question of how an income tax or a public pension system should optimally look like. Chang et al. (2021) set up an incomplete markets economy and insightfully show that the optimal solution for a linear income tax is consistent with the data once one introduces a sufficient degree of heterogeneity in the workers' ability levels coupled with the ability-dependent weights. Their channel is also present in my work, since my calibration implies that a substantial portion of increased residual wage dispersion during 1970–2010 is driven by the ex-ante heterogeneity in abilities, consistent with the empirical evidence by Guvenen et al. (2017). Hosseini and Shourideh (2019) study Pareto optimal public pension reforms with heterogeneous mortality rates and time preferences. Ndiaye (2020) examines lifecycle taxation with endogenous retirement. Moser and Olea de Souza e Silva (2019) construct a model in which Social Security and income taxation arise as the decentralization of an optimal policy that trades off savings adequacy (due to present bias heterogeneity) with income redistribution (due to ability heterogeneity).[9] As opposed to my work, these studies primarily focus on the decentralization of the first-best policies. In a related study, Huggett and Parra (2010) conduct a Social Security reform by optimally choosing the parameters of the existing pension benefit function. While they find a small welfare gain in the model's version without idiosyncratic labor productivity risk, the welfare gain becomes substantial once they add persistent and temporary earnings shocks into the model. The novel feature of my work is to analyze the implications of the rising trend in cross-sectional earnings inequality on the optimal Social Security *and* income tax reform.[10] In Brendler (2020), I restrict the Ramsey government to set the average replacement rate (policy variable α_1) only. Since the model does not feature endogenous human capital accumulation, the government finds it optimal to increase replacement rates in response to higher earnings inequality. In the current paper, I provide a richer answer

[9] Contrary to Tsujiyama and Heathcote (2015), the authors recover a hump-shaped distribution of Pareto weights that puts relatively more weight on the second ability quartile mostly because they match the retirement savings system (which is fairly progressive) jointly with the income tax system (which is fairly regressive). Qualitatively, this finding is similar to Jacobs et al. (2017), who apply the revealed preference approach to the Netherlands' income tax policy.

[10] There is a set of other important studies. Fehr and Habermann (2008) and Fehr et al. (2013) analyze the optimal progressivity of the German pension system and show that progressivity matters quantitatively for households' welfare.

to the question of how the government should respond to inequality by enriching the set of available instruments.

The third strand of related work has analyzed the macroeconomic and welfare consequences of different retirement financing reforms. Conesa and Krueger (1999), Huggett and Ventura (1999), Nishiyama and Smetters (2007), Fuster et al. (2007), Kitao (2014), McGrattan and Prescott (2017), and Nishiyama (2019) have made significant progress in this field. This literature has studied exogenous and arguably politically infeasible reforms (e.g., complete elimination of Social Security). By contrast, my paper rationalizes the existing Social Security system. As argued by Lockwood and Weinzierl (2016) and Stantcheva (2016), the distribution of Pareto weights captures feasibility constraints imposed on the political process. Hence, the weights can be applied in policy analyses to reduce the set of all economically feasible proposals to those that are also implementable from the political standpoint.

3. The model

In this section, I lay out the quantitative life-cycle model that I will employ in the optimal policy analysis.[11]

3.1. Overview

The economy is populated by overlapping generations of agents. Each period a continuum of agents is born. Population grows at an exogenous rate n. Agents enter the economy as workers at age $j = 1$. Agents may live up to a maximum of J periods but may die earlier due to stochastic mortality. At any point in time, the total population size is normalized to 1.[12]

The financial markets are incomplete in that there is no insurance available against idiosyncratic mortality and labor productivity shocks. Agents enter the economy without any assets but can self-insure against these shocks by accumulating two assets: shares in a representative firm and government bonds. Both assets bear no risk and generate a risk-free one-period rate of return equal to r_t. Since agents are indifferent between either asset, I denote the agent's asset holdings using a single variable a. Borrowing is ruled out.

[11] When describing the model, I index the aggregate variables, the prices, the Social Security policy, and the income tax policy by time. It is understood that all the individual variables (e.g., hours worked, consumption) depend on time and the individual characteristics such as age, educational level, etc. To simplify notation, however, I will index the individual variables only by those characteristics that appear necessary in a given context.

[12] This assumption eliminates exogenous growth in the aggregate variables and allows me to define a steady-state equilibrium.

At age 1 and before any decision is made, agents draw an educational level z from the invariant distribution Π_z. Agents can be either college graduates ($z = H$) or non-college graduates ($z = L$). The agent's educational level remains constant throughout life. In the quantitative experiment, I will study the implications of an exogenous increase in the supply of college graduates. I will use Π_z for that purpose.

The educational level plays two important roles in the model. First, it affects the survival rates over the lifecycle. Denote by $\psi_{z,j}$ the probability that the type z agent survives up to age $j+1$, conditional on surviving up to age j. I assume that $\psi_{z,j} \in (0,1)$ for $j = 1, ..., J-1$ and $\psi_{z,J} = 0$ (the maximum age is J). Second, the educational level affects the deterministic and the stochastic components of the worker's labor productivity, as I explain in the next section.

3.2. Households

3.2.1. Preferences

Preferences are assumed to be time-separable, with a constant discount factor β. Agents are endowed with one unit of productive time each period, which they allocate among three activities: leisure ℓ, learning s, and work l. The utility from consumption and leisure in each period is given by the function $u(c, \ell)$. Agents also derive warm-glow utility from bequeathing their wealth. The utility from leaving a bequest a is denoted by $\phi(a)$. Agents face two sources of uncertainty: the survival risk captured by the conditional survival probability rates $\psi_{z,j}$ and the idiosyncratic labor productivity risk that I describe below.

3.2.2. Labor productivity

The worker's wage per unit of time worked is given by

$$w_{t,z} h_{j,z} \nu_z y_{j,z}. \tag{1}$$

The deterministic component of the agent's wage is governed by the wage $w_{t,z}$ (determined in equilibrium), the skill level $h_{j,z}$, and the fixed effect ν_z. The stochastic component of the wage is given by the idiosyncratic productivity shock $y_{j,z}$. Observe that each component of the worker's wage is education-specific.

In a closely related study, Heathcote et al. (2020) find that a dynamic distortion to skill investment is quantitatively important for the optimal policy analysis. To account for this margin in the model, I introduce an endogenous human capital accumulation channel. In particular, each educational level z is associated with a fixed initial amount of skills denoted by $h_{1,z}$ and an immutable learning ability denoted by θ_z. Upon entering the labor market with an initial skill endowment, the agent may choose to acquire new skills. Denote by s the fraction of the unitary time endowment spent on learning. Then, skill accumulation evolves according to

$$h_{j+1,z} = (1 - \delta^h)h_{j,z} + \theta_z(h_{j,z}s)^{\gamma^h}. \tag{2}$$

In this equation, $h_{j,z}s$ is the total effective time the agent spends acquiring new skills. The speed at which the agent builds up human capital is governed by parameter $\gamma^h \in (0,1)$. Skills depreciate at a constant rate δ^h. Since s will be a choice variable, Social Security and income taxation will exert a distortionary impact on the worker's incentives to acquire skills.

The last deterministic component of the worker's wage is the fixed effect ν_z. I assume that the logarithm of the fixed effect is a standard-normally distributed random variable whose variance, denoted by $\sigma^2_{\nu,z}$, depends on the agent's educational level. As I will show in the calibration section, a substantial portion of the rise in residual wage dispersion will be generated through parameter $\sigma^2_{\nu,z}$.

Finally, the stochastic part of the wage is driven by the idiosyncratic productivity shock $y_{j,z}$ that follows an AR(1) process. More specifically, the shock of agent i with educational level z at age $j \geqslant 2$ evolves according to

$$\log(y^i_{j,z}) \;\; = \;\; \rho_z \log(y^i_{j-1,z}) + \epsilon^i_{j,z} \text{ with } y^i_{1,z} = 1 \text{ and } \epsilon^i_{j,z} \sim \mathcal{N}(0, \sigma^2_{\epsilon,z}). \tag{3}$$

I assume away any initial variation in the shock by imposing the condition that $y^i_{1,z} = 1$. In each subsequent period, the agent is hit by a shock of size $\epsilon^i_{j,z}$, which is a standard normally distributed variable with variance $\sigma^2_{\epsilon,z}$ that depends on the agent's educational level. The persistence parameter ρ_z, also education-specific, determines the extent to which the shock is lasting.[13] I assume that $y_{j,z}$ follows a Markov chain with states $y \in \mathcal{Y}_{j,z}$ and transitions $\pi_{j,z}(y_{j+1,z} \mid y_{j,z})$.

3.3. Production technology

I assume that college graduate workers and non-college graduate workers are imperfectly substitutable in production; however, workers of the same education are perfectly substitutable across different ages and skills h.

Let $N_{t,z}$ denote the aggregate labor of type z at time t measured in efficiency units. Then,

[13] As I show in the calibration section, the conditional variance of $y_{j,z}$ increases with the agent's age.

the total effective labor at time t is given by

$$N_t = \left(N_{t,L}^{\rho} + N_{t,H}^{\rho}\right)^{\frac{1}{\rho}}, \tag{4}$$

where $1/(1-\rho)$ is the elasticity of substitution between college graduate workers and non-college graduate workers.

A representative firm produces the final output good according to the production function

$$Y_t = ZK_t^{\varpi} N_t^{1-\varpi}, \tag{5}$$

where K_t denotes the aggregate capital stock, $\varpi \in (0,1)$ measures the elasticity of output with respect to the input of capital services, and Z is a scaling factor. The output can be consumed or invested in capital.

The firm rents capital and hires labor on competitive spot markets at prices $r_t + \delta$ and $w_{t,z}$, where r_t is the rental price of capital, δ – the depreciation rate of capital, and $w_{t,z}$ – the wage per effective unit of labor of education type z. The interest rate and the wage rate follow from the firm's profit maximization and are given by the standard optimality conditions

$$w_{t,z} = (1-\varpi)Z\left(K_t/N_t\right)^{\varpi}\left(N_t/N_{t,z}\right)^{1-\rho}. \tag{6}$$

$$r_t = \varpi Z\left(K_t/N_t\right)^{\varpi-1}. \tag{7}$$

Then, firms pay a wage premium to workers with a college education given by[14]

$$\frac{w_{t,H}}{w_{t,L}} = \left(\frac{N_{t,L}}{N_{t,H}}\right)^{1-\rho}. \tag{8}$$

When $\rho = 1$, college graduates and non-college graduates are perfect substitutes and receive the same wage in equilibrium. As long as $\rho < 1$, both types of workers are imperfectly substitutable in the production process. In the quantitative experiment, I will follow Heathcote et al. (2017) and Abbott et al. (2019) and attribute a portion of the increase in the wage premium during 1970–2010 to a drop in ρ, implicitly assuming that jobs executed

[14]As one can see from (1), a j-year-old college graduate receives, on average, a wage premium over a non-college graduate of the same age given by $(w_{H,t}/w_{L,t}) \times (h_H/h_L)$. The first term is determined by the complementarity between two educational types in the production process, while the second term is driven by the differences in the human capital accumulation between the types.

by college graduate workers become less substitutable in production.[15]

Then, an increase in the supply of college graduates reduces the wage gap.

3.4. Government policies

The government is involved in two activities that I describe below.

3.4.1. Social Security

The government administers the Social Security program by collecting payroll contributions from workers and paying benefits to retirees. Denote the agent's pre-tax earnings by

$$e = w_{t,z} h v y l, \tag{9}$$

where l is the number of hours worked. The individual earnings taxable for the Social Security purpose are then given by

$$\tilde{e} = \min(cap, e), \tag{10}$$

where cap is a maximum taxable earnings threshold. All workers pay a proportional Social Security tax $\tau_{SS,t}$ on their taxable earnings $\tilde{e}$.

Next, I introduce an extensive margin at which income and Social Security taxation will distort the worker's labor supply. More specifically, I assume that workers become eligible for pension benefits already at age J^E before reaching the normal retirement age $J^R > J^E$. If the agent chooses to retire, she can no longer return to the labor force. At the same time, however, retirement is mandatory for all workers once they reach normal retirement age.[16] Since the timing of retirement will affect the amount of the pension benefit (early retirement is penalized through a reduced benefit), I introduce an individual state variable j^R to keep track of the age at which the agent retired. For workers, the convention is to let j^R be equal to 0.

During the working stage, the agent accumulates a history of taxable earnings that determines her pension benefit upon entering retirement. Denote by $\bar{e}_j$ the average earnings that the agent has accumulated up to age j. I will refer to $\bar{e}_j$ as the agent's *average lifetime earnings* below.

[15] Alternatively, Guvenen et al. (2014) set up a framework in which the wage premium is entirely driven by a skill-biased technological change. See Heathcote et al. (2020) for the discussion of two alternative views on technical change.

[16] Since the model is computationally intense, I do not allow workers to delay retirement beyond the normal retirement age J^R. Retirement before the early retirement age J^E is suboptimal since the agent forfeits the pension benefit.

The law of motion for the average lifetime earnings $\bar{e}_j$ is given by

$$\bar{e}_{j+1} = \begin{cases} \left[(j-1)\bar{e} + \tilde{e}\right]/j & \text{if } j^R = 0 \\ \bar{e} & \text{otherwise.} \end{cases} \tag{11}$$

The first line of the expression above refers to workers, for whom the convention was to set j^R to 0. The worker's average lifetime earnings are equal to the simple average of her past taxable earnings $\tilde{e}$. Recall that taxable earnings, defined in (10), include the portion of the worker's pre-tax earnings below the cap. Once the agent retires, in which case $j^R > 0$, her average lifetime earnings remain constant. This case is shown by the second line of (11). All workers enter the model with no prior earnings histories, i.e., $\bar{e}_1 = 0$.

Once the agent retires, she starts receiving a pension benefit $b = b(\bar{b}_t, j^R)$ every period. The benefit depends on two components: 1) the full pension amount $\bar{b}_t$ for which the agent would have qualified had she retired at the normal retirement age J^R and 2) the actual retirement age j^R. I relegate the specification of function $b(\cdot)$ to Section 5.6.1.

Define a replacement rate as the ratio of the agent's full pension benefit to this agent's average lifetime earnings. By definition, the agent's pension then reads

$$\bar{b}_t(\bar{e}; \boldsymbol{\alpha}_t) = \bar{e} \times R_t(\bar{e}, j^R; \boldsymbol{\alpha}_t). \tag{12}$$

In this equation, R_t denotes the replacement rate schedule. It depends on the agent's average lifetime earnings $\bar{e}$, the retirement age j^R, and the Social Security policy $\boldsymbol{\alpha}_t = (\alpha_{1,t}, \alpha_{2,t})$.

Next, I specify the replacement rate schedule. As I will explain in the calibration section, the statutory schedule in the data consists of three brackets with constant marginal replacement rates in each bracket. In the model, I approximate the implied schedule of average replacement rates. To reduce the number of parameters (which will be policy variables), I use a simple parametric class given by[17]

$$R_t(\bar{e}, j^R; \boldsymbol{\alpha}_t) = \begin{cases} \alpha_{1,t}\left(\bar{e}/\tilde{E}_{t-j+j^R}\right)^{\alpha_{2,t}} & \text{if } \bar{e} \geqslant \bar{e}_{\min} \\ \alpha_{1,t}\left(\bar{e}_{\min}/\tilde{E}_{t-j+j^R}\right)^{\alpha_{2,t}} & \text{otherwise.} \end{cases} \tag{13}$$

This expression denotes the replacement rate as of period t for all those agents who are j years

[17] Huggett and Parra (2010) optimize over the full set of parameters of the statutory replacement rate schedule. Their framework, however, is computationally less intense since it does not contain human capital accumulation and retirement decisions. Besides, income taxation is exogenous in their model.

old and who retired at age j^R. $\bar{e}_{\min}$ is a fixed threshold and $\tilde{E}_{t-j+j^R}$ are the economy-wide average taxable earnings at the time when the agent entered retirement.[18]

The first line of (13) refers to all retirees whose average lifetime earnings fall above the minimum threshold $\bar{e}_{\min}$. The reason why I normalize individual average lifetime earnings by the economy-wide average taxable earnings will become evident immediately. The second line shows the replacement rate for all the remaining retirees whose average lifetime earnings are below $\bar{e}_{\min}$. Observe that these agents qualify for the same replacement rate. Introducing this additional case will allow me to fit more accurately the empirical schedule of replacement rates which I will plot in the calibration section.

The replacement rate schedule depends on two variables $(\alpha_{1,t}, \alpha_{2,t})$ that the government will choose optimally as I will explain later. The variable $\alpha_{1,t} \in \mathcal{R}_+$ is equal to the replacement rate of a retiree whose average lifetime earnings are exactly the same as the economy-wide taxable earnings ($\bar{e} = \tilde{E}_{t-j+j^R}$). For the sake of brevity, I will refer to α_1 as the average replacement rate below. When $\alpha_{1,t} = 0$, the government does not pay any benefits. As the government increases $\alpha_{1,t}$, the replacement rate schedule shifts upward, thus raising benefits of all retirees. The left panel of Figure 1 illustrates this point. One can, therefore, think of $\alpha_{1,t}$ as the instrument that controls the overall size and generosity of the pension system.

The variable $\alpha_{2,t} \in \mathcal{R}$ determines the degree of pension system progressivity. When $\alpha_{2,t} < 0$, which will be the case in the calibrated model economy, replacement rates decrease in the agents' average lifetime earnings. The opposite holds when $\alpha_{2,t} > 0$. When the government decreases $\alpha_{2,t}$, which is the case shown in the right panel of Figure 1, replacement rates fall for all agents whose average lifetime earnings are above the economy-wide average taxable earnings ($\bar{e}/\tilde{E}_{t-j+j^R} > 1$), while replacement rates rise for agents with $\bar{e}/\tilde{E}_{t-j+j^R} < 1$. Note that the replacement rate of the *average* agent with $\bar{e} = \tilde{E}_{t-j+j^R}$ remains unaffected by the change in $\alpha_{2,t}$.

Given $\boldsymbol{\alpha}_t$, the Social Security tax rate $\tau_{SS,t}$ balances the government budget constraint[19]

$$\tau_{SS,t}\mu_t^W \tilde{E}_t = B_t, \tag{14}$$

where B_t denotes the aggregate pension benefits, μ_t^W – the population share of working

[18]The definitions of all aggregate variables, including $\tilde{E}_t$, are given in Section 3.6, in which I define the competitive equilibrium.

[19]I solved a version of the model in which Social Security accumulates asset reserves, consistent with the data. Since this feature did not have a significant quantitative effect on the results, I removed it from the model.

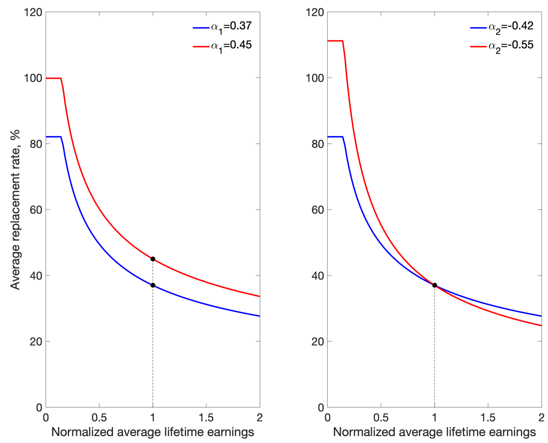


Figure 1: Replacement rate schedule.

