Opportunity & Inclusive Growth Institute
FEDERAL RESERVE BANK of MINNEAPOLIS

Optimal Paternalistic Savings Policies

Christian Moser
Columbia University and Federal Reserve Bank of Minneapolis

Pedro Olea de Souza e Silva
Uber Technologies

Institute Working Paper 17
January 2019

DOI: https://doi.org/10.21034/iwp.17
Keywords: Optimal taxation; Multidimensional screening; Present bias; Preference heterogeneity; Paternalism; Retirement; Savings; Social Security; Active-set algorithm
JEL Codes: H21, E62, H55

Opportunity and Inclusive Growth Institute
Federal Reserve Bank of Minneapolis • Research Division
90 Hennepin Avenue • Minneapolis, MN 55480-0291
https://www.minneapolisfed.org/institute

Optimal Paternalistic Savings Policies*

Christian Moser† Pedro Olea de Souza e Silva‡

January 8, 2019

Abstract

We study optimal savings policies when there is a dual concern about undersaving for retirement and income inequality. Agents differ in present bias and earnings ability, both unobservable to a planner with paternalistic and redistributive motives. We characterize the solution to this two-dimensional screening problem and provide a decentralization using realistic policy instruments: mandatory savings at low incomes but a choice between subsidized savings vehicles at high incomes—resembling Social Security, 401(k), and IRA accounts in the US. Offering more savings choice at higher incomes facilitates redistribution. To solve large-scale versions of this problem numerically, we propose a general, computationally stable, and efficient active-set algorithm. Relative to the current US retirement system, we find significant welfare gains from increasing mandatory savings and limiting savings choice at low incomes.

Keywords: Optimal Taxation, Multidimensional Screening, Present Bias, Preference Heterogeneity, Paternalism, Retirement, Savings, Social Security, Active-Set Algorithm

JEL classification: H21, E62, H55

*We thank the editor and three anonymous referees for thoughtful comments and suggestions. We are grateful to Mike Golosov for invaluable advice and encouragement throughout this project. We also thank Dilip Abreu, Mark Aguiar, Stefania Albanesi, Manuel Amador, Roland Bénabou, V. V. Chari, Olivier Darmouni, Henrique De Oliveira, Fatih Guvenen, Marina Halac, Jonathan Heathcote, Roozbeh Hosseini, Oleg Itskhoki, Larry Jones, Nobu Kiyotaki, Cinthia Konichi Paulo, Wojciech Kopczuk, Dirk Krueger, Felix Kubler, Rasmus Lentz, Ellen McGrattan, Ben Moll, Stephen Morris, Emi Nakamura, Fabrizio Perri, Chris Phelan, Richard Rogerson, Karl Schmedders, Stephanie Schmitt-Grohé, Ali Shourideh, Chris Sleet, Jón Steinsson, Sharon Traiberman, Maxim Troshkin, Aleh Tsyvinsky, Martin Uríbe, Nicolas Werquin, Juan Pablo Xandri, Pierre Yared, Sevin Yeltekin, Pierre Yared, and Steve Zeldes. We benefited from discussions by Youssef Benzarti and Ben Lockwood, and from comments by participants at several seminars and conferences. Salomón García provided excellent research assistance. Moser is grateful to the Opportunity and Inclusive Growth Institute at the Federal Reserve Bank of Minneapolis for hosting him during parts of the duration of this project. The views and opinions expressed are those of the authors and do not necessarily reflect those of any organization with whom the authors have been, are, or may in the future be affiliated.

†Graduate School of Business, Columbia University. Email: c.moser@columbia.edu.

‡Uber Technologies. Email: opedromiguel@uber.com.

1 Introduction

A shared feature of many modern welfare states is limited choice in savings for retirement.[1] These systems commonly force contributions toward old-age benefits on the basis of a paternalistic motive to induce adequate savings for retirement, particularly among low-income groups (Diamond, 1977; Kotlikoff et al., 1982; Feldstein, 1985). Rationales for the paternalistic motive derive both from the behavioral sciences and from a neoclassical economics tradition. Among those rationales, we highlight three. First, individuals may make mistakes when choosing under incomplete information and uncertainty (Tversky and Kahneman, 1974), or they may suffer from time-inconsistent decision making over their life cycle (Laibson, 1997). Second, altruism and lack of government commitment lead individuals to rationally undersave in anticipation of free-riding on public funds during retirement, giving rise to a "Samaritan's dilemma" (Buchanan, 1975; Prescott, 2004; Sleet and Yeltekin, 2006). Third, a planner who directly includes future selves in the welfare function will have a higher discount factor than private agents in the economy (Marglin, 1963; Caplin and Leahy, 2004). In all three environments, individuals exhibit *present bias* and paternalistic savings policies may be welfare improving.

We study optimal retirement savings policies when there is a paternalistic motive to undo individuals' present bias. The central question we ask is: how much choice in savings should be optimally offered throughout the income distribution? To address this question, we integrate a paternalistic savings motive into an optimal taxation framework, which allows us to study the problem of savings adequacy jointly with the issue of income inequality. Our key insight is that a trade-off exists between paternalism and redistribution. As a result, the optimal policy enforces high savings rates at low incomes but offers a choice between various subsidized savings options at high incomes. Qualitatively, the optimal policies in our framework resemble many real-world retirement savings systems, including Social Security and various subsidized savings accounts in the US. Quantitatively, however, we find significant welfare gains relative to current US policies.

In our theoretical framework, the interaction between two ingredients gives rise to a novel trade-off in optimal savings policy design. The first ingredient, motivated by recent experimental

[1] Government-mandated old-age benefits were administered in ancient Rome to prevent revolts by impoverished army veterans (Choi, 2015). German chancellor Otto von Bismarck instituted the Old Age and Disability Insurance Law of 1889 to guarantee adequate incomes for retired workers (Kotlikoff, 1996). The US Old-Age, Survivors, and Disability Insurance program, or Social Security for short, signed into law in 1935 under President Roosevelt to ameliorate the extent of poverty among retirees, is nowadays the nation's largest federal government social policy, with 957 billion US dollars in transfers to 62 million beneficiaries in 2017 (Social Security Administration, 2018).

evidence (Montiel Olea and Strzalecki, 2014; Chan, 2017), is heterogeneity in individuals' present bias. The second ingredient is heterogeneity in earnings ability, as in Mirrlees (1971), Diamond (1998), and Saez (2001). A paternalistic and redistributive planner defines the efficient savings rate according to a single time preference and attaches different welfare weights across ability types. The planner picks an allocation of consumption flows and labor for each type to maximize welfare subject to incentive compatibility and a resource constraint. The theoretical analysis of this problem is complex since, as is well known, multidimensional screening problems can lead to failure of the first-order approach, which the optimal taxation literature usually relies on (Golosov et al., 2003, 2016). We exploit the paternalistic formulation to characterize savings and labor distortions in this setting. Our main theoretical result highlights the trade-off between paternalism and redistribution under progressive social welfare weights: low-ability agents are bunched with savings strictly above the first-best rate, while high-ability agents are separated across present bias levels, with some saving strictly below the first-best rate. Intuitively, the planner offers savings choice as to incentivize work and collect tax revenues, thereby facilitating redistribution.

We show that the optimal allocation can be decentralized as a competitive equilibrium with three realistic policy instruments: first, mandated old-age benefits as a function of lifetime income; second, a number of retirement savings accounts with income-dependent subsidy rates and contribution limits; and third, a nonlinear labor income tax. For high enough forced savings, low-income agents and impatient high-income agents rely only on old-age benefits, whereas patient high-income agents choose to use a subset of the optional retirement accounts. Qualitatively, these policy instruments resemble real-world retirement savings systems, such as Social Security, 401(k), and various individual retirement arrangement (IRA) accounts in the US.

We apply this framework to quantitatively study the current US retirement savings and tax-transfer system vis-à-vis optimal policies in our model. This is a nontrivial task because failure of the linear independence constraint qualification (LICQ) in multidimensional screening problems can render numerical solution methods for these problems unstable (Judd et al., 2018). To overcome this problem, we develop a general, computationally stable, and efficient active-set algorithm that solves large-scale constrained optimization problems by iteratively finding the smallest set of binding constraints at the optimum. We use the algorithm to solve versions of our problem and demonstrate that it performs well in benchmarks vis-à-vis a more direct solution approach.

To calibrate the joint distribution of time preferences and earnings ability, we take an extended

positive version of our model to microdata on lifetime income and retirement wealth from the Health and Retirement Study (HRS) and the Panel Study of Income Dynamics (PSID), which we supplement with data from the US Life Tables and the Health Inequality Project. The extended model allows for several competing savings motives in addition to preference heterogeneity: survival risk, longevity heterogeneity, bequest motives, and medical expenditure shocks. Taking as given the distribution of preferences and ability, we then recover social preferences by extending the inverse-optimum approach (Bourguignon and Spadaro, 2012) to our setting. When allowing for a flexible shape of Pareto weights across earnings ability ranks, we recover a hump-shaped distribution, which puts relatively more weight on the second ability quartile (Jacobs et al., 2017).

We present four main results from our quantitative analysis. First, we find substantial empirical heterogeneity in present bias and hence in implied optimal savings rates throughout the income distribution. The calibration of our richer model delivers estimates of annualized present bias discount factors ranging from 0.915 to 0.998 between the 10th and the 90th percentile of the distribution. We also find a positive correlation of 0.153 between discount factors and ability across workers. Therefore, there remains significant heterogeneity in time preferences after controlling for a host of other factors.

Second, in spite of this heterogeneity, the optimal savings rate at low incomes is set to 46 percent, far above the first-best rate of 18 percent. Savings rates uniformly decline until a threshold of USD 95,000 in income. Above that threshold, savings rates vary across levels of present bias, varying between 3 and 18 percent at the highest income levels.

Third, we analyze properties of the policy instruments in our decentralization. Social Security benefits are hump-shaped, with replacement rates increasing from 68 percent to 145 percent for incomes up to USD 60,000 and declining thereafter. Contribution limits on retirement savings accounts are approximately affine in income. Individuals with annual incomes up to USD 95,000 receive only Social Security payments. Above that threshold, optional retirement savings accounts offer progressive savings subsidies. The optimal income tax schedule features increasing tax levels but decreasing average tax rates across incomes, with little dependence on private savings choices.

Finally, we discuss the welfare implications of policy reforms. On the one hand, the US tax-transfer system is best justified through welfare weights that are less redistributive than utilitarian (Heathcote and Tsujiyama, 2017). On the other hand, our model rationalizes high forced savings and large dispersion in savings rates at high incomes in the US as welfare weights that are more

redistributive than utilitarian. Consequently, our paternalistic framework suggests welfare gains of 8.8 percent in consumption-equivalent terms from increasing mandatory savings and limiting savings choice, particularly at low incomes. In contrast, we find that a nonpaternalistic planner cannot rationalize real-world policies.

Related literature. This paper contributes to three strands of the literature.[2] The first strand is concerned with the optimal taxation of capital. The classic result by Atkinson and Stiglitz (1972) states that with agreement in preferences between the planner and agents, only income—but not savings—should be distorted. Also relying on preference agreement is the zero long-run capital taxation proposition by Judd (1985) and Chamley (1986), subsequently revisited by Atkeson et al. (1999), Lansing (1999), Phelan and Stacchetti (2001), Hassler et al. (2008), Saez (2013), and Straub and Werning (2018). In our framework, paternalism provides an alternative motive for capital taxes or subsidies. Closely related to our work, Saez (2002), Diamond and Spinnewijn (2011), and Golosov et al. (2013) consider heterogeneous time preferences without paternalism and show that the correlation between discount factors and ability matters for the optimal degree of capital taxation. Piketty and Saez (2013) study the related problem of optimal linear or two-bracket inheritance taxation. Hosseini and Shourideh (2018) characterize optimal retirement policy reforms with heterogeneous mortality rates and time preferences. Ndiaye (2018) studies life-cycle taxation with endogenous retirement. A novel aspect of our paper is to consider the interaction between paternalism and redistribution with two dimensions of heterogeneity: present bias and earnings ability. We find that under a strong enough redistributive motive, the optimal dispersion of savings rates is larger at higher incomes.

The second strand of the literature that we relate to is the field of behavioral public finance.[3] Much work focuses on optimal taxation without heterogeneity in behavioral biases or redistribution. O'Donoghue and Rabin (2003, 2006) and Gruber and Köszegi (2004) consider the incidence of linear consumption taxes when certain goods are either overconsumed (e.g., cigarettes) or underconsumed (e.g., retirement savings). Farhi and Werning (2007, 2010), Pavoni and Yazici (2017), and Phelan and Rustichini (2018) study optimal estate taxation with a constant difference between social and private discount factors. Farhi and Werning (2013a) examine heterogeneity in private discount factors but no differences in earnings ability. Lockwood and Taubinsky (2017) allow for

[2] Appendix A.4 discusses our findings in light of some of the most closely related frameworks in the literature.
[3] See Bernheim and Taubinsky (2018) for an excellent survey of optimal policy design with nonstandard preferences.

nonlinear labor earnings taxes and a linear tax on sin goods. Amador et al. (2006) study an optimal delegation problem with uniform present bias and find a minimum savings rule to be optimal.[4] Chetty et al. (2009) and Beshears et al. (2015) consider optimal policy design with behavioral agents but without redistribution. In our environment with transfers, the optimal policy features oversaving at low incomes and differentially distorted savings decisions at high incomes. Amador et al. (2004) study optimal taxation in an environment with transfers and shocks to agents' taste for present consumption and temptation, focusing on the special case when shocks are independently distributed. Our paper also complements recent works including Yu (2016), who studies savings policies as commitment devices when agents are time-inconsistent, and Lockwood (2016), who focuses on implications of present bias for optimal income taxation. Farhi and Gabaix (2018) study optimal taxation in the spirit of Ramsey, Pigou, and Mirrlees under a range of behavioral biases. Complementing this literature, our focus is on characterizing the optimal structure of savings choices throughout the income distribution.

The third strand of related work studies multidimensional screening problems. Rochet (1987), McAfee and McMillan (1988), Armstrong (1996), Rochet and Choné (1998), and Armstrong and Rochet (1999) emphasize challenges in the analysis of optimal contracts with higher-dimensional unobserved heterogeneity. Some important contributions in the field of public finance have made further progress. Kleven et al. (2009) analyze the optimal taxation of couples, while Rothschild and Scheuer (2013, 2015, 2016) characterize optimal income taxes under multidimensional skill heterogeneity. We contribute to this literature by characterizing a two-dimensional screening problem with paternalism and developing a numerical algorithm that efficiently solves more general large-scale nonlinear constrained optimization problems.

Outline. The paper is organized as follows. Section 2 introduces the two-dimensional screening problem. Section 3 characterizes the optimal allocation under paternalism and redistribution. Section 4 provides a decentralization using realistic policy instruments. Section 5 presents an active-set algorithm to solve large-scale nonlinear constrained optimization problems. Section 6 calibrates the model to US microdata on lifetime earnings and retirement savings. Section 7 evaluates optimal retirement savings policies. Finally, Section 8 concludes.

[4] Similar setups have been studied in the context of monetary policy (Athey et al., 2005), sovereign debt dynamics (Aguiar and Amador, 2011), fiscal rules (Halac and Yared, 2014), the market for commitment devices (Galperti, 2015), and parent-child relations (Doepke and Zilibotti, 2017).

2 Two-Dimensional Screening Problem

This section presents our benchmark model for the analysis of optimal savings policies when there is a dual concern about undersaving for retirement and income inequality.

2.1 Model setup

A unit mass of agents live for two periods—working life, $t = 1$, and retirement, $t = 2$—with common discount factor δ between periods.[5] Agents differ in two unobservable attributes. The first attribute is earnings ability, denoted $\theta \in \Theta = \{\theta_1, \ldots, \theta_N\}$, where $0 \leq \theta_1 < \ldots < \theta_N < +\infty$. The second attribute is the degree of *present bias*, denoted $\beta \in B = \{\beta_1, \ldots, \beta_M\}$, where $0 < \beta_1 < \ldots < \beta_M = 1$. Here, β is understood as a reduced-form placeholder for the disagreement between the planner and agents at the time of the savings decision, which may arise from agents' behavioral bias (Laibson, 1997), from lack of government commitment (Sleet and Yeltekin, 2006), or from direct inclusion of future selves in the welfare function (Caplin and Leahy, 2004). We assume that the distribution over two-dimensional types, $\pi(\theta, \beta)$, has full support but do not impose any restrictions on the correlation between θ and β.

Utility is defined over consumption streams, (c_1, c_2), and labor supply $\ell = y/\theta$, where $y \geq 0$ is individual income. The planner evaluates *period 0 or experienced utility* of (θ, β)-types according to

$$V(c_1, c_2, y; \theta) = u(c_1) - v\left(\frac{y}{\theta}\right) + \delta u(c_2),$$

which directly depends on θ but not on β. Consumption utility $u \in \mathbf{C}(\mathbb{R}_+)$ is isoelastic and satisfies $u'(c) > 0$, $u'(0) = +\infty$, $\lim_{c \to +\infty} u'(c) = 0$, and $u''(c) < 0$ for $c \geq 0$. Labor disutility $v \in \mathbf{C}(\mathbb{R}_+)$ is multiplicatively separable as $v(y/\theta) = D(\theta)\tilde{v}(y)$, where $D, \tilde{v} \in \mathbf{C}(\Theta)$ satisfy $D(\theta) < D(\theta')$ for $\theta > \theta'$, $\lim_{\theta \to +\infty} D(\theta) = 0$, $\tilde{v}(0) = 0$, $\tilde{v}'(0) = 0$, $\tilde{v}'(y) > 0$ for $y > 0$, $\lim_{y \to +\infty} \tilde{v}'(y) = +\infty$, $\tilde{v}''(y) > 0$ for $y \geq 0$, $D(0)\tilde{v}(y) = +\infty$ for $y > 0$, and $D(0)\tilde{v}(0) = 0$.[6]

At the time of decision making, (θ, β)-types evaluate *period 1 or decision utility* according to

$$U(c_1, c_2, y; \theta, \beta) = u(c_1) - v\left(\frac{y}{\theta}\right) + \beta\delta u(c_2), \tag{1}$$

[5] In Appendix A.6, we analyze a multiperiod life-cycle model with heterogeneity in hyperbolic discount factors.

[6] These assumptions are common in the public finance literature (see, for example, Heathcote et al., 2017) and include as a special case the power utility formulation that we later use in our quantitative application.

which directly depends on θ and β. Therefore, the planner and (some) agents disagree about intertemporal trade-offs. A storage technology offers fixed gross return R between periods.

As in Mirrlees (1971), the planner observes consumption and labor income but not agents' types directly. Appealing to the revelation principle, the planner designs a direct mechanism, in which an agent is assigned a consumption-income bundle that depends deterministically on the agent's reported type. Such an assignment defines an *allocation*:

$$\mathcal{A} = \{c_1(\theta,\beta), c_2(\theta,\beta), y(\theta,\beta)\}_{(\theta,\beta)\in\Theta\times B}.$$

We write $u_t(\theta,\beta) \equiv u(c_t(\theta,\beta))$ for period t consumption utility, $v(\theta,\beta) \equiv v(y(\theta,\beta)/\theta)$ for work disutility, $V(\theta,\beta) \equiv V(c_1(\theta,\beta), c_2(\theta,\beta), y(\theta,\beta);\theta)$, and $U(\theta,\beta) \equiv U(c_1(\theta,\beta), c_2(\theta,\beta), y(\theta,\beta);\theta,\beta)$. An allocation satisfies *incentive compatibility (IC)* if

$$\forall(\theta,\beta)\in\Theta\times B: \quad (\theta,\beta) \in \underset{(\theta',\beta')\in\Theta\times B}{\arg\max}\ U\left(c_1(\theta',\beta'), c_2(\theta',\beta'), y(\theta',\beta');\theta,\beta\right). \tag{2}$$

An incentive compatible allocation can be implemented with agents truthfully reporting their types as an equilibrium strategy in the direct mechanism. An allocation satisfies *feasibility* if

$$\sum_{(\theta,\beta)\in\Theta\times B} \pi(\theta,\beta)\left[y(\theta,\beta) - c_1(\theta,\beta) - \frac{c_2(\theta,\beta)}{R}\right] \geq 0. \tag{3}$$

A feasible allocation allows for transfers across types but restricts the planner's net budget to be weakly positive. *Welfare* associated with an allocation $\mathcal{A}$ is

$$W(\mathcal{A}) = \sum_{(\theta,\beta)\in\Theta\times B} \pi(\theta,\beta)\lambda(\theta)V(c_1(\theta,\beta), c_2(\theta,\beta), y(\theta,\beta);\theta), \tag{4}$$

with Pareto weights $\lambda(\theta) \geq 0$ normalized such that $\sum_{(\theta,\beta)\in\Theta\times B}\pi(\theta,\beta)\lambda(\theta) = 1$.[7]

The *planner's problem* is to maximize welfare (4) subject to IC constraints (2) and the feasibility constraint (3). In this two-dimensional screening problem, the planner has both a *paternalistic motive* due to the difference between $V(\cdot)$ and $U(\cdot)$ in equations (2) and (4), and a *redistributive motive* due to the relative Pareto weights in (4) and the possibility of transfers across types in (3).

[7] That present bias does not directly enter welfare through either $\lambda(\cdot)$ or $V(\cdot)$ can be motivated by adopting an individual's perspective before the realization of a present bias shock, as in Amador et al. (2006).

Lemma 2 in Appendix A.1 establishes some important properties of the planner's problem, including convexity, existence and uniqueness of a global optimum, applicability of the maximum theorem, and strong duality subject to the Karush-Kuhn-Tucker (KKT) conditions. In Appendix A.2, we characterize benchmark allocations, including the first-best and laissez-faire.

2.2 Savings and Labor Distortions

To describe optimal distortions at the second best, we define three wedges characterizing the solution to the planner's problem. First, the *decision wedge* captures savings distortions in the agent's view, as measured by the deviation from their intertemporal Euler equation:

$$\tau^D(\theta,\beta) = 1 - \frac{u'(c_1(\theta,\beta))}{R\beta\delta u'(c_2(\theta,\beta))}. \tag{5}$$

A positive (negative) decision wedge can be interpreted as an implicit savings tax (subsidy).

Second, the *efficiency wedge* captures savings distortions in the planner's view, as measured by the deviation from the planner's intertemporal Euler equation:

$$\tau^E(\theta,\beta) = 1 - \frac{u'(c_1(\theta,\beta))}{R\delta u'(c_2(\theta,\beta))}.$$

A positive (negative) efficiency wedge can be interpreted as undersaving (oversaving) relative to the welfare-maximizing savings rate.

Third, the *labor wedge* captures labor supply distortions, as measured by the deviation from the intratemporal Euler equation:

$$\tau^L(\theta,\beta) = 1 - \frac{v'(y(\theta,\beta)/\theta)}{u'(c_1(\theta,\beta))\,\theta}. \tag{6}$$

A positive (negative) labor wedge can be interpreted as an implicit labor income tax (subsidy).

Though it may be tempting to interpret the decision wedge (5) and the labor wedge (6) as explicit taxes on savings and income, respectively, this is generally not the case. In Section 4, we provide a decentralization of the solution to the planner's problem using realistic policy tools and discuss the relation between wedges and marginal tax rates in this context.

3 Optimal Paternalistic Savings Policies

We characterize general properties of the optimal allocation before showing how it can be decentralized using a set of realistic policy instruments.

3.1 Simple Environment with 2×2 Types

Consider a simple environment with two levels of earnings ability and two levels of present bias.[8] Let $\theta \in \{\theta_L, \theta_H\}$ with $\theta_L = 0 < \theta_H$, let $\beta \in \{\beta_L, \beta_H\}$ with $0 \leq \beta_L < \beta_H = 1$, and let $R\delta = 1$.

Proposition 1. *If $\lambda(\theta_L) \geq \lambda(\theta_H)$, then the solution to the planner's problem is characterized as follows:*

1. *The first-best allocation is not incentive compatible.*
2. *Low-ability types are bunched across present bias levels: $s(\theta_L, \beta) = s(\theta_L) \quad \forall \beta$.*
3. *High-ability types are separated in savings rates across present bias levels: $s(\theta_H, \beta_L) \neq s(\theta_H, \beta_H)$.*
4. *Low-ability types receive a strictly positive implicit savings subsidy and save strictly above the efficient rate, patient high-ability types save weakly above the efficient rate, and present-biased high-ability types face an implicit savings subsidy or tax but always save strictly below the efficient rate:*

$$\begin{aligned} \tau^D(\theta_L, \beta_L) < \tau^D(\theta_L, \beta_H) < 0 \\ \tau^D(\theta_H, \beta_H) \leq 0 \gtreqless \tau^D(\theta_H, \beta_L) \\ \tau^E(\theta_L) < 0 \\ \tau^E(\theta_H, \beta_H) \leq 0 < \tau^E(\theta_H, \beta_L). \end{aligned}$$

5. *High-ability types face no implicit labor distortions: $\tau^L(\theta_H, \beta_L) = \tau^L(\theta_H, \beta_H) = 0$.*

Proof. It is instructive to flesh out the argument here to illustrate the model mechanics:

1. *(First best violates IC.)* IC for (θ_H, β_H)-types requires $V(\theta_H, \beta_H) \geq \max_\beta V(\theta_L, \beta)$, contradicting the first best for $\lambda(\theta_L) > \lambda(\theta_H)$. For $\lambda(\theta_L) = \lambda(\theta_H)$, then the first-best condition $V(\theta_L) = V(\theta_H)$ combined with the IC constraint from (θ_H, β_L)-types to θ_L-types implies that $s(\theta_H) < s(\theta_L)$, contradicting the fact that all agents save at the first-best rate.

[8] This generalizes the three-type environments contained in Cremer et al. (2009) and Tenhunen and Tuomala (2010).

2. *(Bunching at low ability.)* Note that β does not enter the planner's objective. Since only θ_H-types work and income is observable, the relevant IC constraints in θ-space are the ones from high to low ability. Consider convexifying θ_L-types' allocations in utility space each period by assigning them $\bar{u}_t(\theta_L) = \sum_\beta \pi(\beta|\theta_L) u_t(\theta_L, \beta)$ for $t = 1, 2$. Bunching trivially satisfies θ_L-types' IC constraints, since they cannot deviate elsewhere. As IC constraints are linear in $u_1(\theta_L, \beta)$ and $u_2(\theta_L, \beta)$, the convex combination in utility space also preserves the IC of θ_H-types and leaves welfare unchanged. However, strict concavity of $u(\cdot)$ implies that $\bar{c}_t(\theta_L) = u^{-1}(\bar{u}_t(\theta_L))$ is strictly convex, so bunching θ_L-types is strictly less costly than separating them by β-levels. In summary, bunching low-ability types is optimal as it leaves welfare constant, preserves IC, but saves resources.

3. *(Separation at high ability.)* We first show that high-ability types cannot receive the same consumption stream $(c_1(\theta_H), c_2(\theta_H))$. By way of contradiction, suppose that θ_H-types are bunched in consumption across β-levels. There are two cases to consider.

   First, suppose high-ability agents are bunched with $c_1(\theta_H) > c_2(\theta_H)$, illustrated as point A in Figure 1(a). At this point, the slope of β_L-types' indifference curve is more negative compared to β_H-types. The planner can target β_H-types by offering the welfare-equivalent allocation B in the interior of the budget set, which also preserves IC—a contradiction.

   Second, suppose high-ability agents are bunched with $c_2(\theta_H) \geq c_1(\theta_H)$. Clearly, $c_2(\theta_H) > c_1(\theta_H)$ is dominated, so $c_2(\theta_H) = c_1(\theta_H)$, illustrated as point D in Figure 1(b). From the IC constraint of (θ_H, β_H)-types, we know $V(\theta_H) \geq V(\theta_L)$. In fact, $V(\theta_H) > V(\theta_L)$, or else $V(\theta_H) = V(\theta_L)$ together with IC from (θ_H, β_L)-types to θ_L-types would imply $s^{FB} = s(\theta_H) < s(\theta_L)$. Then we could decrease $s(\theta_H, \beta_L)$ along the budget line and decrease $s(\theta_L)$ along the planner's indifference curve, incurring a second-order welfare loss but a first-order resource gain—a contradiction. Since $V(\theta_H) > V(\theta_L)$ but $\lambda(\theta_L) \geq \lambda(\theta_H)$, the planner can target β_L-types by offering an additional allocation at point E in Figure 1(b), yielding net welfare gains through a first-order transfer from (θ_H, β_L)-types toward θ_L-types but a second-order welfare loss from (θ_H, β_L)-types' deviation—a contradiction.

   Therefore, $(c_1(\theta_H, \beta_L), c_2(\theta_H, \beta_L)) \neq (c_1(\theta_H, \beta_H), c_2(\theta_H, \beta_H))$. Hence, $s(\theta_H, \beta_L) = s(\theta_H, \beta_H)$ would require $c_t(\theta_H, \beta) > c_t(\theta_H, \beta')$ for $t = 1, 2$ and some β, β'. Clearly, this violates IC, given that θ_H-types' labor supply is undistorted (see part 5 of the proposition).

4. *(Oversaving and undersaving.)* Low-ability agents are optimally bunched such that $c_2(\theta_L) \geq c_1(\theta_L)$. If this were not the case, then moving them from point A to point B in Figure 1(a) would preserve IC, leave welfare unchanged, but save resources—a contradiction. In fact, $c_2(\theta_L) > c_1(\theta_L)$. Suppose, by way of contradiction, that $c_2(\theta_L) = c_1(\theta_L)$, illustrated as point F in Figure 1(c). Then we could move θ_L-types toward point G to induce a second-order welfare loss but relax the IC constraint of θ_H-types to a first order, associated with a net welfare gain from improved redistribution or paternalism. To make this point, we consider three exhaustive cases, which we derive in Appendix A.3.1. Case 1 features $V(\theta_H, \beta_H) = V(\theta_H, \beta_L) > V(\theta_L)$ initially, so the planner can use the slackness in IC from the perturbation to transfer resources from (θ_H, β_L)-types to θ_L-types, improving welfare—a contradiction. Cases 2 and 3 feature $V(\theta_H, \beta_H) = V(\theta_L) \geq V(\theta_H, \beta_L)$ initially, so the planner can use the slackness in IC from the perturbation to increase (θ_H, β_L)-types' savings rate toward the first best by moving up along β_H-types' indifference curve, keeping welfare constant, preserving IC, but saving resources—a contradiction. It follows that $s(\theta_L) > s^{FB}$.

   We must have $s(\theta_H, \beta_H) \geq s^{LF}(\beta_H) = s^{FB}$ as there is no reason to distort downward the savings of β_H-types. Furthermore, $s(\theta_H, \beta_H) > s^{LF}(\beta_H)$ if and only if their allocation is envied by (θ_H, β_L)-types, which occurs in Case 2 but not in Case 3.[9] In Cases 1 and 3, conversely, $s(\theta_H, \beta_H) = s^{FB}$ as the IC constraint from (θ_H, β_L)-types to (θ_H, β_H)-types is slack.