Notes: The figure illustrates how the replacement rate schedule $R_t(\bar{e}, j^R; \boldsymbol{\alpha}_t)$, specified in (13), depends on the policy variables $\alpha_{1,t}$ (left panel) and $\alpha_{2,t}$ (right panel). Replacement rates are plotted as a function of the agent's average lifetime earnings normalized by the economy-wide taxable earnings at the time of retirement, $\bar{e}/\tilde{E}_{t-j+j^R}$. The left panel shows the effect of an increase in α_1 from 37% to 45% (α_2 is fixed at -0.42). The right panel shows the effect of decreasing α_2 from -0.42 to -0.55 (α_1 is fixed at 37%). The minimum threshold $\bar{e}_{\min}$ is set to 0.15 in both panels.

agents, and $\tilde{E}_t$ – the economy-wide average taxable earnings.[20] Hence, $\mu_t^W \tilde{E}_t$ are the total taxable earnings in the economy. Intuitively, a change in $\alpha_{1,t}$ will have a strong quantitative effect on $\tau_{SS,t}$ through the government budget constraint (14) because the entire schedule is shifted. By contrast, the impact of $\alpha_{2,t}$ on the Social Security tax rate will depend on the distribution of retirees by their normalized average lifetime earnings.

3.4.2. Income tax program

Besides Social Security, the government needs to finance an exogenous stream of spending G_t and a lump-sum income transfer T_t. The government spending is wasted in the context of the model. Its share in output denoted by $gy = G_t/Y_t$ is a parameter that I will calibrate from the data; at any point in time, the share of wasted spending in GDP remains constant. The lump-sum transfer is paid to both working agents and retirees.

To finance these expenditures, the government imposes a linear consumption tax τ_c and collects income taxes. I restrict attention to an income tax system that discriminates between

[20]Since agents choose optimally when to retire, μ_t^W is an endogenous variable. It is defined in Section 3.6.

the income source (capital versus labor income).[21] In particular, the government taxes capital interest income ra at a proportional rate τ_a.[22] The labor income is taxed at a linear rate denoted by $\tau_{I,t}$. This is the variable that the government will optimally set, jointly with the replacement rate policy $\boldsymbol{\alpha}_t$.

Recent literature on optimal income taxation has devoted considerable attention to the optimal income tax progressivity using non-linear tax functions.[23] Instead, the current paper focuses on a much less explored topic – optimal Social Security as well as the interaction between the public pension system and income taxation. To overcome the computational burden, I therefore choose a linear income tax specification but zoom onto the pension system. Besides, numerous studies in the optimal income taxation literature also assume a linear income tax, see Corbae et al. (2009), Chang et al. (2018), among others. As in their studies, the *effective* income tax rates in my model are progressive, although the income tax $\tau_{I,t}$ is linear. I will return to this point in Section 5.6.2.

Define the agent's taxable labor income as

$$\iota = e - 0.5\tau_{SS,t}\tilde{e}, \tag{15}$$

where e denotes the agent's pre-tax earnings given by (9) and $\tilde{e}$ denotes the earnings taxable for the Social Security purpose given by (10). Since the part of labor income that is paid by the employer as Social Security contributions is not subject to income taxes, a portion $0.5\tau_{SS,t}\tilde{e}$ is deducted from the worker's taxable labor income.

Besides taxing consumption, capital income, and labor income, the government issues government debt $D_t > 0$, and confiscates the wealth left by all agents deceased at the end of the previous period. The initial stock of government debt D_0 is given; the share of government debt in output, $dy = D_t/Y_t$, is a parameter to be calibrated from the data; it remains constant at any point in time.

I assume that the income tax program runs a separate budget.[24] Given $\tau_{I,t}$, the lump-sum

[21]This assumption follows Krueger and Ludwig (2016) and Wu (2021), among others, in the related literature.

[22]The capital income tax τ_a and the consumption tax τ_c are treated as parameters, which is why I drop the time subscript t.

[23]See Section 2 for the literature review.

[24]This assumption reduces the dimensionality of the optimal policy space because for a given choice $\tau_{I,t}$, the government budget constraint pins down the lump-sum transfer T_t.

transfer T_t satisfies the government budget constraint

$$G_t + T_t + (1 + r_t)D_t = \tau_{I,t}\mathcal{I}_t + \tau_a r_t A_t + \tau_c C_t + (1 + (1 - \tau_a)r_t)\Phi_t + D_{t+1}, \tag{16}$$

where A_t, C_t, $\mathcal{I}_t$, and Φ_t denote the aggregate asset holdings, consumption, taxable labor income, and bequests, respectively.

3.5. Dynamic programming problem of households

In this section, I spell out the agent's dynamic programming problems at different stages of her lifecycle. The individual state variables are given by $(z, j, j^R, \nu, y, \bar{e}, h, a)$ with education level z, age j, retirement age j^R, fixed effect ν, stochastic labor productivity shock y, average lifetime earnings $\bar{e}$, skill level h, and assets a.

Problem of workers of age $1, ..., J^E - 2$

Workers become eligible for early retirement benefits once they reach age J^E. I assume that the retirement decision must be made one period in advance. This assumption implies that agents must remain workers during the first $J^E - 2$ periods of their lives and cannot choose to retire early. During this stage, workers choose how much to consume, how much to save, and how to split the unitary time endowment between leisure, skill acquisition, and work. In each period, agents draw the labor productivity shock y before making any decisions. The agent's dynamic programming problem is given by

$$V_t(z, j, j^R = 0, \nu, y, \bar{e}, h, a) = \tag{17}$$
$$\max_{\substack{a' \geqslant 0, l \in [0,1], \\ s \in [0,1]}} \left\{ u(c, 1 - s - l) + \beta\psi_{z,j} \sum_{y'} \pi_{j,z}(y' \mid y) V_{t+1}(z, j+1, j^{R\prime} = 0, \nu, y', \bar{e}', h', a') + (1 - \psi_{z,j})\phi(a') \right\}$$

subject to the budget constraint

$$a' + (1 + \tau_c)c \;=\; (1 + (1 - \tau_a)r_t)a + e - \tau_{SS,t}\tilde{e} - \tau_{I,t}\iota + T_t, \tag{18}$$

where e are pre-tax earnings defined in (9), $\tilde{e}$ – earnings taxable for the Social Security purpose defined in (10), and ι – taxable labor income given by (15).

The variable $V_t(\cdot)$ denotes the agent's discounted indirect utility at time t. The educational level z and the fixed effect ν remain constant throughout the agent's life. Since the agent must remain in the labor force in the next period, $j^{R\prime} = j^R = 0$. The agent who chooses to carry a' assets into the next period but dies unexpectedly (which occurs with probability $1 - \psi_{z,j}$), receives an instantaneous warm-glow utility $\phi(a')$.

The law of motion for the idiosyncratic shock is governed by the age- and education-specific transition matrix $\pi_{j,z}(y' \mid y)$. Given the optimal choice of learning time s, the next period's stock of human capital h' follows from the law of motion in (2). Given the optimal hours worked l and, therefore, the taxable earnings $\tilde{e}$, the next period's average lifetime earnings $\bar{e}'$ are determined by the law of motion in (11).

Problem of workers of age $J^E - 1, ..., J^R - 1$

Once agents reach age $J^E - 1$, they may choose to retire in the next period. Before making the retirement decision, agents draw their labor productivity y. If the agent chooses to retire in the next period, her current welfare is given by the solution to the problem:

$$V_t^R(z, j, j^R = 0, \nu, y, \bar{e}, h, a) = \max_{a' \geqslant 0, l \in [0,1]} \left\{ u(c, 1 - l) + \beta \psi_{z,j} V_{t+1}(z, j+1, j^{R\prime} = j+1, \bar{e}', a') + (1 - \psi_{z,j}) \phi(a') \right\} \tag{19}$$

subject to the working agent's budget constraint (18). Since the agent retires in the next period, $j^{R\prime} = j + 1$. Note that the stock of human capital h, the fixed effect ν, and the productivity shock y do not affect the household's decisions and welfare during retirement. Therefore, these variables become redundant in the next period and do not appear on the right-hand side of the Bellman equation. For the same reason, I drop the expectations operator. Moreover, the agent optimally chooses not to spend any time on learning, i.e., $s = 0$. I continue to keep track of the agent's education level z because it affects the survival probability rate $\psi_{z,j}$.

If instead the agent chooses to work in the next period, her welfare is given by the right-hand side of (17). Denote the welfare associated with this choice by $V_t^W(z, j, j^R = 0, \nu, y, \bar{e}, h, a)$.

The agent decides whether to retire or continue working by comparing the welfare associated with each choice. The retirement decision can then be formalized as:

$$j^{R\prime} = \begin{cases} 0 & \text{if } V_t^W(z, j, j^R = 0, \nu, y, \bar{e}, h, a) > V_t^R(z, j, j^R = 0, \nu, y, \bar{e}, h, a) \\ j + 1 & \text{otherwise.} \end{cases}$$

The agent's welfare, conditional on making the retirement decision, reads

$$V_t(z, j, j^R = 0, \nu, y, \bar{e}, h, a) = \max \left\{ V_t^R(z, j, j^R = 0, \nu, y, \bar{e}, h, a), V_t^W(z, j, j^R = 0, \nu, y, \bar{e}, h, a) \right\}.$$

Problem of retirees of age $J^E, ..., J$

Retired agents receive no labor income and, consequently, do not face labor income risk. Moreover, they devote their unitary time endowment to leisure. Their maximization problem is given by

$$V_t(z, j, j^R, \bar{e}, a) = \max_{c, a' \geqslant 0} \left\{ u(c, \ell = 1) + \beta \psi_{z,j} V_t(z, j+1, j^R, \bar{e}, a') + (1 - \psi_{z,j}) \phi(a') \right\} \quad (20)$$

subject to the budget constraint

$$a' + (1 + \tau_c)c = (1 + (1 - \tau_a) r_t) a + b + T_t. \quad (21)$$

Observe that the agent's average lifetime earnings remain constant during retirement, i.e., $\bar{e}' = \bar{e}$. Similarly, the retirement age remains constant, i.e., $j^{R\prime} = j^R$.

3.6. Definition of equilibrium

To simplify notation, let $\mathbf{x} = (z, j, j^R, \nu, y, \bar{e}, h, a)$ summarize the agent's state variable. Furthermore, let $F_{t,j}(\mathbf{x})$ denote the share of agents of age j with characteristics $\mathbf{x}$ at time t.

Definition 1. Given the initial capital stock K_0 and an initial measure $\{F_{0,j}\}_{j=1}^J$ of households, a competitive equilibrium is sequences of value and policy functions, $\{V_t, c_t, a'_t, \ell_t, l_t, s_t, j_t^R\}_{t=0}^\infty$, production plans for firms $\{K_t, N_t, N_{H,t}, N_{L,t}, Y_t\}_{t=0}^\infty$, prices $\{w_{t,H}, w_{t,L}, r_t\}_{t=0}^\infty$, sequences of government Social Security policies $\{\boldsymbol{\alpha}_t, \tau_{SS,t}\}_{t=0}^\infty$ and income tax policies $\{\tau_{I,t}, T_t\}_{t=0}^\infty$, a sequence of government spending $\{G_t\}_{t=0}^\infty$, a sequence of economy-wide average taxable earnings $\{\tilde{E}_t\}_{t=0}^\infty$, and a sequence of measures $\{F_{t,j}\}_{t=1}^\infty$, such that the following statements hold for all $t \geqslant 0$:

- Given prices and policies, $\{V_t\}$ solves the agent's optimization problem described in Section 3.5 and $\{c_t, a'_t, \ell_t, l_t, s_t, j_t^R\}$ are the associated policy functions.
- Factor prices $\{w_{t,H}, w_{t,L}, r_t\}$ are determined competitively from (6)-(7).
- The Social Security tax rate $\{\tau_{SS,t}\}$ satisfies the government budget constraint

$$\tau_{SS,t} \mu_t^W \tilde{E}_t = B_t,$$

where the population share of working agents μ_t^W, the economy-wide average taxable earnings $\tilde{E}_t$, and the aggregate pension benefits B_t are defined as follows:

$$\mu_t^W = \sum_j \int_{x:j^R=0} dF_{t,j} \tag{22}$$

$$\tilde{E}_t = \frac{\sum_j \int_{x:j^R=0} \min(cap, w_{t,z} y v h l_t(x)) dF_{t,j}}{\mu_t^W} \tag{23}$$

$$B_t = \sum_j \int_{x:j^R \neq 0} b_t(\bar{e}, j^R; \boldsymbol{\alpha}_t) dF_{t,j}. \tag{24}$$

- The government debt and wasted spending $\{D_t, G_t\}$ satisfy constant debt-to-GDP and spending-to-GDP ratios, respectively: $dy = D_t/Y_t$ and $gy = G_t/Y_t$.

- The lump-sum transfer $\{T_t\}$ satisfies the government budget constraint

$$G_t + T_t + (1 + r_t)D_t = \tau_{I,t}\mathcal{I}_t + \tau_a r_t A_t + \tau_c C_t + (1 + (1 - \tau_a)r_t)\Phi_t + D_{t+1},$$

where A_t, C_t, $\mathcal{I}_t$, and Φ_t denote the aggregate asset holdings, consumption, taxable labor income, and bequests, respectively. These variables are defined as

$$A_{t+1} = \sum_j \int a'_t(x) dF_{t,j}, \tag{25}$$

$$C_t = \sum_j \int c_t(x) dF_{t,j}, \tag{26}$$

$$\mathcal{I}_t = \sum_j \int_{x:j^R=0} \iota_t(x) dF_{t,j}, \tag{27}$$

$$\Phi_{t+1} = \sum_j \int (1 - \psi_{j,z}) a'_t(x) dF_{t,j} \tag{28}$$

with the individual taxable income $\iota_t(x)$ given by (15).

- Capital market, labor markets for each skill, and goods market clear:

$$A_{t+1} = K_{t+1} + D_{t+1}, \tag{29}$$

$$N_{H,t} = \sum_j \int_{x:\{j^R=0 \text{ and } z=H\}} y\nu h l_t(x) dF_{t,j}, \tag{30}$$

$$N_{L,t} = \sum_j \int_{x:\{j^R=0 \text{ and } z=L\}} y\nu h l_t(x) dF_{t,j}, \tag{31}$$

$$K_{t+1} = Y_t + (1-\delta)K_t - C_t - G_t.$$

- The aggregate effective labor supply $\{N_t\}$ is determined by (4).

- The next period's distribution of agents is given by $F_{t+1,j+1} = \mathcal{F}_{t,j}(F_{t,j})$, where $\mathcal{F}_{t,j}$ is the law of motion induced by the exogenous Markov process for labor productivity governed by $\pi_{j,z}(y_{j+1,z} \mid y_{j,z})$, the law of motion for skill acquisition in (2), the law of motion for the average lifetime earnings in (11), and the endogenous asset accumulation.

Definition 2. A stationary equilibrium is a competitive equilibrium in which all individual functions and all aggregate variables are constant over time.

Appendix A describes the algorithm to solve the model numerically.

4. Quantitative experiment

4.1. Social Welfare Function

A significant departure from the existing literature is the specification of the social welfare function. I assume that the government maximizes the welfare of *all* agents who are alive at the time when the optimal policy is implemented. Moreover, I allow the government to discriminate agents based on age and education. I introduce this flexibility through an age- and education-specific Pareto weight function $\omega(j, z)$. Hence, my model nests several existing cases in the literature.[25] The motivation behind this departure is twofold. First, it allows me to rationalize the calibrated income tax and Social Security policies as a solution to the government's maximization problem. As I will show later, newborn agents prefer to keep

[25] Among others, Heathcote et al. (2017) and Wu (2021) assume that the government maximizes the welfare of newborn agents and allow the government to discriminate newborns by ability. Brendler (2020) allows the government to attach age-dependent Pareto weights but assumes that both educational groups are treated equally.

the size of the public pension system substantially below the actual one. Second and closely related to the previous point, the distribution of Pareto weights becomes crucial for the optimal policy analysis. I will show below that the agent's preferences over public insurance provision and income redistribution change with age and education. Then, a particular distribution of weights may intensify or mitigate distributional conflicts among agents, thus affecting the optimal policy mix.

Let $\Upsilon = (\alpha, \tau_I)$ denote a constant replacement rate and the income tax policy chosen by the government.[26] The government sets Υ optimally by maximizing the weighted sum of expected discounted lifetime utilities of all generations who are alive at the time of reform.[27] Formally, the social welfare function reads

$$SWF(\Upsilon;\kappa) = \sum_j \int \omega(z,j;\kappa) V_1(z,j,j^R,\nu,y,\bar{e},h,a;\Upsilon) dF_{1,j}, \tag{32}$$

where $V_1(.;\Upsilon)$ is the value function in the first period of the transition induced by the new policy Υ, $F_1 = F_0$ is the initial distribution of households in the stationary equilibrium under the status quo policy, and $\omega(\cdot)$ is the Pareto weight function. Its specification is discussed next.

I assume that Pareto weights depend on the agent's education and age. As I will show, the education dimension alone, which the optimal income taxation literature has focused on so far, is insufficient to rationalize the calibrated income tax and Social Security policies.[28] At the same time, the literature on Social Security has demonstrated that the age dimension can account for the actual average replacement rate level in the U.S. (e.g., Brendler (2020)). My contribution is to merge the two strands of the literature and show that a joint distribution of Pareto weights by age and education can rationalize the observed income tax, average replacement rate, and pension system progressivity.

Another advantage of choosing age and education is because these are observable and measurable characteristics. In the empirical application (Appendix C), I will associate changes in Pareto weights in the model with changes in voter turnout rates in the CPS data. This enables me to contrast model predictions regarding the evolution of Pareto weights with the observed trend in voter turnout rates. Finally, age and education are individual state

[26]All the remaining government variables, including consumption tax τ_c, capital income tax τ_a, maximum taxable earnings threshold cap, and bend point $\bar{e}_{\min}$, are assumed to be exogenous.

[27]In the real world, governments seek reelection and propose policies to gain the support of *current* voters. This idea motivates the assumption that the government maximizes the welfare of currently alive generations.

[28]See Heathcote et al. (2017), Wu (2021), among others.

variables and, therefore, invariant to the optimal policy $\Upsilon^\star$. This significantly simplifies the identification of $\boldsymbol{\kappa}$.[29]

I assume that the Pareto weight of a j-year-old agent, conditional on education, equals $\exp(-\kappa_1 \cdot j)$. The parameter $\kappa_1 \in \mathcal{R}$ that I will refer to as the *age bias* below determines the extent to which the government discriminates against agents of different ages with the same educational background.[30] Figure 2 illustrates the relationship between the age bias parameter κ_1 and the agent's Pareto weight. More specifically, the figure plots the weight of a newborn agent (real-life age 25) relative to the weight of an agent who is at the mandatory retirement age 64 and has the same educational level. When $\kappa_1 = 0$, the government trades off the utility of different age groups at the same rate, so that the relative Pareto weight in the figure equals one. When $\kappa_1 > 0$, the relative Pareto weights are above one because the government assigns a larger weight to younger agents; the opposite is the case when $\kappa_1 < 0$.

Besides age, the government may also discriminate against agents with different educational levels. In particular, the weight assigned to a college graduate relative to a non-college graduate of the same age is given by $\exp(\kappa_2)$, where $\kappa_2 \in \mathcal{R}$ is a parameter that I will refer to as the *educational bias* below. Figure 3 shows the weight of an agent with a college degree relative to the weight of an agent with a non-college degree of the same age as a function of κ_2. When $\kappa_2 = 0$, the government treats both educational groups equally and the relative weight equals one. When $\kappa_2 > 0$, the government favors college graduates over non-college graduates; the opposite is the case when $\kappa_2 < 0$.[31]

Summarizing the above discussion, one can formally write the Pareto weight function as:

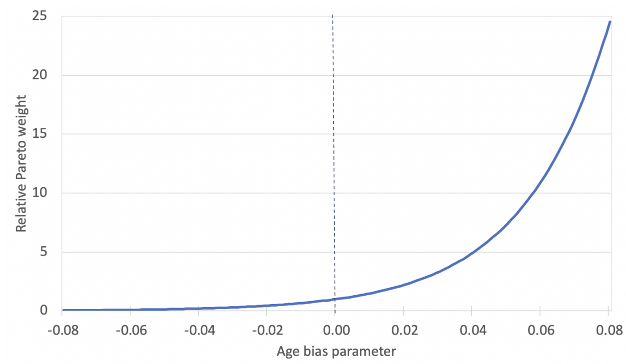
$$\omega(j, z; \boldsymbol{\kappa}) = \exp(-\kappa_1 \cdot j + \kappa_2 \cdot \mathbb{1}_{z=H}), \tag{33}$$

where $\mathbb{1}_{z=H}$ is an indicator function that equals one for agents with college education.

[29] Alternatively, one could let the weight depend on the agent's expected average lifetime earnings at retirement. In this case, however, the weight becomes endogenous since it depends on policy Υ. This significantly complicates identification. Of course, the Pareto weight function can be enriched by other exogenous, i.e., policy-invariant, dimensions of heterogeneity among households (fixed effects, shocks, etc.). However, one obstacle of this approach is computational intensity: identification requires that the number of parameters in the weighting function does not exceed the number of policy instruments that serve as data moments. Hence, mapping Pareto weights to additional characteristics would require expanding the set of instruments available to the government. Besides, it is unclear how to falsify the predictions of the model regarding the evolution of Pareto weights using the data in this case.

[30] Note that the government does not discriminate against *cohorts*. In my quantitative experiments, I will apply the Pareto weights from the 1970s calibration to the 2010s calibration. Agents of the same age in each steady state will be attached the same Pareto weight, even though they belong to different cohorts.

[31] The social welfare function (32) is utilitarian when $(\kappa_1, \kappa_2) = (0, 0)$.



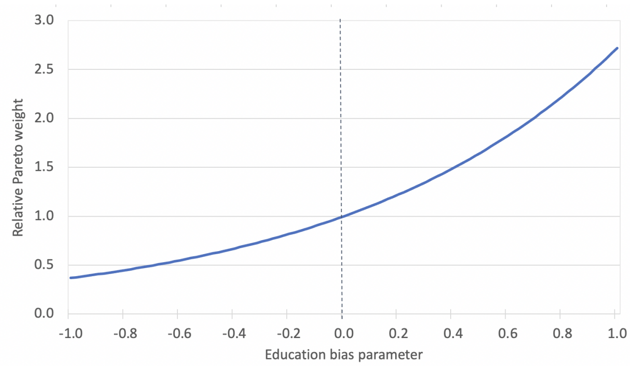


Figure 2: Pareto Weights as a Function Of κ_1 Figure 3: Pareto Weights as a Function Of κ_2

Notes: Figure 2 plots the Pareto weight of a newborn agent (real-life age 25) relative to the Pareto weight of an agent at the mandatory retirement age 64, conditional on the same educational status, as a function of the age bias parameter κ_1 in the Pareto weight function (33). Figure 3 shows the Pareto weight of a college graduate agent relative to a non-college graduate agent of the same age, as a function of the educational bias parameter κ_2.

4.2. Government maximization problem

I assume that the economy is in a steady-state equilibrium which corresponds to period $t = 0$. Given 1) the initial capital stock K_0, 2) the cross-sectional distribution of households F_0 which is determined by the initial policy $\tilde{\Upsilon}$ calibrated from the data, and 3) the Pareto weight parameter vector $\boldsymbol{\kappa}$, the government implements at time $t = 0$ a constant future policy $\Upsilon^\star$ that solves

$$\Upsilon^\star = \arg\max_{\Upsilon} SWF(\Upsilon; \boldsymbol{\kappa}). \tag{34}$$

The optimal policy reform is assumed to be unanticipated. In Section 6, I will explain in detail the key trade-offs faced by the government when choosing Υ.

4.3. Design of the quantitative experiment

The main quantitative experiment consists of two steps. First, I identify the parameter vector $\boldsymbol{\kappa}$ in the Pareto weight function (33), such that the optimal policy $\Upsilon^\star$, that arises as a solution to the government's maximization problem (34), coincides with the actual policy $\tilde{\Upsilon}$ calibrated from the data in the late 1970s and the 2010s. Second, I compute the optimal Social Security and income tax policies under the 2010s calibration of the model, applying the identified Pareto weights from the 1970s. Comparing the estimates of $\boldsymbol{\kappa}$ across the steady states allows me to analyze how the government's preferences over equity and efficiency evolved over time, conditional on the occurred changes in the U.S. economy.[32]

5. Calibration

5.1. Calibration strategy

I calibrate two sets of model parameters. The first set reflects the U.S. economy at the end of the 1970s, whereas the second set characterizes the economy during the 2010s. In both periods, the economy is assumed to be in a steady-state equilibrium, so I drop the time index t throughout this section.[33] One model period equals one year.

I use the Current Population Survey (CPS) data to calibrate most of the model parameters. The harmonized annual CPS extracts are currently available for 1980–2018.[34] I use the 1980–1984 CPS extracts to calibrate the steady state in the 1970s and the 2010–2018 extracts to calibrate the model economy in the 2010s.[35] I assume that an agent in the model corresponds to a household head in the CPS data.[36]

Table 1 shows the parameter values in each steady state. The first part of the table lists the model parameters that I calibrate outside the model. These parameters are divided into 5 groups: demographics, preferences, labor productivity, production, and government policy. The second part of the table lists all the parameters calibrated inside the model. The calibration target for each parameter is shown in brackets. Below I describe the calibration strategy in detail.

5.2. Demographics

Agents enter the model at age 1 corresponding to a real-life age 25. The maximum possible age is $J = 76$ (real-life age 100). Consistent with the Social Security provisions, agents qualify for early retirement at age $J^E = 38$ (real-life age 62). The values of parameters J and J^E are the same in both steady states. I set the normal retirement age, J^R, to 41 (real-life age 65)

[32]Implicitly, the actual policy $\tilde{\Upsilon}$, the optimal policy $\Upsilon^\star$, and the Pareto weight parameter vector $\boldsymbol{\kappa}$ depend on time since I compute them in two distinct steady states. Since I will be explicit about which steady state I mean in each case and to simplify notation, I will omit the time index.

[33]A steady-state equilibrium is defined on page 20.

[34]See Appendix B.1 for sample selection restrictions.

[35]Heathcote et al. (2020) choose similar periods for their analysis.

[36]I use household-level data instead of individual-level data to account for the insurance against idiosyncratic labor productivity risk among household members. The research has demonstrated this insurance to be quantitatively important (see Fuster et al., 2007).

Table 1: Model parameters.

Parameter	Interpretation	Value in 1970s	Value in 2010s
Exogenously calibrated parameters			
Demographics			
J	Maximum life span (age 100)	76	76
J^E	Early retirement age (age 62)	38	38
J^R	Normal retirement age	41 (age 65)	42 (age 66)
$\psi_{z,j}$	Education-specific survival probabilities	Estimates	
Π_z	Distribution of non-college and college graduates, %	$(75, 25)$	$(56, 44)$
n	Population growth rate, %	0.9	1.2
Preferences			
β	Discount factor (capital-to-output ratio)	1	1
σ	Coefficient of relative risk aversion	2	2
(ϕ_1, ϕ_2)	Bequest function	$(-9.5, 11.6)$	$(-9.5, 11.6)$
Labor productivity			
γ^h	Elasticity of human capital production	0.7	0.7
(ρ_H, ρ_L)	Persistence parameter by education	$(0.989, 1.0)$	$(0.993, 0.977)$
$(\sigma^2_{\epsilon,H}, \sigma^2_{\epsilon,L})$	Variance of persistent shock by educ.	$(0.016, 0.006)$	$(0.015, 0.012)$
Production			
ϖ	Capital share	0.43	0.46
δ	Capital depreciation, %	8	6
ρ	Elasticity of substitution is $1/(1-\rho)$	0.75	0.285
Government policy			
α_2	Degree of pension system progressivity	-0.42	-0.43
τ_c	Consumption tax, %	5.3	4.1
τ_a	Capital income tax, %	38.4	33.0
gy	Gov. consumption-to-GDP ratio, %	9.5	7.8
dy	Debt-to-GDP ratio, %	35.0	100.0
φ	EITC-to-GDP, %	0.03	0.31
Parameters calibrated in equilibrium (targets in brackets)			
γ	Weight on consumption (aver. hours worked)	0.47	0.47
(θ_H, θ_L)	Learning ability by educ. (age profile of wages)	$(0.087, 0.086)$	$(0.115, 0.09)$
$(h_{1,H}, h_{1,L})$	Initial skill level by education (age profile of wages)	$(0.8, 0.8)$	$(1.3, 0.6)$
δ^h	Skill depreciation, % (age profile of wages)	0.55	0.35
$(\sigma^2_{\nu,H}, \sigma^2_{\nu,L})$	Variances of fixed effect by educ. (earnings Gini)	$(0.0, 0.006)$	$(0.018, 0.019)$
Z	Scaling factor in production technology (average wage)	0.745	0.28
α_1	Average replacement rate, % (Social Security tax)	37.0	39.0
$\bar{e}_{min}$	Bend point (bend point-to-earnings ratio)	0.06	0.09
δ^p	Penalty for early retirement (share of retired at age 62), %	14.0	19.4
cap	Taxable earnings threshold (share of workers above cap)	0.58	1.07
τ_I	Income tax rate, % (average effective tax rate)	18.6	16.6

in the 1970s calibration. For the 2010s calibration, I increase J^R by one year.[37]

I assume that the agent's education type z in the model corresponds to the household head's educational level in the data. I divide the CPS extract into two subsamples based on the household head's education. I pool together all household heads with a completed college degree or higher and refer to them as *college graduates* below; I refer to all remaining households as *non-college graduates*. The time-invariant distribution of educational types, Π_z, is then given by the population shares of college graduates and non-college graduates in the CPS. In the late 1970s, college graduates constituted 25% of the sample. In the recent data, the share of college graduates grew substantially to 44%.

Rising longevity has been one of the most significant demographic changes in the US during the past four decades. I summarize the occurred changes in longevity in 2 empirical facts that I use as calibration targets in the model. First, the life expectancy of an average worker increased at all ages during 1970–2010, as documented by Bell et al. (1992, Table 6). For example, the life expectancy of a 25-year-old worker rose by 7 years during this period. Second, the rise in longevity was more pronounced for college graduates than for non-college graduates. According to Bound et al. (2014), the life expectancy gap between the two educational groups at age 25 was 4 years in 1990; by 2010, this gap increased to 6 years.[38]

I account for the two empirical facts using the education- and age-specific conditional survival probability rates $\psi_{z,j}$. To estimate them, I specify the mortality rates as a Gompertz force of mortality function.[39] I estimate the parameters of the Gompertz function outside the model matching two empirical targets presented above. Appendix B.2 explains the estimation procedure in greater detail.

Given the estimated probabilities $\psi_{z,j}$, I calibrate the birth rate, n, outside the model to match the dependency ratio. I define the latter as the ratio of old-age households (real-life age 65–100) to working-age households (age 25–64). I target the dependency ratio of 22%

[37]According to the current Social Security legislation, the normal retirement age for all individuals who were born during 1943–1954 is 66 years. This is the value I choose to parametrize J^R in the 2010s calibration. More details on the normal retirement age can be found at `https://www.ssa.gov/benefits/retirement/planner/agereduction.html`.

[38]Elo and Preston (1996) and Meara et al. (2008) also document a rising gap in longevity by education. Meara et al. (2008) proposed one potential explanation for the observed facts. They found that smoking rates and death rates caused by smoking-related diseases dropped more significantly over time for higher socioeconomic groups than for the lower ones.

[39]In the related literature, Hosseini and Shourideh (2019) follow a similar approach.

for the 1970 calibration and 27% for the 2010s calibration, consistent with the CPS.[40]

5.3. Preferences

The instantaneous utility u is a constant relative risk aversion function given by

$$u(c, \ell) = \frac{[c^{\gamma}\ell^{1-\gamma}]^{1-\sigma} - 1}{1-\sigma}, \tag{35}$$

where σ controls the degree of relative risk aversion and γ is the relative weight on consumption. The bequest function is specified as in De Nardi (2004)

$$\phi(a) = \phi_1(1 + a/\phi_2)^{1-\sigma},$$

where ϕ_1 reflects the agent's concern about leaving bequest a, while ϕ_2 measures the extent to which bequests are a luxury good.

I assume that preference parameters $(\sigma, \gamma, \phi_1, \phi_2, \beta)$, where β is the discount factor, remain the same in both steady states.[41] In particular, I set $\sigma = 2.0$, $\beta = 1.0$ and borrow the estimates for the parameters of the bequest function from De Nardi (2004) with $\phi_1 = -9.5$ and $\phi_2 = 11.6$. Given these values, the model matches exactly the capital-to-GDP ratio, K/Y, equal to 3.2 in the 1970s. For the 2010s calibration, the model-implied capital-to-GDP ratio is 3.9 which is fairly close to 3.7 in the data.[42] Finally, I set $\gamma = 0.47$. Given this value, the agents in the model choose to spend on average 39% of their discretionary time endowment on work in both steady states, which is in line with the CPS data for each sample period.[43]

[40] This rise in the dependency ratio implies that the number of working-age households per old-age household declined from 4.5 to 3.7.

[41] In the quantitative experiments below, I will compute welfare measures associated with the optimal policies. The assumption of constant preference parameters allows me to attribute any change in the welfare measures to the underlying change in the fundamental parameters that govern inequality, aging, etc. and not the preference parameters themselves.

[42] I follow the procedure in Hosseini and Shourideh (2019) to compute the share of capital stock in GDP from the National Income and Product Accounts. See their online Supplement, Section S5, which is available at `https://www.econometricsociety.org/sites/default/files/ecta200042-sup-0001-Supplement.pdf`). For the 1970s calibration, I compute the averages of the relevant variables over the period 1970–1980; for the 2010s calibration, I use the period 2010–2018.

[43] See Appendix B.1 for details.

5.4. Labor productivity

In the model, the agent earns an hourly wage equal to

$$w_z h_{j,z} y_{j,z} \nu_z, \tag{36}$$

where w_z is the education-specific wage, $h_{j,z}$ is the skill level, $y_{j,z}$ is the idiosyncratic shock, and ν_z is a fixed effect. w_z is determined in equilibrium through the firm's optimality condition (6). The calibration of the remaining components of the worker's hourly wage is described below.

Agents accumulate skills $h_{j,z}$ according to the deterministic law of motion which I have introduced in the model section:

$$h_{j+1,z} = (1-\delta^h)h_{j,z} + \theta_z (h_{j,z} s)^{\gamma^h},$$

with a given initial skill level $h_{1,z}$ and a fixed learning ability θ_z. I assume that γ^h is the same across the steady states and set it equal to 0.7 following Badel et al. (2020). I calibrate the remaining 5 parameters $(\theta_H, \theta_L, h_{1,H}, h_{1,L}, \delta^h)$ by targeting the profiles of hourly wages by age and education in the CPS.[44]

To obtain hourly wages, I first divide the household's total annual pre-government earnings by the total annual hours worked. Then, I fit a quadratic polynomial in age to the household's log hourly wages by education. Next, I normalize the resulting profiles by the economy-wide average hourly wage to give the profiles a meaningful interpretation. Figure 4 (solid and dashed lines) shows the constructed empirical profiles for both educational groups.

I calibrate the learning ability θ_z and the rate of human capital depreciation δ^h inside the model matching the slope and the curvature of the empirical profiles from Figure 4. The intercept of the profiles pins down the entry skill level $h_{1,z}$.[45] Figure 4 shows that the model fits the empirical age profiles of hourly wages by education in each steady state quite accurately.

The 1970s calibration implies no heterogeneity in the entry skill levels ($h_{1,H} = h_{1,L} = 0.8$) and a slightly higher ability level for college graduates ($\theta_H = 0.087$ and $\theta_L = 0.086$). On the

[44]See Appendix B.1 for sample periods and sample selection criteria.

[45]The intercept of the empirical profile corresponds to $w_z h_{1,z}$ in the model (I normalize the average wage to 1.0 in the model, as I explain in the next section). Hence, the initial skill level $h_{1,z}$ cannot be identified separately from the elasticity parameter ρ that determines the wage premium between college graduates and non-college graduates defined in (8). My calibration strategy, therefore, is to take the estimate of ρ from an external source and calibrate $h_{1,z}$ inside the model.

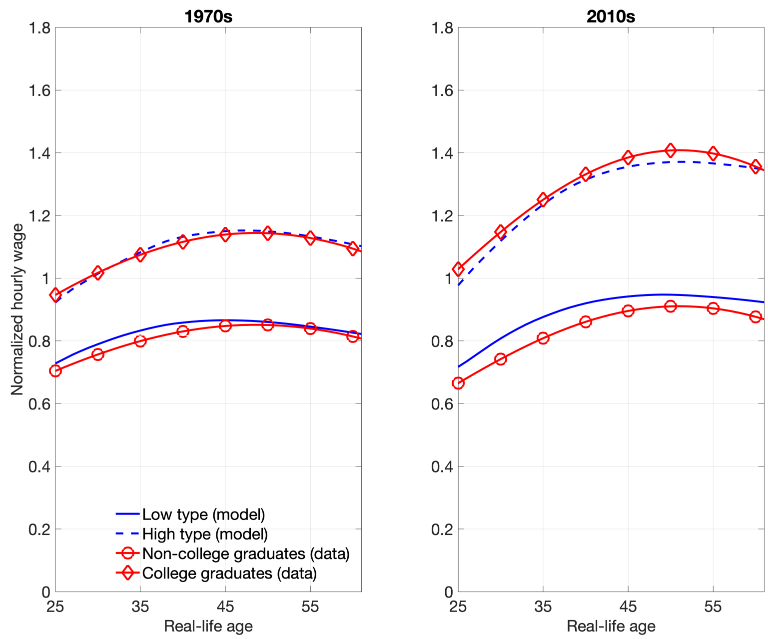


Figure 4: Hourly wages in the model and data, by age and education.

Notes: The figure shows how the model fits the age profiles of hourly wages by education in the 1970s (left panel) and 2010s (right panel). The empirical profiles are constructed using the CPS. See Appendix B.1 for sample periods and sample restrictions. The low (high) type in the model corresponds to the agent with education level $z = \mathsf{L}$ ($z = \mathsf{H}$). The CPS profiles are normalized by the average economy-wide hourly wage; the average wage in the model is normalized to 1.

contrary, the disparity in skills becomes significant in the 2010s calibration with $h_{1,H} = 1.3$, $h_{1,L} = 0.6$, $\theta_H = 0.115$, and $\theta_L = 0.09$.

The idiosyncratic productivity shock $y_{j,z}$ follows an AR(1) process given by (3). With this specification, the conditional variance of y increases with age according to:

$$Var(\log(y^i_{j,z})) = \sigma^2_{\epsilon,z} \times \sum_{m=0}^{j-2} \rho_z^{2m}. \tag{37}$$

Overall, there are 4 parameters to calibrate: $(\rho_H, \rho_L, \sigma^2_{\epsilon,H}, \sigma^2_{\epsilon,L})$. Following Storesletten et al. (2004), I estimate these parameters outside the model fitting (37) to the profiles of wage dispersion by age and education in the CPS.[46,47] As the authors have insightfully pointed out, the conditional variance $\sigma^2_{\epsilon,z}$ can be identified from the slope and the auto-correlation ρ_z from the curvature of the variance profiles in the data.

To construct the empirical variance profiles, I first compute the variance of the logarithm of the households hourly wages by age and cohort, where a cohort is defined to be all households with a head born in the same year. Following Deaton and Paxson (1994), I remove the cohort effects by regressing the computed variances on a set of age and cohort dummies. Figure 5 shows the raw and the fitted coefficients on the age dummies for the 2010s calibration. The coefficients are normalized such that the variance is equal to 0 at age 25.