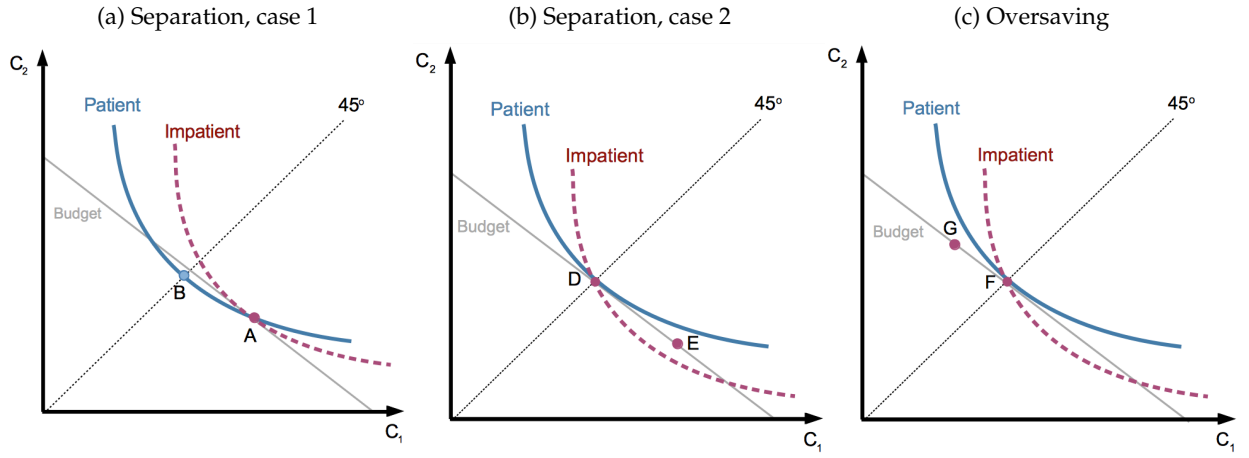
   Finally, (θ_H, β_L)-types' savings may be below or above their laissez-faire rate. On the one hand, Case 1 features a binding IC constraint from (θ_H, β_H)-types to (θ_H, β_L)-types, so $s(\theta_H, \beta_L) < s^{LF}(\beta_L)$ can be optimal to relax that constraint. On the other hand, Cases 2 and 3 feature a slack IC constraint from (θ_H, β_H)-types to (θ_H, β_L)-types, so $s(\theta_H, \beta_L) \geq s^{LF}(\beta_L)$ at the optimum, and $s(\theta_H, \beta_L) > s^{LF}(\beta_L)$ if and only if $\lambda(\theta_H) > 0$ in those cases.

5. *(High-ability types' labor supply is undistorted.)* By way of contradiction, suppose $\tau^L(\theta_H, \beta) \neq 0$ for some $\beta \in \{\beta_L, \beta_H\}$. Then $v'(y(\theta_H, \beta)/\theta_H) \neq u'(c_1(\theta_H, \beta))\theta_H$, and the planner could adjust $y(\theta_H, \beta)$ and $c_1(\theta_H, \beta)$ to reduce $|\tau^L(\theta_H, \beta)|$ while keeping (θ_H, β)-types' period 1 utility and hence welfare constant, preserving IC but saving resources—a contradiction. We conclude that $\tau^L(\theta_H, \beta_L) = \tau^L(\theta_H, \beta_H) = 0$.

□

[9] In a previous version of the paper, we ignored this case.

Figure 1. Consumption perturbations



Proposition 1 shows that paternalism in the face of present bias ($\beta_L < 1$) and the redistributive motive ($\lambda(\theta_L) \geq \lambda(\theta_H)$) interact in nontrivial ways. The planner optimally offers savings choice as a screening device to facilitate redistribution. Savings of high-ability types are differentially distorted to provide incentives, thereby facilitating transfers across types.

Several features of the optimal allocation are noteworthy. At low ability, uniform savings strictly above the efficient rate reflect a combination of paternalism and the desire to deter deviations by high-ability types. In contrast, the redistributive motive dominates at high ability, leading to optimal separation. That is, patient high-ability types optimally save at least at the first-best rate, while impatient high-ability types save strictly below the first best to allow for greater transfers across types. Rational agents' savings may be distorted upward in the presence of behavioral agents. Although all agents in the economy want to save less than the planner, the presence of behavioral agents can also lead savings to be distorted further downward for incentive reasons. Leaving labor supply of the high-ability agents undistorted ensures productive efficiency.[10]

3.2 Characterizing the General Economy

Many insights from the preceding analysis extend to a general economy with N θ-types, where $0 \leq \theta_1 < \ldots < \theta_N < +\infty$, and M β-types, where $0 < \beta_1 < \ldots < \beta_M = 1$.

[10]Variations of this simple environment yield further insights. Appendix A.3.2 shows that there exists a threshold relative Pareto weight above which the result from Proposition 1 continues to hold. Conversely, Appendix A.3.3 shows that there exists a threshold relative productivity level such that the same results apply. Appendix A.3.4 presents conditions under which the optimum features separation of low-ability types and bunching of high-ability types. Appendix A.3.5 characterizes the optimal allocation when all agents have the same ability but different Pareto weights.

Lemma 1. *The following intermediate results hold in the general economy:*

1. *(Monotonicity) For all* (θ, β) *and* (θ, β')*, if* $\beta > \beta'$*, then either* $(c_1(\theta, \beta), c_2(\theta, \beta), y(\theta, \beta)) = (c_1(\theta, \beta'), c_2(\theta, \beta'), y(\theta, \beta'))$*, or* $c_2(\theta, \beta) > c_2(\theta, \beta')$ *and* $u(c_1(\theta, \beta)) - v(y(\theta, \beta)/\theta) < u(c_1(\theta, \beta')) - v(y(\theta, \beta')/\theta)$*. Furthermore, if* $\theta > \theta'$*, then* $y(\theta, \beta) \geq y(\theta', \beta)$ *for all* β*.*

2. *(Single crossing in* β*-dimension) If an allocation satisfies IC of some type* (θ, β_m) *with respect to types* (θ, β_{m-1}) *and* (θ, β_{m+1})*, then it satisfies IC of type* (θ, β_m) *with respect to all* θ*-types.*

Proof. See Appendix A.3.6. □

While global IC constraints may bind in multidimensional screening problems (Rochet and Choné, 1998), Lemma 1 puts some structure on the optimal allocation and pattern of binding IC constraints. Although the single-crossing property implies sufficiency of local constraints in the β-dimension conditional on θ, other global IC constraints still remain relevant.[11]

Going forward, we define $\tilde{\theta}_n^+ \equiv \theta_{n+1}/\theta_n$ as the upward productivity differential and $\tilde{\theta}_n^- \equiv \theta_n/\theta_{n-1}$ as the downward productivity differential. Our first main result establishes sufficient conditions for bunching versus separation of agents across β-levels throughout the θ-distribution.

Theorem 1. *There exist cutoffs* $1 < \overline{\tilde{\theta}_n^+} < +\infty$, $1 < \overline{\tilde{\theta}_n^-} < +\infty$, *and* $0 \leq \underline{\lambda_n} < \overline{\lambda_n} < +\infty$ *such that:*

1. *(i) If* $\tilde{\theta}_1^+ \geq \overline{\tilde{\theta}_1^+}$ *and* $\lambda(\theta_1) > 0$*, then* θ_1*-types are bunched across* β*-levels.*
   (ii) For $n = 2, \ldots, N-1$*, if* $\tilde{\theta}_n^+ \geq \overline{\tilde{\theta}_n^+}$, $\tilde{\theta}_n^- \geq \overline{\tilde{\theta}_n^-}$*, and* $\lambda(\theta_n) \geq \overline{\lambda_n}$*, then* θ_n*-types are bunched across* β*-levels.*
   (iii) If $\tilde{\theta}_N^- \geq \overline{\tilde{\theta}_N^-}$ *and* $\lambda(\theta_N) \geq \overline{\lambda_N}$*, then* θ_N*-types are bunched across* β*-levels.*

2. *If* $\lambda(\theta_n) \leq \underline{\lambda_n}$ *and* $\theta_n > 0$*, then* θ_n*-types are separated across* β*-levels.*

Proof. See Appendix A.3.8. □

Part 1 of Theorem 1 states that a sufficient condition for optimal bunching of θ-types across β-levels is that the planner attaches a high enough welfare weight on those agents while they are relatively isolated in productivity space. If θ-types are not tempted to deviate to another θ'-type's allocation because they find themselves relatively well off, then they pay the planner for this privilege by giving up their preference autonomy. Conversely, part 2 of the theorem states

[11] Leading up to the following results, Lemma 3 in Appendix A.3.7 contains useful insights linking the distribution of abilities, Pareto weights, and the pattern of binding IC constraints.

that a sufficient condition for optimal separation of agents across β-levels of a working θ-type is for the planner to attach a small enough welfare weight on them. From the planner's point of view, granting agents some autonomy helps extract resources that are valued for redistribution.

Our second main result characterizes savings distortions throughout the θ-distribution.

Theorem 2. *There exist cutoffs* $1 < \underline{\tilde{\theta}_n^+} < \overline{\tilde{\theta}_n^+} < +\infty$ *and* $0 \leq \underline{\lambda_n} < \overline{\lambda_n} < +\infty$ *such that:*

1. *If* θ_n*-types are bunched across* β*-levels, then* $\tau^D(\theta_n, \beta_1) < \ldots < \tau^D(\theta_n, \beta_M) = \tau^E(\theta_n) \leq 0$*. If, in addition,* $\tilde{\theta}_1^+ \geq \overline{\tilde{\theta}_1^+}$, $\tilde{\theta}_n^+ \leq \underline{\tilde{\theta}_n^+}$ *for* $n = 2, \ldots, N-1$ *and* $\lambda(\theta_1) \geq \overline{\lambda_1}$*, then* $\tau^D(\theta_1, \beta_1) < \ldots < \tau^D(\theta_1, \beta_M) = \tau^E(\theta_1) < 0$*.*

2. *If* θ_n*-types are separated across* β*-levels, then* $\tau^D(\theta_n, \beta_M) = \tau^E(\theta_n, \beta_M) \leq 0$*. If, in addition,* $\lambda(\theta_n) < \underline{\lambda_n}$*, then* $\tau^E(\theta_n, \beta_1) > \tau^D(\theta_n, \beta_1) \geq 0$*.*

3. *If IC constraints of all* (θ, β)*-types with* $\beta < 1$ *are slack with respect to* (θ_n, β_M)*-types, then* $\tau^D(\theta_n, \beta_M) = \tau^E(\theta_n, \beta_M) = 0$*.*

Proof. See Appendix A.3.9. □

Part 1 of Theorem 2 shows that if agents are bunched, then they must save at least at the first-best rate. Whether bunching occurs strictly above the first-best savings rate depends on the pattern of binding IC constraints. The lowest-productivity agents may be bunched strictly above the first-best savings rate to deter present-biased high-ability agents from shirking. Conversely, part 2 of the theorem shows that among separated agents, the most patient type saves at least at the first-best rate, strictly above the most present-biased type. The planner optimally allows some types to undersave in order to use them as cash cows for redistribution. Finally, part 3 of the theorem shows that the savings of agents without present bias are distorted upward only if they are envied by a present-biased agent.

Our third main result characterizes labor distortions for the highest and lowest ability types.

Theorem 3. *There exist cutoffs* $1 < \overline{\tilde{\theta}_n^+} < +\infty$, $1 < \overline{\tilde{\theta}_n^-} < +\infty$, *and* $0 < \underline{\lambda_n} < +\infty$ *such that:*

1. *(i)* θ_1*-types' labor income is implicitly taxed:* $\tau^L(\theta_1, \beta) \geq 0 \quad \forall \beta$*.*
   (ii) θ_N*-types' labor income is implicitly subsidized:* $\tau^L(\theta_N, \beta) \leq 0 \quad \forall \beta$*.*

2. *(i) If* $\tilde{\theta}_1^+ \geq \overline{\tilde{\theta}_1^+}$ *and* $\lambda(\theta_1) \leq \underline{\lambda_1}$*, then* θ_1*-types' labor margin is undistorted:* $\tau^L(\theta_1, \beta) = 0 \quad \forall \beta$*.*
   (ii) If $\tilde{\theta}_N^- \geq \overline{\tilde{\theta}_N^-}$ *and* $\lambda(\theta_N) \leq \underline{\lambda_N}$*, then* θ_N*-types' labor margin is undistorted:* $\tau^L(\theta_N, \beta) = 0 \quad \forall \beta$*.*

Proof. See Appendix A.3.10. □

Part 1 of Theorem 3 establishes the familiar result that, generally, low-ability types experience an implicit labor tax, while high-ability types experience an implicit labor subsidy (Tuomala, 1990). Part 2 of the theorem shows that labor distortions at the extremes of the ability distribution are used only for incentive reasons. Specifically, it is optimal to leave labor supply undistorted for some θ-type whenever their allocation is not envied by any other θ'-type with $\theta' \neq \theta$.

In summary, Theorems 1–3 highlight the interaction between paternalistic and redistributive motives. On the one hand, relatively unproductive types with high welfare weights are bunched at high savings rates. On the other hand, relatively productive types with low welfare weights are separated, with some of them optimally saving at lower rates.

3.3 Discussion of Assumptions, Generalizability, and Methodology

Sufficiency of conditions on relative Pareto weights and ability levels. Our theory sheds light on the trade-offs faced by a planner in an economy with heterogeneous present bias and earnings ability. These trade-offs are shaped by the relative welfare weights and productivity levels of agents in the economy. Whether the stated sufficient conditions for our theoretical characterization apply is ultimately a quantitative question, which we assess in a calibrated version of the model.

Extension to a continuum of types. Our theoretical results rely only on the pattern of binding IC constraints, for which we provide sufficient conditions in terms of relative Pareto weights and ability levels. Thus, our main result on optimal bunching versus separation of agents in the general economy with $N \times M$ types (Theorem 1) can be readily extended to a continuous-type economy.

Corollary 1. *Let* $\Theta \times B = \left[\underline{\theta}, \overline{\theta}\right] \times \left[\underline{\beta}, \overline{\beta}\right]$ *define a continuous type space. Then:*

1. *For all ability types* $\theta \in \Theta$*, if their IC constraints are slack with respect to all agents* $\theta' \neq \theta$*, then* θ*-types are bunched across* β*-levels.*

2. *For all ability types* $\theta \in \Theta$*, there exists* $\underline{\lambda_\theta} \geq 0$ *such that if* $\lambda(\theta) \leq \underline{\lambda_\theta}$ *and* $\theta > 0$*, then* θ*-types are separated across* β*-levels.*

Proof. See Appendix A.3.11. □

Following the same logic, our results on over- versus undersaving (Theorem 2) and implicit labor income tax rates (Theorem 3) rely only on the pattern of binding IC constraints. In numerical simulations, we find that bunching occurs for a larger share of ability types when the type space is more dense (see Figure 12 of Appendix B.4).

Relation to the first-order approach and resulting optimal tax formulas. Our methodological approach, which relies on global perturbation arguments, differs from many works in the optimal taxation literature. One advantage of our approach is that we are able to characterize properties of the optimal allocation without certain restrictions on the structure of the economy and properties of optimal allocation. A disadvantage of our approach is that some of the common theoretical results derived under these restrictions do not directly carry over to our setting.

For comparison, in a static environment with unidimensional heterogeneity, the Spence-Mirrlees single-crossing property implies that local IC is sufficient for global IC. Guided by this, influential works have derived optimal tax formulas based on a *first-order approach* (Diamond, 1998; Saez, 2001; Golosov et al., 2014, 2016). It is well understood that the validity of these optimal tax formulas rests on three assumptions: local IC constraints are sufficient for global IC, there is no bunching of a mass of agents at the same allocation, and the optimal allocation is continuously differentiable in agents' types.[12] Since, by assumption, this approach rules out the possibility of bunching, which we highlighted to be relevant in our environment, we instead opt for a more general perturbation approach. We verify that bunching is a salient feature in our calibrated economy, that global IC constraints bind, and that, as a result, optimal tax functions can be quite irregular.

4 Decentralization Using Retirement Savings Policies

Consider the general economy with ability types $0 \leq \theta_1 < \ldots < \theta_N < +\infty$ and present bias types $0 < \beta_1 < \ldots < \beta_M = 1$. We now study implications of our characterization of optimal allocations from Section 3 for policy design. To this end, we equip a government with three instruments that share several features with real-world retirement savings systems.

The first instrument consists of old-age transfers as a function of lifetime income, $b(y)$, which agents cannot borrow against. We think of this as resembling *Social Security* in the US.[13]

[12] See Diamond (1998, p. 86), Saez (2001, p. 218), Golosov et al. (2014, p. 13), and Golosov et al. (2016, p. 364).

[13] Indeed, the use of Social Security payment streams as collateral on loans is prohibited by federal law under Title II

The second instrument comprises a set of *retirement savings accounts*, $j = 1, \ldots, J$, with subsidy (or tax) rate $\tau_j(y)$ on agents' savings a_j up to an income-dependent contribution limit $\bar{a}_j(y)$. Among these accounts is one regular savings account with no subsidy or cap. Here we have in mind real-world voluntary retirement plans such as 401(k) and IRA accounts, which feature tax-incentivized employer matching and tax-preferred treatment up to some contribution limits.[14]

The third instrument is a *nonlinear labor income tax schedule*, $T_1(y, \{a_j\}_{j=1}^J)$, which depends on agents' gross earnings and savings in each of the voluntary retirement savings accounts. Indeed, one can interpret the dependence of our income tax schedule on savings as a tax deduction, as for contributions to 401(k) and IRA accounts in the US, though potentially a nonlinear one. This interdependence between taxes and savings is key in our environment.

In the decentralization, agents choose present and future consumption, (c_1, c_2), savings in each retirement savings account, $\{a_j\}_{j=1}^J$, and income, y. Given a set of *retirement savings policies* $(b(\cdot), \{\tau_j(\cdot), \bar{a}_j(\cdot)\}_j, T_1(\cdot))$, a *decentralized allocation* $\mathcal{A} = \{c_1(\theta,\beta), c_2(\theta,\beta), y(\theta,\beta), \{a_j(\theta,\beta)\}_j\}_{(\theta,\beta)}$ constitutes a *competitive equilibrium* if it solves each individual's problem:[15]

$$\forall (\theta,\beta): \quad \left(c_1(\theta,\beta), c_2(\theta,\beta), \{a_j(\theta,\beta)\}_{j=1}^J, y(\theta,\beta)\right) \in \underset{\left(c_1, c_2, \{a_j\}_{j=1}^J, y\right)}{\arg\max} \; u(c_1) - v\left(\frac{y}{\theta}\right) + \beta\delta u(c_2)$$

$$\text{s.t.} \quad c_1 + \sum_{j=1}^J a_j = y - T_1\left(y, \{a_j\}_{j=1}^J\right)$$

$$c_2 = b(y) + R\sum_{j=1}^J \tau_j(y)\, a_j$$

$$0 \leq a_j \leq \bar{a}_j(y) \quad \forall j = 1, \ldots, J.$$

The planner takes as given agents' maximizing behavior and picks parameters on retirement savings and tax-transfer policies to maximize welfare given social preferences δ and $\{\lambda(\theta)\}_\theta$. Proposition 8 in Appendix A.5.1 proves the existence of a general transfer function that implements the second best. The following result shows that an appropriately designed set of retirement savings policies can implement the solution to the planner's problem.

of the Social Security Act, Sec. 207 [42 USC. 407] (a); see also Feldstein and Liebman (2002).

[14] For example, for 401(k) plans, the modal employer match in 2013 was one-for-one for every dollar saved up to 6 percent of the individual's annual earnings (Aon Hewitt, 2015). Similarly, tax-deductible contributions to an IRA account are capped as a function of gross income (Internal Revenue Service, 2018).

[15] We implicitly assume that agents use accounts in order by saving up to the cap on account j before starting to save in account $j+1$, so that $a_{j+1} > 0$ implies $a_j = \bar{a}_j(y)$. This assumption is without loss of generalization as long as the subsidy rate on a given account dominates that on the next account, which we confirm numerically later.

Proposition 2. *There exists a set of retirement savings policies* $(b(\cdot), \{\tau_j(\cdot), \bar{a}_j(\cdot)\}_j, T_1(\cdot))$ *with* $J = M$ *accounts that implements the solution to the planner's problem as a competitive equilibrium.*

Proof. See Appendix A.5. □

The policy instruments in Proposition 2 must be chosen in a way that deters local and global deviations from the optimal allocation by all types. The proof closely follows the argument in Werning (2011), extended to a setting with heterogeneous preferences. That an appropriately chosen set of retirement savings policies replicates the optimal allocation allows us to study properties of specific policy tools rather than characteristics of an abstract allocation.

The following corollary describes how the decentralization mirrors the optimal allocation in that individuals optimally self-select into a subset of the available retirement savings accounts.

Corollary 2. *There exist cutoffs* $1 < \overline{\tilde{\theta}_1^+} < +\infty$ *and* $0 < \overline{\lambda_n} < +\infty$ *such that:*

1. *If* $\tilde{\theta}_1^+ \geq \overline{\tilde{\theta}_1^+}$ *and* $\lambda(\theta_1) > 0$*, then* θ_1*-types receive only old-age benefits and use none of the optional retirement savings accounts;*

2. *If* $\lambda(\theta_n) \leq \underline{\lambda_n}$ *for* $n > 1$*, then some* θ_n*-types use voluntary retirement savings accounts in addition to receiving old-age benefits.*

Proof. See Appendix A.5.3. □

Under the conditions stated in Corollary 2, optimal retirement savings policies replicate two key features of the optimal allocation. First, the lowest-ability types are bunched, with old-age benefits generous enough to force savings off their Euler equation. Second, among agents with low enough welfare weights, some will save in voluntary retirement savings accounts, as old-age benefits phase out toward higher income levels. More patient agents successively exhaust the contribution limits on voluntary savings accounts in descending order of their subsidy rate.

As is customary in the mechanism design literature, there are potentially many ways to decentralize a given allocation. However, any decentralization must spell out how different policy instruments jointly replicate the optimal allocation. In our setting, the interaction between savings and labor income taxes enters through the dependence of the income tax schedule on savings, as well as through subsidy rates and caps of the savings accounts that depend on income. Our policy tools qualitatively resemble many real-world retirement savings systems in that they force

savings at low incomes and offer a choice between subsidized savings accounts at high incomes. Yet there are also important differences. It is not our goal to claim that any real-world policies are optimal. Rather, we want to highlight qualitative features that many actual policies—maybe surprisingly—share with optimal policies as seen through the lens of our model.

5 Active-Set Algorithm

We now present a numerical solution algorithm for general of nonlinear constrained optimization problems, which we apply to solve a large-scale version of our problem.

5.1 Problem Description

Solving multidimensional screening problems has been recognized to be a difficult task both theoretically and numerically (Armstrong, 1996; Rochet and Choné, 1998). This is partly because the techniques that unidimensional optimization problems commonly rely on fail in a multidimensional context. As a result, the curse of dimensionality quickly renders solutions to these problems computationally infeasible. In the general economy with $N = N_\theta \times N_\beta$ types, the planner's problem involves $N^2 - N$ global IC constraints, 1 feasibility constraint, and $3N$ choice variables. This gives rise to the following tension: while larger type spaces are desirable for computational accuracy, the size of the problem may quickly exceed computational capacity.

This tension is less pronounced in environments where the familiar Spence-Mirrlees condition delivers that local IC implies global IC. This is the case in static, unidimensional screening problems such as the optimal nonlinear taxation problem first studied by Mirrlees (1971). However, it is well known that no such condition exists in other important settings, such as in static problems with multiple goods (Mirrlees, 1976), in dynamic problems (Kapička, 2013), or in problems with multidimensional heterogeneity (Rochet and Stole, 2003). Consequently, the first-order approach (Rogerson, 1985), which relies on sufficiency of local IC for global IC, generally fails.

Consider now the economy from Section 3.2, with the associated dual problem in Lemma 2. The complexity of this program depends on the applicability of certain *constraint qualifications*, which ensure that the linearized feasible direction set around an optimum point is an adequate representation of the actual feasible set (Nocedal and Wright, 2006, pp. 338-340). Lemma 2 establishes that a variant of Slater's condition (Boyd and Vandenberghe, 2004, pp. 226-227) applies to

our problem. Furthermore, since our problem is differentiable and convex by part 1 of Lemma 2 and features only inequality constraints, Slater's condition is equivalent to every point in the feasible set satisfying the *Mangasarian-Fromovitz constraint qualification (MFCQ)*, as shown by Solodov (2011). The MFCQ implies the existence of a compact set of Lagrange multipliers satisfying the KKT conditions, but it is not sufficient for uniqueness of these multipliers. The weakest such condition is the *linear independence constraint qualification (LICQ)*, which requires that the set of active constraint gradients is linearly independent at a given point (Wachsmuth, 2013).

In our problem, LICQ may fail when types are bunched and global IC constraints are necessary, which commonly leads to more binding constraints than choice variables.[16] This poses a challenge for Lagrangian optimization routines because LICQ is sufficient for convergence and, in practice, necessary for the stability of many such routines (Judd et al., 2018).[17]

5.2 Algorithm Details

We propose a general, computationally stable, and efficient *active-set algorithm* to solve nonlinear constrained optimization problems.[18] Our algorithm is general in that it relies on neither the paternalistic nor the redistributive formulation of our problem and can be readily extended to other nonlinear (convex) programs. It is computationally stable in that it is designed to find the unique global optimum of our problem with probability one in theory and reliably reaches this in practice. Finally, it is efficient in that it uses a fraction of the computational resources and converges in much shorter time compared to a more naive approach.

The goal of our algorithm is to iteratively determine the smallest subset of global IC constraints necessary to solve the program. To this end, our algorithm solves a large number of small-scale problems instead of directly solving the original large-scale problem. We initiate the algorithm

[16] In our environment with N types, $N^2 - N + 1$ constraints, and $3N$ choice variables, LICQ may fail with as few as $N = 4$ types and in practice fails commonly for larger N. In contrast, in a unidimensional setting, the first-order approach essentially rules out LICQ failure. This is because only $2(N-2)+2$ local IC constraints, instead of $N^2 - N$ global IC constraints, need to be included in the optimization problem. As the number of constraints is always strictly smaller than the number of choice variables in this case, $2(N-2)+3 < 3N$ for all $N \geq 1$, LICQ is guaranteed to hold.

[17] For example, `fsolve` or `fmincon` in MATLAB frequently do not converge to the global optimum for our problem.

[18] Our active-set algorithm derives its name from similar algorithms used to solve convex quadratic programs (Nocedal and Wright, 2006). The active-set architecture is also related to *cutting-plane methods*, *column generation methods*, *constraint reduction methods*, *build-up methods*, and *build-down methods* (Nocedal and Wright, 2006), as well as *ellipsoidal methods* (Hartline, 2013), which have been used in other settings. While potential applications of this type of algorithm span a wide range of nonlinear constrained optimization problems, active-set algorithms are currently not part of the standard toolkit for economists. Economic applications of large-scale nonlinear optimization programs abound, for example, in the context of monopoly pricing, market design, signaling games, screening problems with adverse selection or moral hazard, and other areas of mechanism design.

by selecting a small subset of all global IC constraints. In the beginning of each iteration step, we solve a relaxed problem that includes the current subset of IC constraints, as well as the feasibility constraint. We then process the solution to the relaxed problem (which generally will not be the solution to the full problem) by checking all global IC constraints. Finally, we add a subset of all constraints that were not included but violated, while dropping a subset of all constraints that were included but slack. We repeat this iterative procedure until we find that the solution to the relaxed problem satisfies all global IC constraints.

Formally, consider a general constrained optimization program of the following type:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g_i(x) \geq 0 \text{ for } i \in \mathcal{C}. \tag{7}$$

Here, $f(\cdot)$ is the objective function and $g_i(\cdot)$ are the inequality constraints, where $\mathcal{C} = \{1, \ldots, I\}$ is the collection of constraint indices i.[19] For any subset $\mathcal{W} \subseteq \mathcal{C}$, let $\mathcal{F}_{\mathcal{W}} = \{x \in \mathbb{R}^n | \, g_i(x) \geq 0, \, \forall i \in \mathcal{W}\}$ denote the feasible set with respect to $\mathcal{W}$. We assume that the program satisfies strong duality (as in part 4 of Lemma 2), which delivers existence of a compact set of Lagrange multipliers of the dual problem. Fixing a constraint set $\mathcal{W}$, let ξ_i denote the Lagrange multiplier on constraint $i \in \mathcal{W}$. Given a point $x \in \mathcal{F}_{\mathcal{W}}$ that is feasible with respect to $\mathcal{W}$, let $\mathcal{B}_{\mathcal{W}}(x) = \{i \in \mathcal{W} | \, \xi_i > 0\}$ be the set of binding constraints and let $\mathcal{S}_{\mathcal{W}}(x) = \mathcal{W} \backslash \mathcal{B}(x) = \{i \in \mathcal{W} | \, \xi_i = 0\}$ be the set of slack constraints. Conversely, let $\mathcal{V}(x) = \{i | \, g_i(x) < 0\} \subseteq \mathcal{C} \backslash \mathcal{W}$ be the set of violated constraints at that point. For any set S, we denote its cardinality by $|S|$. Finally, let $x^*_{\mathcal{W}}$ denote the solution to the optimization problem subject to constraints in $\mathcal{W}$.

The most general form of our proposed active-set algorithm can be stated as follows:

Algorithm. *[Active-set]*

Compute a feasible initial point $x_0 \in \mathcal{F}_{\mathcal{C}}$*;*

Select an initial working set $\mathcal{W}_0 \subseteq \mathcal{C}$ *with associated working feasible set* $\mathcal{F}_{\mathcal{W}_0} \supseteq \mathcal{F}_{\mathcal{C}}$*;*

for $k = 0, 1, 2, \ldots$*:*

Given the initial point x_k*, solve for* $x^*_{\mathcal{W}_k} = \min_{x \in \mathcal{F}_{\mathcal{W}_k}} f(x)$*;*

Find the set of violated constraints, $\mathcal{V}\left(x^*_{\mathcal{W}_k}\right)$*;*

if $\left|\mathcal{V}\left(x^*_{\mathcal{W}_k}\right)\right| = 0$*:*

stop *with solution* $x^*_{\mathcal{C}} = x^*_{\mathcal{W}_k}$*;*

[19] The inclusion of only inequality constraints is without loss of generality, since an equality constraint $g(x) = 0$ can simply be written as two inequality constraints: $g(x) \geq 0$ and $-g(x) \geq 0$.

else if $\left|\mathcal{V}\left(x^*_{\mathcal{W}_k}\right)\right| > 0$*:*

Select a random subset of omitted constraints, $\mathcal{O}^R_k \subseteq \mathcal{C}\backslash\mathcal{W}_k$*;*

Select a random subset of included constraints, $\mathcal{I}^R_k \subseteq \mathcal{W}_k$*;*

Update the working set as $\mathcal{W}_{k+1} = \left(\mathcal{W}_k \cup \mathcal{O}^R_k\right)\backslash\mathcal{I}^R_k$*;*

Update the initial point as $x_{k+1} = \chi\left(x_k, x^*_{\mathcal{W}_k}, \mathcal{F}_{\mathcal{W}_{k+1}}\right)$*;*

end (for)

A few comments are in order. First, for our problem a feasible point always exists because a uniform consumption stream $c_1(\theta,\beta) = c_2(\theta,\beta) = \bar{c}$ and income $y(\theta,\beta) = \bar{y} = (1+1/R)\,\bar{c}$ for all agents trivially satisfies feasibility and all IC constraints.