For the 1970s calibration, I estimate $\rho_H = 0.989$ and $\rho_L = 1.0$.[48] Moreover, I find that $\sigma^2_{\epsilon,H} = 0.016$ and $\sigma^2_{\epsilon,L} = 0.006$, i.e., the variance of shocks is 2.6 times higher for college graduates than for non-college graduates. In the 2010s calibration, the persistence parameter slightly increases for college graduates ($\rho_H = 0.993$) and declines for non-college graduates ($\rho_L = 0.977$). The variances of the shocks remains almost unchanged for college graduates ($\sigma^2_{\epsilon,H} = 0.015$) and doubles for non-college graduates $\sigma^2_{\epsilon,L} = 0.012$).

The final component in the agent's hourly wage is the fixed effect ν_z. My strategy is to calibrate the fixed effect inside the model matching the cross-sectional earnings inequality by education that remains after accounting for differences in skill accumulation and the idiosyncratic shock $y_{j,z}$.

I assume that the logarithm of ν_z is a standard-normally distributed variable with an education-specific variance $\sigma^2_{\nu,z}$. For the 1970s, I find that $\sigma^2_{\nu,L} = 0.006$ matches the Gini index for earnings for a sample of non-college graduates equal to 0.294. As for college graduates, I set $\sigma^2_{\nu,H} = 0$, since the model without the fixed effect matches the Gini index of earnings among college graduates of 0.281. For the 2010s calibration, both variances increase substantially: $\sigma^2_{\nu,H} = 0.0176$ matches the Gini index of 0.383 for the sample of college graduates and $\sigma^2_{\nu,L} = 0.0188$ matches the Gini index of 0.377 for the sample non-college graduates.

Overall, the estimates of the idiosyncratic component and the fixed effect imply that a

[46] I also include an intercept term into (37). Since the intercept term contains a measurement error, I do not further use its estimate. Instead, I calibrate the variances of the fixed effects inside the model matching earnings inequality by education (see below).

[47] As opposed to Storesletten et al. (2004) who estimate the parameters using household earnings, I use household hourly wages. In their theoretical framework, the labor supply is exogenous so that the authors are able to estimate the parameters outside the model. In my model, the labor supply is endogenous which is why I take household hourly wages as a calibration target.

[48] A very high value of the persistence parameter is consistent with the literature. For example, Storesletten et al. (2004) use the Panel Study of Income Dynamics during 1969–1992 and estimate a unit root process for a pooled sample of households.

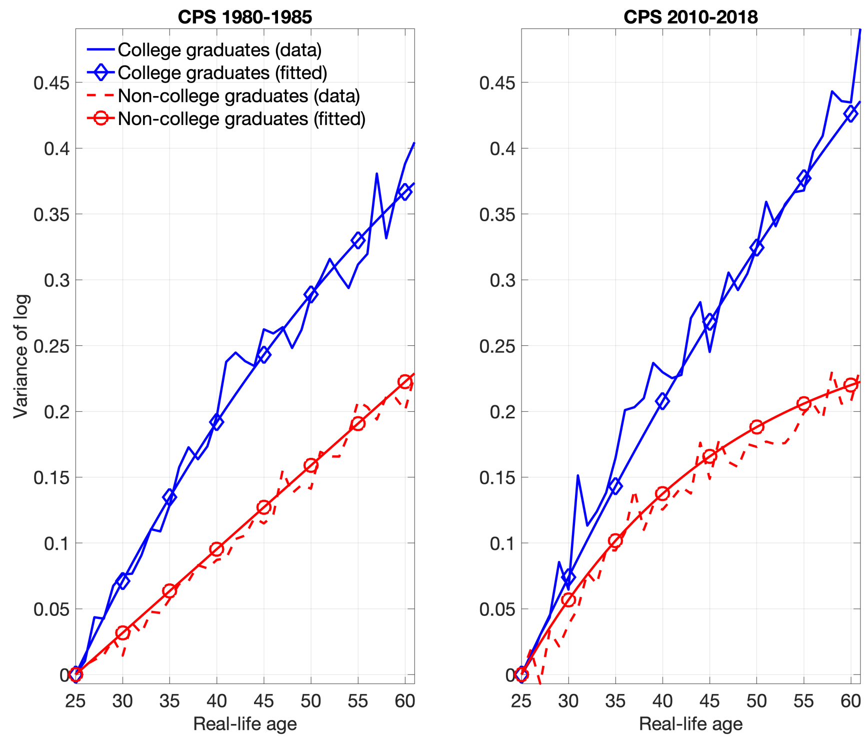


Figure 5: Cross-sectional variance of log hourly wages by education.

Notes: The solid and dashed lines show the variance of the logarithm of cross-sectional household hourly wages by age and education. The profiles are constructed using cross-sectional CPS data. See Appendix B.1 for sample period and sample restrictions. The variances are computed net of cohort effects via a cohort and age dummy-variable regression as in Deaton and Paxson (1994). The solid and the dashed lines are the coefficients on the age dummies, normalized to 0 at age 25 for each educational group. The solid line with circles and the solid line with diamonds represent the population cross-sectional variances associated with the process formulated in (3), with parameter values chosen to best match the slope and the curvature of the empirical age profiles.

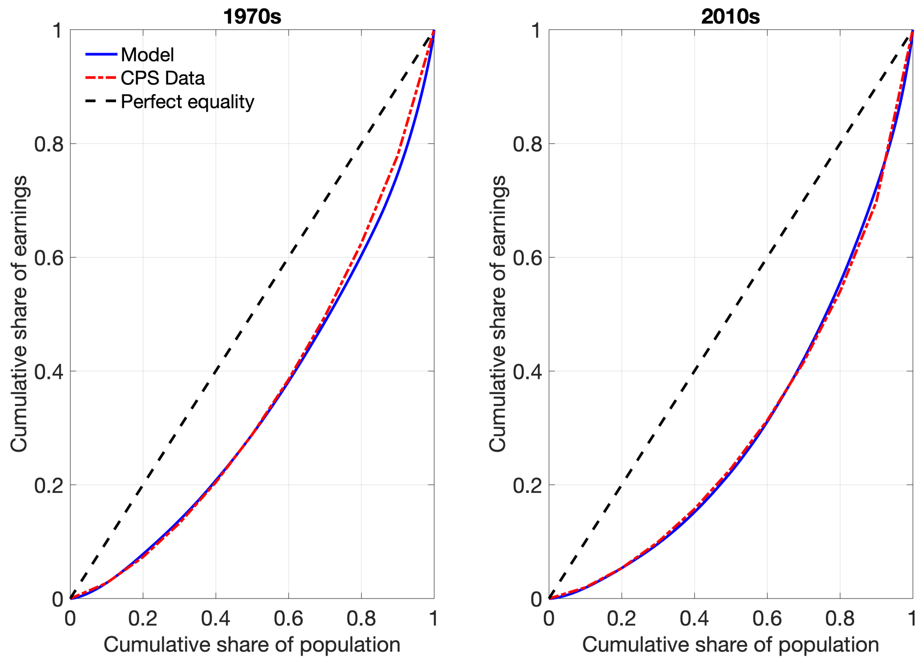


Figure 6: Earnings inequality in the model and data.

Notes: The figure shows how the model fits the Lorenz curves for the earnings distribution in the 1970s (left panel) and 2010s (right panel). Earnings are measured before taxes and government transfers. In the model, the Lorenz curve is constructed using the cross-sectional distribution of earnings in the respective steady state. The empirical curves are computed using the CPS data. See Appendix B.1 for sample period and sample restrictions.

substantial portion of the residual earnings inequality is due to larger dispersion in initial conditions at labor market entry. This viewpoint is supported by recent evidence from administrative data (Guvenen et al., 2017).

Figure 6 compares the model-based Lorenz curves for pre-government earnings in each steady state to the data.[49] The left panel corresponds to the 1970s calibration, while the right panel shows the fit for the 2010s. As one can see, the model achieves a surprisingly good fit in both steady states.

5.5. Production

The capital share ϖ in the production function (5) is set outside the model to match the average ratio of capital income in GDP. The depreciation rate of capital δ in the expression for the interest rate in (7) is calibrated outside the model to match the average ratio of investment

[49] The empirical Lorenz curve for earnings was not targeted during the calibration.

in GDP.[50] I obtain $\varpi = 0.43$ and $\delta = 8\%$ for the 1970s calibration; the corresponding values are 0.46 and 6% for the 2010s calibration. The scaling factor Z in the production function is calibrated inside the model such that the average wage in the economy is equal to 1.0 in each steady state.

Finally, I parameterize ρ that governs the elasticity of substitution between college and non-college graduates in production (see eq. 4). In line with Heathcote et al. (2017) and Abbott et al. (2019), my theoretical framework attributes the rise in the wage premium $w_{H,t}/w_{L,t}$ to stronger complementarity between college graduate workers and non-college graduate workers during 1970–2010. Following Katz and Murphy (1992), I set ρ to 0.75 for the 1970s and 0.285 for the 2010s. These values imply that the elasticity of substitution declines from 4.0 to 1.4.[51]

5.6. Government policy

5.6.1. Social Security

In this section, I describe briefly the calibration of the Social Security parameters. All the details can be found in Appendix B.3.

The pension benefit of a worker who retires at the normal retirement age is determined by a statutory replacement rate schedule that is a function of the worker's average lifetime earnings. The empirical schedule comprises three brackets with constant marginal replacement rates of 90%, 32%, and 15% in the lowest, intermediate, and highest bracket, respectively.[52] Using this information, I construct the implied schedule of average replacement rates. To bring the latter schedule to the model, I apply several data transformations. First, I annualize the brackets multiplying each by 12 since one period in my model corresponds to one year, while the statutory schedule is based on monthly data. Second, I adjust the brackets to the average number of earners in a household because the observation unit in my model is a household, whereas the statutory replacement rate schedule applies to individuals. Third, I normalize the workers' annual average lifetime earnings and the brackets by the economy-wide average taxable earnings computed from the CPS. Finally, I estimate the parameter vector $\boldsymbol{\alpha}$ by fitting the replacement rate function $R(\bar{e}, j^R; \boldsymbol{\alpha})$ in (13) to the empirical schedule

[50] The respective targets were computed using the NIPA data. See Footnote 42 for the procedure and the data source.

[51] Heathcote et al. (2017) choose similar estimates of the elasticity of substitution: 3.3 for 1980 and 2.4 for 2016.

[52] Table 2.A11 in Social Security Administration (2019) shows the parameters of the statutory pension benefit formula.

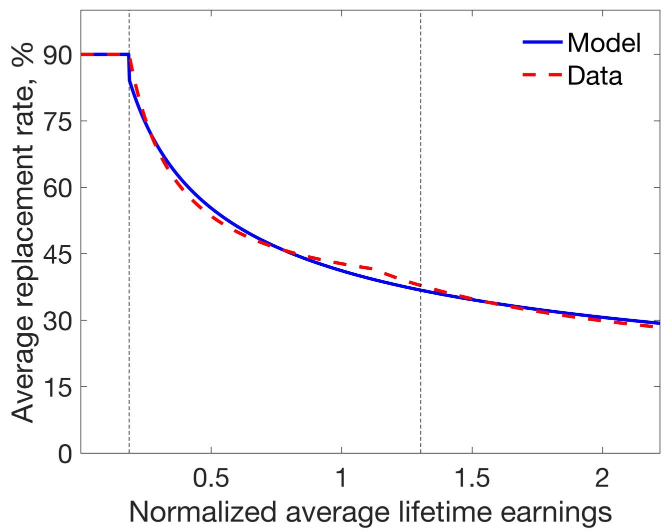


Figure 7: Fit of the replacement rate schedule for the 2010s calibration.

Notes: The figure shows how the model fits the statutory schedule of average replacement rates for the 2010s calibration. The dashed vertical line marks the first bracket of the statutory replacement rate schedule and corresponds to the parameter $\bar{e}_{\text{min}}$ in the replacement rate schedule (13). The average lifetime earnings in the model and data are normalized by the economy-wide average taxable earnings.

of average replacement rates just constructed. That is, I keep the first bend point as it is and fit a power function to the remaining part of the schedule.

I obtain very similar estimates of $(\alpha_1, \alpha_2, \bar{e}_{\text{min}})$ in both steady states.[53] For the 1970s calibration, $\alpha_1 = 0.39$, $\alpha_2 = -0.42$, while the lower bracket $\bar{e}_{\text{min}}$ is equal to 15% of the economy-wide average taxable earnings $\tilde{E}$.[54] For the 2010s calibration, the estimates of $\boldsymbol{\alpha}$ are $(0.42, -0.43)$, while $\bar{e}_{\text{min}}$ is equal to 19% of $\tilde{E}$.

Figure 7 shows the empirical and the fitted schedules of the average replacement rates for the 2010s calibration. The vertical dashed line corresponds to the lower bend point. The fit of the schedule for the 1970 calibration is qualitatively very similar.

To improve the fit of the model, I slightly adjust the obtained estimate of α_1, keeping α_2 fixed, such that the model matches exactly the Social Security tax rate in the data. More specifically, I reduce the average replacement rate, α_1, from 39% to 37% to match $\tau_{SS} = 8.9\%$

[53]This result is expected because the statutory marginal replacement rates in the pension benefit formula have been unchanged since adoption of the Social Security Amendments of 1979.

[54]The economy-wide average taxable earnings $\tilde{E}$ are defined in (23).

in the 1970s and from 42% to 39% to match $\tau_{SS} = 10.6\%$ in the 2010s.[55]

The statutory replacement rate schedule determines the pension benefit of an individual who chooses to retire at the normal retirement age J^R. If the individual chooses to retire earlier, her benefit is reduced according to a penalty function. The age of first eligibility for early retirement benefits is 62 years. which is the variable J^E in the model. When the agent retires, she receives a pension benefit given by $b = b(\bar{b}, j^R)$, where $\bar{b} = \bar{b}(\bar{e}, \boldsymbol{\alpha})$ is the full pension amount determined by the replacement rate schedule $R(\cdot)$ calibrated above. I specify the penalty function as:[56]

$$b(\bar{b}, j^R) = (1 - \delta^p) \cdot \bar{b}(\bar{e}; \boldsymbol{\alpha}) + \left(\frac{j^R - J^E}{J^R - J^E}\right) \cdot \delta^p \cdot \bar{b}(\bar{e}; \boldsymbol{\alpha}). \tag{38}$$

When the agent retires at the earliest possible age ($j^R = J^E$), the expression above boils down to $(1 - \delta^p)\bar{b}(\bar{e}; \boldsymbol{\alpha})$. Hence, the parameter δ^p denotes the fraction of the full benefit that the agent loses due to early retirement. On the contrary, if the agent retires at the normal retirement age ($j^R = J^R$), she receives the full amount $\bar{b}$.

I estimate the penalty δ^p inside the model targeting the share of retired households at age 62 in the CPS. I assume that a household head is retired if she reports a positive Social Security income. In the 1970s calibration, the fraction of retired household heads is 39% and the estimated penalty is $\delta^p = 14.0\%$. The latter equilibrium value comes surprisingly close to the statutory penalty of 20% in the data. Moreover, the model achieves a good fit of the distribution of early retirees by education. In particular, the model predicts that a vast majority of agents who retire at real-life age 62 are non-college graduates (80%); qualitatively, this finding is consistent with the CPS, though the data moment is higher (93%). In the 2010s calibration, there are 26% of retired households at age 62. I obtain $\delta^p = 19.4\%$ which is again fairly close to the statutory penalty of 25%. Surprisingly, the model matches almost exactly the share of college graduates among retired agents at age 62 equal to ca. 20%.

Finally, I calibrate cap inside the model matching the fraction of households in the CPS whose total pre-tax earnings exceed the maximum taxable earnings threshold adjusted for

[55] I take the average over the same periods as in the CPS data, see Appendix B.1. The time-series of the Social Security tax rate is taken from Table 2.A3 in Annual Statistical Supplement to Social Security Administration (2017) available at https://www.ssa.gov/policy/docs/statcomps/supplement/2017/index.html.

[56] For simplicity, I assume a proportional benefit reduction in the number of years before the normal retirement age. In the data, the relationship is piecewise linear with a kink. More details on the early retirement legislation, including the maximum penalty, can be found at https://www.ssa.gov/oact/quickcalc/early_late.html.

the number of earners (see above). In the 1970s calibration, there are 13% of such households leading to $\mathsf{cap} = 0.58$. In the 2010s calibration, there are 8% of these households leading to $\mathsf{cap} = 1.07$. Endogenously, the model matches fairly well the share of college graduates among workers with earnings above the cap. In the 1970s calibration, this number is 40% compared to 48% in the CPS data; in the recent calibration, this number is 76% compared to 80% in the data.

5.6.2. Income taxation

I calibrate the income tax rate, τ_I, inside the model matching the average effective income tax rate estimated by the Congressional Budget Office (CBO).[57] The CBO defines the average effective income tax rate as the amount of total income tax liability divided by the total pre-tax and after-transfer income. I exclude from the CBO's definition of transfers the Social Security payments because my model accounts for them explicitly and Medicare transfers because they are not part of the model. The pre-tax income comprises labor income and business income; I exclude capital income because my model accounts for it explicitly. Under these assumptions, the average effective tax rate in the data drops from 14.4% in the 1970s to 12.4% in the 2010s.[58,59]

The CBO reports the total tax liability net of the Earned Income Tax Credit (EITC); neither does it include EITC in the transfer measure. To bring the empirical effective tax rate to the model, I will assume that from the total transfer T in the model a given fraction $\varphi \in [0, 1]$ is spent as a credit in income taxes and the rest, $1-\varphi$, is distributed as a lump-sum transfer.[60] Thus, for accounting reasons, let $\Psi = \varphi T$ denote the EITC and $T^f = (1-\varphi)T$ – the lump-sum transfer. Then, the average effective income tax rate in the model is given by

$$\frac{\tau_I \mathcal{I} - \Psi}{\mathcal{I} + T^f}, \tag{39}$$

where $\mathcal{I}$ is the aggregate taxable income defined in (27). The expression above shows that the effective tax rate in my model increases with income. In other words, the effective income

[57] The detailed data are publicly available at `https://www.cbo.gov/publication/56575`, see *Additional data for researchers*.

[58] Throughout this section, I use the same sample periods to compute statistics as in the CPS, see Appendix B.1.

[59] Wu (2021) uses the CPS data and estimates a sizeable decline in the income tax progressivity during 1970–2010. At the same time, Heathcote et al. (2020) find that the income tax progressivity has changed only insignificantly during a similar period. In the quantitative experiments, I will discuss a counterfactual in which income tax progressivity remains constant over time.

[60] I borrow this approach from Corbae et al. (2009).

taxes are progressive although the income tax τ_I is linear.

I set parameter φ such that the ratio of total EITC expenditures to GDP is equal to 0.03% in the 1970s calibration and 0.31% in the 2010s calibration, consistent with the data.[61] I calibrate τ_I inside the model such that the average effective income tax rate defined in (39) is equal to the CBO estimates of 14.4% in the 1970s and 12.4% in the 2010s calibration. As a result, τ_I drops from 18.6% in the 1970s to 16.6% in the 2010s.

I borrow the estimates for consumption tax, τ_c, and capital income tax τ_a from Wu (2021) who applies the methodology developed by Mendoza et al. (1994) and Trabandt and Uhlig (2011). According to his results, τ_c declines from 5.3% in the 1970s to 4.1% in the 2010s. Congruently, τ_a falls from 38.4% to 33.0%.

To calibrate the share of wasted government spending in GDP, gy, which pins down G in the government budget constraint (16), I proceed as follows. I sum up federal mandatory and discretionary government outlays, net of: 1) outlays on income security, which comprises food and nutrition assistance, family assistance, and EITC, and 2) outlays on Social Security and Medicare.[62] The resulting fraction of wasted spending in GDP is equal to 9.5% in the 1970s and 7.8% in the 2010s.

Finally, the share of federal government debt in GDP, dy, is set to 35% for the 1970s calibration and 100% for the 2017 calibration.[63]

6. Findings

I present the findings in the following order. Section 6.1 uncovers distributional conflicts between agents and sheds light on the trade-offs faced by the government. I exploit these distributional conflicts to identify the parameters of the Pareto weight distribution. Section 6.2 rationalizes the change in the observed income tax and Social Security policies during 1970–2010 into the impact of demographic and economic changes, on the one hand, and the impact of government preferences, on the other hand. Finally, Section 6.3 takes a granular

[61] The data are provided by the Office of Management and Budget at the White House and is publicly available at `https://www.whitehouse.gov/omb/historical-tables/`. For the EITC expenditures, see their Table 8.5. The GDP time series is reported in their Table 10.1.

[62] Since I assume that the EITC is part of the transfer income, I exclude it from the calculation of the government spending.

[63] The data is provided by the Office of Management and Budget at the White House and is publicly accessible at `https://www.whitehouse.gov/omb/historical-tables/`. Total federal outlays are reported as the sum of: 1) outlays for discretionary programs (their Table 8.7) and 2) outlays for mandatory and related programs (their Table 8.5). The share of gross federal government debt in GDP is reported in Table 7.1.

approach on rising earnings inequality by disentangling its driving forces and analyzing the mechanisms through which they affect the optimal outcome.

6.1. Pareto weight distribution and optimal policy

The impact of demographic and economic changes on the optimal policy will depend on the distribution of age- and education-specific Pareto weights. The current section explains how these weights are obtained. More specifically, Section 6.1.1 uncovers heterogeneity in the agents' preferences by age and education, while Section 6.2.2 exploits this heterogeneity to identify the Pareto weight parameter $\boldsymbol{\kappa}$.

6.1.1. Preferences over income taxation and Social Security

Consider the government's maximization problem in (34). Denote the i-th policy instrument in the set $\{\tau_I, \alpha_1, \alpha_2\}$ by Υ_i. Then, the first-order optimality condition with respect to policy Υ_i is given by

$$\sum_j \int \omega(j, z; \boldsymbol{\kappa}) \frac{\partial V_1(\boldsymbol{x}; \Upsilon)}{\partial \Upsilon_i} dF_{1,j} = 0. \tag{40}$$

Recall that $V_1(\boldsymbol{x}; \Upsilon)$ is the value function of an agent in state $\boldsymbol{x} = (z, j, j^R, \nu, y, \bar{e}, h, a)$ in the first period of the transition from a steady state associated with the actual policy $\tilde{\Upsilon}$ to a new steady state associated with the policy Υ; F_1 is the pre-reform stationary distribution of households. This optimality condition shows that the government's optimal choice of Υ_i depends on three model ingredients: 1) the Pareto weights $\omega(j, z; \boldsymbol{\kappa})$, 2) the population density F_1, and 3) the partial derivative of the agent's value function with respect to policy Υ_i. The last object is the focus of the current section.

I evaluate the partial derivatives $\partial V_1(\boldsymbol{x}; \Upsilon)/\partial \Upsilon_i$ at the true policy $\tilde{\Upsilon}$ because, by construction of Pareto weights, the optimal solution *must* coincide with the policy in the data.[64] Furthermore, I analyze the marginal welfare effects by the agent's age and education because these are the arguments of the Pareto weight function (33). I average out any differences in welfare along other dimensions of $\boldsymbol{x}$ using the stationary distribution of agents in the pre-reform steady state, F_1. To further simplify the exposition of results, I pool agents into 5-year bins by age: 25–29, 30–34, etc. and report the average derivative within each age/education bin.

In all conducted experiments, the partial derivatives are concave in each policy instru-

[64] Appendix A.3 explains the computational procedure to obtain the derivatives.

ment.[65] Hence, the most preferred policy Υ_i of an agent, who is located in state $\mathbf{x}$ in the pre-reform distribution, must lie below the observed policy $\tilde{\Upsilon}_i$ if the derivative of the agent's value function with respect to Υ_i, is strictly negative; the opposite is true if the derivative is strictly positive.

Suppose that, conditional on age, the derivatives of the value function with respect to policy Υ_i for college graduates and non-college graduates have opposing signs, meaning that one educational group is in favor of increasing Υ_i above the actual level in the data, while the other group prefers to decrease it. I will refer to such situations as a *major* disagreement below. In this case, the model will be able to match the actual policy $\tilde{\Upsilon}_i$ by properly setting the educational bias κ_2. Similarly, suppose that, conditional on education, the sign of the derivatives changes as agents become older, meaning that young and old agents have opposing views on how to change the existing policy. Again, the model will be able to rationalize the true policy $\tilde{\Upsilon}_i$ by properly choosing the age bias κ_1. The existence of such major disagreements is what ensures the identification in the first place.

While the derivative's sign identifies major disagreements, the derivative's *magnitude* shows the size of the welfare gain (if the derivative is positive) or the size of the welfare loss (if the derivative is negative) induced by a marginal increase in policy Υ_i. All else equal, the higher the magnitude of a derivative of an agent in a given state, the closer will the optimal policy be located to the most preferred policy of that agent.

After making these preliminary points, I plot in Figure 8 the age- and education-specific welfare profiles induced by a marginal change in the income tax policy τ_I (left panel), the average replacement rate α_1 (middle panel), and pension system progressivity α_2 (right panel). The figure is constructed based on the 1970s calibration of the model. Below, I discuss the effects of each policy in turn.

Income taxation

Consider first the welfare implications of a marginal increase in the income tax rate τ_I. One can see that income taxation generates a two-dimensional distributional conflict. On the one hand, there is an inter-generational conflict. All agents above age 50 prefer to increase τ_I above the actual level, regardless of their educational status. By contrast, young workers (age 25–34) are unanimously in favor of lowering the tax. On the other hand, there is a distributional conflict between college graduate workers and non-college graduate workers

[65] I do not have a formal proof of concavity; however, I check that the indirect utility function satisfies this property for every state $\mathbf{x} = (z, j, j^R, \nu, y, \bar{e}, h, a)$ in the calibrated model. Corbae et al. (2009) take a similar approach when proving the existence of a median voter in the economy à la Aiyagari (1994).

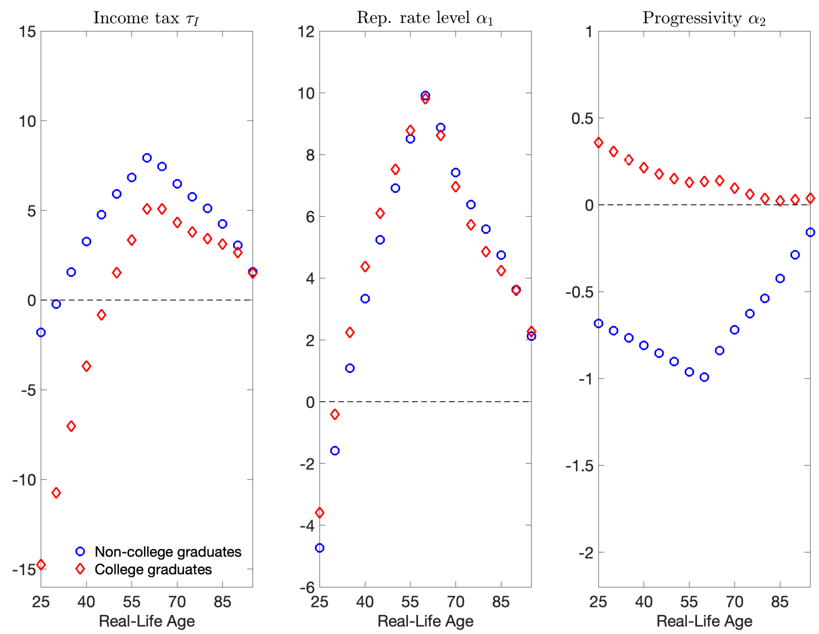


Figure 8: Age-profile of marginal welfare effects of policy $\Upsilon_i = \{\tau_I, \alpha_1, \alpha_2\}$ in the 1970s calibration, by education.

Notes: The figure plots the partial derivative of the agent's value function with respect to income tax rate τ_I (left panel), average replacement rate α_1 (middle panel), and pension system progressivity α_2 (right panel). Derivatives are computed for pre-defined groups of agents by age and education, where I use the stationary distribution of agents in the pre-reform steady state, F_1, to compute averages. Derivatives are evaluated at the actual policy $\tilde{\Upsilon}$. All results are obtained under the 1970s calibration. Appendix A.3 explains the computational procedure to obtain the results.

aged 35–50. While college graduates favor more taxation compared to the data, the opposite is true for non-college graduates. Below, I explain both findings.

Recall that only workers' earnings are subject to income taxation in the model, while capital income is taxed at an exogenous rate. Since retired agents do not bear the cost of taxation, they prefer to increase τ_I above the prevailing level to enjoy a larger lump-sum income transfer T_t.[66] By contrast, a working-age agent faces a trade-off. Due to the permanent nature of reform, she must trade-off lower after-tax earnings during the remaining stage of her working career against the benefit of receiving a higher lump-sum transfer throughout her remaining life. Beside lower after-tax wages, workers suffer welfare losses because income taxation distorts their optimal labor supply, skill acquisition, and retirement decisions. According to the figure, the welfare cost of income taxation dominates for young workers of both educational levels. However, as workers age, the benefit starts to dominate. For non-college graduates, this happens already at age 35; for college graduates, this occurs much later in life, at age 50.

The finding that workers aged 25–34 of both educational groups prefer to reduce τ_I below the actual level can be explained by relatively small differences in hourly wages between the two groups.[67] As one can see from Table 1, which shows the calibrated model parameters, both types of workers enter the labor market with the same stock of skills, $h_{1,z}$, and almost identical innate ability levels θ_z. The major driver of wage disparity between the two groups is the wage premium $w_{t,H}/w_{t,L}$ in (8) which results from the complementarity of both educational types in the production sector. In the 1970s calibration, the model endogenously generates a skill premium of 36%, consistent with the CPS data.[68] Figure 8 suggests that the extent of the premium is insufficient for non-college graduate entrants to favor, on average, a permanent rise in income taxation.

Average replacement rate

Consider next the welfare implications of a marginal change in the average replacement rate α_1 (Figure 8, middle panel). One can immediately notice that there is a distributional conflict between workers aged 25–34 who prefer to reduce α_1 below status quo and the remaining population who favor a rise in replacement rates.

[66]By definition of a partial derivative, this result assumes that remaining policies (α_1, α_2) are fixed at the true levels. Hence, Figure 8 does not say that retired agents would want to increase τ_I above status quo *if* they could also set other policies. I will study the optimal policy in Section 6.2.

[67]The expression for the hourly wage is shown in (36).

[68]In the data, the wage premium is defined as the ratio of average hourly wages between college graduate and non-college graduate workers.

To understand this finding, consider how an increase in α_1 affects the agents' welfare. As the replacement rate schedule shifts upward permanently, all current and future retirees enjoy higher pension benefits.[69] As the aggregate amount of pension entitlements rises, the Social Security tax rate must adjust upward to satisfy the government budget constraint in (14), thus reducing after-tax wages, $(1-\tau_{SS,t})w_{t,z}$. Since retired agents do not bear the cost of Social Security taxation, they unanimously opt to raise α_1 above the status-quo level.

As opposed to retirees, workers face a trade-off. Similar to the above-discussed effect of income taxation, they suffer welfare losses from falling after-tax wages due to the rise in the Social Security tax and the distortionary effects of taxation on their optimal decision to work, learn, and retire. However, contrary to income taxation, they must trade off these costs not against higher lump-sum transfers but higher replacement rates during their retirement stage. Figure 8 suggests that the cost of a marginal and permanent increase in α_1 dominates for young workers aged 25–34. By contrast, all workers aged 35 and above choose to enlarge the existing public pension system.

Finally, the figure suggests that there is little disagreement over α_1 between college graduates and non-college graduates of the same age. On the one hand, college graduates should choose a smaller public pension system because they are persuaded to allocate a larger share of their earnings to the pension fund. On the other hand, college graduates can expect to receive pension benefits for a longer period because they face lower mortality rates.[70] Besides, there is a positive general equilibrium effect on the interest rate because a larger pension system discourages workers from saving privately.

To summarize, the middle panel of Figure 8 reveals a major disagreement over the average replacement rate α_1 between workers aged 25–34 and the remaining population.

Pension system progressivity

In the final step, consider the marginal welfare impact of pension system progressivity α_2 (Figure 8, right panel). It is immediately noticeable that there is a major disagreement between the two educational groups over policy α_2, regardless of age. Particularly, non-college graduates prefer to reduce α_2 below the actual level and make the pension system more progressive; the opposite is true for college graduates.

According to the replacement rate specification in (13), a rise in α_2 reduces pension benefits of all agents whose average lifetime earnings are below the economy-wide average

[69]The reader may view Figure 1 to see graphically the impact of a rise in α_1.

[70]The calibrated life expectancy gap between the two educational groups at age 25 is 4 years in the 1970s calibration.

taxable earnings $\tilde{E}_t$ and increases pension benefits for all other retirees. Since the model does not exhibit aggregate risk, all agents know with certainty the aggregate taxable earnings $\tilde{E}_t$. Hence, all retired agents know exactly their relative position in the distribution of lifetime earnings. Contrary to retirees, workers face idiosyncratic labor productivity risk, which makes their relative position at retirement uncertain, so their most preferred policy α_2 depends on the *expected* relative position at the time when these agents retire.

Recall that educated workers face a deterministic wage premium, $w_{H,t}/w_{L,t}$, during the entire working stage. Moreover, the 1970s calibration implies no heterogeneity in fixed effects among college graduates, while the variance of the idiosyncratic shock is fairly small. For these reasons, college graduates can expect, on average, to enter retirement with average lifetime earnings above the economy-wide average taxable earnings and, therefore, prefer to increase α_2 above the actual level. Ex-post, i.e., after the idiosyncratic risk has been realized, this is indeed what happens, such that retired agents with a college degree prefer to increase α_2. By contrast, non-college graduate workers expect their average earnings to fall below the economy-wide average earnings and, therefore, favor a reduction in α_2 below the status-quo level.

6.1.2. Identification of Pareto weights in the 1970s

Armed with intuition about the agents' preferences over policies $(\tau_I, \alpha_1, \alpha_2)$, I proceed to explain the identification of the Pareto weight parameter $\boldsymbol{\kappa}$. The results of this section are summarized in Table 2. All results are based on the 1970s calibration of the model.

Consider first a special case of the model when a utilitarian government maximizes the welfare of newborn agents only (upon observing their state $\mathbf{x}$). This is the prevailing assumption in the income taxation literature. The second column (*Newborns*) shows the optimal policy under this scenario. For comparison, the fourth column (*Baseline*) records the actual policy in the 1970s. As we can see, the optimal solution for the income tax is fairly close to the one in the data (17.8% versus 18.6%). Heathcote et al. (2020) have previously pointed out that a utilitarian approach can indeed rationalize the empirically observed income tax policy in the U.S., once one introduces endogenous skill acquisition, which is also part of my model.

However, the same table shows that the utilitarian solution implies a substantially lower average level of replacement rates (7% as opposed to 37% in the data). Moreover, the government chooses a more progressive pension system with $\alpha_2 = -2.3$ (compared to -0.42 in the data). To put the solution for (α_1, α_2) into perspective, note that the implied replacement rate of all agents whose average lifetime earnings fall below the bend point $\bar{e}_{\min}$ is 3.6 times

Table 2: Utilitarian policies versus actual policies in the 1970s.

	Equal weights		Baseline
	Newborns	All alive	
Optimal policy:			
Income tax $\tau_I^\star$, %	17.8	5.27	18.6
Average rep. rate $\alpha_1^\star$, %	7.0	91.0	37.0
Progressivity $\alpha_2^\star$	−2.3	−1.79	−0.42
Equilibrium variables:			
Average eff. income tax, %	X	X	14.4
Soc.Sec. tax $\tau_{SS,t}$, %	2.48	27.88	8.9

Notes: The table contrasts the actual policy (*Baseline*) with the utilitarian policies (*Equal weights*) under the 1970s calibration. The utilitarian solutions assumes that the government assigns the same weights to: 1) all newborn agents upon observing their initial conditions (column *Newborns*) and 2) all alive agents (*All alive*). The reported values of the Social Security tax rate $\tau_{SS,t}$ and the effective income tax are taken from the final steady state associated with the corresponding policy. Average effective income tax is defined in (39).

larger than in the data.[71] Essentially, the utilitarian government wants to shuts down the pension system for all retirees, except those unlucky workers who enter retirement with very low average lifetime earnings.

To sum up, a standard model in which the government maximizes the welfare of newborn agents fails to rationalize the actual income tax *and* Social Security policies. Furthermore, note that introducing education-specific Pareto weights would not resolve the issue. As we have seen in Figure 8 (middle panel), all newborn agents, regardless of their educational status, prefer to trim down the existing pension system.

Next, suppose that a utilitarian government maximizes the welfare of *all* alive agents. This case is documented in column *All alive*.[72] As we can see, the model performance becomes even worse than in the previously discussed scenario. Compared to the data, the utilitarian solution implies a significantly lower income tax (5.27% versus 18.6%), a substantially higher average replacement rate (91.0% versus 37.0%), and a more progressive system (−1.79 versus

[71] Based on my calibration strategy for the 1970s, the target value for the ratio $\bar{e}_{\min}/\mathtt{tilde}E$ is 0.18. Plugging the optimal solution for (α_1, α_2) from column *Newborns* in Table 2 into replacement rate schedule (13), one obtains $0.07 \times (0.18)^(- 2.3) = 3.6$. In the data, the corresponding value is only 90%. See Section 5.6.1 for details.

[72] This solution arises as a special case of the Pareto weight function in (33) with the age bias parameter κ_1 and the educational bias parameter κ_2 both equal to zero.

-0.42) than in the data. To quantify the total tax burden borne by the working-age population in this case, the table also reports the contribution rate $\tau_{SS,t}$.[73] Due to a generously high average replacement rate level, the equilibrium value of $\tau_{SS,t}$ is 19 percentage points higher compared to the status quo. Overall, the utilitarian government prefers income taxation to Social Security taxation, while we see the opposite composition in the data.

By construction, any deviation of the actual policy from the utilitarian solution must be captured by Pareto weights $\omega(\cdot;\boldsymbol{\kappa})$.

As I have just established, the utilitarian government chooses α_1, which is substantially above the status quo. Since the only group of agents who dislike Social Security are young workers (recall Figure 8, middle panel), the age-distribution of Pareto weights must be skewed toward younger workers. Consistent with this intuition, I obtain $\kappa_1 = 0.065$, where the positive sign implies that weights decline in the agents' age, conditional on their education.[74] To put this value into perspective, note that the implied weight on a newborn agent relative to an agent of the normal retirement age with the same education is 13.5. This finding is in stark contrast to the optimal income taxation literature, in which the age distribution locates its entire mass on newborn agents.

Since the mass of Pareto weights must be centered around younger workers, my further argument will focus on the preferences of young and middle-aged workers. In the previous section, I have detected a major disagreement over the income tax policy τ_I and pension system progressivity α_2 between college graduate and non-college graduate workers. All else equal, a larger weight on educated agents leads to a lower income tax. Similarly, a larger weight on educated agents results in a less progressive pension system. The estimated value of the educational bias κ_2 must, therefore, strike a balance between matching the income tax policy, on the one hand, and the degree of pension system progressivity, on the other hand. I obtain $\kappa_2 = -0.998$, which means that non-college graduates receive a $1/\exp(-0.998) = 2.7$ times larger weight than college graduates of the same age.