Second, an initial working set of smaller cardinality leads the relaxed problem at each iteration to be solved more quickly but requires a larger number of iterations. The working set must always contain the feasibility constraint, which we omit from the iterative constraint set selection.

Third, we efficiently solve the relaxed problem in each iterative step of the algorithm by using the Interior Point Optimizer (IPOPT; see Wächter and Biegler, 2006) library for large-scale nonlinear optimization problems, written in C++, via a Python interface.

Fourth, we update the working set by adding a subset of omitted constraints, $\mathcal{O}^R_k \subseteq \mathcal{C}\backslash\mathcal{W}_k$, and dropping a subset of included constraints, $\mathcal{I}^R_k \subseteq \mathcal{W}_k$, through a stochastic selection criterion. Randomly adding and dropping constraints avoids infinite loops between constraint sets.

Finally, we update the initial point, x_{k+1}, at the end of each iteration according to a function $\chi(\cdot)$ that depends on the previous initial point, x_k, the solution to the relaxed problem with respect to the previous working set, $x^*_{\mathcal{W}_k}$, and the feasible set with respect to the new working set, $\mathcal{F}_{\mathcal{W}_{k+1}}$. Updating the initial point is costly but can save valuable time in the next iteration of the algorithm.

5.3 Discussion of the Active-Set Algorithm

Global convergence. Once we find a solution that satisfies global IC and feasibility, convexity of the problem guarantees that this is the unique global solution (see parts 1 and 2 of Lemma 2). A fundamental property that our algorithm preserves is convergence to the global optimum.

Proposition 3. *The active-set algorithm converges with probability one to the unique global solution of the*

planner's problem:

$$\operatorname*{plim}_{k\to\infty} x^*_{\mathcal{W}_k} = x^*$$

Proof. See Appendix B.1. □

Recall that both $\mathcal{W}_k$ and $x^*_{\mathcal{W}_k}$ are random variables since the active-set algorithm selects random subsets of constraints at each iteration. Proposition 3 guarantees that the algorithm converges to the unique global solution almost surely. Selecting an efficient schedule of probabilities $\{p_i\}_i$ is an art, but adding some stochasticity is key in order for the algorithm to avoid infinite loops.[20]

Further efficiency gains. Our active-set algorithm applies to general nonlinear constrained optimization problems. We customize this algorithm, guided by economic theory and mathematical properties specific to our problem, to further enhance its performance in our application.

First, we increase computational speed through a change of variables from consumption-labor-space to utility-disutility-space, reducing the number of nonlinear constraints from $N^2 - N$ to 1.[21]

Second, we pick as an initial point, x_0, the laissez-faire solution, for which, under the assumption of power utility later adopted in our quantitative exercise, we have closed-form expressions.[22] This starting point guarantees that all IC constraints and the feasibility constraint are satisfied and leads to more efficient solutions than a uniform allocation across types.

Third, for the initial working set, we pick the set of "local" IC constraints:

$$\left\{ IC\left(\theta_{i_1}, \beta_{i_2}, \theta_{j_1}, \beta_{j_2}\right) \middle| \, i_1 = j_1 + 1, |i_2 - j_2| \leq 1 \right\},$$

which consists of all IC constraints from type i to type j such that θ_{i_1} is the ability level above θ_{j_1} and β_{i_2} is adjacent to β_{j_2}.[23] Guided by the usual Mirrleesian intuition that downward-binding IC constraints in ability space are relevant, this initial guess (although generally not sufficient) allows for further efficiency gains relative to more naive initial guesses we tried.

[20] Global convergence would also be guaranteed by, for example, brute force iteration through all possible subsets of IC constraints. Our algorithm retains this important property while—compared to the brute-force method—converging much faster and sidestepping LICQ failure by including only a small subset of all constraints.

[21] See the proof of Lemma 2 for details regarding this change of variables. We describe the semi-matrix representation of the transformed planner's problem in Appendix B.2.

[22] We provide further details and derive closed-form expressions for the alternative initializations in Appendix B.3.

[23] Note that in the planner's problem, there are $3N_\theta N_\beta - 2N_\theta - 3N_\beta + 2 < 3N_\theta N_\beta$ such constraints, guaranteeing that the initial working set contains fewer constraints than choice variables, so LICQ always holds in the initial step.

Fourth, to select a random subset of omitted constraints, $\mathcal{O}_k^R \subseteq \mathcal{C}\backslash\mathcal{W}_k$, we form a constraint index list in descending order of the absolute distance of the constraint violation:

$$\left(i_1,\ldots,i_{\left|\mathcal{V}\left(x^*_{\mathcal{W}_k}\right)\right|},\ldots,i_{|\mathcal{C}\backslash\mathcal{W}_k|}\right), \quad -g_{i_1}\left(x^*_{\mathcal{W}_k}\right) \geq \ldots \geq -g_{i_{\left|\mathcal{V}\left(x^*_{\mathcal{W}_k}\right)\right|}}\left(x^*_{\mathcal{W}_k}\right) > 0 \geq \ldots \geq -g_{i_{|\mathcal{C}\backslash\mathcal{W}_k|}}\left(x^*_{\mathcal{W}_k}\right),$$

from which we include in $\mathcal{O}_k^R$ the n-th element with probability p_n, such that $p_n \geq p_{n+1}$ for each $n = 1,\ldots,|\mathcal{C}\backslash\mathcal{W}_k|$. Analogously, to select a random subset of included constraints, $\mathcal{I}_k^R \subseteq \mathcal{W}_k$, we form a constraint index list of, first, in ascending order the Lagrange multiplier and, second, in descending order the absolute distance of the constraint slackness:

$$\left(i_1,\ldots,i_{\left|\mathcal{S}\left(x^*_{\mathcal{W}_k}\right)\right|},\ldots,i_{|\mathcal{W}_k|}\right), \quad \xi_{i_1} = \ldots = \xi_{\left|\mathcal{S}\left(x^*_{\mathcal{W}_k}\right)\right|} = 0 < \ldots \leq \xi_{i_{|\mathcal{W}_k|}}, \quad g_{i_1}\left(x^*_{\mathcal{W}_k}\right) \geq \ldots \geq g_{i_{\left|\mathcal{S}\left(x^*_{\mathcal{W}_k}\right)\right|}}\left(x^*_{\mathcal{W}_k}\right),$$

from which we include in $\mathcal{I}_k^R$ the n-th element with probability q_n, such that $q_n \geq q_{n+1}$ for each $n = 1,\ldots,|\mathcal{W}_k|$.

Fifth, when updating the initial point, our choice of function $\chi(\cdot)$ either reuses the globally feasible initial point, $x_{k+1} = x_0$, or, alternatively, determines x_{k+1} through an *alternating projection algorithm* (Cheney and Goldstein, 1959), which finds the point closest to the previous relaxed problem's solution, $x^*_{\mathcal{W}_k}$, that lies in the feasible set with respect to the updated working set, $\mathcal{F}_{\mathcal{W}_{k+1}}$.

Finally, we make use of our theory by imposing single-crossing in the β-dimension conditional on θ (part 2 of Lemma 1): whenever the IC constraint of some type (θ,β) is satisfied with respect to type (θ,β') for some $\beta' < \beta$ (or $\beta' > \beta$), then all IC constraints of type (θ,β) with respect to type (θ,β'') are also satisfied for $\beta'' < \beta'$ (or $\beta'' > \beta'$).

Benchmarking. Appendix B.4 presents benchmarks from a range of practical applications of the active-set algorithm, showcasing its computational efficiency.

Potential limitations. While the active-set algorithm has many advantages and performs well in practical applications, we want to highlight three potential limitations.

First, although Proposition 3 guarantees convergence in theory, it may take a long time for the algorithm to converge in practice. Proofs concerning the rate of convergence of numerical algorithms are often challenging, and we have no such result for our algorithm. Luckily, we find that in practice, the algorithm converges quickly even for large-scale problems of the type that we study.

Second, LICQ failure may still be an issue when there are actually more binding constraints than choice variables at the optimum. In cases in which the algorithm builds up to a working set with cardinality exceeding the number of choice variables, using an optimization routine robust to mild LICQ failure, such as IPOPT, yields stable results in a reasonable amount of time.

Third, solving the global problem through an active-set algorithm may be redundant under certain conditions.[24] However, our active-set algorithm could still be useful. For certain problems it may not be known whether or not the first-order approach applies. Even if the first-order approach applies, defining the appropriate notion of "local" may be complex (see, e.g., Armstrong, 1996). Finally, even if the first-order approach applies and the notion of "local" is well defined in theory, the first-order approach may still fail in applications using numerical approximations.[25]

6 Calibration

We now operationalize our model by calibrating its key parameters to US microdata on lifetime income and savings decisions. Our calibration proceeds in two steps. In the first step—the positive model calibration—we compute the distribution of earnings ability and present bias that matches the empirical lifetime incomes and conditional lifetime savings rates. In the second step—the normative model calibration—we use the inverse-optimum approach to infer social preferences that best match existing tax-transfer and retirement savings policies.[26]

6.1 Positive Model Calibration

The goal of the positive model calibration is to infer the joint distribution of ability and present bias from empirical lifetime earnings and conditional lifetime savings rates.

[24] For example, this may be the case when local IC constraints are sufficient for global optimality (Carroll, 2012; Battaglini and Lamba, 2018), when a multidimensional screening problem can be rewritten as a unidimensional one (Rothschild and Scheuer, 2013), when the type space dimension is larger than the allocation space dimension (Pass, 2012), for problems with random participation (Rochet and Stole, 2002; Jacquet et al., 2013), or when screening optimally occurs along each unobserved type component separately (Carroll, 2017).

[25] For example, Proposition 5 of Carroll (2012) shows that the first-order approach applies in a setting with transfers and interdependent utility that is linear in types, if all agents have a convex type space. But Proposition 6 of the same paper shows that the first-order approach does not apply under a strong form of nonconvexity of the type space, which may be relevant when approximating a convex continuous type space with a (nonconvex) set of discrete points.

[26] Appendix C.1 describes the data we use for the calibration of our model, which we draw from the Health and Retirement Study (HRS), the Panel Study of Income Dynamics (PSID), the US Life Tables (Arias, 2010), the Health Inequality Project (Chetty et al., 2016), and other data sources. Appendix C.2 details our data cleaning and sample selection procedures. Appendix C.3 documents how we construct key variables.

6.1.1 Identification

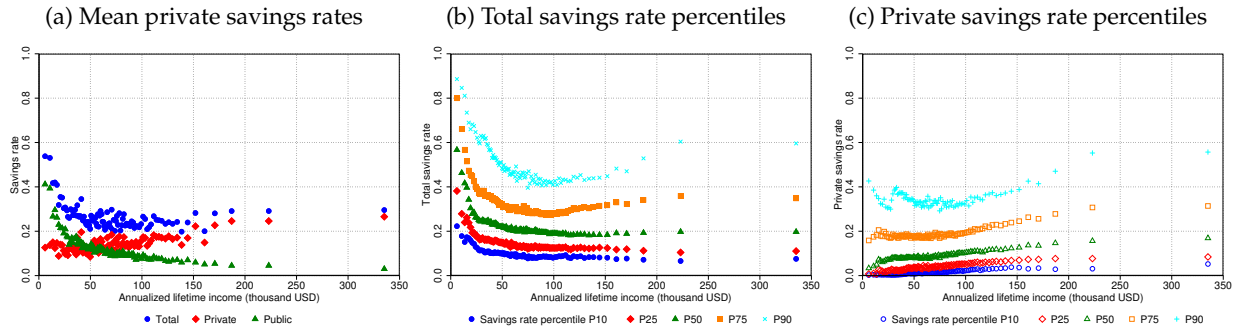
We first present some moments of the distributions of savings rates in the HRS data. Let us define the *total lifetime savings rate* as the ratio of wealth at retirement to lifetime income. We further split the total lifetime savings rate into *private savings* (net value of real estate, businesses, IRA accounts, stocks, etc.) and *public savings* (predicted Social Security wealth). Through the lens of our model, these moments will be informative about an individual's time preference conditional on other factors that may affect savings decisions. We make two observations on the distribution of retirement savings rates within and across income groups.

Our first observation is that total savings rates are swoosh-shaped across income levels, with private savings rates increasing but public savings rates decreasing in income. Figure 2(a) plots the three lifetime savings rates as a function of annualized lifetime income. The total lifetime savings rate declines for the first USD 50,000 in annual income, from around 53 percent to 23 percent, before stabilizing and slowly increasing again up to 31 percent. This is largely accounted for by declining public savings rates over this income range, due to the decreasing replacement ratio of Social Security benefits. In contrast, mean private savings rates increase monotonically from 13 percent to 27 percent throughout this income range.

Our second observation is that there is substantial dispersion of savings rates within income levels. Panel (b) of Figure 2 shows percentiles of total savings rates, while panel (c) shows percentiles of private savings rates throughout the income distribution. We find substantial heterogeneity in both savings rates conditional on income. Total savings rates vary by more than 30 percentage points between the 10th and 90th percentiles of the savings distribution. While the 25th percentile of private savings rates is close to zero at most income levels, the 90th percentile ranges from 30 to 56 percent.

To isolate the role of present bias heterogeneity in determining lifetime savings rates, we bring to the data a much richer version of our framework, allowing for many factors other than present bias heterogeneity to affect savings decisions: survival risk, longevity heterogeneity, bequest motives, and medical expenditure shocks. It is important to note that, naturally, we adopt only a subset of all possible savings motives in our framework. For instance, we abstract from life-cycle uninsurable income shocks due to the technical difficulty of handling dynamics in our already complex model. To the extent that we ignore other important factors driving empirical savings

Figure 2. Moments of the savings rate distributions conditional on lifetime income



Notes: Annualized lifetime income is the net present discounted value of net income during working life, divided by the number of years worked. Private savings are defined as private wealth at retirement age. Public savings are defined as predicted Social Security wealth at retirement age. Total savings is the sum of private and public savings. Savings rates are defined as the share of savings out of total net income, both in net present discounted value terms. Source: Authors' calculations based on HRS.

behavior, our procedure will misinterpret these omitted factors as preferences. In this sense, preferences in our model pick up the residual of savings rates—conditional on other factors that we account for—in the data.

6.1.2 Implementation

We now introduce the richer model that we use to infer heterogeneity in θ and β from the data.[27]

Model setup. At the beginning of their retirement life, agents are characterized by their net income after the realization of a medical expenditure shock $m_2(h)$, their warm-glow bequest parameter ϕ_1, and their retirement life length T_R. Old agents then choose how much to consume, c_2, and how much to bequeath, b^{late}. Conditional on surviving and the old-age medical shock realization, an old agent solves the following problem:

$$\begin{aligned} U_2(y_2;\phi_1,T_R) &= \max_{c_2,b^{late}} \left\{ u(c_2) + \Phi\left(b^{late};\phi_1\right) \right\} \\ \text{s.t.} \quad c_2 + \tilde{b} &= \max\{y_2, \underline{c}_2\} \\ b^{late} &= \tilde{b} - T_b\left(\tilde{b}\right), \end{aligned}$$

where y_2 is retirement wealth net of taxes and old-age medical expenditures, $\Phi\left(b^{late};\phi_1\right)$ is warm-glow utility from bequeathing b^{late}, $\tilde{b}$ is the gross bequest amount, R is the annual gross interest

[27] We here present the agent's problem in a two-period formulation. Appendix C.4 details the complete life-cycle model with annual computations underlying our calibration.

rate, $\underline{c}_2 > 0$ is a minimum consumption floor, and $T_b\left(\tilde{b}\right)$ is the estate tax schedule.

At the beginning of their working life, agents are characterized by their earnings ability θ and present bias level β, a health type h, warm-glow bequest motive ϕ_1, their working life length T_W, their retirement life length T_R, and their probability of dying before retirement, $P\left[death\right]$. Young agents then choose how much to consume, c_1, how much to save in 401(k), IRA, and regular taxable accounts, $\left(s^{401k}, s^{IRA}, s^{taxable}\right)$, and how much labor to supply, ℓ, given medical expenditures $m_1\left(h\right)$. Given the continuation problem when old, a young agent solves the following problem:

$$
\begin{aligned}
U_1\left(\theta,\psi;h,\phi_1,T_W,T_R,P\left[death\right]\right) = & \max_{c_1,s^{401k},s^{IRA},s^{taxable},y_1} \left\{ \begin{array}{c} u\left(c_1\right) - v\left(y_1/\theta\right) \\ +\psi \left[\begin{array}{c} \left(1-P\left[death\right]\right) \times \mathbb{E}_{m_2} U_2\left(y_2\left(m_2\right);\phi_1,T_R\right) \\ +P\left[death\right] \times \Phi\left(b^{early};\phi_1\right) \end{array} \right] \end{array} \right\} \\
\text{s.t.} \quad c_1 + s^{401k} + s^{IRA} + s^{taxable} = & \max\left\{ y_1 - T_y\left(\max\left\{ y_1 - s^{401k} - s^{IRA} - m_1, 0 \right\}\right) - m_1, \underline{c}_1 \right\} \\
\tilde{y}_2^{pre-tax} = & \ 0.85 \times SS\left(y_1\right) + R \times match\left(s^{401k}\right) + Rs^{IRA} + \left(R-1\right)s^{taxable} \\
\tilde{y}_2^{post-tax} = & \ 0.15 \times SS\left(y_1\right) + s^{taxable} \\
y_2\left(m_2\right) = & \ \tilde{y}_2^{pre-tax} - T_y\left(\max\left\{ \tilde{y}_2^{pre-tax} - m_2, 0 \right\}\right) - m_2 + \tilde{y}_2^{post-tax} \\
\tilde{\tilde{b}} = & \ R \times match\left(s^{401k}\right) + Rs^{IRA} \\
\tilde{b} = & \ \tilde{\tilde{b}} - T_y\left(\tilde{\tilde{b}}\right) + Rs^{taxable} \\
b^{early} = & \ \tilde{b} - T_b\left(\tilde{b}\right) \\
s^{401k} \leq & \ \overline{s}^{401k}\left(y_1\right) \\
s^{IRA} \leq & \ \overline{s}^{IRA},
\end{aligned}
$$

where $\psi = \beta\delta$ is the compound discount factor between periods, y_1 is income during working life, b^{early} denotes accidental bequests before retirement, $T_y\left(\cdot\right)$ is the income tax schedule, $\underline{c}_1 > 0$ is a minimum consumption floor, $\tilde{y}_2^{pre-tax}$ and $\tilde{y}_2^{post-tax}$ are pre- and post-tax retirement wealth, and $y_2\left(\cdot\right)$ is retirement wealth net of taxes and out-of-pocket old-age medical expenditures, $m_2 \sim H\left(\cdot;h\right)$. Furthermore, $SS\left(\cdot\right)$ denotes Social Security old-age benefits, $match\left(\cdot\right)$ is the inclusive-of-match wealth in 401(k) accounts, $\tilde{\tilde{b}}$ denotes accidental bequests gross of income and estate taxes, $\tilde{b}$ is the accidental bequest gross of estate taxes, and $\overline{s}^{401k}$ and $\overline{s}^{IRA}\left(\cdot\right)$ are the limits for tax-exempt contributions to 401(k) and IRA accounts, respectively.

Preferences. We parameterize household preferences by constant relative risk aversion (CRRA) utility of consumption and power disutility of labor effort:

$$u(c) = \begin{cases} \frac{c^{1-1/\sigma}-1}{1-1/\sigma} & \text{for } \sigma \neq 1 \\ \ln(c) & \text{for } \sigma = 1 \end{cases}, \quad v(\ell) = \kappa \frac{\ell^{1+1/\gamma}}{1+1/\gamma},$$

where $\ell = y/\theta$ is labor supply. We fix values for the intertemporal elasticity of substitution, $\sigma = 1.5$, labor disutility intercept, $\kappa = 1$, and Frisch elasticity of labor supply, $\gamma = 1$.

Following De Nardi (2004), warm-glow utility from a net bequest of b function is $\Phi(b;\phi_1) = \phi_1 (1 + b/\phi_2)^{1-\phi_3}$, where (ϕ_1, ϕ_2, ϕ_3) are parameters guiding, respectively, the scale, nonhomotheticity, and income elasticity of the bequest motive. We adopt $\phi_2 = 11.6$ and $\phi_3 = 1.5$ from De Nardi (2004) and calibrate ϕ_1 individual by individual to match the empirical bequest expectation at retirement in the data and in the model.

As Dynan et al. (2004), we choose a consumption floor of $\underline{c} = 10,000$ US dollars to capture home production and informal safety nets such as insurance within the family.

Survival risk and longevity. We estimate $P[death]$ as the probability of death before retirement in the HRS, combining mean survival hazards from the US Life Tables with the survival probability gradient across income deciles from the Health Inequality Project. We account for longevity differences by summing years of working life and retirement in the combined data.

Income tax and transfer function. We adopt a reduced-form tax function commonly used in the literature (Feldstein, 1969; Bénabou, 2002): $T(y) = y - \lambda y^{1-\tau}$, where $T(y)$ is net transfer as a function of income, τ is a tax-progressivity parameter, and λ is a tax-level parameter. Following Heathcote et al. (2017, henceforth HSV), we estimate τ and λ by regressing log net income on log gross income in the RAND HRS Tax Calculations data, yielding $\tau = 0.197$ and $\lambda = 5.114$.[28]

Estate tax function. We approximate the estate tax as $T_b(b) = \tau_b \times \max(0, b - \bar{e})$, where $\tau_b = 0.1$ is the marginal tax on bequests and the estate tax exemption level is $\bar{e} = 2,000,000$ dollars, corresponding to roughly 40 years of average earnings (De Nardi, 2004).

[28] While Heathcote et al. (2017) estimate this tax function on the PSID data, we here use the HRS tax data to be consistent with the HRS sample of households at the center of our analysis. Using their data, they report a tax progressivity parameter of $\tau^{HSV} = 0.181$, slightly below our point estimate.

Social Security. We approximate the Social Security benefit schedule as a function of average annual working life income, $y_1 \geq 0$, using the real-world 2014 policy parameters:

$$SS(y_1) = \begin{cases} 0.9y_1 & \text{for } y_1 < 9,912 \\ 0.9 \times 9,912 + 0.32(y_1 - 9,912) & \text{for } 9,912 \leq y_1 < 59,760 \\ 0.9 \times 9,912 + 0.32(59,760 - 9,912) + 0.15(y_1 - 59,760) & \text{for } 59,760 \leq y_1 < 118,500 \\ 0.9 \times 9,912 + 0.32(59,760 - 9,912) + 0.15(118,500 - 59,760) & \text{for } y_1 \geq 118,500 \end{cases}$$

Savings accounts. We model three optional savings accounts. The first account is a 401(k) account with pre-tax contributions and an employer matching rate of 50 percent up to a cap of $\bar{s}^{401k}(y_1) = 0.06y_1$ (Financial Engines, 2015). The second account is an IRA account with tax-deferred contributions up to $\bar{s}^{IRA} = 5,500$ dollars but no matching (Internal Revenue Service, 2018). The third account is a regular savings account offering a real annual interest rate of 3.44 percent (Gourinchas and Parker, 2002). In the second period, withdrawals from all three accounts are taxed at the statutory income tax rate that applies to the individual.

Medical expenditures. We denote by $h \in \{h_1, \ldots, h_{N_H}\}$ the agent's health risk type, which is associated with medical expenditures during working life, $m_1(h)$, and during retirement, $m_w(h) \sim H(\cdot; h)$. We model $H(\cdot; h)$ as a lognormal distribution with mean μ_m^h and standard deviation σ_m^h estimated separately for groups by education and income decile. We approximate $H(\cdot; h)$ on a two-point grid, using the 10th and 90th percentiles of the group-specific medical expenditure distribution. This simple but flexible specification captures some key features of the empirical income-health gradient and the wealth-health gradient (De Nardi et al., 2018).

Calibration targets and parameters. Appendix C.4.1 describes details of how we calibrate the distribution over (θ, ψ, ϕ_1)—earnings ability, compound discount factors, and strength of bequest motive—to minimize the distance between individual lifetime income, savings rates, and expected bequest in the model and in the data. Through this, we recover a parameter vector consisting of the mean and standard deviation of a lognormal approximation to individual earnings ability $(\mu_\theta, \sigma_\theta)$, and intercepts and slopes of the shape parameters for a beta distribution for compound discount factors, $(a_{\psi,0}, a_{\psi,1}, b_{\psi,0}, b_{\psi,1})$, estimated as affine functions of ability quantiles.

6.1.3 Results

Table 1 summarizes our positive calibration results.[29] The lognormal distribution of earnings abilities has mean 9.659 and standard deviation 0.420. The mean compound (annualized) discount factor is 0.103 (0.941). The distribution of compound (annualized) discount factors is heavily left-skewed, ranging from 0.009 (0.885) to 0.214 (0.965) between the 10th and 90th percentiles. Compound (annualized) discount factors and earnings ability are positively correlated, with a correlation coefficient of 0.163 (0.153). Thus, our results reflect significant variation in time preferences, both within and across ability levels.

Table 1. Positive calibration results

PANEL A. CALIBRATED PARAMETERS	Values		
Mean of lognormal earnings abilities, μ_θ	9.659		
St.d. of lognormal earnings abilities, σ_θ	0.420		
Intercept of first beta shape parameter of discount factors, $a_{\psi,0}$	13.997		
Slope of first beta shape parameter of discount factors, $a_{\psi,1}$	8.840		
Intercept of second beta shape parameter of discount factors, $b_{\psi,0}$	2.083		
Slope of second beta shape parameter of discount factors, $b_{\psi,1}$	0.384		
PANEL B. IMPLIED MOMENTS	**Mean**	**P10**	**P90**
Earnings ability, θ	16,379	10,706	22,924
Compound discount factor, ψ	0.103	0.009	0.214
Annualized discount factor, ψ_{annual}	0.941	0.885	0.965
Corr. between ability and compound discount factor, $Corr\,(\theta, \psi)$	0.163		
Corr. between ability and annualized discount factor, $Corr\,(\theta, \psi_{annual})$	0.153		

Note: See text for details. Source: Authors' calculations based on model.

Table 2 shows that the positive model matches salient features of empirical income and savings rates. In the model, as in the data, savings rates are convex across savings percentiles for two reasons. First, individuals may choose to rely on the minimum consumption floor rather than saving privately (Buchanan, 1975). Second, warm-glow bequests are a luxury good, with a small share of households leaving large bequests (De Nardi, 2004).

Table 3 shows savings rates across income and wealth groups. In the model, as in the data, there are large differences in savings rates conditional on income (Venti and Wise, 1998). Our model generates this through present bias dispersion within earnings ability. The calibrated model matches the positive gradient of savings rates across lifetime earnings levels (Dynan et al., 2004).

[29] Figure 17 in Appendix C.4 shows the implied marginal and average tax rates under the HSV tax function, as well as the fit of the estimated tax function to pre- versus post-tax income in the HRS data.

Table 2. Income and savings in data vs. calibrated model

	Annual income (USD)		Savings rate	
	Data	Model	Data	Model
Mean	97,685	93,461	0.118	0.111
10th pctl	47,194	43,266	0.000	0.003
25th pctl	63,266	63,089	0.028	0.020
50th pctl	85,093	72,908	0.101	0.050
75th pctl	120,197	116,153	0.181	0.139
90th pctl	162,127	163,207	0.262	0.243

Notes: Annual income and savings rates are computed at the level of the household. See text for details. Source: Authors' calculations based on model, HRS, and PSID.

Table 3. Savings rates in data vs. calibrated model, by lifetime income and retirement wealth quartiles

Lifetime income	Q1 (lowest)		Q2		Q3		Q4 (highest)	
	Data	Model	Data	Model	Data	Model	Data	Model
Wealth Q1 (lowest)	0.000	0.025	0.000	0.023	0.035	0.046	0.043	0.032
Wealth Q2	0.011	0.031	0.045	0.081	0.102	0.114	0.095	0.066
Wealth Q3	0.074	0.102	0.143	0.185	0.165	0.123	0.149	0.122
Wealth Q4 (highest)	0.295	0.222	0.260	0.226	0.265	0.184	0.217	0.203

Notes: Lifetime income, retirement wealth, and mean savings rates are computed at the level of the household. See text for details. Source: Authors' calculations based on model, HRS, and PSID.

6.2 Normative Model Calibration

Taking as given the positive calibration results, we now calibrate social preferences.

6.2.1 Identification

In theory, we could use any social preferences for the optimal policy analysis. Following a growing strand of the literature, we use the inverse optimum approach (Bourguignon and Spadaro, 2012; Lockwood and Weinzierl, 2016; Heathcote and Tsujiyama, 2017) to select Pareto weights that yield the optimal allocation closest to that under current real-world policies.[30] In our baseline normative calibration, we fix $\delta = 1/R = 0.214$, where R is the compounded gross interest rate between periods that corresponds to an annual net interest rate of 3.44 percent (Gourinchas and Parker, 2002).[31]

[30] Stantcheva (2016) discusses some of the advantages and drawbacks of the inverse optimum approach.

[31] Fixing $\delta = 1/R$ is without much loss of generality. In an extended version of the normative calibration, we add the planner's discount factor, δ, to the optimization procedure and find an optimal compound value of $\delta = 0.261$, which constitutes a small difference in annualized discount factors.

6.2.2 Implementation

We allow for nonmonotonic welfare weights, $\lambda(\theta; \lambda_w, \lambda_1, \lambda_2) = \lambda_w \times beta(\rho(\theta), \lambda_1, \lambda_2) + (1 - \lambda_w)$, where $beta(\cdot, \lambda_1, \lambda_2)$ is a beta distribution with shape parameters λ_1 and λ_2, $\rho(\theta) \in [0, 1]$ is the rank of earnings ability, with $\rho = 0$ ($\rho = 1$) representing the lowest (highest) ability, and λ_w is the relative weight on the beta distribution versus uniform weights. This specification is flexible enough to encompass utilitarian, monotonically upward- or downward-sloping Pareto weights, and various hump shapes.[32] Appendix C.4.2 details how we pick $(\lambda_1, \lambda_2, \lambda_w)$ to minimize the distance between the optimal allocation and that under real-world policies. We use a grid over 1000 θ-types and 6 β-types for this calibration and all results. We impose exogenous government spending of USD 40,058 per capita to match the difference between aggregate gross earnings and consumption from our positive calibration.

6.2.3 Results

Table 4 summarizes our normative calibration results. We find that $\lambda_1 = 1.743$, $\lambda_2 = 9.999$, and $\lambda_w = 0.358$ brings the normative model allocation closest to that under current policies.[33] We find that the implied distribution of Pareto weights, shown in Figure 3, is indeed hump-shaped, putting relatively lower weight on agents at the top and the bottom of the ability distribution relative to the center. A compound discount factor of $\delta = 0.214$ between working life and retirement, together with the values of ψ from the positive calibration in Section 6.1, implies an annualized mean level of present bias of $\mathbb{E}[\beta_{annual}] = 0.973$, corresponding to a mean compound present bias level of $\mathbb{E}[\beta] = 0.482$ between two periods. We find substantial dispersion in our present bias estimates, ranging from annualized values of 0.915 to 0.998 between the 10th and 90th percentiles of the distribution.

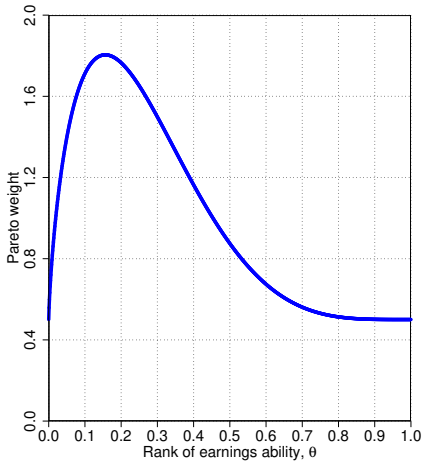
[32] Figure 18 in Appendix C.4.2 demonstrates the flexibility of this welfare weight specification. Heathcote and Tsujiyama (2017) parameterize Pareto weights across ability levels as $\lambda(\theta) = \exp(-\alpha\theta) / (\sum_{\theta', \beta'} \pi(\theta', \beta') \exp(-\alpha\theta'))$, where $\alpha \in \mathbb{R}$ indexes the government's redistributive motive. This specification encompasses decreasing ($\alpha > 0$), utilitarian ($\alpha = 0$), and increasing ($\alpha < 0$) welfare weights but may be too restrictive for certain applications, as it constrains Pareto weights to be monotonic in θ. For example, recent findings by Jacobs et al. (2017) suggest that Pareto weights under the inverse optimum approach are hump-shaped in the Netherlands.

[33] Our estimated Pareto weights are more progressive relative to those in Heathcote and Tsujiyama (2017) because our paternalistic model rationalizes high levels of forced savings through Social Security at low through relatively high Pareto weights on low-ability types.

Table 4. Parameter values and implied moments

PANEL A. EXTERNALLY SET PARAMETERS	Values		
Compound social discount factor, δ	0.214		
PANEL B. CALIBRATED PARAMETERS	**Values**		
First beta shape param. of Pareto weights, λ_1	1.751		
Second beta shape param. of Pareto weights, λ_2	5.050		
Weight on beta dist. of Pareto weights, λ_w	0.499		
PANEL C. IMPLIED MOMENTS	**Mean**	**P10**	**P90**
Annualized social discount factor, δ_{annual}	0.967		
Compound present bias, β	0.482	0.042	0.981
Annualized present bias, β_{annual}	0.973	0.915	0.998

Figure 3. Calibrated Pareto weights



Notes: Table and figure show best fit of Pareto weights under the inverse optimum approach with $\lambda(\theta) = \lambda_w \times beta(\rho(\theta), \lambda_1, \lambda_2) + (1 - \lambda_w)$, where λ_w is the relative weight on a beta distribution versus a uniform weight on all agents, $beta(\cdot, \lambda_1, \lambda_2)$ is a beta distribution with shape parameters λ_1 and λ_2, and $\rho(\theta) \in [0,1]$ is the rank of earnings ability, with $\rho = 1$ representing the highest ability and $\rho = 0$ representing the lowest ability. See text for details. Source: Authors' calculations based on model.

6.3 Interpretation of Results and Context

In the preceding analysis, we found substantial dispersion in savings rates conditional on a range of other factors, which included survival risk, longevity heterogeneity, bequest motives, and medical expenditure shocks. Our enriched model rationalized this dispersion through heterogeneity in the compound discount factor $\psi = \beta\delta$. The interpretation of this heterogenity is a pivotal issue. One interpretation is that all heterogeneity in time preferences is in line with the planner's evaluation of time preferences, maybe because observed differences in behavior are based on individuals' rational, altruistic, and farsighted optimization, and the government respects their decisions. A diametrically opposed interpretation is that the government respects a unique time preference, and agents to varying degrees deviate from that preference.[34]

Given the large and growing body of evidence for behavioral biases shaping household finances, the behavioral interpretation of our evidence seems like a natural starting point.[35] The question then arises: how much of the observed heterogeneity in time preferences reflects present

[34] At this point, it is helpful to recall the three justifications for paternalism that motivated our analysis: behavioral biases such as hyperbolic discounting (Laibson, 1997), the Samaritan's dilemma (Sleet and Yeltekin, 2006), and positive social welfare weight on future selves (Caplin and Leahy, 2004).

[35] See Beshears et al. (2018) for a comprehensive recent survey.

bias heterogeneity, as opposed to differences in geometric discount factors? With observational data on retirement wealth and lifetime income such as that in the HRS, it is all but impossible for us to provide a definitive answer to this question. Thus, we take a four-pronged approach.

First, interpreting ψ-heterogeneity as present bias heterogeneity is internally consistent. In the HRS microdata, households with below-median savings rates are 73 percent more likely to have thought "hardly at all" about retirement, and 25 percent less likely to have thought "a lot" about retirement (Engen et al., 2005). This suggests that the nature of the decision process differs across savings groups, consistent with experimental evidence by Burks et al. (2009).

Second, this interpretation is also externally consistent with a range of laboratory and field estimates, which we summarize in Appendix C.5. Bernheim et al. (2001) use income and consumption data to evaluate competing explanations for savings differences, including heterogeneity in geometric discount factors, risk tolerances, uncertainty, and age-dependent tastes for work and leisure. They find that consumption growth rates near retirement do not vary systematically with wealth at retirement, leading them to reject a model with differences in exponential discount factors in favor of a model with present bias heterogeneity.

Third, we extend the inverse optimum approach to infer the implied degree of paternalism from current policies. To this end, we add as an additional parameter to the normative calibration the degree of paternalism, $p \in [0,1]$, which guides the planner's evaluation of intertemporal trade-offs of β-types by applying the effective discount factor $\delta\left(\tilde{\delta}, \beta; p\right) = p\tilde{\delta} + (1-p)\,\beta\tilde{\delta}$, where $\tilde{\delta} = 1/R$ is the benchmark discount factor. If $p = 1$, as studied so far, then the planner is perfectly paternalistic and ignores agents' behavioral discount factor. If $p = 0$, then the planner fully respects every agent's behavioral discount factor. We find that the degree of paternalism that most closely approximates current policies is $p = 0.633$.[36] Therefore, to rationalize current real-world policies, the planner must be mostly paternalistic, which justifies our benchmark assumption.

Fourth, in Section 7.3.1, we explore optimal policies under alternative interpretations of the estimated discount factor heterogeneity. We find that welfare calculations are similar for degrees of paternalism between our benchmark and the calibrated value. In contrast, optimal savings rates are starkly different from reality under no paternalism. Therefore, we conclude that justifying real-world retirement savings policies requires a substantial degree of paternalism.

[36] Our estimates of the other parameters remain qualitatively unchanged. In a full normative calibration, where we estimate the planner's discount factor in addition to the degree of paternalism, we find $\delta = 0.208$ and $p = 0.666$.

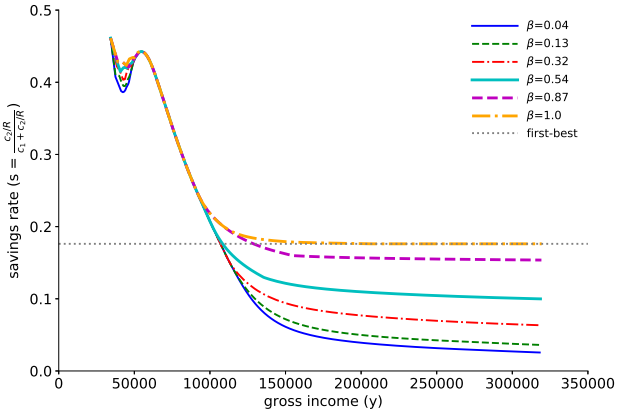
7 Optimal Policies and Reforms to the Current System

In this section, we use the calibrated model to evaluate optimal paternalistic savings policies and compare them to the current US retirement savings system.

7.1 Properties of the Optimal Allocation

Dispersion in Savings Rates. How much savings choice is optimal throughout the income distribution? Figure 4 plots optimal retirement savings rates from our calibrated normative model as a function of income and present bias levels. The lowest income levels optimally save at a uniform rate of 46.2 percent—more than twice the first-best savings rate of 17.7 percent. Savings rates remain essentially uniform until around USD 95,000 in income, when they start fanning out across present bias levels. At an income of USD 320,000, agents with $\beta = 1$ save 17.7 percent, while agents with $\beta = 0.04$ save as little as 2.7 percent. Compared to the empirical total savings rates in Figure 2, optimal savings rates start out at a similarly high level but decline more slowly between USD 50,000 and USD 95,000 in income. However, there is more dispersion in empirical savings rates than judged optimal by our framework, particularly among households earning between USD 50,000 and USD 95,000.

Figure 4. Savings rates across incomes and present bias levels

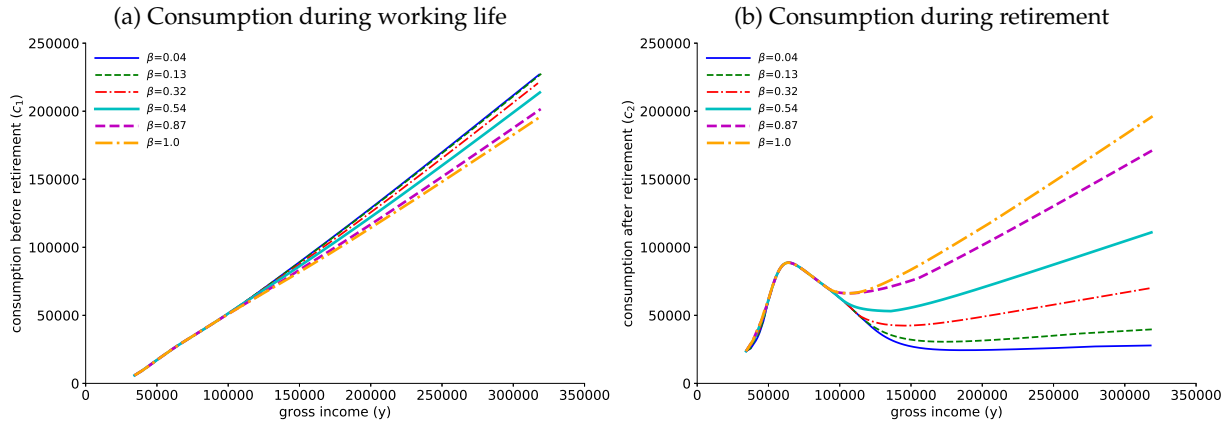


Note: Savings rate is defined as $s = (c_2/R) \,/\, (c_1 + c_2/R)$. Source: Authors' calculations based on model.

Consumption Inequality during Working Life and Retirement. Figure 5(a) shows that, among the young, optimal consumption inequality conditional on income is relatively small at all incomes, suggesting that it is costly to separate agents across present bias levels during working

life. Therefore, most consumption inequality among the young is between income groups. Panel (b) shows that, among the old, consumption inequality conditional on income is substantial above USD 95,000 in working life income.[37] Optimal old-age consumption differs by a factor of 7 between agents with $\beta = 0.04$ and those with $\beta = 1$ at an income of USD 320,000, suggesting that most of the screening of between present bias levels occurs on the intertemporal margin. Consequently, higher old-age consumption inequality within lifetime income groups is optimal.

Figure 5. Consumption streams across incomes and present bias levels



Note: Figure shows $c_1\left(y\left(\theta,\beta\right),\beta\right)$ in panel (a) and $c_2\left(y\left(\theta,\beta\right),\beta\right)$ in panel (b). Source: Authors' calculations based on model.

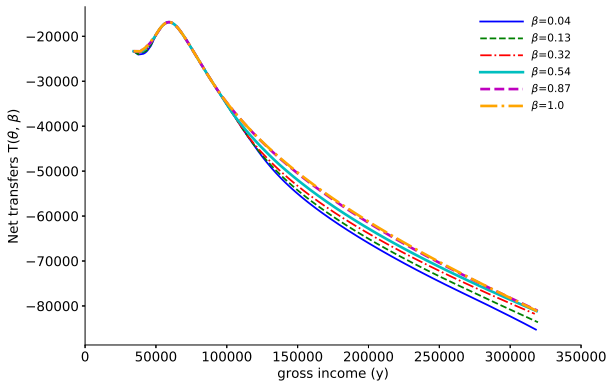
Net Transfers. Figure 6 shows the optimal net transfer schedule.[38] Below a threshold of USD 100,000 in income, net transfers essentially do not vary across present bias levels. Above that threshold, agents with $\beta = 0.04$ pay up to USD 5,700 higher net taxes than agents with $\beta = 1$. That net taxes decrease with β is the result of two opposing forces. On the one hand, lower β-types have lower (higher) absolute (marginal) welfare than agents without present bias, so the planner would like to provide them with more resources.[39] On the other hand, the planner can extract resources from present-biased high-ability types by offering them the choice of paying higher net taxes in exchange for a lower savings rate. Quantitatively, we find that the second motive dominates.

[37] It is worth noting the salient nonmonotonicity in period-2 consumption as a function of income and also—it turns out—as a function of ability due to the hump-shaped Pareto weight distribution (Myerson, 1981).

[38] Recall that these are means transfers over the lifetime. That all agents pay net taxes is due to the need to finance government expenditures in our normative model. While in reality temporary low-income households receive positive net transfers, they may reach higher income levels and become net tax payers later in life.

[39] This follows directly from IC of agents with $\beta = 1$, whose utility evaluation agrees with that of the planner.

Figure 6. Net transfers across incomes and present bias levels



Note: Net transfers are defined as $T(\theta,\beta) = c_1(\theta,\beta) + c_2(\theta,\beta)/R - y(\theta,\beta)$. Source: Authors' calculations based on model.

7.2 Optimal Policy Tools in the Decentralization

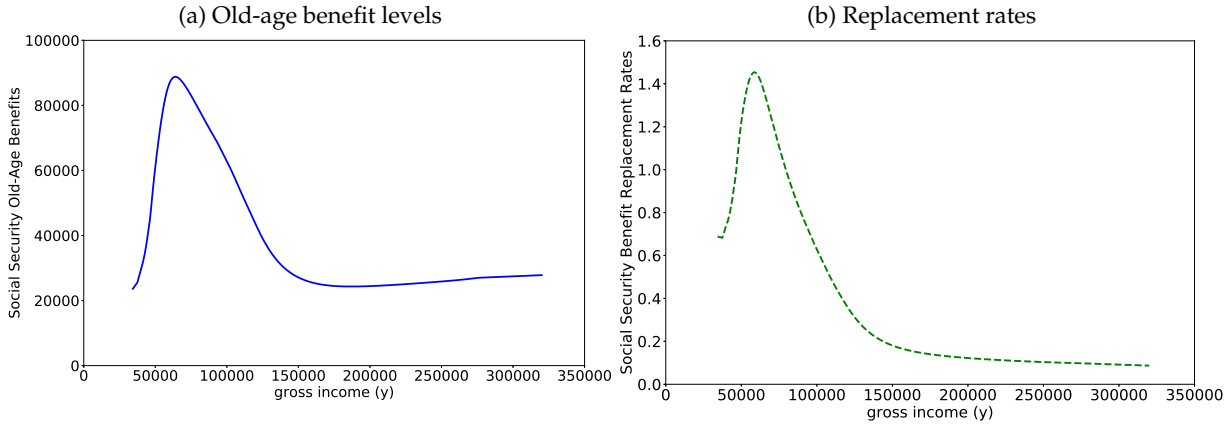
We now turn to the analysis of the decentralization we proposed in Section 4, which consists of Social Security old-age benefits, a set of retirement savings accounts with subsidies and caps, and a working life income tax schedule: $(b(\cdot), \{\tau_j(\cdot), \bar{a}_j(\cdot)\}_j, T_1(\cdot))$.

Social Security Old-Age Benefits. Figure 7 presents the optimal Social Security old-age benefits schedule as a function of lifetime income. Panel (a) of the figure shows optimal old-age benefit levels, which inherit a hump shape from the calibrated social welfare weights. Optimal benefit levels increase from USD 22,000 in benefits at USD 33,000 in household income to USD 88,000 in benefits at USD 68,000 in income. Benefit levels then decrease again until USD 150,000 in income before stabilizing at around USD 27,000. Panel (b) of Figure 7 shows the optimal replacement rates, defined as benefits relative to preretirement earnings. As a result of benefits increasing faster than income, replacement rates increase from 68 percent to 145 percent leading up to USD 60,000 in household income. Thereafter, due to declining benefit levels, the replacement rate decreases at a decreasing rate, reaching 20 percent at around USD 150,000 in income.

Compared to the current US Social Security schedule, the optimal schedule shows some similarities but also some important differences. The current schedule has more intricacies but—similar to the optimal schedule schedule—offers decreasing replacement rates for annualized average indexed monthly earnings (AIME) above USD 68,000 and constant benefits above some income threshold. One difference is that optimal replacement rates are increasing in income up to around

USD 60,000, whereas current replacement rates are a decreasing step function of income. A second difference is that optimal benefit levels decrease between USD 68,000 and USD 150,000, whereas they are weakly increasing in income in the current system.[40]

Figure 7. Social Security schedule



Note: See Section 4 for details of the decentralization component shown in the figure. Source: Authors' calculations based on model.

Retirement Savings Accounts. Our decentralization features as many savings accounts as there are present bias types. For our benchmark calibration with six β-types, there are five retirement savings accounts with subsidies and caps, in addition to a regular bank account at free disposal. We omit the regular bank account from our analysis and label the subsidized savings vehicles "Account 1" to "Account 5." These accounts mirror real-world retirement savings vehicles such as 401(k) and IRA accounts with employer matching and tax-preferred treatment up to some cap.

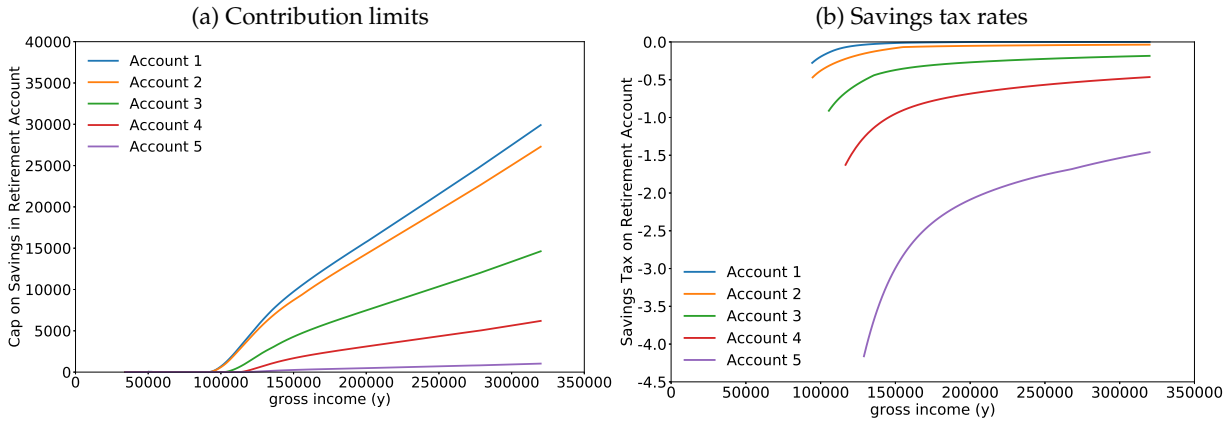
Figure 8(a) presents optimal contribution limits on the five accounts as a function of income. Individuals earning less than USD 95,000 do not have access to any subsidized savings accounts, hence they rely only on Social Security benefits when old.[41] Around that income threshold, the decentralization starts offering five retirement savings accounts with caps that are approximately affine in income. Account 1 has a limit of approximately 13.2 percent on income above USD 95,000. The contribution limits on Accounts 2–5, as a share of income above USD 95,000, are 12.2, 6.6, 2.8, and 0.6 percent, respectively. Panel (b) shows that all accounts offer a negative savings tax—that

[40] While statutory replacement rates are bounded above by 90 percent of AIME, which is calculated over a subset of the highest incomes during one's working life, Social Security replacement rates actually make up 173 percent of annualized present value career earnings for retired beneficiaries in the lowest individual lifetime earnings quintile (Biggs and Springstead, 2008). As in Hosseini and Shourideh (2018), for computational reasons we do not distinguish between AIME and lifetime earnings in our life-cycle model.

[41] Equivalently, they may have access to such accounts but receive no subsidy or tax on their savings in any of them.

is, a savings subsidy—that tends toward zero at higher incomes. Account 1, which has the most generous contribution limit, offers a 29 percent subsidy rate for individuals with USD 95,000 in income, which declines to zero at around USD 150,000 in income. The other four accounts have more generous subsidy rates, which remain positive at all income levels.

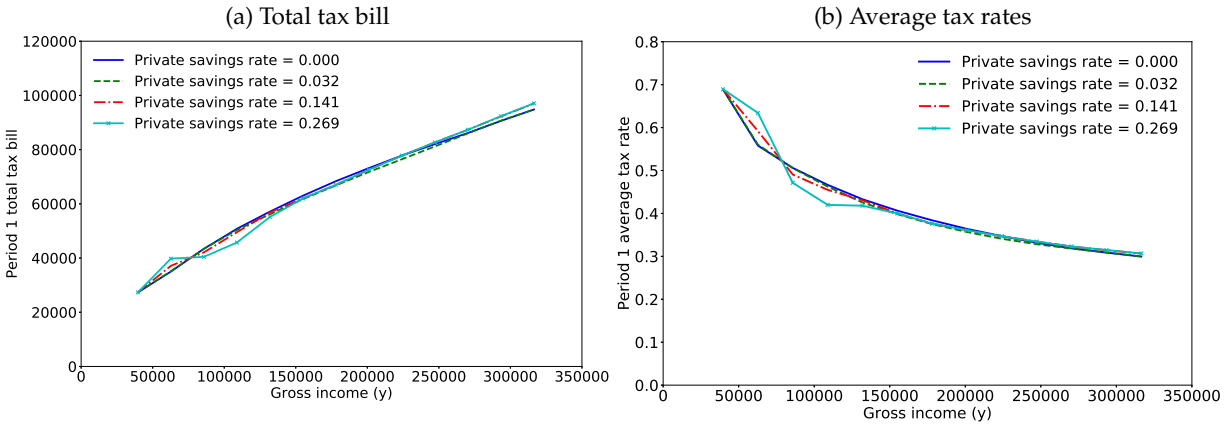
Figure 8. Retirement savings accounts across incomes and present bias levels



Notes: See Section 4 for details of the decentralization component shown in the figure. Optimal savings tax rates (if positive, or subsidy rates if negative) are shown over the range where the respective accounts have nonzero contribution limits, where contribution limits below USD 100 are classified as zero. Source: Authors' calculations based on model.

Income Tax Schedule. The optimal tax schedule as a function of income and private savings, ranging from zero to the 99th percentile, is presented in Figure 9. Panel (a) shows that the optimal tax bill rises from around USD 27,000 at an income of USD 40,000 to a tax bill of USD 97,000 at an income of USD 320,000. Consequently, as shown in Panel (b), the average rate decreases from around 69.1 percent at the lowest income to 30.0 percent at the highest income. Although taxes are relatively high at the bottom, the government offers those households generous Social Security benefits at old age. A notable feature of the decentralized income tax schedule is that it does not vary significantly with individuals' private savings rates conditional on income. Consequently, most of the interaction between savings and taxes plays out through the nonlinear retirement savings accounts described above.

Figure 9. Income tax schedule across incomes and private savings rates



Note: See Section 4 for details of the decentralization component shown in the figure. Source: Authors' calculations based on model.

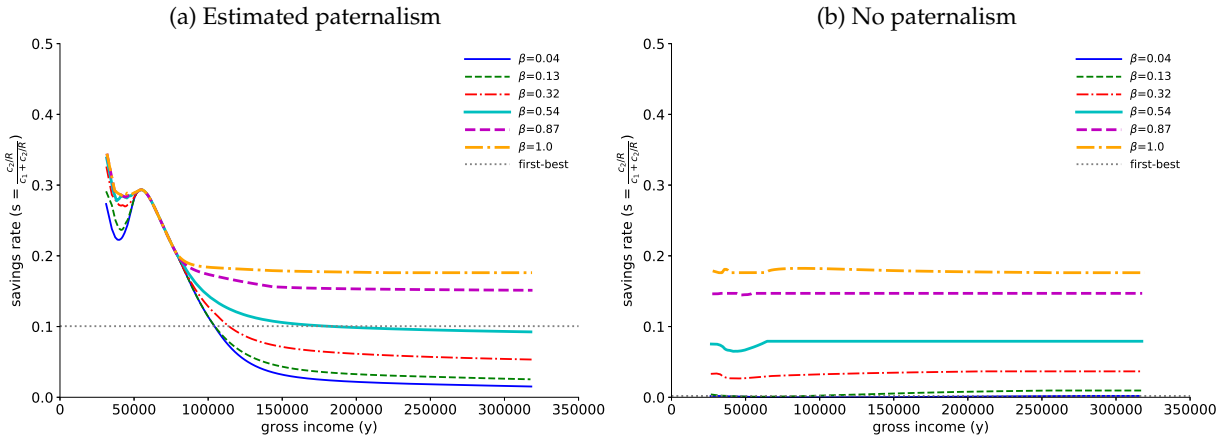
7.3 Comparative Statics

7.3.1 The Role of Paternalism

How do optimal savings rates depend on the strength of the paternalistic motive? To answer this question, we index the planner's degree of paternalism as $p \in [0, 1]$ and compute optimal policies under the social discount factor $\delta\left(\tilde{\delta}, \beta; p\right) = p\tilde{\delta} + (1-p)\,\beta\tilde{\delta}$, where $\tilde{\delta} = 1/R$ is the benchmark discount factor. While we assumed $p = 1$ in our benchmark calibration, we find that $p = 0.633$ provides the best fit in an extended normative calibration (see Section 6.3 for details).

Under this estimated degree of paternalism, Figure 10(a) shows that the resulting optimal savings rates are broadly in line with our benchmark results. One important difference is that, under partial paternalism, savings rates are significantly lower for agents earning less than USD 100,000. In contrast, panel (b) shows that optimal savings rates without paternalism $(p = 0)$ are close to constant across incomes. Some agents are allowed to save at vanishingly small savings rates, which implies that there is little scope for Social Security to provide a floor on old-age consumption. This suggests that a high degree of paternalism is needed to rationalize real-world retirement savings policies.

Figure 10. Optimal savings rates under various degrees of paternalism



Note: Savings rate is defined as $s = (c_2/R) \,/\, (c_1 + c_2/R)$. Estimated degree of paternalism is $p = 0.633$, where the planner's discount factor is $\delta\left(\tilde{\delta}, \beta; p\right) = p\tilde{\delta} + (1-p)\,\beta\tilde{\delta}$, and $\tilde{\delta} = 1/R$ is the benchmark discount factor. Source: Authors' calculations based on model.

7.3.2 The Role of Redistribution

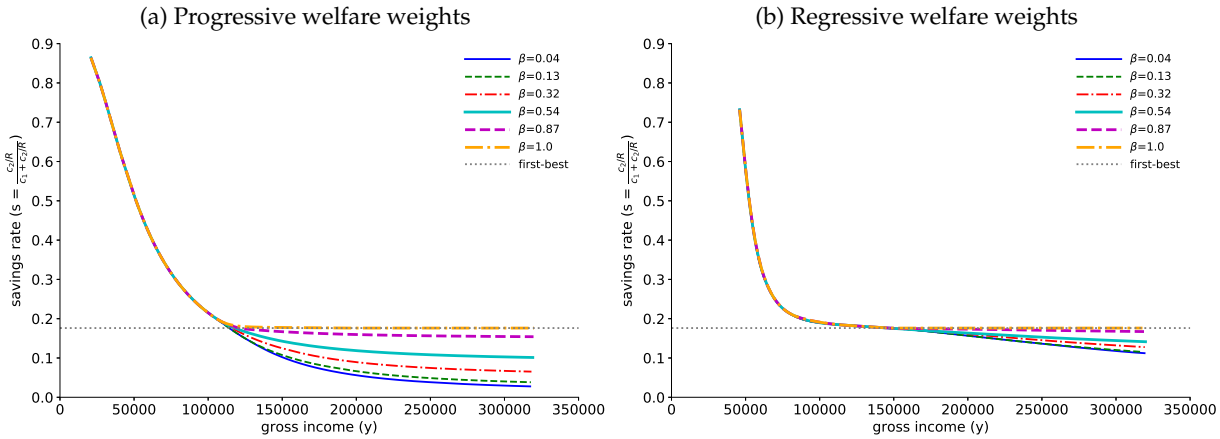
How do optimal savings rates depend on the strength of the planner's redistributive motive? To answer this question, we solve for optimal savings rates under a range of different Pareto weights.[42] Figure 11(a) plots optimal retirement savings rates for a progressive planner with monotonically decreasing Pareto weights $(\lambda_1 = 1.000, \lambda_2 = 2.000, \lambda_w = 0.499)$. Under these more progressive social preferences, savings rates are bunched at higher rates for incomes up to USD 85,000. Moving consumption into the second period is a relatively cheap way for the planner to redistribute resources toward low-ability types in the presence of present-biased agents.[43]

Panel (b) shows optimal savings rates for a regressive planner with monotonically increasing Pareto weights $(\lambda_1 = 2.000, \lambda_2 = 1.000, \lambda_w = 0.499)$. As previously, low-income agents are bunched, but now savings rates at high incomes are also less dispersed. Among high-ability types, a regressive planner does not need to raise as much resources but rather ensures that they save close to the socially optimal rate.

[42] In Appendix D.1.1, we consider further Pareto weight shapes, including the utilitarian benchmark. In Appendix D.1.2, we compute optimal savings rates under different levels of exogenous government spending.