Table 3 (column *Baseline (1970s)*) summarizes the estimates of $\boldsymbol{\kappa}$ that support the actual policy as an optimal outcome. The table also puts the obtained estimates of $\boldsymbol{\kappa}$ into perspective. In particular, it shows the implied Pareto weight of a 25-year-old agent relative

[73] Note that the optimal policy $\Upsilon^\star$ induces an entire transitional path of all endogenous variables, including $\tau_{SS,t}$, from the pre-reform steady state associated with policy $\tilde{\Upsilon}$ to the final steady state associated with $\Upsilon^\star$. In the table, I report the value of $\tau_{SS,t}$ from the *final* steady state. Appendix A.2 explains the numerical algorithm to compute transitional dynamics.

[74] The reader might find it useful to return to Figures 2–3 which plot Pareto weights as a function of κ_1 and κ_2.

Table 3: Estimated parameters of the Pareto weight function and the implied weights.

	Baseline (1970s)	Baseline (2010s)
Age bias, κ_1	0.065	0.078
Implied weight on age 25 / age 64:		
– Pareto weight	13.46	22.64
– Effective weight	25.6	42.77
Educational bias, κ_2	-0.998	0.971
Implied weight on col./ non-col.:		
– Pareto weight	0.37	2.64
– Effective weight	0.13	2.14

Notes: The table shows the estimates of the Pareto weight parameter κ in the Pareto weight function (33) under which the actual policy $\tilde{\Upsilon}$ arises as a solution to the government's maximization problem in (34). The table also reports the implied relative Pareto weight and the relative effective weight. For the age bias κ_1, the numbers are computed for a newborn agent (real-life age 25) relative to the agent of (real-life) age 66 with the same educational level. For the educational bias κ_2, the ratios are computed for a college graduate relative to a non-college graduate of the same age. The effective weight is defined as the product of the Pareto weight and the population share of agents in a given age-education group. Appendix A.3 explains the computational procedure to estimate κ.

to a 64-year-old agent with the same education and the implied weight on a college graduate relative to a non-college graduate of the same age. Note, however, that these are the weights assigned to a *single* agent located in a given age/education bin. The total weight on *all* agents located in a given (j, z) bin is given by a product of the Pareto weight $\omega(z, j; \kappa)$ and the total share of these agents in the population. I will refer to it as the *effective weight* below and report it in the table. As we can see, the effective weight on newborn agents is higher because young agents constitute a larger share in the population. Similarly, the effective weight on college graduates is smaller since their population share is only 25% in the 1970s.

To sum up, the distribution of Pareto weights must be skewed towards younger, uneducated agents in the 1970s for the model to be consistent with the actual policy in the data.

6.2. Evolution of the income tax and Social Security policies during 1970–2010

Return to Table 1 in the calibration section and compare the calibrated values of policies $(\tau_I, \alpha_1, \alpha_2)$ in each steady state. One can see that the policies have changed fairly little over time. In particular, the income tax rate τ_I has slightly declined (from 18.6% to 16.6%), the average replacement rate α_1 has slightly increased (from 37.0% to 39.0%), while pension system progressivity α_2 has remained almost unchanged during 1970–2010.

Through the lens of the model, two distinct forces affect the optimal solution. First, the government may find it optimal to adjust both programs in the 2010s because it faces new economic conditions, including increased earnings inequality, as well as new demographic challenges. Section 6.2.1 discusses this force. Second, the government may want to adjust the system because its tastes over equity and efficiency have changed over the past four decades. Section 6.2.2 quantifies this effect. As I will show, both forces roughly cancel out each other, which explains why policies changed insignificantly over time.

6.2.1. Impact of economic and demographic changes on the optimal policy

To quantify the impact of economic and demographic changes on the optimal income tax and Social Security programs, I will conduct the following thought experiment. Suppose that the U.S. economy is in a steady-state equilibrium in the 2010s. Except for some parameters that control a specific development in the U.S. economy (e.g., inequality), the fundamental parameters of this economy are otherwise identical to those in the 1970s. The income tax and Social Security policies in this economy are also assumed to be the same as in the 1970s. By construction of the Pareto weights, these policies must be optimal under the 1970s calibration, but will unlikely remain so due to the change in the fundamentals. Given this setting, I ask: What is the optimal government response?

I split all parameters whose values vary between the 1970s and 2010s calibrations into five groups which capture a certain development in the U.S. economy: 1) Rising earnings inequality, 2) Population aging, 3) Advances in production technology, 4) Changes in Social Security, and 5) Changes in other government policies. Table 4 lists the parameters in each group, while the respective values can be inferred from Table 1 in the calibration section.

In each experiment, I solve the government's maximization problem in (34) setting the model parameters from a given group to their 2010s values, while keeping the remaining parameters at their 1970s levels. To tease out potential complementarities between the three policy instruments, I solve for two special cases of the government's maximization problem. In particular, I first solve for the optimal income tax rate $\tau_1^\star$, assuming that the government cannot adjust Social Security, which I parameterize following the 1970s calibration. This scenario tells me how much redistribution the government is able to achieve with income taxation, only. Next, I resolve the government's problem allowing the policymaker to optimally choose $\tau_1^\star$ *and* the average replacement rate, $\alpha_1^\star$, while keeping pension system progressivity α_2 fixed. This experiment introduces a trade-off between paying transfers to all households or

Table 4: Parameters adjusted in counterfactual experiments.

Experiment	Parameters changed
1. Rising earnings inequality	$(\theta_z, h_{1,z}, \delta^h, \rho, Z, \Pi_z, \rho_z, \sigma^2_{\epsilon,z}, \sigma^2_{\nu,z})$
2. Population aging	$(\psi_{z,j}, n)$
3. Production technology	(ϖ, δ)
4. Social Security	$(J^R, \bar{e}_{\min}, \delta^p, cap)$
5. Other government policies	(τ_c, τ_a, gy, dy)

Notes: The table shows the parameters that change their values between the 1970s and the 2010s calibrations of the model. Table 1 in the calibration section explains the parameter meaning and shows the initial and updated values.

rather targeting benefits at the retired population. In the final step, I solve for the full policy $\Upsilon^\star$. The comparison of the results of this experiment with those in the previous step allows me to understand how the sizes of the income tax and Social Security programs change when the government can choose how to target pension benefits. In all the experiments above, I maintain the assumption that the government preferences captured by the Pareto weight parameter κ remain fixed at the level calibrated in the 1970s. Such an experimental design disentangles the impact of varying government preferences on the optimal policy.

Table 5 presents the findings. To ease comparison across experiments, the optimal policies in each experiment are reported relative to the respective optimal policy under the 1970s calibration. The values for $(\tau_I^\star, \alpha_1^\star, \tau_{SS,t})$ are given in percentage points, while $\alpha_2^\star$ is expressed in percent deviations.

Consider the first special case when the government can set τ_I, only. As one can see, the government optimally increases the level of taxation in response to rising inequality and population aging, while in all other experiments the income tax goes down compared to the optimal policy in the 1970s. Among all experiments, inequality has the most pronounced effect on τ_I (11.4 pt.pt.).

Compare these results to the second special case, when the government is allowed to also set the average replacement rate, α_1. The previously described positive responses of $\tau_I^\star$ are mitigated in the inequality and aging counterfactuals, while other three experiments exhibit insignificant adjustments. In response to rising inequality, the government shrinks the Social Security system by reducing α_1 by 9.3%. Recall that the optimal policies are computed under the identified distribution of Pareto weights which puts a relatively high mass on workers in early stages of their working careers. As inequality rises, these workers reduce demand for a publicly provided pension system in exchange for higher after-tax wages. The reduced

Table 5: Optimal policy.

	All Changes	Decomposition: Ineq.	Aging	Prod.	Soc.Sec.	Other
I. Only income taxation:						
– Income tax $\tau_I^\star$, pt.pt.		+11.4	+2.8	−0.9	−1.3	−5.1
– Soc.Sec. tax $\tau_{SS,t}$, pt.pt.						
II. Income taxation and average replacement rate:						
– Income tax $\tau_I^\star$, pt.pt.		+5.2	+0.2	−1.5	−0.5	−5.4
– Aver. rep. rate $\alpha_1^\star$, pt.pt.		−9.3	+23.2	+7.2	+22.4	−1.4
– Soc.Sec. tax $\tau_{SS,t}$, pt.pt.						
III. Full model:						
– Income tax $\tau_I^\star$, pt.pt.		+8.8	−3.6	−3.3	−2.4	−5.2
– Aver. rep. rate $\alpha_1^\star$, pt.pt.		+14.0	+79.8	+42.9	+57.6	+42.9
– Progressivity $\alpha_2^\star$, %		−17.8	−16.1	−6.5	−20.3	−16.1
– Soc.Sec. tax $\tau_{SS,t}$, pt.pt.						

Notes: The table shows the optimal policy response to the economic and demographic changes which I obtain by solving the government maximization problem in (34). The Social Security tax rate $\tau_{SS,t}$ is taken from the final steady state associated with the corresponding policy. All variables are reported relative to the respective optimal solution under the 1970s calibration (in percentage point deviations for $(\tau_I^\star, \alpha_1^\star, \tau_{SS,t})$ and in percent deviations for $\alpha_2^\star$). *Only income taxation* shows the optimal income tax $\tau_I^\star$, assuming that (α_1, α_2) are fixed at the calibrated 1970s levels. *Income taxation and average replacement rate* shows optimal τ_I and α_1, assuming that α_2 is fixed. *Full model* presents the results when all three policies are endogenous. Column All changes reports the results when all model parameters are at their 2010s values. Columns *Decomposition* show the optimal policy when only a subset of the model parameters is updated to their 2010s values, while the remaining parameters stay constant at the calibrated 1970s levels. Table 4 lists the parameters updated in each counterfactual. See Table 1 for respective parameter values. All policies are computed under the Pareto weights from the 1970s (Table 3).

distortionary pressure on labor supply generates higher tax revenues used to pay for lump-sum transfers, so that the government does not require to raise income taxes as much as it had to do when it was persuaded to keep the old size of the pension system.

On the contrary, α_1 increases by 23.2% when the economy is confronted with population aging. There are several effects that operate in the same direction. First, the distribution of effective Pareto weights shifts towards older, more educated, workers who prefer Social Security to income taxation. Second, as retired agents are expected to live longer, the marginal gains from the public pension system increase.

In the final experiment (*Full model*), I allow the government to additionally choose how to allocate pension benefits depending on the retired agents' average lifetime earnings. One can see immediately that the results across all experiments change substantially. However, there is a common pattern in the responses that highlights how the three policy instruments interact. In all experiments, the government opts for higher replacement rates for earnings-poor agents by reducing α_2. As pension benefits become targeted more precisely at earnings-poor agents, the government raises the average replacement rate α_1 so as to intensify the flow of resources towards the poor. To dampen the distortionary pressure of Social Security taxation on workers, the government reduces τ_I in all experiments, except inequality, by more than it did in the previous two cases. In response to rising inequality, income taxes increase by 8.8 pt. pt. In Section 6.3, I will zoom onto the optimal response of income inequality.

Finally, I examine the implications of the optimal policies for the households' welfare. Consider a counterfactual U.S. economy in which the new economic and demographic conditions have already taken place, but the government has not yet responded and adjusted the income tax and Social Security programs. How much would the households value the optimal system? I find that the agents in the 2010s would be willing to pay 1.16% in consumption equivalent terms.[75]

6.2.2. Impact of government preferences on the optimal policy

In the previous section, I analyzed the optimal government response to the economic and demographic changes, assuming government preferences from the 1970s. Comparing the optimal solution to the actual policy in the 2010s, we see that. Through the lens of the model, the reason why the government in the 2010s chose to deviate from the optimal outcome is

[75]For both questions, I compute by how much percent the consumption of all living agents in the economy has to increase in all future periods and contingencies (keeping their leisure and bequests unchanged) so that their expected utility equals that under the alternative policy? I weigh all households equally when computing both welfare measures.

because its tastes over equity and efficiency had changed over the past four decades.

Return to Table 3 that summarizes the estimates of $\boldsymbol{\kappa}$ which support the actual policy as an optimal outcome in each steady state. Compared to the 1970s calibration, the age bias parameter κ_1 has increased. The implied Pareto weight on a 25-year-old agent relative to a 64-year-old agent with the same education rises from 13.5 in the 1970s to 22.6 in the 2010s. At the same time, the educational bias parameter κ_2 has increased, too, switching its sign to positive. The estimates imply that the relative weight on a college graduate increased substantially from 0.4 in the 1970s to 2.6 in the 2010s.

Similar to the previous section, I examine the implications of changed government's preferences for the households' welfare. Consider the current U.S. economy, in which the new economic and demographic conditions have already taken place and the government has implemented the actual policy calibrated for the 2010s. As we have already established, the actual policy reflects the government's new tastes over efficiency and redistribution. I ask: How much would currently lived U.S. households value a counterfactual economy with new economic and demographic conditions but the policy optimally chosen by the government with old preferences? I find that the agents in the 2010s would be willing to pay 0.6% consumption equivalent terms.

6.3. Impact of rising earnings inequality on the optimal policy

In the previous section, I quantified the effect of earnings inequality on the optimal policy. The purpose of the current section is to zoom onto this finding finding. In particular, I will decompose the drivers of inequality and detect those forces that have the most pronounced effect on the optimal solution quantitatively. Also, I will uncover and discuss the mechanisms through which these forces affect the optimal solution.

To this end, I split all model parameters that control inequality and whose calibrated values vary between the 1970s and the 2010s into five groups: 1) Supply of college graduates, 2) Human capital accumulation, 3) Fixed effects, 4) Skill complementarity, and 5) Idiosyncratic labor productivity risk. Then I analyze the optimal policy, proceeding in exact same way as I did in Section 6.2.1. I first solve the model for a steady-state equilibrium, updating the parameters of a given group to their 2010s values and fixing all the remaining parameters at the 1970s levels. In this setting, I solve the government maximization problem in (34). In all experiments, I assume that the Pareto weights, governed by the parameter $\boldsymbol{\kappa}$, stay at their identified levels from the 1970s calibration.

Table 6 shows the parameters of each group that I update in the experiments. The numerical values of these parameters can be taken from Table 1 in the calibration section.

Table 6: Parameters adjusted in counterfactual experiments with earnings inequality.

Counterfactual	Parameters updated
1. Supply of college graduates	Π_z
2. Human capital accumulation	$(\theta_z, h_{1,z}, \delta^h)$
3. Fixed effects	$\sigma^2_{\nu,z}$
4. Skill complementarity	(ρ, Z)
5. Idiosyncratic risk	$(\rho_z, \sigma^2_{\epsilon,z})$

Notes: The table shows the parameters of earnings inequality that change their values between the 1970s and the 2010s calibrations of the model. This table accompanies the results displayed in Tables 7. Table 1 in the calibration section shows the initial and updated parameter values.

Table 5 displays the optimal policy results. Given that the set of instruments is three-dimensional, it appears instructive to compute the optimal policy in steps. In particular, I first solve for the optimal income tax $\tau_I^\star$, assuming that the government takes Social Security, calibrated to the 1970s, as given. Next, I allow the government to optimally choose $\tau = I^\star$ and the average replacement rate, $\alpha_1^\star$, keeping pension system progressivity α_2 fixed. These two experiments allow me to quantify the distributional conflict between young workers and the remaining population. Then, I compare the optimal policy response allowing the government to also choose α_2. This experiment sheds light on the distributional conflict between college graduates and non-college graduates.

Below I discuss each counterfactual in order.

1) Supply of college graduates

In the first counterfactual experiment, I study the implications of an increased supply of college graduates. To this end, I update the distribution of newborn agents by education denoted by Π_z. Based on my calibration results, the share of college graduates almost doubles, from 25% to 44% during 1970–2010s.

Consider first a special case of the model when the government has access to the income tax instrument only. Despite a relatively large perturbation of the model, the government's response is to increases $\tau_I^\star$ by 0.5 pt.pt., as we can see from Table 7. The reason for this modest response is because there are opposing forces at work.

First, there is a direct effect operating through effective Pareto weights. As the supply of college graduates increases, so does their effective weight, even though the Pareto weight parameter κ are assumed to stay fixed. Since college graduates prefer, on average, lower income taxes that non-college graduates (see Figure 8, left panel), one should see a decline

Table 7: Driving forces of earnings inequality and the optimal policy.

	Inequality	Decomposition:				
		Supply	HCA	Fixed	Compl.	Risk
I. Only income taxation:						
– Income tax $\tau_I^\star$, pt.pt.	+11.4	+0.5	+12.5	+8.0	+9.0	−0.8
– Soc.Sec. tax $\tau_{SS,t}$, pt.pt.						
II. Income taxation and average replacement rate:						
– Income tax $\tau_I^\star$, pt.pt.	+5.2	+0.4	+8.1	+7.9	+18.6	−1.3
– Aver. rep. rate $\alpha_1^\star$, pt.pt.	−9.3	+0.1	−0.1	−0.3	−0.1	+0.1
– Soc.Sec. tax $\tau_{SS,t}$, pt.pt.						
III. Full model:						
– Income tax $\tau_I^\star$, pt.pt.	+8.8	−4.4	+7.2	+0.2	+5.8	−5.0
– Aver. rep. rate $\alpha_1^\star$, pt.pt.	+14.0	+39.9	+84.9	−12.3	+27.2	+66.5
– Progressivity $\alpha_2^\star$, %	−17.8	−21.2	−32.0	−11.6	−15.0	−21.8
– Soc.Sec. tax $\tau_{SS,t}$, pt.pt.						

Notes: The table shows the optimal policy response to the underlying driving forces of rising earnings inequality which I obtain by solving the government maximization problem in (34). The Social Security tax rate $\tau_{SS,t}$ is taken from the final steady state associated with the corresponding policy. All variables are reported relative to the respective optimal solution under the 1970s calibration (in percentage point deviations for $(\tau_I^\star, \alpha_1^\star, \tau_{SS,t})$ and in percent deviations for $\alpha_2^\star$). *Only income taxation* shows the optimal income tax $\tau_I^\star$, assuming that (α_1, α_2) are fixed at the calibrated 1970s levels. *Income taxation and average replacement rate* shows optimal τ_I and α_1, assuming that α_2 is fixed. *Full model* presents the results when all three policies are endogenous. In each experiment, only a subset of the model parameters is updated to their 2010s values, while the remaining parameters stay constant at the calibrated 1970s levels. Tables 4 and 6 list the parameters updated in each counterfactual. See Table 1 for respective parameter values. All policies are computed under the Pareto weights from the 1970s (Table 3).

Table 8: Key variables of the steady-state equilibrium in the baseline model and counterfactual experiments.

	w_H/w_L	Above cap	$\text{Corr}(\bar{e}, y)$	Gini Earn.	Gini Inc.	Gini Inc.	
						L	H
Supply	-22.34	+1.78	-4.73	-0.32	-3.87	+ 0.22	+6.99
HCA	-19.09	+10.97	+12.98	+25.24	+1.51	+0.22	-27.72
Fixed	-0.15	+6.02	+10.69	+25.56	+10.54	+13.67	+12.95
Compl.	+57.43	-5.17	+13.59	+16.29	+19.14	+9.76	-0.52
Risk	-0.14	-1.46	+0.76	-7.67	+0.65	-0.22	+0.78

in $\tau_I^\star$. Second, recall that, even though both educational types enter the labor market with fairly similar initial conditions based on the 1970s calibration, college graduates receive a wage premium of 27% due to complementarities in the production sector. Therefore, a rise in the share of college graduates should amplify earnings inequality between college graduates and non-college graduate workers and give incentives to the government to increase τ_I, all else equal. However, there is an opposing general equilibrium effect because a higher supply of highly educated workers reduces the wage premium, w_H/w_L, given by (8). Indeed, consider Table 8 which records some of the key variables in the pre-reform steady-state equilibrium. According to the table, the wage premium drops from 27% to 4%. As the wage gap declines, the redistributive incentives become weaker. Finally, notice from Table 1 that educated workers face a higher dispersion of the idiosyncratic labor productivity risk than uneducated workers. When I shift the initial composition of workers towards educated workers, these agents opt for a higher for insurance purpose (recall that the Pareto weight distribution puts a larger weight on younger workers). When we combine all of the described effects, we obtain a small increase in $\tau_I^\star$.