[43] For this reason, combined with the fact that empirical mean savings rates are significantly below 83 percent at low incomes, our normative calibration infers relatively low welfare weights at the bottom of the ability distribution.

Figure 11. Optimal savings rates under various redistributive preferences



Note: Savings rate is defined as $s = (c_2/R) \,/\, (c_1 + c_2/R)$. Source: Authors' calculations based on model.

7.4 Welfare Gains from Reforms

Tension between tax-transfer and retirement savings policies. Our framework lends itself to jointly analyzing the current US retirement savings and tax-transfer systems. Our main finding is that no social preferences $(\lambda_1, \lambda_2, \lambda_w)$ exist that jointly rationalize both systems.[44] On the one hand, the US tax-transfer system is best justified through welfare weights that are less redistributive than utilitarian (Heathcote and Tsujiyama, 2017). On the other hand, rationalizing the current retirement savings system requires welfare weights that put relatively high weight on the second quartile of earnings ability. Consequently, the two parts of the current policy system—the tax-transfer system and the retirement savings system—are mutually inconsistent.

This implies that, qualitatively, current US policies are off the Pareto frontier. How much off are they, quantitatively? Under our benchmark calibration, we find wefare gains of 8.8 percent in consumption-equivalent terms associated with moving from the current system to the optimum.[45] The potential welfare gains are large in comparison to those found when evaluating reforms to only the tax-transfer system (Heathcote and Tsujiyama, 2017). This suggests that changes to the savings system must be an important component of any optimal reform.

[44] We come to a similar conclusion when extending our normative calibration to search over the parameter vector $(\lambda_1, \lambda_2, \lambda_w, \delta, p)$, which includes the planner's discount factor δ and degree of paternalism p as additional arguments. This result can be rationalized by two insights. First, the calibrated value of the planner's discount factor, $\delta = 0.261$, is quite close to $1/R$ (see Section 6.2). Second, the optimal allocation under the calibrated degree of paternalism, $p = 0.633$, is quite similar to that under complete paternalism (see Figure 7.3.1).

[45] We compute consumption-equivalent welfare gains as a uniform change in consumption during working life and retirement, holding fixed labor supply.

Savings distortions facilitate redistribution. Changing the savings system may result in welfare gains through two separate channels. First, it may directly improve the intertemporal consumption allocation. Second, it may facilitate redistribution by improving incentives. Quantitatively, we find that essentially all welfare gains derive from the second source. To convey this point, Table 5 shows consumption-equivalent welfare gains from moving either savings or net income or both from the current system to the optimum. Around 10.6 percent consumption-equivalent welfare gains—more than the total welfare gain—would accrue from implementing the optimal tax-transfer policies while keeping savings at their empirical rates. This is, of course, not a feasible reform as it violates IC. The savings reform, which by itself would lead to a 1.5 percent consumption-equivalent reduction in welfare, renders the optimal income reform feasible, resulting in net welfare gains of 8.8 percent.

Table 5. Consumption-equivalent welfare gains from changes in savings and net income

	Savings rates as in...	
	...current system	...optimal policies
Net income as in current system	0.000	-0.015
Net income as in optimal policies	0.106	0.088

Note: Consumption-equivalent welfare gains are computed as additional consumption while young and old, holding constant labor supply, required to achieve the welfare of the respective allocation, relative to that under the current system (positive model). Source: Authors' calculations based on model.

Welfare under different degrees of paternalism. The welfare gains we identified under complete paternalism are robust to the planner's degree of paternalism. Table 6 presents welfare calculations under different degrees of paternalism. Our benchmark result of 8.8 percent welfare gains is comparable to those found under the calibrated degree of paternalism, $p = 0.633$, which amount to 8.4 percent. In contrast, a planner without paternalism perceives current policies to be 31.0 percent below the optimal welfare level. Current savings policies yield savings rates (Figure 2) that are far from the nonpaternalistic optimum (Figure 10(b)), and the current tax-transfer system is far from the progressive social preferences estimated to match the system holistically. The optimal allocations under complete paternalism $(p = 1)$ and the calibrated degree of paternalism $(p = 0.633)$ are close to one another in consumption-equivalent welfare terms.

Table 6. Consumption-equivalent welfare gains under various degrees of paternalism

	Planner's degree of paternalism		
	$p = 0.000$	$p = 0.633$	$p = 1.000$
Current system	0.000	0.000	0.000
Optimal policies under $p = 0.000$	0.310	0.003	-0.069
Optimal policies under $p = 0.633$	0.293	0.084	0.077
Optimal policies under $p = 1.000$	0.270	0.074	0.088

Notes: Consumption-equivalent welfare gains are computed as additional consumption while young and old, holding constant labor supply, required to achieve the welfare of the respective allocation, relative to that under the current system (positive model). Estimated degree of paternalism is $p = 0.633$, where the planner's discount factor is $\delta\left(\tilde{\delta}, \beta; p\right) = p\tilde{\delta} + (1-p)\beta\tilde{\delta}$, and $\tilde{\delta} = 1/R$ is the benchmark discount factor. Source: Authors' calculations based on model.

8 Conclusion

In this paper, we have developed a theory of optimal paternalistic savings design. Our main insight is that a redistributive planner optimally restricts choice at low incomes but offers various distorted choices at higher incomes. Intuitively, the planner offers choice as a carrot and stick to incentivize work effort among high-ability individuals, thereby facilitating redistribution. We apply this insight to study optimal retirement savings policies. The optimum can be implemented through Social Security old-age benefits and a number of savings accounts with subsidies and caps, such as 401(k) and IRA accounts in the US. To solve large-scale versions of this two-dimensional screening problem, we propose a general, computationally stable, and efficient active-set algorithm. We find significant variation in time preferences underlying the empirical dispersion in savings rates. Interpreting this heterogeneity as due to present bias helps to rationalize many qualitative features of real-world policies. Quantitatively, we find find significant welfare gains from reforms to the current system, due to a tension between redistributive preferences embedded in US retirement savings and tax-transfer policies.

The theoretical insights and numerical solution method we develop in this paper open the door to studying a large class of multidimensional screening problems used in public finance, contract theory, and industrial organization. Our work also points to interesting avenues for future research. First, it would be interesting to explore to what extent other fiscal and social policies can be rationalized with a paternalistic motive. Second, while our paper explores the implications of paternalism for optimal policy design in the US, future work could employ a similar framework to estimate the implied degree of paternalism inherent in other countries' policies.

References

Aguiar, Mark and Erik Hurst, "Consumption versus Expenditure," *Journal of Political Economy*, 2005, *113* (5), 919–948.

_ and Manuel Amador, "Growth in the Shadow of Expropriation," *Quarterly Journal of Economics*, 2011, *126* (2), 651–697.

Amador, Manuel, George-Marios Angeletos, and Iván Werning, "Redistribution and Corrective Taxation," *Working Paper*, 2004.

_ , Iván Werning, and George-Marios Angeletos, "Commitment vs. Flexibility," *Econometrica*, 2006, *74* (2), 365–396.

Angeletos, George-Marios, David Laibson, Andrea Repetto, Jeremy Tobacman, and Stephen Weinberg, "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation," *Journal of Economic Perspectives*, 2001, *15* (3), 47–68.

Aon Hewitt, "2015 Trends & Experience in Defined Contribution Plans," 2015.

Arias, Elizabeth, "United States Life Tables, 2006," *National Vital Statistics Reports*, 2010, *58* (21).

Armstrong, Mark, "Multiproduct Nonlinear Pricing," *Econometrica*, 1996, *64* (1), 51–75.

_ and Jean-Charles Rochet, "Multi-Dimensional Screening: A User's Guide," *European Economic Review*, 1999, *43* (4), 959–979.

Ashraf, Nava, Dean Karlan, and Wesley Yin, "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *Quarterly Journal of Economics*, 2006, *121* (2), 635–672.

Athey, Susan, Andrew Atkeson, and Patrick J Kehoe, "The Optimal Degree of Discretion in Monetary Policy," *Econometrica*, 2005, *73* (5), 1431–1475.

Atkeson, Andrew, V. V. Chari, and Patrick J. Kehoe, "Taxing Capital Income: A Bad Idea," *Federal Reserve Bank of Minneapolis Quarterly Review*, 1999, *23* (3), 3–17.

Atkinson, Anthony B and Joseph E Stiglitz, "The Structure of Indirect Taxation and Economic Efficiency," *Journal of Public Economics*, 1972, *1* (1), 97–119.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger, "Working over Time: Dynamic Inconsistency in Real Effort Tasks," *Quarterly Journal of Economics*, 2015, *130* (3), 1067–1115.

Battaglini, Marco and Rohit Lamba, "Optimal Dynamic Contracting: The First-Order Approach and Beyond," *Working Paper*, 2018.

Bénabou, Roland, "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?," *Econometrica*, 2002, *70* (2), 481–517.

Bernheim, B. Douglas and Dmitry Taubinsky, "Behavioral Public Economics," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics*, Vol. 1, Amsterdam: Elsevier B.V, 2018, chapter 5, pp. 381–516.

_ , Jonathan Skinner, and Steven Weinberg, "What Accounts for the Variation in Retirement Wealth among U.S. Households?," *American Economic Review*, 2001, *91* (4), 832–857.

Beshears, John, James J. Choi, Christopher Harris, David Laibson, Brigitte C. Madrian, and Jung Sakong, "Self Control and Commitment: Can Decreasing the Liquidity of a Savings Account Increase Deposits?," *NBER Working Paper No. 21474*, 2015.

_ , _ , David Laibson, and Brigitte C. Madrian, "Behavioral Household Finance," in B. Dou-

glas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics*, Amsterdam: Elsevier B.V, 2018, chapter 3, pp. 177–276.

Biggs, Andrew G. and Glenn R. Springstead, "Alternate Measures of Replacement Rates for Social Security Benefits and Retirement Income," *Social Security Bulletin*, 2008, *68* (2).

Bourguignon, François and Amedeo Spadaro, "Tax-Benefit Revealed Social Preferences," *The Journal of Economic Inequality*, 2012, *10* (1), 75–108.

Boyd, Stephen and Lieven Vandenberghe, *Convex Optimization*, Cambridge: Cambridge University Press, 2004.

Browning, Martin and Annamaria Lusardi, "Household Saving: Micro Theories and Micro Facts," *Journal of Economic Literature*, 1996, *34* (4), 1797–1855.

Buchanan, J. M., "The Samaritan's Dilemma," in E.S. Phelps, ed., *Altruism, Morality and Economic Theory*, New York: Russel Sage Foundation, 1975, pp. 71–85.

Burks, Stephen V., Jeffrey P. Carpenter, Lorenz Goette, and Aldo Rustichini, "Cognitive Skills Affect Economic Preferences, Strategic Behavior, and Job Attachment," *Proceedings of the National Academy of Sciences of the United States of America*, 2009, *106* (19), 7745–7750.

Caplin, Andrew and John Leahy, "The Social Discount Rate," *Journal of Political Economy*, 2004, *112* (6), 1257–1268.

Carroll, Christopher, Jiri Slacalek, Kiichi Tokuoka, and Matthew N. White, "The Distribution of Wealth and the Marginal Propensity to Consume," *Quantitative Economics*, 2017, *8* (3), 977–1020.

Carroll, Gabriel, "When Are Local Incentive Constraints Sufficient?," *Econometrica*, 2012, *80* (2), 661–686.

_ , "Robustness and Separation in Multidimensional Screening," *Econometrica*, 2017, *85* (2), 453–488.

Chamley, Christophe, "Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives," *Econometrica*, 1986, *54* (3), 607–622.

Chan, Marc K., "Welfare Dependence and Self-Control: An Empirical Analysis," *Review of Economic Studies*, 2017, *84* (4), 1379–1423.

Cheney, Ward and Allen A. Goldstein, "Proximity Maps for Convex Sets," *Proceedings of the American Mathematical Society*, 1959, *10*, 448–450.

Chetty, Raj, Adam Looney, and Kory Kroft, "Salience and Taxation: Theory and Evidence," *American Economic Review*, 2009, *99* (4), 1145–1177.

_ **, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler**, "The Association between Income and Life Expectancy in the United States, 2001-2014," *JAMA*, 2016, *315* (16), 1750.

Choi, James J., "Contributions to Defined Contribution Pension Plans," *Annual Review of Financial Economics*, 2015, *7*, 161–178.

Cremer, Helmuth, Philippe De Donder, Dario Maldonado, and Pierre Pestieau, "Forced Saving, Redistribution, and Nonlinear Social Security Schemes," *Southern Economic Journal*, 2009, *76* (1), 86–98.

Cronqvist, Henrik and Stephan Siegel, "The Origins of Savings Behavior," *Journal of Political Economy*, 2015, *123* (1), 123–169.

De Nardi, Mariacristina, "Wealth Inequality and Intergenerational Links," *Review of Economic*

Studies, 2004, *71* (3), 743–768.

_ **and Giulio Fella**, "Saving and Wealth Inequality," *Review of Economic Dynamics*, 2017, *26*, 280–300.

_ **, Svetlana Pashchenko, and Ponpoje Porapakkarm**, "The Lifetime Costs of Bad Health," *NBER Working Paper No. 23963*, 2018.

Diamond, Peter A., "A Framework for Social Security Analysis," *Journal of Public Economics*, 1977, *8* (3), 275–298.

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

Diamond, Peter and Johannes Spinnewijn, "Capital Income Taxes with Heterogeneous Discount Rates," *American Economic Journal: Economic Policy*, 2011, *3* (4), 52–76.

Doepke, Matthias and Fabrizio Zilibotti, "Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission," *Econometrica*, 2017, *85* (5), 1331–1371.

Dynan, Karen E., Jonathan Skinner, and Stephen P. Zeldes, "Do the Rich Save More?," *Journal of Political Economy*, 2004, *112* (2), 397–444.

Engen, Eric M., William G. Gale, and Cori E. Uccello, "Lifetime Earnings, Social Security Benefits, and the Adequacy of Retirement Wealth Accumulation," *Social Security Bulletin*, 2005, *66* (1), 38.

Esteban, Susanna and Eiichi Miyagawa, "Optimal Menu of Menus with Self-Control Preferences," *Working Paper*, 2004.

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde, "Global Evidence on Economic Preferences," *Quarterly Journal of Economics*, 2018, *133* (4), 1645–1692.

Farhi, Emmanuel and Iván Werning, "Inequality and Social Discounting," *Journal of Political Economy*, 2007, *115* (3), 365–402.

_ **and** _ , "Progressive Estate Taxation," *Quarterly Journal of Economics*, 2010, *125* (2), 635–673.

_ **and** _ , "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation," *Journal of Political Economy*, 2012, *120* (3), 398–445.

_ **and** _ , "Estate Taxation with Altruism Heterogeneity," *American Economic Review*, 2013, *103* (3), 489–495.

_ **and** _ , "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 2013, *80* (2), 596–635.

_ **and Xavier Gabaix**, "Optimal Taxation with Behavioral Agents," *Working Paper*, 2018.

Feenberg, Daniel and Elisabeth Coutts, "An Introduction to the TAXSIM Model," *Journal of Policy Analysis and Management*, 1993, *12* (1), 189–194.

Feldstein, Martin and Jeffrey B. Liebman, "Social Security," in Alan J. Auerbach and Martin Feldstein, eds., *Handbook of Public Economics*, Vol. 4, Amsterdam, North Holland: Elsevier, 2002, chapter 32, pp. 2245–2324.

Feldstein, Martin S., "The Effects of Taxation on Risk Taking," *Journal of Political Economy*, 1969, pp. 755–764.

_ , "The Optimal Level of Social Security Benefits," *Quarterly Journal of Economics*, 1985, *100* (2), 303–320.

Fernandes, Ana and Christopher Phelan, "A Recursive Formulation for Repeated Agency with History Dependence," *Journal of Economic Theory*, 2000, *91* (2), 223–247.

Financial Engines, "Missing Out: How Much Employer 401(K) Matching Contributions Do Employees Leave on the Table?," Technical Report, Financial Engines 2015.

Friedman, Milton, "Choice, Chance, and the Personal Distribution of Income," *Journal of Political Economy*, 1953, *61* (4), 277–290.

Galperti, Simone, "Commitment, Flexibility, and Optimal Screening of Time Inconsistency," *Econometrica*, 2015, *83* (4), 1425–1465.

Golosov, Mikhail, Aleh Tsyvinski, and Nicolas Werquin, "A Variational Approach to the Analysis of Tax Systems," *Working Paper*, 2014.

_ **and** _ , "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy*, 2006, *114* (2), 257–279.

_ **, Maxim Troshkin, Aleh Tsyvinski, and Matthew Weinzierl**, "Preference Heterogeneity and Optimal Capital Income Taxation," *Journal of Public Economics*, 2013, *97*, 160–175.

_ **,** _ **, and** _ , "Redistribution and Social Insurance," *American Economic Review*, 2016, *106* (2), 359–386.

_ **, Narayana Kocherlakota, and Aleh Tsyvinski**, "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 2003, *70* (3), 569–587.

Gourinchas, Pierre-Olivier and Jonathan A. Parker, "Consumption over the Life Cycle," *Econometrica*, 2002, *70* (1), 47–89.

Gruber, Jonathan and Botond Köszegi, "Tax Incidence When Individuals Are Time-Inconsistent: The Case of Cigarette Excise Taxes," *Journal of Public Economics*, 2004, *88* (9-10), 1959–1987.

Halac, Marina and Pierre Yared, "Fiscal Rules and Discretion under Persistent Shocks," *Econometrica*, 2014, *82* (5), 1557–1614.

_ **and** _ , "Fiscal Rules and Discretion in a World Economy," *American Economic Review*, 2018, *108* (8), 2305–2234.

Hartline, Jason D., "Bayesian Mechanism Design," *Foundations and Trends in Theoretical Computer Science*, 2013, *8* (3), 143–263.

Hassler, John, Per Krusell, Kjetil Storesletten, and Fabrizio Zilibotti, "On the Optimal Timing of Capital Taxes," *Journal of Monetary Economics*, 2008, *55* (4), 692–709.

Heathcote, Jonathan and Hitoshi Tsujiyama, "Optimal Income Taxation: Mirrlees Meets Ramsey," *Working Paper*, 2017.

_ **, Kjetil Storesletten, and Giovanni L. Violante**, "Optimal Tax Progressivity: An Analytical Framework," *The Quarterly Journal of Economics*, 2017, *132* (4), 1693–1754.

Hendricks, Lutz, "How Important Is Discount Rate Heterogeneity for Wealth Inequality?," *Journal of Economic Dynamics and Control*, 2007, *31* (9), 3042–3068.

Hosseini, Roozbeh and Ali Shourideh, "Retirement Financing: An Optimal Reform Approach," *Working Paper*, 2018.

Internal Revenue Service, "Publication 590-A (2017), Contributions to Individual Retirement Arrangements (IRAs)," 2018.

Jacobs, Bas, Egbert L. W. Jongen, and Floris T. Zoutman, "Revealed Social Preferences of Dutch

Political Parties," *Journal of Public Economics*, 2017, *156*, 81–100.

Jacquet, Laurence, Etienne Lehmann, and Bruno Van der Linden, "Optimal Redistributive Taxation with Both Extensive and Intensive Responses," *Journal of Economic Theory*, sep 2013, *148* (5), 1770–1805.

Jones, Damon and Aprajit Mahajan, "Time-Inconsistency and Saving: Experimental Evidence from Low-Income Tax Filers," *NBER Working Paper No. 21272*, 2015, (21272).

Judd, Kenneth L., "Short-Run Analysis of Fiscal Policy in a Simple Perfect Foresight Model," *Journal of Political Economy*, 1985, *93* (2), 298–319.

_ , Ding Ma, Michael A. Saunders, and Che-Lin Su, "Optimal Income Taxation with Multidimensional Taxpayer Types," *Working Paper*, 2018.

Kapička, Marek, "Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach," *Review of Economic Studies*, 2013, *80*, 1027–1054.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan, "Self-Control at Work," *Journal of Political Economy*, 2015, *123* (6), 1227–1277.

Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez, "The Optimal Income Taxation of Couples," *Econometrica*, 2009, *77* (2), 537–560.

Kotlikoff, Laurence J., "Privatizing Social Security at Home and Abroad," *American Economic Review*, 1996, *86* (2), 368–372.

_ , Avia Spivak, and Lawrence H. Summers, "The Adequacy of Savings," *American Economic Review*, 1982, *72* (5), 1056–1069.

Krusell, Per and Anthony A. Smith, Jr., "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 1998, *106* (5), 867–896.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 1997, *112* (2), 443–478.

Laibson, David I., Andrea Repetto, and Jeremy Tobacman, "Self-Control and Saving for Retirement," *Brookings Papers on Economic Activity*, 1998, *29* (1), 91–196.

Laibson, David, Peter Maxted, Andrea Repetto, and Jeremy Tobacman, "Estimating Discount Functions with Consumption Choices over the Lifecycle," *Working Paper*, 2017.

Lansing, Kevin J., "Optimal Redistributive Capital Taxation in a Neoclassical Growth Model," *Journal of Public Economics*, 1999, *73* (3), 423–453.

Lockwood, Benjamin B., "Optimal Income Taxation with Present Bias," *Working Paper*, 2016.

_ and Dmitry Taubinsky, "Regressive Sin Taxes," *NBER Working Paper No. 23085*, 2017.

_ and Matthew Weinzierl, "Positive and Normative Judgments Implicit in U.S. Tax Policy, and the Costs of Unequal Growth and Recessions," *Journal of Monetary Economics*, 2016, *77*, 30–47.

Marglin, Stephen A., "The Social Rate of Discount and the Optimal Rate of Investment," *Quarterly Journal of Economics*, 1963, *77* (1), 95–111.

McAfee, R. Preston and John McMillan, "Multidimensional Incentive Compatibility and Mechanism Design," *Journal of Economic Theory*, 1988, *46* (2), 335–354.

Meghir, Costas and Luigi Pistaferri, "Income Variance Dynamics and Heterogeneity," *Econometrica*, 2004, *72* (1), 1–32.

Meier, Stephan and Charles D. Sprenger, "Temporal Stability of Time Preferences," *Review of*

Economics and Statistics, 2015, *97* (2), 273–286.

Mirrlees, J. A., "Optimal Tax Theory: A Synthesis," *Journal of Public Economics*, 1976, *6* (4), 327–358.

Mirrlees, James A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38* (2), 175–208.

Mitchell, Olivia S., James M. Poterba, Mark J. Warshawsky, and Jeffrey R. Brown, "New Evidence on the Money's Worth of Individual Annuities," *American Economic Review*, 1999, *89* (5), 1299–1318.

Montiel Olea, José Luis and Tomasz Strzalecki, "Axiomatization and Measurement of Quasi-Hyperbolic Discounting," *Quarterly Journal of Economics*, 2014, *129* (3), 1449–1499.

Myerson, Roger B., "Optimal Auction Design," *Mathematics of Operations Research*, 1981, *6* (1), 58–73.

Ndiaye, Abdoulaye, "Flexible Retirement and Optimal Taxation," *Federal Reserve Bank of Chicago Working Paper No. 2018-18*, 2018.

Nocedal, Jorge. and Stephen J. Wright, *Numerical Optimization*, 2 ed., New York: Springer-Verlag, 2006.

O'Donoghue, Ted and Matthew Rabin, "Studying Optimal Paternalism, Illustrated by a Model of Sin Taxes," *American Economic Review*, 2003, *93* (2), 186–191.

_ **and** _ , "Optimal Sin Taxes," *Journal of Public Economics*, 2006, *90* (10-11), 1825–1849.

Ok, Efe A., *Real Analysis with Economic Applications*, Princeton, NJ: Princeton University Press, 2007.

Olafsson, Arna and Michaela Pagel, "The Retirement-Consumption Puzzle: New Evidence from Personal Finances," *NBER Working Paper No. 24405*, 2018.

Paserman, M Daniele, "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation," *Economic Journal*, 2008, *118* (531), 1418–1452.

Pass, Brendan, "Convexity and Multi-Dimensional Screening for Spaces with Different Dimensions," *Journal of Economic Theory*, 2012, *147* (6), 2399–2418.

Pavoni, Nicola and Hakki Yazici, "Intergenerational Disagreement and Optimal Taxation of Parental Transfers," *Review of Economic Studies*, 2017, *84* (3), 1264–1305.

Phelan, Christopher and Aldo Rustichini, "Pareto Efficiency and Identity," *Theoretical Economics*, 2018, *13* (3), 979–1007.

_ **and Ennio Stacchetti**, "Sequential Equilibria in a Ramsey Tax Model," *Econometrica*, 2001, *69* (6), 1491–1518.

Phelps, E. S. and R. A. Pollak, "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies*, 1968, *35* (2), 185.

Piketty, Thomas and Emmanuel Saez, "A Theory of Optimal Inheritance Taxation," *Econometrica*, 2013, *81* (5), 1851–1886.

Prescott, Edward C., "It's Irrational to Save," *Wall Street Journal*, 2004.

Rochet, Jean-Charles, "A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context," *Journal of Mathematical Economics*, 1987, *16* (2), 191–200.

_ **and Lars A. Stole**, "Nonlinear Pricing with Random Participation," *Review of Economic Studies*, 2002, *69* (1), 277–311.

_ **and** _ , "The Economics of Multidimensional Screening," in Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, eds., *Advances in Economics and Econometrics*, Cambridge: Cambridge University Press, 2003, pp. 150–197.

_ **and Philippe Choné**, "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 1998, *66* (4), 783–826.

Rogerson, William P., "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 1985, *53* (6), 1357–1367.

Rothschild, Casey and Florian Scheuer, "Redistributive Taxation in the Roy Model," *Quarterly Journal of Economics*, 2013, *128* (2), 623–668.

_ **and** _ , "A Theory of Income Taxation under Multidimensional Skill Heterogeneity," *CESifo Working Paper Series*, 2015.

_ **and** _ , "Optimal Taxation with Rent-Seeking," *Review of Economic Studies*, 2016, *83* (3), 1225–1262.

Rudin, Walter, *Principles of Mathematical Analysis*, 3 ed., McGraw-Hill, Inc., 1976.

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

_ , "The Desirability of Commodity Taxation under Non-Linear Income Taxation and Heterogeneous Tastes," *Journal of Public Economics*, 2002, *83* (2), 217–230.

_ , "Optimal Progressive Capital Income Taxes in the Infinite Horizon Model," *Journal of Public Economics*, 2013, *97*, 61–74.

Sleet, Christopher and Sevin Yeltekin, "Credibility and Endogenous Societal Discounting," *Review of Economic Dynamics*, 2006, *9* (3), 410–437.

Social Security Administration, "Annual Statistical Supplement, 2018 - Summary of OASDI Benefits in Current-Payment Status (Table 5.A1)," 2018.

Solodov, Mikhail V., "Constraint Qualifications," in "Wiley Encyclopedia of Operations Research and Management Science," Hoboken, NJ: John Wiley & Sons, Inc., 2011.

Stantcheva, Stefanie, "Optimal Taxation and Human Capital Policies over the Life Cycle," *NBER Working Paper No. 21207*, 2015.

_ , "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*, 2016, *77*, 48–52.

Straub, Ludwig and Iván Werning, "Positive Long Run Capital Taxation: Chamley-Judd Revisited," *NBER Working Paper No. 20441*, 2018.

Strotz, R. H., "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 1955, *23* (3), 165–180.

Tanaka, Tomomi, Colin F. Camerer, and Quang Nguyen, "Risk and Time Preferences: Linking Experimental and Household Survey Data from Vietnam," *American Economic Review*, 2010, *100* (1), 557–571.

Tenhunen, Sanna and Matti Tuomala, "On Optimal Lifetime Redistribution Policy," *Journal of Public Economic Theory*, 2010, *12* (1), 171–198.

Tuomala, Matti, *Optimal Income Tax and Redistribution*, Oxford: Oxford University Press, 1990.

Tversky, A. and D. Kahneman, "Judgment under Uncertainty: Heuristics and Biases," *Science*, 1974, *185* (4157), 1124–1131.

U.S. Bureau of Labor Statistics, "Consumer Price Index for All Urban Consumers: All Items [CPIAUCSL]," 2018.

Venti, Steven and David Wise, "The Cause of Wealth Dispersion at Retirement: Choice or Chance?," *American Economic Review*, 1998, *88* (2), 185–191.

Wachsmuth, Gerd, "On LICQ and the Uniqueness of Lagrange Multipliers," *Operations Research Letters*, 2013, *41* (1), 78–80.

Wächter, Andreas and Lorenz T. Biegler, "On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming," *Mathematical Programming*, 2006, *106* (1), 25–57.

Werning, Iván, "Nonlinear Capital Taxation," *Working Paper*, 2011.

Yu, Pei Cheng, "Optimal Retirement Policies," *Working Paper*, 2016.

Online Appendix (Not for Publication)

The Online Appendix is organized as follows: Appendix A presents further theoretical results and proofs for the two-dimensional screening problem that we present in Section 2, characterize in Section 3, and decentralize in Section 4. Appendix B provides further details of the active-set algorithm introduced in Section 5. Appendix C describes the data and calibration procedure used in Section 6. Appendix D contains further results on optimal policies.

A Theory

A.1 Properties of the planner's problem

Lemma 2. *The planner's problem has the following properties:*

1. *(Convexity) The planner's problem is a convex program.*

2. *(Existence and uniqueness of a global optimum) There exists a unique allocation $\mathcal{A}$ that solves the planner's problem. This allocation satisfies $0 < c_t(\theta,\beta) < +\infty$ and $0 \leq y(\theta,\beta) < +\infty$ for $t = 1,2$ and all $(\theta,\beta) \in \Theta \times B$.*

3. *(Maximum theorem) The solution to the planner's problem, $\{\mathcal{A}^{SB}, W(\mathcal{A}^{SB})\}$, is continuous in $(\theta,\beta,\pi,\lambda) \in \mathbb{R}_{++} \times \mathbb{R}_{+} \times [0,1] \times \mathbb{R}_{+}$.*

4. *(Strong duality) Define the Lagrangian $\mathcal{L} : \mathbb{R}_{+}^{3(N\cdot M)} \times \mathbb{R}^{(N\cdot M)^2-(N\cdot M)} \times \mathbb{R} \times \mathbb{R}^{3(N\cdot M)} \to \mathbb{R}$ associated with the planner's problem as*

$$\begin{aligned}\mathcal{L}(\mathcal{A},\xi,\mu,\nu) = W(\mathcal{A}) &+ \sum_{(\theta,\beta)\neq(\theta',\beta')} \xi(\theta,\beta,\theta',\beta') \left\{ \begin{array}{c} U(c_1(\theta,\beta), c_2(\theta,\beta), y(\theta,\beta);\theta,\beta) \\ -U(c_1(\theta',\beta'), c_2(\theta',\beta'), y(\theta',\beta');\theta,\beta) \end{array} \right\} \\ &+ \mu \left\{ \sum_{(\theta,\beta)} \pi(\theta,\beta) \left[y(\theta,\beta) - c_1(\theta,\beta) - \frac{c_2(\theta,\beta)}{R} \right] \right\} \\ &+ \sum_{(\theta,\beta)} \nu_1(\theta,\beta) c_1(\theta,\beta) + \sum_{(\theta,\beta)} \nu_2(\theta,\beta) c_2(\theta,\beta) + \sum_{(\theta,\beta)} \nu_3(\theta,\beta) y(\theta,\beta).\end{aligned}$$

Then an allocation $\mathcal{A}^* = \{c_1^*(\theta,\beta), c_2^*(\theta,\beta), y^*(\theta,\beta)\}_{(\theta,\beta)}$ is optimal if and only if there exist Lagrange multiplieres on the IC constraints, $\xi = \{\xi(\theta,\beta,\theta',\beta')\}_{(\theta,\beta)\neq(\theta',\beta')} \in \mathbb{R}_{+}^{(N\cdot M)^2-(N\cdot M)}$, a

Lagrange multiplier on the feasibility constraint, $\mu \in \mathbb{R}_+$, and Lagrange mutlipliers on the variable domain constraints, $\nu = \{(\nu_1(\theta,\beta), \nu_2(\theta,\beta), \nu_3(\theta,\beta))\}_{(\theta,\beta)} \in \mathbb{R}_+^{3(N\cdot M)}$ that satisfy the following KKT conditions:

$$
\begin{aligned}
U\left(c_1^*(\theta,\beta), c_2^*(\theta,\beta), y^*(\theta,\beta);\theta,\beta\right) - U\left(c_1^*(\theta',\beta'), c_2^*(\theta',\beta'), y^*(\theta',\beta');\theta,\beta\right) &\geq 0, \quad \forall (\theta,\beta), (\theta',\beta') \in \Theta \times B \\
\sum_{\theta,\beta} \pi(\theta,\beta)\left[y^*(\theta,\beta) - c_1^*(\theta,\beta) - \frac{c_2^*(\theta,\beta)}{R}\right] &\geq 0 \\
c_1^*(\theta,\beta) &\geq 0, \quad \forall (\theta,\beta) \in \Theta \times B \\
c_2^*(\theta,\beta) &\geq 0, \quad \forall (\theta,\beta) \in \Theta \times B \\
y^*(\theta,\beta) &\geq 0, \quad \forall (\theta,\beta) \in \Theta \times B \\
\xi(\theta,\beta,\theta',\beta') &\geq 0, \quad \forall (\theta,\beta), (\theta',\beta') \in \Theta \times B \\
\mu &\geq 0 \\
\nu_1(\theta,\beta) &\geq 0, \quad \forall (\theta,\beta) \in \Theta \times B \\
\nu_2(\theta,\beta) &\geq 0, \quad \forall (\theta,\beta) \in \Theta \times B \\
\nu_3(\theta,\beta) &\geq 0, \quad \forall (\theta,\beta) \in \Theta \times B \\
\xi(\theta,\beta,\theta',\beta') \left\{ \begin{array}{c} U\left(c_1^*(\theta,\beta), c_2^*(\theta,\beta), y^*(\theta,\beta);\theta,\beta\right) \\ -U\left(c_1^*(\theta',\beta'), c_2^*(\theta',\beta'), y^*(\theta',\beta');\theta,\beta\right) \end{array} \right\} &= 0, \quad \forall (\theta,\beta), (\theta',\beta') \in \Theta \times B \\
\mu \sum_{\theta,\beta} \pi(\theta,\beta)\left[y^*(\theta,\beta) - c_1^*(\theta,\beta) - \frac{c_2^*(\theta,\beta)}{R}\right] &= 0 \\
\nu_1(\theta,\beta)\, c_1^*(\theta,\beta) &= 0, \quad \forall (\theta,\beta) \in \Theta \times B \\
\nu_2(\theta,\beta)\, c_2^*(\theta,\beta) &= 0, \quad \forall (\theta,\beta) \in \Theta \times B \\
\nu_3(\theta,\beta)\, y^*(\theta,\beta) &= 0, \quad \forall (\theta,\beta) \in \Theta \times B
\end{aligned}
$$

$$
\left\{ \begin{array}{c} \nabla W(\mathcal{A}^*) \\ + \sum_{(\theta,\beta)\neq(\theta',\beta')} \xi(\theta,\beta,\theta',\beta') \nabla \left[\begin{array}{c} U\left(c_1^*(\theta,\beta), c_2^*(\theta,\beta), y^*(\theta,\beta);\theta,\beta\right) \\ -U\left(c_1^*(\theta',\beta'), c_2^*(\theta',\beta'), y^*(\theta',\beta');\theta,\beta\right) \end{array} \right] \\ + \mu \sum_{\theta,\beta} \pi(\theta,\beta) \nabla \left[y^*(\theta,\beta) - c_1^*(\theta,\beta) - \frac{c_2^*(\theta,\beta)}{R} \right] \\ + \sum_{\theta,\beta} \nu_1(\theta,\beta) + \sum_{\theta,\beta} \nu_2(\theta,\beta) + \sum_{\theta,\beta} \nu_3(\theta,\beta) \end{array} \right\} = 0
$$

Proof. We start by applying a convenient change of variables. Any results that we prove for the transformed problem directly carry over to the original problem, but exposition of the proofs is facilitated through a convenient choice of transformation. Let $u_t(\theta,\beta) \equiv u(c_t(\theta,\beta))$ for $t = 1, 2,$

and $\tilde{v}(\theta,\beta) \equiv \tilde{v}(y(\theta,\beta))$, so $v(\ell(\theta,\beta)) = D(\theta)\tilde{v}(\theta,\beta)$. Write $c_t(\theta,\beta) = C(u_t(\theta,\beta))$, where $C(\cdot) = u^{-1}(\cdot)$, and $y(\theta,\beta) = Y(\tilde{v}(\theta,\beta))$, where $Y(\cdot) = v^{-1}(\cdot)$. Define the transformed allocation as $\tilde{\mathcal{A}} = \{u_1(\theta,\beta), u_2(\theta,\beta), \tilde{v}(\theta,\beta)\}_{(\theta,\beta)}$. Let $N = N_\theta \times N_\beta$. Then we can reformulate the planner's problem as a nonlinear program with $3N$ choice variables, a linear objective, $N^2 - N$ linear IC constraints, and 1 nonlinear feasibility constraint:

$$
\begin{aligned}
& -\left[\min_{\tilde{\mathcal{A}}} -\sum_{\theta,\beta} \pi(\theta,\beta)\lambda(\theta)\left[u_1(\theta,\beta) - D(\theta)\tilde{v}(\theta,\beta) + \delta u_2(\theta,\beta)\right]\right] \qquad (8)\\
\text{s.t.} \quad & \forall(\theta,\beta),(\theta',\beta'): -\infty \leq \begin{bmatrix} u_1(\theta',\beta') - D(\theta)\tilde{v}(\theta',\beta') \\ +\beta\delta u_2(\theta',\beta') \end{bmatrix} - \begin{bmatrix} u_1(\theta,\beta) - D(\theta)\tilde{v}(\theta,\beta) \\ +\beta\delta u_2(\theta,\beta) \end{bmatrix} \leq 0\\
& -\infty \leq \sum_{\theta,\beta} \pi(\theta,\beta)\left[C(u_1(\theta,\beta)) + \frac{C(u_2(\theta,\beta))}{R} - Y(\tilde{v}(\theta,\beta))\right] \leq 0\\
& \forall t: \underline{u} \leq u_t(\theta,\beta) \leq \overline{u}\\
& \underline{\tilde{v}} \leq \tilde{v}(\theta,\beta) \leq \overline{\tilde{v}}
\end{aligned}
$$

We now proceed stepwise to prove parts 1–4 of the lemma: □

1. *Clearly, the objective and IC constraints are linear, hence continuous, in* $\{u_1, v, u_2\}$. *Next, we show that the feasibility constraint function* $F(\cdot) = \sum \pi\left[C(\cdot) + C(\cdot)/R + G - Y(\cdot)\right]$ *defines a strictly convex set. Since* $u(\cdot)$ *is increasing and strictly concave, then* $C(\cdot) = u^{-1}(\cdot)$ *is strictly convex. Similarly, since* $v(\cdot)$ *is increasing and strictly convex, then* $Y(\cdot) = v^{-1}(\cdot)$ *is strictly concave. Clearly, the negative of a strictly concave function is strictly convex. Since the sum of strictly convex functions is strictly convex and multiplication by or adding a scalar preserves strict convexity, we know that the feasibility constraint* $F(\cdot)$ *is strictly convex. We conclude that the planner's problem is a convex problem.*

2. *The result follows from a an application of the Extreme Value Theorem—see, e.g., Theorem 4.16 on pp.89–90 of Rudin (1976)—to a modified version of the planner's problem. We have already shown in part 1 of the lemma that the IC constraints, feasibility constraint, and choice variable bounds describe a strictly convex set, which we denote S. Since* $\lim_{c\to+\infty} u'(\cdot) = 0$ *implies* $\lim_{u\to\overline{u}} C(u) = +\infty$, *we know that* $u_t = \overline{u}$ *cannot be optimal for any t. Hence, we can find a* $\tilde{\overline{u}} < \overline{u}$ *such that if an optimum exists, then* $u_t^* \leq \tilde{\overline{u}}$ *for* $t = 1, 2$. *Similarly, since* $u'(0) = +\infty$ *implies* $\lim_{u\to\underline{u}} C(u) = 0$, *we know*

that $u_t = \underline{u}$ cannot be optimal for any t. Hence, we can find a $\tilde{\underline{u}} > \underline{u}$ such that if an optimum exists, then $u_t^* \geq \tilde{\underline{u}}$ for $t = 1, 2$. Finally, since $\lim_{\ell \to +\infty} \tilde{v}'(\cdot) = +\infty$ implies $\lim_{v \to +\infty} Y(\cdot) = 0$, we know that $v = +\infty$ cannot be optimal. Hence, we can find a $\tilde{\bar{v}} < +\infty$ such that if an optimum exists, then $v^* \leq \tilde{\bar{v}}$. Now take the subset $S' \subset S$ defined by $S' = S \cap \left\{ (u_1, u_2, v) \middle| u \in \left[\tilde{\underline{u}}, \tilde{\bar{u}}\right], v \in \left[0, \tilde{\bar{v}}\right] \right\}$, which is closed and bounded, hence compact in $\mathbb{R}$, and it inherits the strict convexity property from S. The Extreme Value Theorem states that a continuous real function on a compact metric space attains its minimum and maximum. Hence a solution to the planner's problem exists. By strict convexity of S', the optimum must also be unique. To see this, suppose this were not the case, then there exist two optimal points (u_1, u_2, v) and (u_1', u_2', v') that solve the planner's problem. Then one could construct a third optimal point $(u_1'', u_2'', v'') = \mu(u_1, u_2, v) + (1 - \mu)(u_1', u_2', v')$ for $\mu \in (0, 1)$, which by strict convexity lies in the strict interior of the feasible set, contradicting optimality of the original two points as the objective function is linear.

3. *The result follows from an application of the maximum theorem—see, e.g., p.306 of Ok (2007)—to the same modified version of the planner's problem as above. The statement of the maximum theorem is as follows. Let $P \subseteq \mathbb{R}^d$ for some integer d be the compact metric space containing all possible vectors of model parameters $p = (\theta, \beta, \pi, \lambda) \in P = \mathbb{R}_{++} \times \mathbb{R}_+ \times [0, 1] \times \mathbb{R}_+$. Let $\tilde{S} \subseteq \mathbb{R}^{3N \cdot M}$ be the metric space containing all possible allocations. Let $\Gamma : P \rightrightarrows \tilde{S}$ be the constraint set correspondence defined by IC constraints, the feasibility constraint, and reformulated choice variable bound constraints as defined above. Γ is continuous, that is both upper hemicontinuous and lower hemicontinuous, due to continuity of all constraint functions. Γ is also compact-valued, that is $\Gamma(p)$ is closed and bounded at any $p \in P$, because all constraint functions are continuous over a compact space P, so the graph of $\Gamma(\cdot)$ attains a maximum and a minimum in each of its dimensions. Denote the linear objective function in the reformulated planner's problem (8) by $\tilde{W} : P \times \tilde{S} \to \mathbb{R}$. Since $\tilde{W}$ is linear in $s \in \tilde{S}$ and well-behaved around $\theta = 0$ by our assumption that $D(0)\tilde{v}(0) = 0$, then $\tilde{W} \in \mathbf{C}\left(P \times \tilde{S}\right)$. Define the unique (by part 2 of the lemma) maximizer of the planner's problem given model parameters p as*

$$\sigma(p) \equiv \arg\max_{\zeta} \left\{ \tilde{W}(p, \zeta) \middle| \zeta \in \Gamma(p) \right\} \quad \forall p \in P$$

and define the optimal welfare level given model parameters p as

$$\varphi^*(p) \equiv \max_{\zeta} \left\{ \tilde{W}(p, \zeta) \middle| \zeta \in \Gamma(p) \right\} \quad \forall p \in P$$

Under these conditions, the maximum theorem implies that:

- $\sigma : P \to \tilde{S}$ *is compact-valued, upper hemicontinuous and closed at* p.
- $\varphi^* : P \to \mathbb{R}$ *is continuous at* p.

Because $\sigma(\cdot)$ *is nonempty and single-valued for all* $p \in P$, *by part 2 of the lemma, then* $\sigma : P \to \tilde{S}$ *must also be lower hemicontinuous. Hence* σ *is continuous.*

4. *Note that the reformulated planner's problem (8) features only one nonaffine constraint, namely feasibility. A refined variant of Slater's condition or constraint qualification for nonaffine functions (see, e.g., pp.226–227 of Boyd and Vandenberghe (2004)), requiring the existence of a feasible point in the interior of the nonaffine constraint set, is satisfied for our problem:*

$$\begin{aligned}
&\exists \tilde{\mathcal{A}} = \left(u_1(\theta,\beta), u_2(\theta,\beta), \tilde{v}(\theta,\beta)\right)_{(\theta,\beta)} \\
s.t. \quad &\forall (\theta,\beta) : (\theta,\beta) = \arg\max_{(\theta',\beta')} u_1(\theta,\beta) - D(\theta)\tilde{v}(\theta,\beta) + \beta\delta u_2(\theta,\beta) \\
-\infty < &\sum_{\theta,\beta} \pi(\theta,\beta)\left[C(u_1(\theta,\beta)) + \frac{C(u_2(\theta,\beta))}{R} - Y(D(\theta)\tilde{v}(\theta,\beta))\right] + G < 0 \\
&\forall t, (\theta,\beta) : u_t(\theta,\beta) > 0 \\
&\forall (\theta,\beta) : \tilde{v}(\theta,\beta) > 0
\end{aligned}$$

where the existence of a strict inequality for nonaffine feasibility constraint remains to be demonstrated. Note that we can easily construct such a point by allocating to all agents positive consumption utility $u_t(\theta,\beta) = \tilde{u} > \underline{u}$ *in each period* $t = 1, 2$ *and labor disutility* $\tilde{v}(\theta,\beta) = \tilde{v}\left((1 + 1/R)C(\tilde{u}) + G + \varepsilon\right)$ *for some* $\varepsilon > 0$ *so that*

$$\sum_{\theta,\beta} \pi(\theta,\beta)\left[C(u_1(\theta,\beta)) + \frac{C(u_2(\theta,\beta))}{R} - Y(D(\theta)\tilde{v}(\theta,\beta))\right] + G < 0$$

Slater's Theorem (see, e.g., p.226 and pp.234–236 of Boyd and Vandenberghe (2004)) states that Slater's condition and convexity of the primal problem imply strong duality, meaning that the duality gap is zero. Therefore, the KKT conditions are necessary and sufficient for primal and dual optimality of the planner's problem.

The results in Lemma 2 will guide our further analysis. Convexity provides a strong motive for pooling agents in the economy. Existence and uniqueness of the optimum are important for the theoretical characterization and also for designing a numerical solution algorithm for the problem. The maximum theorem will allow us to characterize a subset of all interior points of the type space by using perturbation arguments. Strong duality is used primarily in the numerical solution algorithm, but some theoretical proofs will also make use of it, as it implies existence of a bounded shadow price of resources. Going forward, with a slight abuse of language, we will say that a constraint of the planner's problem is *binding* if there exists a Lagrange multiplier on the constraint that is strictly positive, and conversely we say that a constraint is *slack* if all possible Lagrange multipliers on the constraint equal zero.

A.2 Benchmark allocations

We define an agent's implicit savings rate as their share of old-age consumption out of lifetime consumption, in discounted present value terms, $s(\theta,\beta) \equiv (c_2(\theta,\beta)/R)/(c_1(\theta,\beta)+c_2(\theta,\beta)/R)$. We define an agent's implicit transfer receipt as the difference between the discounted present value of their consumption and their labor income, $T(\theta,\beta) \equiv c_1(\theta,\beta)+c_2(\theta,\beta)/R-y(\theta,\beta)$.

Agents' *laissez-faire (LF)* allocation maximizes their decision utility (1) subject to individual feasibility:

$$\forall(\theta,\beta): \quad \left(c_1^{LF}(\theta,\beta), c_2^{LF}(\theta,\beta), y^{LF}(\theta,\beta)\right) = \underset{(c_1,c_2,y)\in\{(c_1,c_2,y)|y-c_1-c_2/R\geq 0\}}{\arg\max} U(c_1,c_2,y;\theta,\beta)$$

Under laissez-faire, agents' savings rates only depend on β, labor supply equates the marginal rates of substitution and transformation between consumption and labor, and there are no transfers:

$$\begin{aligned} s^{LF}(\beta) &= \arg\max_s U\left((1-s)y^{LF}(\theta,\beta), Rsy^{LF}(\theta,\beta), y^{LF}(\theta,\beta);\theta,\beta\right) \\ v'\left(\frac{y^{LF}(\theta,\beta)}{\theta}\right) &= u'\left(\left(1-s^{LF}(\beta)\right)y^{LF}(\theta,\beta)\right)\theta \\ T^{LF}(\theta,\beta) &= 0 \end{aligned}$$

The *first-best (FB)* allocation maximizes welfare (4) subject to feasibility (3):

$$\left\{c_1^{FB}(\theta,\beta), c_2^{FB}(\theta,\beta), y^{FB}(\theta,\beta)\right\}_{(\theta,\beta)} = \underset{\{c_1(\theta,\beta)c_2(\theta,\beta),y(\theta,\beta)\}_{(\theta,\beta)}\in F}{\arg\max} \sum_{(\theta,\beta)} \pi(\theta,\beta)\lambda(\theta) V(\theta,\beta),$$

where F is the set of feasible allocations satisfying (3). Since β does not directly enter welfare, the first-best savings rate is constant, labor supply equates the marginal rates of substitution and transformation between consumption and labor, and transfers equate marginal welfare across types:

$$\begin{aligned} s^{FB} = \arg\max_s V\left((1-s)\left(y^{FB}(\theta)+T^{FB}(\theta)\right), Rs\left(y^{FB}(\theta)+T^{FB}(\theta)\right), y^{FB}(\theta);\theta\right) \\ v'\left(\frac{y^{FB}(\theta)}{\theta}\right) = u'\left(\left(1-s^{FB}\right)\left(y^{FB}(\theta)+T^{FB}(\theta)\right)\right)\theta \\ \forall\theta,\theta': \quad \lambda(\theta)\frac{\partial\tilde{V}\left(T^{FB}(\theta);\theta\right)}{\partial T} = \lambda(\theta')\frac{\partial\tilde{V}\left(T^{FB}(\theta');\theta'\right)}{\partial T}, \end{aligned}$$

where $\tilde{V}(T;\theta) \equiv V\left(\left(1-s^{FB}\right)\left(y^{FB}(\theta)+T\right), Rs^{FB}\left(y^{FB}(\theta)+T\right), y^{FB}(\theta);\theta\right)$ is the indirect utility from transfers under the first-best consumption and labor supply policies.

The laissez-faire allocation satisfies IC and feasibility, providing a lower bound on welfare. The first-best allocation generally features higher welfare but it may violate IC. Therefore, we are interested in characterizing the constrained optimum, or *second-best*.

A.3 Proofs for the general economy

A.3.1 Additional result: Relevant IC constraints with 2×2 types

In this section, we analyze which IC constraints bind in the simple production economy with 2×2 types presented in Section 3.1.

First, IC constraints of θ_L-types are trivially satisfied. The IC constraints between (θ_L, β_L)-types and (θ_L, β_H)-types hold with equality since they are bunched, and they are slack with respect to both (θ_H, β_L)-types and (θ_H, β_L)-types because θ_L-types cannot work.

Second, the IC constraint of type (θ_H, β_L) with respect to θ_L-types binds. Suppose, by way of contradiction, that it does not. Then there is no reason to distort savings among low-ability types,

and thus $c_1(\theta_L) = c_2(\theta_L)$ at the solution. We have already shown that $c_2(\theta_H, \beta_H) \geq c_1(\theta_H, \beta_H)$, therefore $c_2(\theta_H, \beta_H) > c_2(\theta_L)$ to preserve IC between working (θ_H, β_H)-types and idle θ_L-types. We want to show that $\lambda(\theta_L) \geq \lambda(\theta_H)$ implies that the IC constraint of (θ_H, β_H)-types with respect to θ_L-types binds in this case.

By way of contradiction, suppose that IC from (θ_H, β_H) to θ_L also does not bind. There are two cases to consider. Case 1: IC from (θ_H, β_H) to (θ_H, β_L) binds, i.e. they have the same welfare, which together with the fact that θ_H-types are separated implies

$$c_2(\theta_H, \beta_L) < c_2(\theta_H, \beta_H)$$
$$\text{and} \quad u(c_1(\theta_H, \beta_L)) - v(y(\theta_H, \beta_L)/\theta_H) > u(c_1(\theta_H, \beta_H)) - v(y(\theta_H, \beta_H)/\theta_H)$$

From this and the fact that $\tau(\theta_H, \beta) = 0$ for $\beta \in \{\beta_L, \beta_H\}$, it follows that $s(\theta_H, \beta_L) < s(\theta_H, \beta_H)$ and the IC from (θ_H, β_L) to (θ_H, β_H) is slack. Since $s(\theta_H, \beta_L) < s(\theta_H, \beta_H) < s^{FB}$ and $s^{FB} < s(\theta_H, \beta_L) < s(\theta_H, \beta_H)$ cannot be optimal, then two subcases. Subcase 1A: $s(\theta_H, \beta_L) = s^{FB} < s(\theta_H, \beta_H)$, then since IC from (θ_H, β_L) to (θ_H, β_H) does not bind, we can decrease $s(\theta_H, \beta_H)$ by moving down along β_H-types' indifference curve to keep welfare constant, preserve IC, but save resources—a contradiction. Subcase 1B: $s(\theta_H, \beta_L) < s^{FB} \leq s(\theta_H, \beta_H)$, then since all ICs of (θ_H, β_L) are slack, we can increase (θ_H, β_L)-types' savings rate along β_H-types' indifference curve while keeping welfare constant, preserving IC, but saving resources—a contradiction. This rules out Case 1. Case 2: IC from (θ_H, β_H) to (θ_H, β_L) does not bind, i.e. welfare is strictly higher for (θ_H, β_H) than for (θ_H, β_L). Since $\lambda(\theta_L) \geq \lambda(\theta_H)$, the fact that no IC constraints bind from θ_H-types to θ_L-types implies that we could transfer some period 2 consumption from (θ_H, β_H)-types to θ_L-types to improve welfare but preserve IC—a contradiction. This rules out Case 2. Therefore, we have shown that the IC constraint from (θ_H, β_H)-types to θ_L-types must bind.

Together with $c_2(\theta_H, \beta_H) > c_2(\theta_L)$, this implies that (θ_H, β_L)-types strictly prefer θ_L-types' allocation over that of (θ_H, β_H)-types. Combining this with the IC constraint from (θ_H, β_L)-types

to θ_L-types, we infer that the IC constraint between (θ_H, β_L)-types and (θ_H, β_H)-types is slack:

$$\begin{aligned}
&u\left(c_1\left(\theta_H, \beta_L\right)\right) - v\left(\frac{y\left(\theta_H, \beta_L\right)}{\theta_H}\right) + \beta_L \delta u\left(c_2\left(\theta_H, \beta_L\right)\right) \\
\geq &u\left(c_1\left(\theta_L\right)\right) + \beta_L \delta u\left(c_2\left(\theta_L\right)\right) \\
> &u\left(c_1\left(\theta_H, \beta_H\right)\right) - v\left(\frac{y\left(\theta_H, \beta_H\right)}{\theta_H}\right) + \beta_L \delta u\left(c_2\left(\theta_H, \beta_H\right)\right)
\end{aligned}$$

Consequently, all of (θ_H, β_L)-types' IC constraints are slack, so it must be the case that all agents save at the first-best rate. But we already showed that (θ_H, β_L)-types and (θ_H, β_H)-types cannot have the same consumption bundle, so for IC to hold one of the types must consume more in both periods and work more relative to the other type—contradicting the fact that $\tau^L\left(\theta_H, \beta\right) = 0$ for $\beta \in \{\beta_L, \beta_H\}$. Hence, (θ_H, β_L)-types' IC constraint must bind with respect to θ_L-types.

Third, the pattern of other θ_H-types' IC constraints depends on parameter values. We consider three cases. Case 1 features a binding IC constraints from (θ_H, β_H)-types to (θ_H, β_L)-types but not the other way around. This case obtains when $\beta_L \approx 1$, $0 \approx \lambda\left(\theta_H\right) < \lambda\left(\theta_L\right)$, and $\pi\left(\theta_L\right) \approx 1$. Since $c_2\left(\theta_H, \beta_H\right) > c_2\left(\theta_L\right)$, if $\beta_L = 1 - \varepsilon$ for small enough $\varepsilon > 0$, by the maximum theorem (part 3 of Lemma 2) we still have $c_2\left(\theta_H, \beta_L\right) > c_2\left(\theta_L\right)$. Combined with the fact that the IC constraint of (θ_H, β_L)-types binds with respect to θ_L-types, this implies that the IC constraint of (θ_H, β_H)-types with respect to θ_L-types is slack. Then it must be that the IC constraint of (θ_H, β_H)-types with respect to (θ_H, β_L)-types binds, or else none of (θ_H, β_H)-types' IC constraints would bind and we could improve welfare by transfering period 2 consumption from (θ_H, β_H)-types to (θ_H, β_L)-types. Hence, $V\left(\theta_H, \beta_H\right) = V\left(\theta_H, \beta_L\right) > V\left(\theta_L\right)$. In this case, as $\pi\left(\theta_L\right) \to 1$, then $\tau^E\left(\theta_L\right) \to 0$ and $\tau^D\left(\theta_H, \beta_L\right) \to +\infty$, so for high enough values of $\pi\left(\theta_L\right)$ it must be that the IC constraint from (θ_H, β_L)-types to (θ_H, β_H)-types becomes slack, while the IC constraint from (θ_H, β_H)-types to (θ_H, β_L)-types binds. As a result, Case 1 features $s\left(\theta_H, \beta_H\right) = s^{FB} > s^{LF}\left(\beta_L\right) > s\left(\theta_H, \beta_L\right)$.

Case 2 features a binding IC constraints from (θ_H, β_L)-types to (θ_H, β_H)-types but not the other way around. This case obtains for intermediate values of β_L, $0 \approx \lambda\left(\theta_H\right) < \lambda\left(\theta_L\right)$, and $\pi\left(\theta_L\right) \approx 0$. In this case, as $\pi\left(\theta_L\right) \to 0$, then $\tau^E\left(\theta_L\right) \to -\infty$ and $\tau^D\left(\theta_H, \beta_L\right) \to 0$, so for low enough values of $\pi\left(\theta_L\right)$ it must be that the IC constraint from (θ_H, β_H)-types to (θ_H, β_L)-types becomes slack, while the IC constraint from (θ_H, β_L)-types to (θ_H, β_H)-types binds. Consequently, we also must have $V\left(\theta_H, \beta_H\right) = V\left(\theta_L\right) > V\left(\theta_H, \beta_L\right)$. As a result, Case 2 features $s\left(\theta_H, \beta_H\right) > s^{FB} > s\left(\theta_H, \beta_L\right) \geq$

$s^{LF}(\beta_L)$.

Case 3 features no binding IC constraints between (θ_H, β_L)-types and (θ_H, β_H)-types. This case obtains when $\beta_L \approx 0$, $\lambda(\theta_H) \approx 0 < \lambda(\theta_L)$, and intedmediate values of $\pi(\theta_L)$. As (θ_H, β_L)-types with $\lambda(\theta_H) = 0$ and $\beta_L = 0$ optimally save at their laissez-faire rate, $s(\theta_H, \beta_L) = s^{LF}(\beta_L) = 0 < s(\theta_L)$, then $c_2(\theta_H, \beta_L) = 0 < \min\{c_2(\theta_L), c_2(\theta_H, \beta_H)\}$ implies that the IC constraint from (θ_H, β_H)-types to (θ_H, β_L)-types does not bind if period utility is unbounded from below. If $\beta_L = \varepsilon$ and $\lambda(\theta_H) = \eta$ for small enough $\varepsilon > 0$ and $\eta > 0$, by the maximum theorem (part 3 of Lemma 2) we also have $0 < c_2(\theta_H, \beta_L) < \min\{c_2(\theta_L), c_2(\theta_H, \beta_H)\}$ and the IC constraint from (θ_H, β_H)-types to (θ_H, β_L)-types still does not bind. Then it must be the case that the IC constraint from (θ_H, β_H)-types to θ_L-types binds, or else none of (θ_H, β_H)-types' IC constraints would bind and we could improve welfare by transfering period 2 consumption from (θ_H, β_H)-types to (θ_H, β_L) types. Hence, $V(\theta_H, \beta_H) = V(\theta_L) > V(\theta_H, \beta_L)$. Note that as $\pi(\theta_L) \to 1$, then $\tau^E(\theta_L) \to 0$ and $\tau^D(\theta_H, \beta_L) \to +\infty$, while as $\pi(\theta_L) \to 0$, then $\tau^E(\theta_L) \to -\infty$ and $\tau^D(\theta_H, \beta_L) \to 0$, so for intermediate values of $\pi(\theta_L)$, since not both constraints can bind between θ_H-types since they are separated, it must be that the IC constraint from (θ_H, β_L)-types to (θ_H, β_H)-types becomes slack. As a result, Case 3 features $s(\theta_H, \beta_H) = s^{FB} > s(\theta_H, \beta_L) \geq s^{LF}(\beta_L)$.

In summary, the bindingness of (θ_H, β_H)-types' IC constraints is a function of model parameters. This indeterminacy of which IC constraints bind is precisely what renders solutions to multidimensional screening problems elusive (Armstrong, 1996; Rochet and Choné, 1998). Luckily, our characterization of bunching versus separation at the bottom versus the top of the ability distribution does not depend on this particular feature of the solution.