According to Table 7, the optimal income tax increases compared to the baseline model in the 1970s, though only slightly (0.5 pt.pt.) due to the offsetting nature of the effects described above. Indeed, Table 8 confirms that inequality in earnings and incomes remains virtually unchanged compared to the 1970s.

As opposed to the previous two counterfactuals, the mechanism in this experiment operates through the density function in the government's optimality condition. As I have already established in Section 6.1, high-ability agents across all age groups prefer a less progressive system than in the data. Hence, as their population size increases, the government's optimal response should be to reduce progressivity to cater to these agents' preferences. Indeed, Table 7 shows that α_2 goes down by 5.3% compared to the 1970s. The government complements this choice by reducing the average replacement rate by 8.9%. As the same table reveals,

Table 9: Earnings inequality and estimated Pareto weight parameters.

	Baseline (1970s)	Ineq.	Decomposition: Hum.cap.	Compl.	Supply	Fix.eff.	Risk
Age bias, κ_1	0.065	0.098	0.093	0.075	0.059	0.107	0.059
– Weight age 25/66	13.5	50.4	41.3	20.1	10.6	72.2	10.6
Educ. bias, κ_2	-0.998	1.822	0.603	0.488	-2.208	1.608	-1.582
– Weight col./non-col.	0.4	6.2	1.8	1.6	0.1	5.0	0.2

Notes: The table shows the estimates of the Pareto weight parameter κ in the Pareto weight function (33) in counterfactual experiments when only a subset of model parameters is updated to their 2010s values. Column *Baseline (1970s)* shows the estimates consistent with the actual policy under the 1970s calibration. Column *Ineq.* shows the total impact of rising earnings inequality on κ. There are five forces analyzed: 1) human capital accumulation, 2) skill complementarity, 3) supply of college graduates, 4) fixed effects, and 5) idiosyncratic labor productivity risk. Table 6 lists the parameters updated in each counterfactual; the respective parameter values can be retrieved from Table 1. All remaining parameters stay constant at the calibrated 1970s levels. Row *Weight age 25/66* shows the implied weight of a newborn agent (real-life age 25) relative to the weight of an agent at (real-life) age 66 with the same educational level. Row *Weight col./non-col.* reports the implied weight on a college graduate relative to a non-college graduate of the same age. Appendix A.3 explains the computational procedure to estimate κ.

this measure dampens the distortionary pressure of labor taxation by reducing the Social Security tax rate by 0.8 percentage points in the long run. All working-age agents benefit, especially those with high-ability because their pre-tax wages are higher, on average.

Table 9 decomposes the impact of the driving forces behind earnings inequality on the identified shift in the parameter vector κ.

7. Outlook

The analysis in this paper is subject to several critique points. I mention two of them below. Addressing each of these points opens exciting new avenues for future research.

First, I identify the shift in government preferences over insurance and redistribution but abstain from exploiting this information in further analysis. There could be a benefit from doing so. As already mentioned in the introduction, a large strand of economic literature studies the macroeconomic and welfare consequences of different retirement financing reforms. It remains unclear, however, to what extent the discussed reforms are feasible from a political standpoint. The identified distribution of Pareto weights can be applied in policy analyses to restrict all economically feasible proposals to those that are also politically viable.

Second, the shift in government preferences is *one* explanation for why Social Security has not adjusted to rising earnings inequality. Admittedly, there are alternative justifications. One such explanation is political gridlock. It describes a situation in which politicians fail to reach an agreement during the post-election bargaining stage. As a consequence, the policy remains at status-quo. Piguillem and Riboni (2016) have an interesting application of this mechanism to capital taxation. Their paper's mechanism might provide micro-foundations for the persistence of particular policies, including Social Security, despite the change in fundamentals.

References

Abbott, B., Gallipoli, G., Meghir, C., Violante, G.L., 2019. Education policy and intergenerational transfers in equilibrium. Journal of Political Economy 127, 2569–2624.

Aiyagari, S.R., 1994. Uninsured Idiosyncratic Risk and Aggregate Saving. The Quarterly Journal of Economics 109, 659–684.

Badel, A., Huggett, M., Luo, W., 2020. Taxing Top Earners: a Human Capital Perspective. The Economic Journal 130, 1200–1225.

Bartels, L., 2009. Unequal Democracy: The Political Economy of the New Gilded Age. Princeton University Press.

Bell, F.C., Wade, A.H., Goss, S.C., 1992. Life Tables for the United States Social Security Area. Actuarial Study 116, 1900–2100.

Benabou, R., 2000. Unequal Societies: Income Distribution and the Social Contract. American Economic Review , 96–129.

Bound, J., Geronimus, A., Rodriguez, J., Waidmann, T., 2014. The Implications of Differential Trends in Mortality for Social Security Policy. Working Papers 314. University of Michigan, Michigan Retirement Research Center.

Brendler, P., 2020. Why Hasn't Social Security Changed since 1977? Review of Economic Dynamics 36, 134 – 157.

Carceles-Poveda, E., Abraham, A., Unpublished. Tax Reform with Endogenous Borrowing Limits and Incomplete Asset Markets.

Chang, B.H., Chang, Y., Kim, S.B., 2018. Pareto Weights in Practice: A Quantitative Analysis Across 32 OECD Countries. Review of Economic Dynamics 28, 181 – 204.

Chang, B.H., Chang, Y., Kim, S.B., 2021. Ex-Ante Heterogeneity and Equilibrium Redistribution. Working Paper.

Conesa, J.C., Krueger, D., 1999. Social Security Reform with Heterogeneous Agents. Review of Economic Dynamics 2, 757–795.

Corbae, D., D'Erasmo, P., Kuruscu, B., 2009. Politico-economic consequences of rising wage inequality. Journal of Monetary Economics 56, 43–61.

De Nardi, M., 2004. Wealth Inequality and Intergenerational Links. The Review of Economic Studies 71, 743–768.

Deaton, A., Paxson, C., 1994. Intertemporal choice and inequality. Journal of Political Economy 102, 437–467.

Elo, I.T., Preston, S.H., 1996. Educational Differentials in Mortality: United States, 1979–1985. Social Science & Medicine 42, 47 – 57.

Fehr, H., Habermann, C., 2008. Risk Sharing and Efficiency Implications of Progressive Pension Arrangements. The Scandinavian Journal of Economics 110, 419–443.

Fehr, H., Kallweit, M., Kindermann, F., 2013. Should Pensions Be Progressive? European Economic Review 63, 94 – 116.

Fuster, L., Imrohoroglu, A., Imrohoroglu, S., 2007. Elimination of Social Security in a Dynastic Framework. The Review of Economic Studies 74, 113–145.

Guvenen, F., Kaplan, G., Song, J., Weidner, J., 2017. Lifetime Incomes in the United States over Six Decades. NBER Working Papers 23371. National Bureau of Economic Research, Inc.

Guvenen, F., Kuruscu, B., Ozkan, S., 2014. Taxation of human capital and wage inequality: A cross-country analysis. The Review of Economic Studies 81, 818–850.

Heathcote, J., Perri, F., Violante, G.L., 2010a. Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006. Review of Economic Dynamics 13, 15–51.

Heathcote, J., Storesletten, K., Violante, G.L., 2010b. The Macroeconomic Implications of Rising Wage Inequality in the United States. Journal of Political Economy 118, 681–722.

Heathcote, J., Storesletten, K., Violante, G.L., 2017. Optimal Tax Progressivity: An Analytical Framework. The Quarterly Journal of Economics 132, 1693–1754.

Heathcote, J., Storesletten, K., Violante, G.L., 2020. How should tax progressivity respond to rising income inequality? Journal of the European Economic Association 18.

Hosseini, R., Shourideh, A., 2019. Retirement financing: An optimal reform approach. Econometrica 87, 1205–1265.

Huggett, M., 1996. Wealth distribution in life-cycle economies. Journal of Monetary Economics 38, 469–494.

Huggett, M., Parra, J.C., 2010. How Well Does the U.S. Social Insurance System Provide Social Insurance? Journal of Political Economy 118, 76–112.

Huggett, M., Ventura, G., 1999. On the Distributional Effects of Social Security Reform. Review of Economic Dynamics 2, 498–531.

Jacobs, B., Jongen, E.L., Zoutman, F.T., 2017. Revealed Social Preferences of Dutch Political Parties. Journal of Public Economics 156, 81–100.

Katz, L.F., Murphy, K.M., 1992. Changes in relative wages, 1963-1987: Supply and demand factors. The Quarterly Journal of Economics 107, 35–78.

Kitao, S., 2014. Sustainable Social Security: Four Options. Review of Economic Dynamics 17, 756–779.

Krueger, D., Ludwig, A., 2016. On the optimal provision of social insurance: Progressive taxation versus education subsidies in general equilibrium. Journal of Monetary Economics 77, 72–98.

Lindbeck, A., Weibull, J.W., 1987. Balanced-Budget Redistribution as the Outcome of Political Competition. Public Choice 52, 273–297.

Lockwood, B.B., Weinzierl, M., 2016. Positive and Normative Judgments Implicit in U.S. Tax Policy and the Costs of Unequal Growth and Recessions. Journal of Monetary Economics 77, 30–47.

McGrattan, E.R., Prescott, E.C., 2017. On Financing Retirement with an Aging Population. Quantitative Economics 8, 75–115.

Meara, E.R., Richards, S., Cutler, D.M., 2008. The Gap Gets Bigger: Changes In Mortality And Life Expectancy, By Education, 1981–2000. Health Affairs 27, 350–360.

Mendoza, E.G., Razin, A., Tesar, L.L., 1994. Effective Tax Rates in Macroeconomics: Cross-country Estimates of Tax Rates on Factor Incomes and Consumption. Journal of Monetary Economics 34, 297 – 323.

Moser, C., Olea de Souza e Silva, P., 2019. Optimal Paternalistic Savings Policies. Working Paper 17-51.

Ndiaye, A., 2020. Flexible Retirement and Optimal Taxation. Federal Reserve Bank of Chicago Working Paper 2018-18.

Nishiyama, S., 2019. The Joint Labor Supply Decision of Married Couples and the U.S. Social Security Pension System. Review of Economic Dynamics 31, 277 – 304.

Nishiyama, S., Smetters, K., 2007. Does Social Security Privatization Produce Efficiency Gains? The Quarterly Journal of Economics 122, 1677–1719.

Piguillem, F., Riboni, A., 2016. Dynamic Bargaining over Redistribution in Legislatures. Working Papers 2016-15. Center for Research in Economics and Statistics.

Rosenstone, S., Hansen, J., 1993. Mobilization, Participation, and Democracy in America. Macmillan, New York.

Social Security Administration, 2019. Annual Statistical Supplement To The 2019 Annual Report Of The Board Of Trustees Of The Federal Old-Age And Survivors Insurance And Federal Disability Insurance Trust Fund. Technical Report. The Board Of Trustees.

Stantcheva, S., 2016. Comment on Positive and Normative Judgments Implicit in U.S. Tax Policy and the Costs of Unequal Growth and Recessions by Benjamin Lockwood and Matthew Weinzierl. Journal of Monetary Economics 77, 48–52.

Storesletten, K., Telmer, C.I., Yaron, A., 2004. Consumption and risk sharing over the life cycle. Journal of Monetary Economics 51, 609–633.

Trabandt, M., Uhlig, H., 2011. The Laffer Curve Revisited. Journal of Monetary Economics 58, 305 – 327.

Tsujiyama, H., Heathcote, J., 2015. Optimal Income Taxation: Mirrlees Meets Ramsey. Staff Report 507. Federal Reserve Bank of Minneapolis.

Wu, C., 2021. More unequal income but less progressive taxation. Journal of Monetary Economics 117, 949–968.

Appendix A. Computational algorithm

Appendix A.1. Steady-state equilibrium

Follow the steps below to compute a steady-state equilibrium defined in Section 3.6, see Definition 2.

1. Make initial guesses of the steady-state values of aggregate capital stock K, aggregate effective labor supplies of each skill $\{N_H, N_L\}$, lump-sum transfer T, Social Security tax rate τ_{SS}, and average taxable earnings $\tilde{E}$.
2. Given the guesses on K and $\{N_H, N_L\}$, compute the total effective labor supply N from (4), total output Y from (5), the skill-specific wages $\{w_H, w_L\}$ from (6), and the interest rate r from (7).
3. Given Y, compute the government debt D and wasted spending G using the constant debt-to-GDP and spending-to-GDP ratios $dy = D/Y$ and $gy = G/Y$.
4. Compute household's optimal labor, leisure, skill acquisition, consumption, savings, and retirement decisions $\{l, \ell, s, c, a', j^R\}$ starting at the final age J and proceeding backwards.
5. Initialize the time-invariant distribution of agents at age 1, $F_1(\mathbf{x})$, given that agents enter the model with zero assets and zero average lifetime earnings. Iterate the distribution forward using the computed decision rules in the previous step to compute the cross-sectional distribution $\{F_j\}_{j=1}^{j=J}$.
6. Compute the following aggregate quantities: pension benefits B from (24), asset supply A from (25), consumption C from (26), taxable income $\mathcal{I}$ from (27), and bequests Φ from (28).
7. Compute the new values of the guessed variables: Social Security tax rate τ_{SS} from (14), lump-sum transfer T from (16), average taxable earnings $\tilde{E}$ from (27), aggregate physical capital K from (29), and skill-specific total effective labor supplies $\{N_H, N_L\}$ from (30)-(31).
8. If the newly computed values of $\{K, N_H, N_L, T, \tau_{SS}, \tilde{E}\}$ are sufficiently close to the guesses from step 1, we have found a steady-state equilibrium. Otherwise, update the guesses, return to step 1, and repeat the steps until convergence.

Appendix A.2. Transitional dynamics

The economy is initially a steady-state equilibrium with policy $\tilde{\Upsilon} = (\boldsymbol{\alpha}_0, \tau_{I,0})$. At time $t = 0$, the government makes an unanticipated announcement that it will implement a constant future policy $\Upsilon = (\boldsymbol{\alpha}, \tau_I)$ in the next period. To solve for the transitional dynamics from the initial steady state under policy $\tilde{\Upsilon}$ to the final steady state under policy Υ proceed as follows.

1. Compute the initial steady state associated with policy $\tilde{\Upsilon}$ and the final steady state associated with policy Υ following the steps described in Appendix A.1. Denote the initial steady-state variables with a lower bar, e.g. $\{\underline{K}, \underline{V}\}$, and the final steady-state variables with an upper bar, e.g. $\{\overline{K}, \overline{V}\}$.
2. Assume that the transition from the initial steady state to the final steady state is completed within T periods.
3. Guess the paths of $\{K_t, N_{H,t}, N_{L,t}, T_t, \tau_{SS,t}, \tilde{E}_t\}_{t=0}^{T}$ with $K_0 = \underline{K}$, $N_{H,T} = \overline{N}_H$, $N_{L,T} = \overline{N}_L$, and $\tilde{E}_T = \overline{\tilde{E}}$.
4. Given the guessed paths of $\{K_t, N_{H,t}, N_{L,t}\}$, compute the path of total effective labor supply N_t from (4), total output Y_t from (5), skill-specific wages $\{w_{H,t}, w_{L,t}\}$ from (6), interest rate from (7), government debt D_t and wasted spending G_t using the constant debt-to-GDP and spending-to-GDP ratios $dy = D_t/Y_t$ and $gy = G_t/Y_t$ for all t.
5. Given the paths of $\{w_{H,t}, w_{L,t}, \tau_{SS,t}, T_t, \tilde{E}_t\}$,[76] solve the household's optimal labor, leisure, skill acquisition, consumption, savings, and retirement decisions during periods $1, ..., T-1$. To do so, proceed backwards from period $T-1$ to period 1. Note that the continuation value in the household's Bellman equation at time $T-1$ is known and given by $V_T = \overline{V}$.
6. Starting from the cross-sectional distribution of agents in the initial steady state, $\underline{F}$, apply the decision rules computed in the previous step to find the path of distributions, $\{F_{j,t}\}_{t=1}^{T}$.
7. Compute the paths of the following aggregate quantities: pension benefits B_t from (24), asset supply A_t from (25), consumption C_t from (26), taxable income $\mathcal{I}_t$ from (27), and bequests Φ_t from (28).
8. Compute the new paths of the guessed variables: Social Security tax rate $\tau_{SS,t}$, lump-sum transfer T_t from (16), average taxable earnings $\tilde{E}_t$ from (27), total capital K_t from (29), and education-specific total effective labor supplies $\{N_{H,t}, N_{L,t}\}$ from (30)-(31).
9. If the newly computed paths are sufficiently close to the guessed ones in each period, we have found the solution. Otherwise, update the guesses and return to step 3. Proceed until convergence.

[76] The path of $\tilde{E}_t$ determines the retired household's pension benefit from (12).

10. Once the sequence has converged, check whether the transition length T guessed in step 2 is sufficient by increasing T and checking whether the equilibrium paths are affected.

Appendix A.3. Estimating Pareto weights

This section explains how to estimate a vector of Pareto weight parameters κ under which the optimal policy is equal to the actual policy, i.e., $\Upsilon^\star = \tilde{\Upsilon}$. Before I lay out the computational algorithm, it is instructive to write down the first-order optimality conditions to the government optimization problem in (34). These conditions are given by:

$$\mathcal{A}_i \equiv \sum_j \int \omega(j, z; \kappa) \frac{\partial V_1(z, j, j^R, \nu, y, \bar{e}, h, a; \Upsilon)}{\partial \Upsilon_i} dF_{1,j} \Bigg|_{\Upsilon=\tilde{\Upsilon}} = 0, \quad \text{(A.1)}$$

where Υ_i refers to the i-th policy in the set $\{\tau_I, \alpha_1, \alpha_2\}$.

Observe that I evaluate the first-order derivative at the true policy $\tilde{\Upsilon}$ because this is what the solution to the government's optimization problem has to be, by construction. Furthermore, note that the initial distribution of agents over the state space, $F_{1,t}$, does not depend on policy Υ. Hence, once one obtains the derivatives of the value function and the initial distribution of agents, the parameter vector κ can be computed by solving the system of 3 equations given by (A.1) for $i = \{1, 2, 3\}$.

The necessary steps to compute κ can be summarized as follows:

1. Solve the model for the initial steady state equilibrium, as described in Appendix A.1, and obtain the stationary distribution of agents, F_1.
2. Following the algorithm described in Appendix A.2, solve for transitional dynamics from the initial steady state with $\Upsilon = \tilde{\Upsilon}$ to a new steady state associated with policy $\Upsilon_1^1 = (\tau_{I,0} - h_1, \alpha_{1,0}, \alpha_{2,0})$, where $h_1 > 0$ is a sufficiently small step. Denote the resulting value function by $V_1(z, j, j^R, \nu, y, \bar{e}, h, a; \Upsilon_1^1)$.
3. Repeat the previous step to solve for transitional dynamics from the initial steady state to a new steady state associated with policy $\Upsilon_1^2 = (\tau_{I,0} + h_1, \alpha_{1,0}, \alpha_{2,0})$. Denote the resulting value function by $V_1(z, j, j^R, \nu, y, \bar{e}, h, a; \Upsilon_1^2)$.
4. Compute numerically the derivative:

$$\frac{\partial V_1(z, j, j^R, \nu, y, \bar{e}, h, a; \Upsilon)}{\partial \tau_I} \Bigg|_{\Upsilon=\tilde{\Upsilon}} = \frac{V_1(z, j, j^R, \nu, y, \bar{e}, h, a; \Upsilon_1^2) - V_1(z, j, j^R, \nu, y, \bar{e}, h, a; \Upsilon_1^1)}{2h_1}.$$

5. Repeat steps (2)-(4) to compute the derivatives of the value function with respect to the policy instruments α_1 and α_2.