A.3.2 Additional result: Threshold relative Pareto weight for fixed abilities

The results in Proposition 1 characterize the constrained optimum of the simple environment in Section 3.1 under the two assumptions that $\lambda(\theta_L) \geq \lambda(\theta_H)$ and $\theta_L = 0 < \theta_H$. It will be instructive to revisit these results while stepwise relaxing each of the two assumptions.

Let us now discuss our first generalization. By restricting welfare weights to be weakly more redistributive than utilitarian, $\lambda(\theta_L) \geq \lambda(\theta_H)$, our analysis exploited the planner's motive to redistribute from high-ability toward low-ability types. More generally, the previous results continue to hold under redistributive enough social preferences.

Proposition 4. *Define* $\tilde{\lambda} \equiv \lambda(\theta_H)/\lambda(\theta_L)$. *There exists a unique cutoff* $1 < \overline{\tilde{\lambda}} < +\infty$ *such that:*

1. *If* $\tilde{\lambda} < \overline{\tilde{\lambda}}$*, then a second-best allocation is obtained, with* θ_L*-types bunched across* β*-levels, with* θ_H*-types separated, and with savings and labor distortions as in Proposition 1.*

2. *If* $\tilde{\lambda} \geq \overline{\tilde{\lambda}}$*, then the first-best allocation is obtained.*

Proof. We proceed in steps:

1. Indeed, the proof for bunching of θ_L-types does not rely on any assumptions on Pareto weights, hence goes through for any $\tilde{\lambda} \in \mathbb{R}_+$. In Section 3.1 we have shown that separation among θ_H-types obtains for $\tilde{\lambda} \leq 1$. Hence, by continuity of the optimal allocation in model parameters, due to the maximum theorem (part 3 of Lemma 2), we know that there exists a $\overline{\tilde{\lambda}} \in (1, +\infty)$ with the desired properties. The implied savings levels and labor distortions follow immediately from the previous analysis.

2. For $\lambda(\theta_L) = 0 < \lambda(\theta_H)$, that is $\tilde{\lambda} = +\infty$, clearly it is optimal to set $c_1(\theta_L) = c_2(\theta_L) = 0 < \min_{(t,\beta)\in\{1,2\}\times\{\beta_L,\beta_H\}} \{c_t(\theta_H, \beta)\}$, so the IC constraint from θ_H-types to θ_L-types cannot bind, hence θ_H-types are optimally bunched with $s(\theta_H) = s^{FB}$. By continuity of the optimal allocation in $(\lambda(\theta_L), \lambda(\theta_H))$, due to the maximum theorem (part 3 of Lemma 2), there exists a finite $\overline{\tilde{\lambda}} \in (1, +\infty)$ such that separation of high-ability types is optimal for all $\tilde{\lambda} < \overline{\tilde{\lambda}}$ but high ability types are bunched for $\tilde{\lambda} = \overline{\tilde{\lambda}}$. Consider the case when $\tilde{\lambda} = \overline{\tilde{\lambda}}$, so θ_H-types are bunched. Then the only reason that the planner would allow (θ_H, β_L)-types to decrease their savings rate along β_L-types indifference curve, thereby decreasing welfare from θ_H-types but saving resources, is to increase transfers to θ_L-types. As the net effect on welfare of such a perturbation is strictly decreasing in $\lambda(\theta_L)$, it follows that the optimal rule has a threshold property. Thus, θ_H-types are bunched for some $\tilde{\lambda} = \overline{\tilde{\lambda}}$, then bunching must also occur for $\tilde{\lambda} > \overline{\tilde{\lambda}}$. That all agents must save at the first-best rate follows from the fact that bunching of θ_H-types together with no binding IC constraint from θ_L-types to θ_H-types implies $s(\theta_H) = s^{FB}$, while no IC constraint binding from (θ_H, β_L)-types to θ_L-types implies $s(\theta_L) = s^{FB}$. That the labor margin remains undistorted follows from the previous analysis.

□

Part 1 of Proposition 4 nests as a special case our characterization of the simple environment in Section 3.1. Part 2 of the proposition states that the first-best is obtained for a regressive enough social planner. The threshold on relative Pareto weights, $\overline{\tilde{\lambda}}$, is defined as the point at which all IC

constraints from θ_H-types stop binding with respect to θ_L-types. Absent binding IC constraints across θ-levels, there is no reason to distort savings or labor supply of any agents, so the first-best is obtained.

A.3.3 Additional result: Threshold relative ability level for fixed Pareto weights

Let us now discuss our second generalization. We have so far studied the case when $\theta_L = 0 < \theta_H$. The assumption that θ_L-types cannot work implied that no IC constraint could bind from θ_L-types to θ_H-types, thereby reducing the complexity of the incentive problem. An analogous result is obtained when we allow both ability levels to be strictly positive as long as the productivity differential between them remains large enough.

Proposition 5. *Define $\tilde{\theta} \equiv \theta_H/\theta_L$. If $\lambda(\theta_L) \geq \lambda(\theta_H)$, then there exists a cutoff $1 < \overline{\tilde{\theta}} < +\infty$ such that if $\tilde{\theta} \geq \overline{\tilde{\theta}}$, then a second-best is obtained, where θ_L-types are bunched across β-levels and θ_H-types are separated. In this case, savings and labor distortions are as in Proposition 1 and, in addition, θ_L-types face a strictly positive implicit labor income tax: $\tau^L(\theta_L) > 0$.*

Proof. First, a special case of part 3 of Lemma 3 for the environment with 2×2 types implies that there exists a finite $1 < \tilde{\theta} < +\infty$ such that if $\theta_H/\theta_L \geq \tilde{\theta}$, then all IC constraints from θ_L-types to θ_H-types are slack. If in addition $\lambda(\theta_L) \geq \lambda(\theta_H)$, then an IC constraint from some θ_H-type to some θ_L-type must bind, or else the planner could redistribute resources from θ_H-types towards θ_L-types to improve welfare. Then all of our previous arguments apply and the results about decision wedges and efficiency wedges follow. That the labor wedge on θ_H-types stays at zero follows since IC constraints from θ_L-types to θ_H-types remain slack. That the labor wedge on θ_L-types is positive for finite values of $\tilde{\theta}$ follows from the fact that the IC constraint from (θ_H, β_L)-types and potentially also from (θ_H, β_H)-types binds with respect to θ_L-types, so a small increase in $\tau^L(\theta_L)$ from zero results in a second-order welfare loss but a first-order relaxation of θ_H-types' incentive compatibility constraints with respect to θ_L-types. Relaxing this IC constraint facilitates redistribution toward θ_L-types, which is desirable under the assumption that $\lambda(\theta_L) \geq \lambda(\theta_H)$. □

Proposition 5 shows that if the ability differential is large enough, then we obtain results analogous to those from Proposition 1. Intuitively, under these conditions, high-ability types are treated as cash cows, while low-ability types are subjected to a strong form of paternalism.

A.3.4 Additional result: Separation at low ability, bunching at high ability

So far, we have relied on the intuition that the interaction between paternalism and redistribution leads the optimal allocation to feature bunching of low-abiltiy types across present bias levels, but separation of high-ability types. Proposition 6 shows that the reverse can be the optimal for very regressive preferences, that is when the planner attaches much higher weight on high-ability types than on low-ability types.

Proposition 6. *Fix all parameters in the production economy with 2 θ-types, $\theta \in \{\theta_L, \theta_H\}$ with $0 < \theta_L < \theta_H$ and $\lambda(\theta_L) = 0 < \lambda(\theta_H)$, and 2 β-types, $0 \leq \beta_L < \beta_H = 1$. Define $\tilde{\theta} \equiv \theta_H/\theta_L$. Then there exists a cutoff $1 < \overline{\tilde{\theta}} < +\infty$ such that if $\overline{\tilde{\theta}} \leq \tilde{\theta} < +\infty$, then a second-best obtains, where θ_L-types are separated and θ_H-types are bunched, with the following savings and labor distortions:*

$$\begin{aligned} \tau^D(\theta_L, \beta_L) = 0 > \tau^D(\theta_L, \beta_H) \\ \tau^D(\theta_H, \beta_L) < \tau^D(\theta_H, \beta_H) = 0 \\ \tau^E(\theta_L, \beta_L) > \tau^E(\theta_H) = 0 \geq \tau^E(\theta_L, \beta_H) \\ \tau^L(\theta_L, \beta_L) = \tau^L(\theta_L, \beta_H) = 0 > \tau^L(\theta_H) \end{aligned}$$

Proof. Let $\lambda(\theta_L) = 0 < \lambda(\theta_H)$. Clearly, for large enough $\tilde{\theta}$, all IC constraints are slack from θ_H-types to θ_L-types. Therefore, θ_H-types must be bunched, or else we could partially convexify their allocation in utility space to maintain constant welfare, preserve IC, but save resources—a contradiction. Because for large enough $\tilde{\theta}$ and $\lambda(\theta_L) = 0 < \lambda(\theta_H)$ we must also have $c_2(\theta_H) > \max\{c_2(\theta_L, \beta_L), c_2(\theta_L, \beta_H)\}$, then only the IC from (θ_L, β_H)-types to θ_H-types binds, but not that of (θ_L, β_L)-types to θ_H-types. Hence, θ_H-types must be bunched at the first-best savings rate. Furthermore, θ_L-types must be separated across β-levels since $s(\theta_L, \beta_L) \leq s^{LF}(\beta_L) < s^{LF}(\beta_H) \leq s(\theta_L, \beta_H)$ or else we could save resources by pushing the respective type along their indifference curve toward their laissez-faire savings rate. Note that exactly one of the IC constraints between θ_L-types must bind. If non were binding, then (θ_L, β_L)-types would have no binding IC constraints, so the planner could take some resources from them without affecting welfare. If both were binding, then θ_L-types would have to be bunched at the first-best rate, which cannot be optimal. Now, if the IC constraint were to bind from (θ_L, β_H)-types to (θ_L, β_L)-types, then all of (θ_L, β_L)-types' IC constraints would be slack—a contradiction. Thus, the IC constraint must binds

from (θ_L, β_L)-types to (θ_L, β_H)-types, so $s(\theta_L, \beta_L) = s^{LF}(\beta_L)$ and $s(\theta_L, \beta_H) > s^{LF}(\beta_H)$. Finally, that θ_L-types' labor margin is undistorted follows directly from the fact that no IC constraint binds from θ_H-types to θ_L-types, and that θ_H-types experience an implicit labor subsidy follows from the fact that some IC constraint from θ_L-types must bind with respect to θ_H-types. □

A.3.5 Additional result: Threshold relative Pareto weight for uniform ability

In Proposition 4, we fixed productivities such that only high-ability types could work and considered all possible relative Pareto weights. Complementary to that analysis, the following proposition traces out the range of all possible relative Pareto weights when all agents have the same, strictly positive ability level.

Proposition 7. *Consider the production economy with 2 θ-types, $\theta_L = \theta_H > 0$, and 2 β-types, $0 \leq \beta_L < \beta_H = 1$. Define $\tilde{\lambda} \equiv \lambda(\theta_H)/\lambda(\theta_L)$. There are three cases:*

1. *If $\tilde{\lambda} = 1$, then the first-best obtains.*

2. *If $\tilde{\lambda} < 1$, then a second-best obtains, with types $(\theta, \beta) \in \{(\theta_L, \beta_L), (\theta_L, \beta_H), (\theta_H, \beta_H)\}$ bunched, while (θ_H, β_L)-types are separated from the rest, with the following savings and labor distortions:*

$$\tau^D(\theta_L, \beta_L) < \tau^D(\theta_L, \beta_H) = \tau^D(\theta_H, \beta_H) < 0 \begin{cases} = \tau^D(\theta_H, \beta_L) & \text{if } \tilde{\lambda} = 0 \\ > \tau^D(\theta_H, \beta_L) & \text{if } \tilde{\lambda} > 0 \end{cases}$$

$$\tau^E(\theta_L) = \tau^E(\theta_H, \beta_H) < 0 < \tau^E(\theta_H, \beta_L)$$

$$\tau^L(\theta_L) = \tau^L(\theta_H, \beta_L) = \tau^L(\theta_H, \beta_H) = 0$$

3. *If $\tilde{\lambda} > 1$, a second-best obtains, with types $(\theta, \beta) \in \{(\theta_L, \beta_H), (\theta_H, \beta_L), (\theta_H, \beta_H)\}$ bunched, while (θ_L, β_L)-types are separated from the rest, with the following savings and labor distortions:*

$$\tau^E(\theta_H, \beta_L) < \tau^D(\theta_L, \beta_H) = \tau^E(\theta_H, \beta_H) < 0 \begin{cases} = \tau^D(\theta_L, \beta_L) & \text{if } \tilde{\lambda} = +\infty \\ > \tau^D(\theta_L, \beta_L) & \text{if } \tilde{\lambda} < +\infty \end{cases}$$

$$\tau^E(\theta_L, \beta_H) = \tau^E(\theta_H) < 0 < \tau^E(\theta_L, \beta_L)$$

$$\tau^L(\theta_L, \beta_L) = \tau^L(\theta_L, \beta_H) = \tau^L(\theta_H) = 0$$

Proof. Let $\theta_L = \theta_H > 0$. All agents' labor margins must be undistorted, or else one could equate $u'(\cdot)$ to $v'(\cdot)/\theta$ in order to maintain period 1 utility constant for all agents but save resources. IC constraints then imply that for any $\theta, \theta' \in \{\theta_L, \theta_H\}$ with $\theta_L = \theta_H$:

$$\begin{aligned} & u\left(c_1\left(\theta, \beta_L\right)\right) - v\left(\frac{y\left(\theta, \beta_L\right)}{\theta}\right) + \beta_L \delta u\left(c_2\left(\theta, \beta_L\right)\right) \geq u\left(c_1\left(\theta', \beta_H\right)\right) - v\left(\frac{y\left(\theta', \beta_H\right)}{\theta}\right) + \beta_L \delta u\left(c_2\left(\theta', \beta_H\right)\right) \\ \wedge \quad & u\left(c_1\left(\theta, \beta_L\right)\right) - v\left(\frac{y\left(\theta, \beta_L\right)}{\theta'}\right) + \delta u\left(c_2\left(\theta, \beta_L\right)\right) \leq u\left(c_1\left(\theta', \beta_H\right)\right) - v\left(\frac{y\left(\theta', \beta_H\right)}{\theta'}\right) + \delta u\left(c_2\left(\theta', \beta_H\right)\right) \\ \implies \quad & c_2\left(\theta, \beta_L\right) \leq c_2\left(\theta', \beta_H\right) \end{aligned}$$

so either (θ, β_L)-types are bunched with (θ', β_H)-types, or else $s(\theta, \beta_L) < s(\theta', \beta_H)$. From here, we proceed in steps:

1. Let $\theta_L = \theta_H$. If $\lambda(\theta_L) = \lambda(\theta_H)$, then the first-best assigns the same allocation to all agents, making it trivially incentive compatible.

2. Consider the case when $\tilde{\lambda} < 1$, so the planner would like to make a net resource transfer to θ_L-types. Note that (θ_L, β_H)-types and (θ_H, β_H)-types must be bunched since IC between them implies $V(\theta_L, \beta_H) = V(\theta_H, \beta_H) = V(\beta_H)$, so the cost of their allocation is the same, which by strict concavity of consumption utility and strict convexity of labor disutility defines a unique consumption-labor bundle. Similarly, $U(\theta_L, \beta_L) = U(\theta_H, \beta_L)$, since those types have identical preferences and ability, hence $V(\theta_L, \beta_L) \geq V(\theta_H, \beta_L)$, or else the planner could switch their allocations to preserve IC but increase welfare. Combined with IC from β_H-types to β_L-types, it follows that $V(\beta_H) \geq V(\theta_L, \beta_L) \geq V(\theta_H, \beta_L)$.

   Then it must be that $V(\theta_L, \beta_L) > V(\theta_H, \beta_L)$ and/or $V(\theta_H, \beta_H) > V(\theta_L, \beta_L)$, as otherwise we would have $V(\beta_H) = V(\theta_L, \beta_L) = V(\theta_H, \beta_L)$, which to be optimal requires that all agents are bunched at the first-best savings rate, but this scenario could be improved upon by first allowing (θ_H, β_L)-types to dissave along their indifference curve and then transferring those resources to other agents.

   Suppose that $V(\beta_H) > V(\theta_L, \beta_L) \geq V(\theta_H, \beta_L)$. Then we could convexify the allocation of β_H-types with that of (θ_L, β_L)-types in utility space to preserve IC and save a strictly positive amount of resources. But this convex combination would also improve welfare because the Pareto weight on (θ_L, β_L)-types, $\lambda(\theta_L)$, is strictly greater than the average Pareto weight on β_H-types, but $V(\beta_H) > V(\theta_L, \beta_L)$ to begin with—a contradiction. Hence $V(\theta_L, \beta_L) =$

$V(\beta_H)$ at the optimum. Then (θ_L, β_L)-types must be bunched with β_H-types, or else $s(\beta_H) = s^{FB}$ and we could save resources by increase β_L-types' savings rate to move them into the interior of the budget set along β_H-types' indifferences curves through their respective allocations. Therefore, we have bunching of types $(\theta, \beta) \in \{(\theta_L, \beta_L), (\theta_L, \beta_H), (\theta_H, \beta_H)\}$.

Next, $V(\beta_H) = V(\theta_L, \beta_L) > V(\theta_H, \beta_L)$, or else if $V(\beta_H) = V(\theta_L, \beta_L) = V(\theta_H, \beta_L)$, then $s(\theta_H, \beta_L) < s(\beta_H) = s(\theta_L, \beta_L) = s^{FB}$ and we could move (θ_H, β_L)-types up along β_H-types' indifference curve to keep welfare constant and preserve IC but save resources—a contradiction.

Furthermore, the IC constraint from (θ_H, β_L)-types to the bunched types must bind, or else we would have $s(\theta_H, \beta_L) < s(\theta, \beta) = s^{FB}$ for $(\theta, \beta) \in \{(\theta_L, \beta_L), (\theta_L, \beta_H), (\theta_H, \beta_H)\}$ and we could increase (θ_H, β_L)-types' savings rate along the budget constraint to increase welfare and preserve IC at no cost—a contradiction. The savings rate of the bunched types must be strictly above the first-best, while that of (θ_H, β_L)-types must be strictly below the first-best, by the usual argument.

Finally, for $\lambda(\theta_H) > 0$ it is optimal to set $s(\theta_H, \beta_L) > s^{LF}(\beta_L)$, or else we could increase their savings rate along β_L-types' indifference curve to improve welfare to a first order, preserve IC, but incur only a second-order resource cost.

In summary, the optimal allocation features $(c_1(\theta, \beta), c_2(\theta, \beta), y(\theta, \beta)) = (c_1, c_2, y)$ such that $s(\theta, \beta) = s > s^{FB}$ and $V(\theta, \beta) = V$ for bunched types $(\theta, \beta) \in \{(\theta_L, \beta_L), (\theta_L, \beta_H), (\theta_H, \beta_H)\}$, but (θ_H, β_L)-types are separated with $s^{LF}(\beta_L) \leq s(\theta_H, \beta_L) < s^{FB}$ and $V(\theta_H, \beta_L) < V$.

3. An analogous argument to the proof of part 2 of the proposition shows that if $\tilde{\lambda} > 1$, then there will be bunching across types $(\theta, \beta) \in \{(\theta_L, \beta_H), (\theta_H, \beta_L), (\theta_H, \beta_H)\}$, while (θ_L, β_L)-types are separated from the rest, with savings and labor distortions satisfying the desired properties.

□

Proposition 7 shows that if agents share the same productivity level, then it becomes hard to separate agents by Pareto weights. A utilitarian planner is, of course, perfectly happy to have agents bunched across θ-levels as well as across β-levels, which gives rise to the first-best. If the planner wants to redistribute from some type of agents to another, however, it becomes subopti-

mal to separate β_H-types across θ-levels. As a result, only β_L-types of the θ-level that the planner cares least about are separated and allowed to save below the efficient rate in exchange for resources used for redistribution.

A.3.6 Proof of Lemma 1

Proof. We proceed in steps:

1. Let $\beta > \beta'$. Combining IC constraints of types (θ, β) and (θ, β') for any θ:

$$\begin{aligned}
& u\left(c_1\left(\theta,\beta\right)\right) - v\left(y\left(\theta,\beta\right)/\theta\right) + \beta\delta u\left(c_2\left(\theta,\beta\right)\right) \geq u\left(c_1\left(\theta,\beta'\right)\right) - v\left(y\left(\theta,\beta'\right)/\theta\right) + \beta\delta u\left(c_2\left(\theta,\beta'\right)\right) \\
\wedge \quad & u\left(c_1\left(\theta,\beta'\right)\right) - v\left(y\left(\theta,\beta'\right)/\theta\right) + \beta'\delta u\left(c_2\left(\theta,\beta'\right)\right) \geq u\left(c_1\left(\theta,\beta\right)\right) - v\left(y\left(\theta,\beta\right)/\theta\right) + \beta'\delta u\left(c_2\left(\theta,\beta\right)\right) \\
\implies \quad & \left(\beta' - \beta\right)\delta u\left(c_2\left(\theta,\beta'\right)\right) \geq \left(\beta' - \beta\right)\delta u\left(c_2\left(\theta,\beta\right)\right) \\
\implies \quad & u\left(c_2\left(\theta,\beta'\right)\right) \leq u\left(c_2\left(\theta,\beta\right)\right) \\
\implies \quad & u\left(c_1\left(\theta,\beta'\right)\right) - v\left(y\left(\theta,\beta'\right)/\theta\right) \geq u\left(c_1\left(\theta,\beta\right)\right) - v\left(y\left(\theta,\beta\right)/\theta\right)
\end{aligned}$$

   There are two cases. Case 1 is when

$$\begin{aligned}
& u\left(c_2\left(\theta,\beta\right)\right) = u\left(c_2\left(\theta,\beta'\right)\right) \\
\text{and} \quad & u\left(c_1\left(\theta,\beta\right)\right) - v\left(y\left(\theta,\beta\right)/\theta\right) = u\left(c_1\left(\theta,\beta'\right)\right) - v\left(y\left(\theta,\beta'\right)/\theta\right).
\end{aligned}$$

   Since types (θ, β) and (θ, β') are indistinguishable in terms of welfare and IC, then their allocation must cost the same, which by convexity of the problem implies that that their consumption in both periods is the same and that they work the same amount. This proves the first statement of this part of the lemma. Case 2 is the complement of Case 1 and congruent to the second statement of this part of the lemma.

   For the second part, fix β and $\theta > \theta'$, then combining IC constraints between (θ', β)-types

and (θ, β)-types:

$$\begin{aligned}
& u\left(c_1\left(\theta', \beta\right)\right) - v\left(y\left(\theta', \beta\right)/\theta'\right) + \beta\delta u\left(c_1\left(\theta', \beta\right)\right) \geq u\left(c_1\left(\theta, \beta\right)\right) - v\left(y\left(\theta, \beta\right)/\theta'\right) + \beta\delta u\left(c_1\left(\theta, \beta\right)\right) \\
\wedge \quad & u\left(c_1\left(\theta', \beta\right)\right) - v\left(y\left(\theta', \beta\right)/\theta\right) + \beta\delta u\left(c_1\left(\theta', \beta\right)\right) \leq u\left(c_1\left(\theta, \beta\right)\right) - v\left(y\left(\theta, \beta\right)/\theta\right) + \beta\delta u\left(c_1\left(\theta, \beta\right)\right) \\
\Longrightarrow \quad & v\left(y\left(\theta', \beta\right)/\theta'\right) - v\left(y\left(\theta', \beta\right)/\theta\right) \leq v\left(y\left(\theta, \beta\right)/\theta'\right) - v\left(y\left(\theta, \beta\right)/\theta\right) \\
\Longrightarrow \quad & \left[D\left(\theta'\right) - D\left(\theta\right)\right]\tilde{v}\left(y\left(\theta', \beta\right)\right) \leq \left[D\left(\theta'\right) - D\left(\theta\right)\right]\tilde{v}\left(y\left(\theta, \beta\right)\right) \\
\Longrightarrow \quad & y\left(\theta', \beta\right) \leq y\left(\theta, \beta\right)
\end{aligned}$$

where the last line follows from the fact that $D\left(\theta'\right) > D\left(\theta\right)$ and the fact that $\tilde{v}\left(\cdot\right)$ is a strictly increasing function.

2. Let

$$\begin{aligned}
G\left(b\right) = & \left[u\left(c_1\left(\theta, \beta\right)\right) - v\left(y\left(\theta, \beta\right)/\theta\right) + b\delta u\left(c_2\left(\theta, \beta\right)\right)\right] \\
& - \left[u\left(c_1\left(\theta', \beta'\right)\right) - v\left(y\left(\theta', \beta'\right)/\theta\right) + b\delta u\left(c_2\left(\theta', \beta'\right)\right)\right]
\end{aligned}$$

for any pair of types (θ, β) and (θ', β') with $\beta > \beta'$, where

$$G'\left(b\right) = \delta\left[u\left(c_2\left(\theta, \beta\right)\right) - u\left(c_2\left(\theta', \beta'\right)\right)\right] \geq 0,$$

with the inequality following from the above monotonicity result. IC from (θ', β') to (θ, β) implies $G\left(\beta'\right) \leq 0$. Then for any $\beta'' < \beta'$ we have $G\left(\beta''\right) \leq 0$, i.e. (θ'', β'') weakly prefers the allocation of (θ', β') over that of (θ, β). Hence, if type (θ'', β'') prefers their own allocation weakly (strictly) over that of type (θ', β') and type (θ', β') prefers their own allocation weakly (strictly) over that of type (θ, β), then type (θ'', β'') also prefers their own allocation weakly (strictly) over that of (θ, β), which proves the first statement of this part of the lemma. Conversely, to show sufficiency of local IC constraints downwards in β-space, let

$$\begin{aligned}
H\left(b\right) = & \left[u\left(c_1\left(\theta', \beta'\right)\right) - v\left(y\left(\theta', \beta'\right)/\theta'\right) + b\delta u\left(c_2\left(\theta', \beta'\right)\right)\right] \\
& - \left[u\left(c_1\left(\theta'', \beta''\right)\right) - v\left(y\left(\theta'', \beta''\right)/\theta'\right) + b\delta u\left(c_2\left(\theta'', \beta''\right)\right)\right]
\end{aligned}$$

for any pair of types (θ', β') and (θ'', β'') with $\beta'' < \beta'$, where

$$H'(b) = \delta \left[u\left(c_2\left(\theta', \beta'\right)\right) - u\left(c_2\left(\theta'', \beta''\right)\right) \right] \geq 0$$

IC from (θ', β') to (θ'', β'') implies $H(\beta') \geq 0$. Then for any $\beta > \beta'$ we have $H(\beta) \geq 0$, i.e. (θ, β) prefers the allocation of (θ', β') over that of (θ'', β''). Hence, if type (θ, β) prefers their own allocation weakly (strictly) over that of type (θ', β') and type (θ', β') prefers their own allocation weakly (strictly) over that of type (θ'', β''), then type (θ, β) also prefers their own allocation weakly (strictly) over that of (θ'', β''), which proves the second statement of this part of the lemma.

□

A.3.7 Statement and proof of Lemma 3, leading up to Theorems 1–3

Lemma 3. *The following results hold in the general economy with N θ-types and M β-types:*

1. *If $\lambda(\theta) \geq \lambda(\theta')$ for some $\theta < \theta'$, then the first-best allocation is not incentive compatible.*
2. *Agents with $\beta = 1$ save weakly above the efficient rate, and strictly above iff. the IC constraint of some (θ', β)-type with $\beta' < 1$ binds with respect to their allocation.*
3. *Define $\tilde{\theta}_n^+ \equiv \theta_{n+1}/\theta_n$. For each $n = 1, \ldots, N-1$, there exists a cutoff $1 < \overline{\tilde{\theta}_n^+} < +\infty$ such that if $\tilde{\theta}_n^+ \geq \overline{\tilde{\theta}_n^+}$, then for any $\theta' \leq \theta_n < \theta''$ with $\lambda(\theta') > 0$ we have $y(\theta', \beta') < y(\theta'', \beta'')$ and $T(\theta', \beta') > T(\theta'', \beta'')$ for any β', β'', and all IC constraints from θ'-types to θ''-types are slack.*
4. *Define $\tilde{\theta}_n^- \equiv \theta_n/\theta_{n-1}$. For each $n = 2, \ldots, N$, there exist cutoffs $1 < \overline{\tilde{\theta}_n^-} < +\infty$ and $1 < \overline{\lambda_n} < +\infty$ such that if $\tilde{\theta}_n^- \geq \overline{\tilde{\theta}_n^-}$ and $\lambda(\theta_n) \geq \overline{\lambda_n}$, then for any $\theta' < \theta_n \leq \theta''$ we have that all IC constraints from θ''-types to θ'-types are slack.*
5. *For a given θ, if IC constraints from θ-types to θ'-types are slack for all $\theta' \neq \theta$, then θ-types are bunched across β-levels.*
6. *For a given θ, if IC constraints from θ'-types to θ-types are slack for all $\theta' \neq \theta$, then θ-types' labor margin is undistorted.*

Proof. We proceed in steps: □

1. *Let $\theta < \theta'$ and $\lambda(\theta) \geq \lambda(\theta')$, and consider the first-best allocation, then the IC constraint from $(\theta', \beta_M = 1)$ to θ-types is:*

$$\begin{aligned}(1+\delta)\, u\left(c\left(\theta'\right)\right) - v\left(y\left(\theta'\right)/\theta'\right) &\geq (1+\delta)\, u\left(c\left(\theta\right)\right) - v\left(y\left(\theta\right)/\theta'\right) \\ &> (1+\delta)\, u\left(c\left(\theta\right)\right) - v\left(y\left(\theta\right)/\theta\right),\end{aligned}$$

which states that $V(\theta') > V(\theta)$, contradicting the first-best condition for $\lambda(\theta) \geq \lambda(\theta')$.