6. In a software such as Matlab create a function that executes the following steps:
   (a) The function takes as arguments some arbitrary value of $\boldsymbol{\kappa}$, the three derivatives of the value function computed in the previous steps, as well as the distribution of agents F_1 computed in step 1.
   (b) Given $\boldsymbol{\kappa}$, the function computes the set of Pareto weights $\omega(j, z; \boldsymbol{\kappa})$ using XXX.
   (c) Given $\omega(j, z; \boldsymbol{\kappa})$, the function evaluates the terms $\mathcal{A}_i$ in (A.1) and returns as output the sum of squared deviations $\sum_{i=1}^{3} \mathcal{A}_i^2$.
7. Pass the function constructed in the previous step to a standardized function minimization routine (such as *fmincon* in Matlab) and let the routine optimize over $\boldsymbol{\kappa}$.

Appendix B. Calibration

Appendix B.1. Sample selection

The Current Population Survey (CPS) data are available for the time period 1980–2018.[77] Unless otherwise stated, I use the 2010–2018 extracts for the 2010s calibration and the 1980–1985 extracts to calibrate the parameters in the initial steady state. Note that questions regarding income and earnings are retrospective (e.g., the 1980–1985 extracts contain the earnings and income data during 1979–1984).

The unit of observation is a household. I distinguish two levels of the household head's education level: 1) household heads with a completed college degree or higher, and 2) all remaining households with non-missing values of education. My CPS sample includes both male- and female-headed households of age 25–64. Following Heathcote et al. (2010a), I define earnings as wage income plus 2/3 of self-employment income. The household's total earnings are given by the sum of earnings of all its members. I drop a household if at least one member reports strictly positive earnings but zero hours worked. I drop all observations with non-positive household earnings.

When calibrating the parameters of the idiosyncratic AR(1) shock, I restrict the sample to household heads of age 25–61. As in Storesletten et al. (2004, Figure 3), the variance of log hourly wages (which is the calibration target) increases linearly during this part of the lifecycle, which is why I use it for estimation. For consistency reasons, I make the same selection restriction when estimating the parameters of the human capital accumulation process. To obtain hourly wages from the CPS, I divide the household's total earnings by the

[77]The CPS extracts are publicly available at http://ceprdata.org/cps-uniform-data-extracts/march-cps-supplement/march-cps-data/. The variables in the extracts are harmonized across years.

sum of annual hours worked among all of its members. I keep only those households whose head works at least 260 hours annually and whose hourly wage rate is above the minimum wage.[78]

I apply the last sample selection restriction regarding the minimum hours worked in the model when computing the moments of earnings and income inequality. More specifically, I drop all agents in the model who supply less than 5% of their unitary time endowment.[79]

Appendix B.2. Education-specific mortality differences

I assume that mortality rates are age- and education specific and follow a Gompertz force of mortality (i.e., mortality hazard) function

$$M_{z,j} = \mu_{1,z} \times \frac{\exp(\mu_2(j-1)) - 1}{\mu_2}. \tag{B.1}$$

The parameter $\mu_{1,z}$ shows how mortality varies with the agent's educational level z. The second term governed by parameter μ_2 controls how mortality changes with age j, conditional on education. Given this specification, the probability that the type z agent of age 1 survives up to age $j > 1$ is then given by

$$P_{z,j} = \Pi_{i=1}^{j-1} \psi_{z,i} = \exp(-M_{z,j}). \tag{B.2}$$

Thus, the *conditional* survival probability rate, i.e., the parameter $\psi_{z,j}$ in my model, can be computed recursively as: $\psi_{z,j} = P_{z,j+1}/P_{z,j}$ with $P_{z,1} \equiv 1$.

There are 3 parameters to calibrate: $(\mu_{1,L}, \mu_{1,H}, \mu_2)$. I calibrate them outside the model matching 2 targets. First, I require that the age profile of survival probability rates averaged among the two educational groups in the model matches the empirical counterpart provided by Bell et al. (1992). The authors report life expectancy separately for female and male Social Security covered workers. I compute the weighted average between the reported life expectancy using the population share of female household heads in the respective CPS sample. For the 1970s calibration, I use their data from 1970; for the 2010s calibration, I take their data from 2010.

[78] The time-series of the minimum wage can be found at https://www.dol.gov/agencies/whd/minimum-wage/history/chart.

[79] Carceles-Poveda and Abraham (Unpublished) analyze the Time Use Surveys of 2003–2005 and find that the household's disposable time is 97 hours per week after deducting sleep and personal care. This sums up to roughly 5,096 hours annually. The threshold of 5% is then computed as 260/5,096, which corresponds to supplying fewer than 260 hours annually.

Second, I target the age profiles of life expectancy by education reported by Bound et al. (2014) (see their Appendix 1). The authors use the data from the National Vital Statistics System and the Census from 1990–2010. They compute survival rates for a 25 year-old individual by education, gender, and race for 1990, 2000 and 2010. Their data are tabulated at 5-year intervals (age 30, 35, etc.). Since I am unaware of any other consistent data source on mortality by education, I take their data from 1990 to calibrate the 1970s steady state and the data from 2010 to calibrate the 2010s steady state. Bound et al. (2014) report survival rates for 4 educational groups: less than a high school degree, a completed high school degree, some years of college, and a complete college degree or higher. I pool the first three groups into a group that I defined as the non-college graduates in the paper. To average out differences in morality by gender, race, and education, I take the respective population shares from the CPS.

For the 1970s calibration, the parameters that result in the best fit are: $\mu_{1,L} = 0.001$, $\mu_{1,H} = 0.0007$, and $\mu_2 = 0.082$. For the 2010s calibration, the obtained estimates are: $\mu_{1,L} = 0.0006$, $\mu_{1,H} = 0.0003$, and $\mu_2 = 0.0855$.

Figure B.9 illustrates the fit of the two empirical targets for the 2010s calibration. The top panel shows the fit of the survival probability rates for an average 25-year-old individual. The solid line corresponds to the empirical data from Bell et al. (1992) that was the first calibration target. The bottom panel shows the fit of the age profiles of survival rates by education. The circle and diamond markers correspond to the data from Bound et al. (2014) that was the second calibration target. Overall the model achieves an acceptable fit. The performance of the model is qualitatively similar for the 1970s calibration.

Appendix B.3. Social Security

This section explains how to bring the empirical schedule of replacement rates to the model.

According to the Social Security legislation, an individual's pension benefit depends on their average monthly earnings. The monthly pension benefit of a worker who retires at the normal retirement age is determined by a statutory replacement rate schedule which is a function of the individual's average monthly earnings. The schedule comprises three brackets with a constant marginal replacement rate of 90%, 32%, and 15% in the lowest, intermediate, and highest bracket, respectively.

To bring the empirical schedule of replacement rates to the model, I conduct several transformations. I explain the procedure for the 2010s calibration, but the same steps apply to the 1970s calibration.

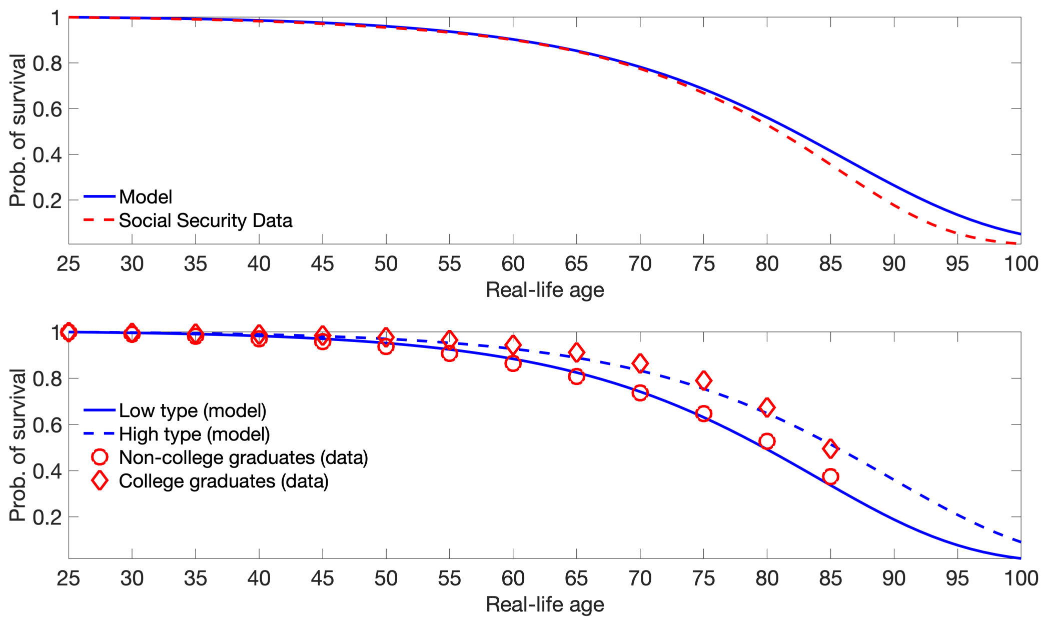


Figure B.9: Survival probability rates for a 25-year-old individual in the model and data.

Notes: Both panels show how the model calibrated to the 2010s fits the empirical survival probability rates. The top panel shows the fit of survival probability rates for an average 25-year old worker. The empirical rates (dashed line) were constructed based on the data by Bell et al. (1992) for 2010. The bottom panel compares the fit of survival probability rates for a 25-year-old worker by education. The low (high) type in the model corresponds to the agent with education level $z = \mathsf{L}$ ($z = \mathsf{H}$). The empirical moments (tabulated at 5-year intervals) are taken from Bound et al. (2014).

First, I convert the worker's average monthly earnings and the brackets into their annual counterparts (multiplying each by 12) since one period in my model is equal to one year. Figure B.10 shows the implied schedule of average replacement rates obtained after this step. The two vertical dashed lines correspond to the annualized bend points of \$10,620 and \$64,032.[80,81]

Second, I adjust the brackets to the average number of earners in a household, since the observation unit in my model is a household, while the statutory schedule applies to individuals. In the 2018 CPS extract, less than 3% of households consist of 4 or more earners. I disregard these households in the calculations. In the remaining sample, there are 47% of single-earner, 46% of two-earner, and 7% of three-earner households. Adjusting the first annualized bend point of 10,620 to the number of earners, I obtain: $10{,}620 \times [1 \times 0.47 + 2 \times 0.46 + 3 \times 0.07] = \$16{,}992$. I proceed similarly with the second annualized bend point of 64,032 to get \$102,451.

Third, I normalize the household's average annual earnings and the bend points by the economy-wide average taxable earnings. Figure 7 in the main text shows the replacement rate schedule after this transformation.

Finally, I estimate the parameter vector $\boldsymbol{\alpha}$ by fitting the replacement rate function $R(\bar{e}, j^R; \boldsymbol{\alpha})$ in (13) to the empirical schedule of average replacement rates depicted in the figure.

During the calibration, I also estimate the cap inside the model by targeting the share of households in the CPS whose total pre-tax earnings exceed the maximum taxable earnings threshold adjusted for the number of earners. The statutory value of the cap is $127,200$ in 2017.[82] Adjusting this value for the number of earners in a household computed above, I obtain $127{,}200 \times [1 \times 0.47 + 2 \times 0.46 + 3 \times 0.07] = 203{,}520$. Hence, I calibrate cap targeting the share of households with earnings above this threshold. I proceed similarly in the 1970s calibration.

[80] All dollar amounts in this section are given in 2017 U.S. dollars.

[81] The parameters of the statutory pension benefit formula can be found in Social Security Administration (2019, Table 2.A11).

[82] The data on the maximum taxable earnings threshold are taken from Table 2.A3 in Annual Statistical Supplement to Social Security Administration (2017) available at https://www.ssa.gov/policy/docs/statcomps/supplement/2017/index.html.

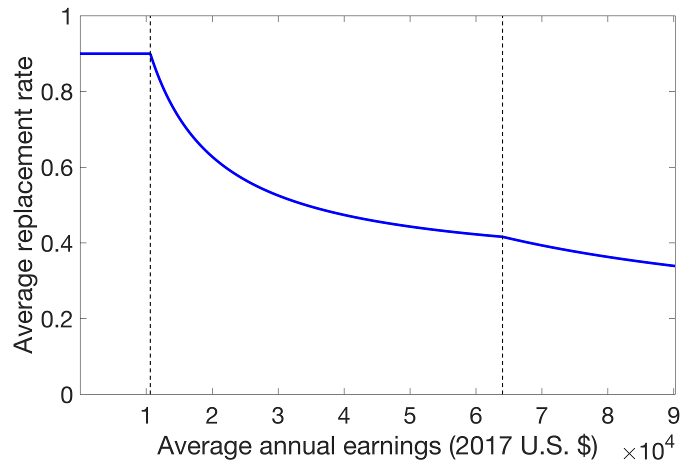


Figure B.10: Statutory Replacement Rate Schedule.

Notes: The figure visualizes the statutory relationship between the individual's average earnings and their average replacement rate. The figure is constructed based on the statutory pension benefit formula from 2017 that applies to all individuals who retired at the normal retirement age. In the data, the formula is expressed in monthly terms, while the figure shows the annualized values. The dashed lines correspond to the bend points equal to $10,620$ and $64,032$ (in 2011 U.S. dollars). The statutory marginal replacement rates in the lowest, intermediate, and highest brackets are 90%, 32%, and 15%, respectively.

Appendix C. Identified shift in the Pareto weights

Section **??** showed that Pareto weights in the model must shift toward older agents, on the one hand, and college graduate agents, on the other hand, for the model to rationalize the calibrated policy in the 2010s. The current section provides empirical evidence in favor of this model prediction.

First of all, note that the social welfare function in (34) is equivalent to the micro-founded probabilistic voting environment introduced by Lindbeck and Weibull (1987). In this environment, two candidates maximize the probability of winning an election by proposing simultaneously and independently a (potentially, multidimensional) policy. Voters differ in their most preferred policies and other exogenous characteristics independent of the citizens' most preferred policies. The weight that both candidates attach to a given group is given by a product of the group's population size and its degree of homogeneity regarding exogenous characteristics. One such important characteristic is the group's turnout rate. The higher the group's propensity to vote, the higher the candidate's incentive to shift the policy closer to this group's most preferred policy. In equilibrium, both candidates propose the same policy.

In my model, the Pareto weights in the social welfare function can be interpreted as the weights that the candidates assign to electoral groups in the probabilistic voting environment.

With this interpretation of the Pareto weights, the qualitative predictions of my model are as follows:

1. Conditional on age, the turnout rate of a college graduate relative to the high-school graduate's turnout must increase during 1970–2010.
2. Conditional on education, the turnout rate of older households relative to the young households' turnout must increase during 1970–2010.

To test these model predictions, I merge the CPS March Supplement with the survey data on voting behavior from the *Voting and Registration Supplement*, which is part of the CPS. It is most convenient to access the data through the *Integrated Public Use Microdata Series* (IPUMS) project webpage.[83]

In line with the sample selection criteria explained in the calibration section, a household head is a college graduate if she/he has a completed college degree. Otherwise, the household head is a high school graduate. I drop all households whose educational level is missing. Furthermore, I restrict the CPS sample to include only household heads aged 25–85, consistent with the model. Finally, I restrict the same to those households who answer *Yes* or *No* to the question: "*Have you voted in the most recent November election?*" (variable *VOTED*). Thus, I remove all those households from the sample who refuse to answer the question or claim they do not know the answer. When computing voter turnout rates, I weight observations using the variable *VOSUPPWT* from the *Voting and Registration Supplement.*

I report voter turnout statistics for Congressional elections because Congress would implement a Social Security reform. The first available Congressional election is from 1978, while the latest is from 2018. I split the data set into two subsamples: 1978–1986 and 2010–2018, each comprising three Congressional election cycles. In each subsample, I split households into four groups by household head's age (25–44, 45–54, 55–64, and 65–85) and two groups by household head's education.

To test the first model prediction, I proceed in three steps. First, I compute the average turnout rate for every age and education subgroup in each subsample period. For the sake of brevity, I do not report these results in the paper. Second, I ask: By how much percent does the computed turnout rate for college degree graduates exceeds that for high school graduates in a given age group in a given period? Table C.10 (second and third columns) reports the results. Finally, the last column shows the percentage change in the relative turnout rate

[83] See `https://cps.ipums.org/cps-action/data_requests/download`. To assemble the raw data set, select the *Basic March CPS* data set and the *Voting and Registration Supplement*.

Table C.10: Turnout rates of college graduates relative to high-school graduates, by age.

Age	1978–1986	2010–2018	% change
25–44	60.8	67.9	4.5
45–54	38.4	43.9	4.0
55–64	30.7	38.3	5.8
65–85	34.1	30.2	−2.9

Notes: The central two columns of the table show the turnout rates of college graduates normalized by the turnout rates of high school graduates, by age (in %). The last column shows the percentage change in relative voter turnout rates between 1978–1986 and 2010–2018. The table is constructed based on the CPS March Supplement merged with the survey data on voting behavior from the Voting and Registration Supplement. The turnout rates are computed at the household level. Each subsample comprises three Congressional election cycles.

across the two subsamples.

As one can see from Table C.10, college graduates vote at higher rates than high school graduates, since all numbers in the second and third columns are positive. This is true for all age groups and both subsample periods. Numerous empirical studies have already documented that participation among households in almost any form of political activity (including voting) rises with the households' level of education in the U.S.[84] I add to this finding in the literature how relative turnout rates have changed over time. The last column of the table displays the percentage change in the relative turnout rate. When comparing the numbers in the second and third columns, one can see that the relative turnout rates increase for all age groups over time, except for the 65–85-year-olds. Overall, the empirical evidence largely supports the model's first qualitative prediction regarding the shift in Pareto weights toward high-ability agents, conditional on their age.

To test the second model prediction, I start with the same data set constructed in the first step above. Next, I ask: Conditional on household's education, by how much percent does the turnout rate of older households differ from the youngest (25–44) group? Table C.11 reports the results (columns `Col` and `HS`) for each time frame. Finally, I calculate the percentage change in the relative turnout rates across time within each education sample.

According to Table C.11, all three age groups (45–54, 55–64, and 65–85) voted at higher rates than the youngest group (25–44). This observation holds at each education level and in

[84]See, among many others, Rosenstone and Hansen (1993), Benabou (2000), Bartels (2009).

Table C.11: Turnout rates relative to 25–44-year-old households, by age and education.

	1978–1986		2010–2018		% change	
	COL	HS	COL	HS	COL	HS
45–54	22.8	42.6	23.5	43.9	0.5	0.9
55–64	30.7	60.7	35.8	64.7	3.9	2.5
65–85	33.5	60.0	41.6	83.2	6.1	14.5

Notes: The table shows the normalized turnout rates of college graduates (COL) and high-school graduates (HS), by age. Within each education type and time period, the turnout rates are normalized by the turnout rate of 25–44-year-old individuals of the same education level. All numbers are in percent. The last two columns show the percentage change in relative voter turnout rates between 1978–1986 and 2010–2018. The table is constructed based on the CPS March Supplement merged with the survey data on voting behavior from the Voting and Registration Supplement. The turnout rates are computed at the household level. Each subsample comprises three Congressional election cycles.

each time frame. These facts are well-known in the literature (see the sources cited above). Next, observe that the relative turnout rates rise at a higher rate for older groups within each education type over time. This empirical evidence supports the model's second prediction regarding the shift in Pareto weights toward older agents, conditional on their ability.[85]

[85]In Brendler, 2020, I documented a rise in relative turnout rates by age using a *pooled* sample of college and high-school graduates. In the current paper, I show that the results also hold conditioning on the household head's education.