2. *Suppose $\beta = 1$ but $s(\theta, \beta) < s^{FB}$ for some θ, then we could increase (θ, β)-types' savings rate along the planner's indifference curve to leave welfare constant, leave slack the IC constraints of all agents with $\beta' < 1$, but save resources—a contradiction. "$\Longrightarrow$" Suppose $s(\theta, \beta) > s^{FB}$. If the IC constraints of all β'-types, for $\beta' < 1$, of were slack with respect to (θ, β)-types' allocation, then we could decrease $s(\theta, \beta)$ along β_H-types' indifference curve to maintain constant welfare, preserve IC, but save resources—a contradiction. "$\Longleftarrow$" Suppose the IC constraint of some (θ', β')-types, for $\beta' < 1$, binds with respect to the allocation of (θ, β)-types. If $s(\theta, \beta) = s^{FB}$, then we could increase $s(\theta, \beta)$ along the budget constraint to incur a second-order welfare loss, preserve IC, but get a first-order welfare gain from relaxing the binding IC—a contradiction.*

3. *Consider $\theta' \leq \theta_n < \theta''$ for some $n = 1, \ldots, N-1$. From our assumption that $\lim_{\theta'' \to +\infty} D(\theta'') = 0$ it follows that $\lim_{\theta'' \to +\infty} y(\theta'', \beta'') = +\infty$ is optimal for any β''. Thus, for fixed $\theta_1, \ldots, \theta_n$, a sufficiently high θ'' guarantees that θ'-types' perceived disutility level and marginal disutility of mimicking one of θ''-types' labor effort grow arbitrarily large: $\lim_{\theta'' \to +\infty} v\left(y\left(\theta'', \beta''\right)/\theta'\right) = +\infty$ and $\lim_{\theta'' \to +\infty} v'\left(y\left(\theta'', \beta''\right)/\theta'\right) = +\infty$. At the same time, a fixed level of $\lambda(\theta') > 0$ guarantees that optimally, under the same limit, low-ability types' consumption tends to infinity: $\lim_{\theta'' \to +\infty} c_t(\theta', \beta') = +\infty$ for any β' and $t = 1, 2$. This requires that θ'-types' labor effort tends to zero: $\lim_{\theta'' \to +\infty} y(\theta', \beta') = 0$ for any β', or else the planner could reduce θ'-types' consumption and lower $y(\theta', \beta')$ in tandem to keep welfare constant but save resources. IC constraints for θ''-types also ensure that $\lim_{\theta'' \to +\infty} c_t(\theta'', \beta'') = +\infty$ for any β'' and $t = 1, 2$. Now define the following object:*

$$\begin{aligned}G(\theta) \equiv &\left[u\left(c_1\left(\theta', \beta'\right)\right) - v\left(y\left(\theta', \beta'\right)/\theta'\right) + \beta'\delta u\left(c_2\left(\theta', \beta'\right)\right)\right] \\ &- \left[u\left(c_1\left(\theta, \beta''\right)\right) - v\left(y\left(\theta, \beta''\right)/\theta'\right) + \beta'\delta u\left(c_2\left(\theta, \beta''\right)\right)\right]\end{aligned}$$

where $c_1(\cdot), c_2(\cdot), y(\cdot)$ *are all implicit functions of* θ*. Note that the IC constraint from* (θ', β')*-types to* (θ'', β'')*-types for any* β', β'' *holds iff.* $G(\theta'') \geq 0$*. We want to show that* $G(\theta'') > 0$ *for large enough* θ''*. Suppose, by way of contradiction, that* $G(\theta'') = 0$*. Now consider an increase in* θ *associated with an increase in* $y(\theta, \beta'')$ *of size* $\Delta y > 0$*. For small* Δy*, this leads to an increase in* $G(\cdot)$ *of* $G' \geq -\left[\max\left\{u'\left(c_1\left(\theta, \beta''\right)\right), \beta' \delta u'\left(c_2\left(\theta, \beta''\right)\right)\right\} - v'\left(y\left(\theta, \beta''\right)/\theta'\right)\right]$*. Because* $\lim_{\theta \to +\infty} u'\left(c_t\left(\theta, \beta''\right)\right) = 0$ *for* $t = 1, 2$ *but* $\lim_{\theta \to +\infty} v'\left(y\left(\theta, \beta''\right)/\theta'\right) = +\infty$*, for high enough* θ *it must be that* $G' > 0$*, so the IC constraint from* (θ', β')*-types to* (θ'', β'')*-types becomes slack. Thus, by continuity of the optimal allocation in* θ''*, due to the maximum theorem (part 3 of Lemma 2), there exists a finite* $\theta'' < +\infty$ *such that* $0 < y(\theta', \beta) < y(\theta'', \beta')$ *and all IC constraints from* θ'*-types to* θ''*-types are slack.*[46] *Similarly, since* $\lim_{\theta'' \to +\infty} y(\theta', \beta') = 0$*, then* $y(\theta', \beta') < y(\theta'', \beta'')$ *and* $T(\theta', \beta') > T(\theta'', \beta'')$ *for any* β', β'' *for relatively high enough* θ''*.*

4. *Consider* $\theta' < \theta_n \leq \theta''$ *for some* $n = 2, \ldots, N$*. Let* $\lambda(\theta_1) = \ldots = \lambda(\theta_{n-1}) = 0 < \lambda(\theta_n)$*, so the planner wants to transfer additional resources to* θ_n*-types from all* θ'*-types, for* $\theta' < \theta_n$*. Following an argument similar to that in part 4 of the lemma, for fixed* $\theta_1, \ldots, \theta_{n-1}$*, a sufficiently high* θ'' *guarantees that* θ'*-types' perceived disutility level and marginal disutility of mimicking one of* θ''*-types' labor effort grow arbitrarily large:* $\lim_{\theta'' \to +\infty} v\left(y\left(\theta'', \beta''\right)/\theta'\right) = +\infty$ *and* $\lim_{\theta'' \to +\infty} v'\left(y\left(\theta'', \beta''\right)/\theta'\right) = +\infty$*. Thus, for arbitrarily small* $\varepsilon > 0$ *and* $\delta > 0$ *there exists* $\overline{\tilde{\theta}_n^-}$ *such that for* $\tilde{\theta}_n^- \geq \overline{\tilde{\theta}_n^-}$ *we have* $y(\theta', \beta') \geq y(\theta_n, \beta) - \varepsilon$ *for all* β*, and* $\max\left\{c_1(\theta', \beta'), c_2(\theta', \beta')\right\} \leq \delta$*, since the disutility gap* $v\left(y(\theta', \beta')/\theta'\right) - v\left(y(\theta_n, \beta)/\theta'\right)$ *grows arbitrarily big for large enough* $\tilde{\theta}_n^-$*. We conclude that for* $\lambda(\theta_1) = \ldots = \lambda(\theta_{n-1}) = 0 < \lambda(\theta_n)$ *and* $\tilde{\theta}_n^- \geq \overline{\tilde{\theta}_n^-}$*, the IC constraint from* θ_n*-types must be slack with respect to all* θ'*-types, for* $\theta' < \theta_n$*. By continuity of the optimal allocation in* $\lambda(\theta)$*, due to the maximum theorem (part 3 of Lemma 2), then there exists a cutoff* $1 < \overline{\lambda_n} < +\infty$ *such that if* $\tilde{\theta}_n^- \geq \overline{\tilde{\theta}_n^-}$ *and* $\lambda(\theta_n) \geq \overline{\lambda_n}$*, then the desired result obtains.*

5. *Suppose, by way of contradiction, that the IC constraints from* θ*-types to* θ'*-types are slack for all* $\theta' \neq \theta$*, but not all* θ*-types are bunched. Take the original allocation of all* θ*-types,* $\left\{\left(c_1(\theta, \beta), c_2(\theta, \beta), y(\theta, \beta)\right)\right\}_{\beta \in B}$*, with associated utilities* $\left\{\left(u_1(\theta, \beta), u_2(\theta, \beta), v(\theta, \beta)\right)\right\}_{\beta \in B}$*. Now consider shifting* (θ, β)*-types toward a convex combination in utility space between their own allocation and all other* θ*-types' allo-*

[46] Note that the maximum theorem applies here because there is no discontinuity in either the objective or the IC constraints as $\theta'' \to +\infty$.

cations:

$$\left(\tilde{u_1}(\theta,\beta),\tilde{u_2}(\theta,\beta),\tilde{v}(\theta,\beta)\right) = \xi \underbrace{\left(u_1(\theta,\beta),u_2(\theta,\beta),v(\theta,\beta)\right)}_{(\theta,\beta)\text{-types' original allocation, depends on } \beta} + (1-\xi)\underbrace{\sum_{\beta'}\pi\left(\beta'|\theta\right)\left(u_1\left(\theta,\beta'\right),u_2\left(\theta,\beta'\right),v\left(\theta,\beta'\right)\right)}_{\text{convex combination, does not depend on } \beta}$$

for some $\xi \in [0,1]$. Since IC was satisfied before convexifying, then shifting (θ,β)-types from $\left(u_1(\theta,\beta),u_2(\theta,\beta),v(\theta,\beta)\right)$ to $\left(\tilde{u_1}(\theta,\beta),\tilde{u_2}(\theta,\beta),\tilde{v}(\theta,\beta)\right)$ preserves IC for all types other than (θ,β). But if only (θ,β)-types were shifted then this could violate their own IC in case some of their IC constraints with respect to other θ-types were originally slack while some of their other IC constraints with respect to anyone were originally binding. So we also need to shift agents other than (θ,β)-types toward the convex combination. Consider some (θ,β')-type with respect to whom (θ,β)-types' IC constraint was initially satisfied:

$$\underbrace{u_1(\theta,\beta)-v(\theta,\beta)+\beta u_2(\theta,\beta)}_{(\theta,\beta)\text{-types' original allocation, depends on } \beta} \geq \underbrace{u_1\left(\theta,\beta'\right)-v\left(\theta,\beta'\right)+\beta u_2\left(\theta,\beta'\right)}_{(\theta,\beta')\text{-types' original allocation, depends on } \beta'}$$

$$\Longrightarrow \quad \xi\underbrace{\left[u_1(\theta,\beta)-v(\theta,\beta)+\beta u_2(\theta,\beta)\right]}_{(\theta,\beta)\text{-types' original allocation, depends on } \beta} + (1-\xi)\underbrace{\sum_{\beta''}\pi\left(\beta'|\theta\right)\left[u_1\left(\theta,\beta''\right)-v\left(\theta,\beta''\right)+\beta u_2\left(\theta,\beta''\right)\right]}_{\text{convex combination, does not depend on } \beta \text{ or } \beta'}$$

$$\geq \xi\underbrace{\left[u_1\left(\theta,\beta'\right)-v\left(\theta,\beta'\right)+\beta u_2\left(\theta,\beta'\right)\right]}_{(\theta,\beta')\text{-types' original allocation, depends on } \beta'} + (1-\xi)\underbrace{\sum_{\beta''}\pi\left(\beta''|\theta\right)\left[u_1\left(\theta,\beta''\right)-v\left(\theta,\beta''\right)+\beta u_2\left(\theta,\beta''\right)\right]}_{\text{convex combination, does not depend on } \beta \text{ or } \beta'}$$

Therefore, the new, partially convexified allocation also satisfies IC of (θ,β)-types with respect to other θ-types. With respect to other θ'-types, since θ-types' IC constraints with respect to all other θ'-types were slack to begin with, for small enough $\xi > 0$ they remain slack. Thus IC is preserved. Welfare is kept constant by construction. But due to strict concavity of $u(\cdot)$ such a (partial) convexification will save a strictly positive amount of resources—a contradiction.

6. *The statement follows directly from the fact that imposing $\tau^L(\theta) \neq 0$ has negative consequences for efficiency but no incentive effects absent any binding constraints from other θ'-types to θ-types, so $\tau(\theta) = 0$ is optimal.*

A.3.8 Proof of Theorem 1

Proof. We proceed in steps:

1. (i) For high enough $\tilde{\theta}_1^+$, by part 3 of Lemma 3, all IC constraints of θ_1-types are slack with respect to θ''-types, for $\theta'' > \theta_1$. By part 5 of Lemma 3, then θ_1-types are bunched across β-levels.
   (ii) For high enough $\tilde{\theta}_n^+$, by part 3 of Lemma 3, all IC constraints of θ_n-types are slack with respect to θ''-types, for $\theta'' > \theta_n$. Furthermore, for high enough $\tilde{\theta}_n^-$, and $\lambda(\theta_n)$, by part 4 of Lemma 3, also all IC constraints of θ_n-types are slack with respect to θ'-types, for $\theta' < \theta_n$. By part 5 of Lemma 3, then θ_n-types are bunched across β-levels.
   (iii) For high enough $\tilde{\theta}_N^-$ and $\lambda(\theta_N)$, by part 4 of Lemma 3, all IC constraints of θ_N-types are slack with respect to θ'-types, for $\theta' < \theta_N$. By part 5 of Lemma 3, then θ_N-types are bunched across β-levels.

2. Suppose $\lambda(\theta_n) = 0$. If θ_n-types were bunched across β-levels, which by part 2 of Lemma3 must occur weakly above the first-best savings rate, then we could allow (θ_n, β_1)-types to reduce their savings along their indifference curve, thereby maintaining constant welfare, preserving IC, but saving a strictly positive amount of resources—a contradiction. By an application of the maximum theorem (part 3 of Lemma 2), we know that for small enough $\lambda(\theta_n) > 0$ the desired result obtains.

□

A.3.9 Proof of Theorem 2

Proof. We proceed in steps:

1. The first statement follows from an application of part 2 of Lemma 3. For the second statement, by part 2 of Lemma 3, it suffices to show that some β-type, for $\beta < 1$, must have an IC constraint binding with respect to θ_1-types. If not, then θ_1-types would have $s(\theta_1) = s^{FB}$ and since β_M-types have $s(\theta, \beta_M) \geq s^{FB}$ (recalling that $\beta_M = 1$), this implies that $c_2(\theta, \beta_M) > c_2(\theta_1)$ for all θ. Since $\lambda(\theta_1) = \max_\theta \{\lambda(\theta)\}$, then the IC constraint of some (θ, β_M)-type, for some θ, must bind with respect to θ_1-types, or else no IC constraints would be binding with respect to θ_1-types and we could transfers resources from all agents towards θ_1-types. Then

IC from (θ, β_M)-types to θ_1-types together with the fact that $c_2(\theta, \beta_M) \geq c_1(\theta, \beta_M)$ implies that IC from all (θ, β)-types, for $\beta < 1$, is slack with respect to (θ, β_M)-types. If $\theta = \theta'$ for all $\theta, \theta' > \theta_1$, then we could transfer resources from all types towards θ_1-types to improve welfare—a contradiction. This proves that if θ_1-types are bunched and $\theta \approx \theta'$ for all $\theta, \theta' > \theta_1$, then $s(\theta_1) > s^{FB}$.

2. The first statement follows from part 2 of Lemma 3. The second statement follows from the fact that for $\lambda(\theta_n) = 0$ it is optimal to have (θ_n, β_1)-types save at or below their laissez-faire savings rate, which lies strictly below the first-best savings rate. If this were not the case, then the planner could allow (θ_n, β_1)-types to decrease their savings rate along β_1-types' indifference curve to maintain constant welfare, preserve IC, but save a strictly positive amount of resources—a contradiction. Therefore, $\underline{\lambda_n} = 0$ always delivers the desired result. If $\tau^D(\theta_n, \beta_1) > 0$ initially, then since the optimal allocation is continuous in $\lambda(\theta)$, by an application of the maximum theorem (part 3 of Lemma 2), we know that for small enough $\lambda(\theta_n) > 0$ the same result obtains.

3. Parts 1 and 1 of the theorem together already imply that $\tau^D(\theta_n, \beta_M) = \tau^E(\theta_n, \beta_M) \leq 0$. Suppose now that $\tau^D(\theta_n, \beta_M) = \tau^E(\theta_n, \beta_M) < 0$ but that no IC constraint binds from any (θ, β)-type with respect to type (θ_n, β_M) for $\beta < \beta_M = 1$. Then the planner could reduce (θ_n, β_M)-types' savings rate along their indifference curve while preserving welfare and IC but saving a strictly positive amount of resource—a contradiction.

□

A.3.10 Proof of Theorem 3

Proof. We proceed in steps:

1. (i) Suppose $\tau^L(\theta_1, \beta) = 1 - v'(y(\theta_1, \beta)/\theta_1) / [u'(c_1(\theta_1, \beta))\theta_1] < 0$ for some β, so (θ_1, β)-types' labor is implicitly subsidized. Some IC constraint must bind from θ-types, for $\theta > \theta_1$, to θ_1-types, or else distorting θ_1-types' labor would not be optimal. Therefore:

$$u(c_1(\theta, \beta')) - v(y(\theta, \beta')/\theta) + \beta'\delta u(c_2(\theta, \beta')) = u(c_1(\theta_1, \beta)) - v(y(\theta_1, \beta)/\theta) + \beta'\delta u(c_2(\theta_1, \beta))$$

for some β, β'. Then the planner could increase $\tau^L(\theta_1, \beta)$ toward zero by decreasing $y(\theta_1, \beta)$

and $c_1(\theta_1, \beta)$ in a way to hold fixed (θ_1, β)-types' period 1 utility, save resources, but leave slack the IC constraints from θ-types to θ_1-types. To see this, note that the change in utility due to a perturbation (dc_1, dy) that keeps period 1 utility constant for θ_1-types is

$$dc_1(\theta_1, \beta) u'(c_1(\theta_1, \beta)) - dy(\theta_1, \beta) v'(y(\theta_1, \beta)/\theta_1)/\theta_1 = 0$$

Rearranging, this yields

$$\underbrace{dc_1(\theta_1, \beta)}_{<0} = \underbrace{dy(\theta_1, \beta)}_{<0} \underbrace{\frac{v'(y(\theta_1, \beta)/\theta_1)}{u'(c_1(\theta_1, \beta))\theta_1}}_{>1},$$

which means that for $dc_1(\theta_1, \beta), dy(\theta_1, \beta) < 0$ we save resources: $dc_1(\theta_1, \beta) < dy(\theta_1, \beta) < 0$. What is the associated change in utility for other agents? Because all θ_1-types have the same preference over period 1 consumption and labor, their IC remains unaffected. For other θ-types, the perturbation results in the following change in utility:

$$\begin{aligned} & dc_1(\theta_1, \beta) u'(c_1(\theta_1, \beta)) - dy(\theta_1, \beta) v'(y(\theta_1, \beta)/\theta)/\theta \\ < & dc_1(\theta_1, \beta) u'(c_1(\theta_1, \beta)) - dy(\theta_1, \beta) v'(y(\theta_1, \beta)/\theta_1)/\theta_1 = 0 \\ & \implies dy(\theta_1, \beta) < 0, \end{aligned}$$

where the inequality follows from the fact that $dy(\theta_1, \beta) < 0$ and $\theta > \theta_1$. Hence, the utility change for θ-types is negative, so their IC constraint with respect to (θ_1, β)-types becomes slack. In summary, the perturbation has kept welfare constant, preserved IC, but saved resources—a contradiction. We conclude that $\tau^L(\theta_1, \beta) \geq 0$ for all β.
(ii) An exactly analogous argument shows that $\tau^L(\theta_N, \beta) \leq 0$ for all β.

2. We prove parts (i) and (ii) jointly. It is readily seen that a zero labor distortion is optimal when the only relevant IC constraints are within a set of agents that share the same θ. Suppose this were not the case, then the planner could undo the distortion and keep all θ-types' period 1 utility constant but save resources while preserving IC—a contradiction. The remaining argument is symmetric between θ_1-types and θ_N-types, so we will state it only for the former. Suppose that an IC constraint binds from some θ-type to some (θ_1, β)-type but $\tau(\theta_1, \beta) = 0$. Then the planner could decrease $y(\theta_1, \beta)$ and $c(\theta_1, \beta)$ by a small amount, keep

resources constant, reduce welfare of (θ_1, β)-types to a second order, but relax θ-types' IC constraint to a first order, thereby facilitating redistribution and improving welfare to a first order—a contradiction. Hence if an IC constraint binds from some θ-type to some (θ_1, β)-type, then $\tau(\theta_1, \beta) > 0$ is optimal. An analogous argument shows that if an IC constraint binds from some θ_1-type to some (θ, β)-type, then $\tau(\theta, \beta) < 0$ is optimal. Finally, that the conditions stated in the theorem are sufficient for the above argument to apply follows from part 2 of Lemma 3.

□

A.3.11 Proof of Corollary 1

Proof. We proceed in two steps:

1. Suppose that for some $\theta \in \Theta$ the IC constraints are slack with respect to all agents $\theta' \neq \theta$, but that there exist $\beta, \beta' \in B$ such that $(c_1(\theta, \beta), c_2(\theta, \beta), y(\theta, \beta)) \neq (c_1(\theta, \beta'), c_2(\theta, \beta'), y(\theta, \beta'))$. Then a direct application of part 5 of Lemma 3 in Appendix A.3.7, with summation signs replaced by integrals, proves the desired result.

2. The proof is exactly identical to that of part 2 of Theorem 1 in Appendix A.3.8.

□

A.4 Comparison of theoretical findings to most related results in the literature

It is instructive to relate our model to three influential results in the literature. First, the intermediate goods taxation result of Atkinson and Stiglitz (1972) implies that under nonlinear income taxation intertemporal consumption decisions are optimally undistorted even in the presence of a redistributive motive. Their result can be illustrated in our simple environment with $\beta_L = \beta_H = 1$ and $\theta_L = 0 < \theta_H$. The only relevant IC constraint is then $u(c_1(\theta_H)) - v(y(\theta_H))/\theta_H + \delta u(c_2(\theta_H)) \geq u(c_1(\theta_L)) + \delta u(c_2(\theta_L))$. Since agents agree with the planner on the intertemporal rate of substitution, we can rewrite the planner's problem in its dual form as a resource cost minimization problem for each ability type: $\min_{c_1, c_2} \{c_1 + c_2/R\}$ s.t. $u(c_1) + \delta u(c_2) = \overline{U}(\theta)$, where $\overline{U}(\theta)$ depends on the optimal transfers across ability types subject to IC. Taking first-order conditions, we get $u'(c_1(\theta)) = R\delta u'(c_1(\theta))$ and therefore $\tau^D(\theta) = \tau^E(\theta) = 0 \quad \forall \theta$. Hence, redistribution without paternalism leads to undistorted savings.

Second, Farhi and Werning (2010) show that with redistribution and a constant level of present bias the optimal efficiency wedge monotonically increases in ability. Their setup maps into our simple environment with $\beta_L = \beta_H = \beta < 1$ and $\theta \in \{\theta_L, \theta_H\}$. The relevant IC constraint is then $u(c_1(\theta_H)) - v(y(\theta_H))/\theta_H + \beta\delta u(c_2(\theta_H)) \geq u(c_1(\theta_L)) + \beta\delta u(c_2(\theta_L))$. Given $\lambda(\theta_L) \geq \lambda(\theta_H)$ and $0 = y(\theta_L) < y(\theta_H)$, the IC constraint must bind at the optimum, so more redistribution toward θ_L-types is desirable. The planner can improve upon the efficient savings rate by increasing θ_L-types' savings rate and decreasing θ_H-types' savings rate. Both perturbations incur a second-order welfare loss but facilitate redistribution by relaxing the relevant IC constraint, which leads to a first-order net welfare gain. Hence, at the optimum θ_L-types strictly oversave while θ_H-type agents strictly undersave: $\tau^E(\theta_L) < 0 < \tau^E(\theta_H)$. In summary, the interaction between redistribution and a constant level of present bias yields a savings wedge that is increasing in ability.

Third, Amador et al. (2006)'s model without redistribution but with heterogeneity in present bias is also relevant to our analysis. They demonstrate that the optimal policy in this framework takes the form of a minimum savings threshold, which leaves patient agents' savings undistorted. Although for a different reason—namely to facilitate redistribution—the optimal policy in our environment also entails greater dispersion in savings at higher ability levels. A distinguishing feature of our environment relative to theirs is that at low ability, bunching occurs above the first-best savings rate, while implied savings rates are differentially distorted at high ability.

Our model combines the two ingredients of redistribution and present bias heterogeneity. The forces in Atkinson and Stiglitz (1972), Farhi and Werning (2010), and Amador et al. (2006) are also present in our model and partially characterized by Theorems 1–2. As in Atkinson and Stiglitz (1972), the savings of agents with $\beta = 1$ are undistorted if their allocation is not envied by other agents with $\beta' < 1$. As in Farhi and Werning (2010), under the conditions stated in Theorem 2, savings of low-ability types are distorted strictly above the first-best, while at least some high-ability types save strictly below the first-best. And as in Amador et al. (2006), under the conditions stated in Theorem 1, low-ability types are optimally bunched regardless of their present bias level. The planner balances the redistributive motive with productive efficiency by picking optimal labor distortions as characterized in Theorem 3.

A.5 Proofs for the decentralization

A.5.1 Statement and proof of Proposition 8, leading up to Proposition 2

Our argument follows closely that in Werning (2011), extended to a setting with heterogeneous present bias levels. We first show that a working life tax function alone can be used to decentralize an incentive compatible allocation. Let $\{c_1(\theta_n, \beta_m), c_2(\theta_n, \beta_m), y(\theta_n, \beta_m)\}_{n=1,...,N;m=1,...,M}$ be the solution to the planner's problem.

Proposition 8. *Consider an optimal allocation $\{c_1(\theta_n, \beta_m), c_2(\theta_n, \beta_m), y(\theta_n, \beta_m)\}_{n=1,...,N;m=1,...,M}$ from the planner's problem. Then:*

1. *There exists a working life tax function T_1 that implements the optimal allocation as a competitive equilibrium with one retirement savings account, without the need for retirement benefits, and without retirement savings withdrawal taxes $(J = 1, T_2(y, Ra) = 0)$;*

2. *There exists a policy $(J = 1, T_1(y, a), b(y))$ that implements the optimal allocation as a competitive equilibrium with one retirement savings account, and without retirement savings withdrawal taxes.*

Proof. We proceed in steps:

1. We proceed in a sequence of three Lemmas to prove part 1 of Proposition 8.

   Lemma 4. *We assume $\beta > 0$. For any allocation $x^\star = (c_1^\star, c_2^\star, y^\star)$ with strictly positive consumption and for any type (θ, β), there exists an affine tax function*

   $$T_1^{(x^\star;\theta,\beta)}(y, a) = t_0 + t_y y + t_a a$$

   such that

   $$(c_1^\star, c_2^\star, y^\star) \in \arg\max_{c_1,c_2,y,a} u(c_1) - v\left(\frac{y}{\theta}\right) + \beta\delta u(c_2)$$
   $$s.t. \quad c_1 + \frac{c_2}{R} = y - T_1^{(x^\star;\theta,\beta)}\left(y, \frac{c_2}{R}\right)$$

   Proof. Fix a type (θ, β). Then consider the set

   $$\bar{U}(\theta, \beta) = \{(c_1, c_2, y) : U(c_1, c_2, y; \theta, \beta) \geq U(c_1^\star, c_2^\star, y^\star; \theta, \beta)\}$$

Since the utility function is strictly concave and differentiable, by the supporting hyperplane theorem there exists a unique vector $\left(\alpha_0, \alpha_1, \alpha_2, \alpha_y\right)$ such that

$$\alpha_0 + \alpha_1 c_1^\star + \alpha_2 c_2^\star + \alpha_y y^\star = 0$$

and

$$\alpha_0 + \alpha_1 c_1 + \alpha_2 c_2 + \alpha_y y > 0$$

for all $(c_1, c_2, y) \in \bar{U}(\theta, \beta) / (c_1^\star, c_2^\star, y^\star)$. Moreover, since $\beta > 0$, $u'(\cdot) > 0$ and $v'(\cdot) > 0$, we have $\alpha_1 > 0$, $\alpha_2 > 0$ and $\alpha_y > 0$. We can let $t_0 = -\frac{\alpha_0}{\alpha_1}$, $t_a = \frac{R\alpha_2}{\alpha_1} - 1$ and $t_y = \frac{\alpha_y}{\alpha_1} - 1$, then we can write the hyperplane found above equivalently as

$$c_1 + \frac{c_2}{R} = y - t_0 - t_y y - t_a \frac{c_2}{R}$$

and since the hyperplane is separating, any allocation with higher utility is outside this budget set. □

Lemma 5. *Assume $\beta_1 > 0$. Consider an incentive compatible allocation $\mathcal{A}$, and a fixed type (θ, β). Then there exists a piecewise linear tax function $T_1^{(\mathcal{A};\theta,\beta)}(y, a)$ such that if we denote the allocation for each type by*

$$x\left(\theta', \beta'\right) = \left(c_1\left(\theta', \beta'\right), c_2\left(\theta', \beta'\right), y\left(\theta', \beta'\right)\right)$$

for all (θ', β'). Then there exists a piecewise linear function $T_1^{(\theta,\beta)}(y, a)$ such that

$$\begin{aligned} x(\theta, \beta) \in \arg \max_{c_1, c_2, y, a} & \; u(c_1) - v\left(\frac{y}{\theta}\right) + \beta\delta u(c_2) \\ s.t. \quad & c_1 + \frac{c_2}{R} = y - T_1^{(\theta,\beta)}\left(y, \frac{c_2}{R}\right) \end{aligned}$$

and

$$\begin{aligned} U\left(x(\theta, \beta); \theta', \beta'\right) \geq \max_{c_1, c_2, y, a} & \; u(c_1) - v\left(\frac{y}{\theta'}\right) + \beta'\delta u(c_2) \\ s.t. \quad & c_1 + \frac{c_2}{R} = y - T_1^{(\theta,\beta)}\left(y, \frac{c_2}{R}\right) \end{aligned}$$

for all $(\theta', \beta') \neq (\theta, \beta)$. *Moreover we can write this function as*

$$T_1(y, a) = \max_{i=1,\ldots,N\times M} \left\{ t_i^0 + t_i^y y + t_i^a a \right\}$$

Proof. Fix the type (θ, β). Then from Lemma 4 we have that there exists a tax function defined by

$$t(\theta, \beta) = \left(t_0(\theta, \beta), t_y(\theta, \beta), t_a(\theta, \beta) \right)$$

such that the best choice for type (θ, β) on the associated budget set is

$$x(\theta, \beta) = \left(c_1(\theta, \beta), c_2(\theta, \beta), y(\theta, \beta) \right)$$

Now we want to find analogous tax functions for types $(\theta', \beta') \neq (\theta, \beta)$ that also contain $x(\theta, \beta)$ but do not contain any allocation with higher utility than $x(\theta', \beta')$ from the point of view of type (θ', β'). This is to say that type (θ', β') weakly prefers their own allocation to any allocation in this alternative budget set that contains $x(\theta, \beta)$.

Fix a type (θ', β'). There are two cases depending on whether the incentive constraint of (θ', β') to (θ, β) is binding or not. Let us first consider the case in which this incentive constraint is binding, that is we have

$$U\left(x(\theta, \beta); \theta', \beta'\right) = U\left(x\left(\theta', \beta'\right); \theta', \beta'\right)$$

In this case, from Lemma 4 there exists $t(\theta', \beta')$ with associated tax function that makes $x(\theta, \beta)$ the optimal choice within the respective budget set. Therefore, by transitivity agent (θ', β') weakly prefers $x(\theta', \beta')$ to any available choice in this budget set. Lastly let us consider the case of a strictly slack incentive constraint, that is

$$U\left(x(\theta, \beta); \theta', \beta'\right) < U\left(x\left(\theta', \beta'\right); \theta', \beta'\right)$$

There are two sub-cases in this scenario. On one hand, if under the tax function associated with $t(\theta, \beta)$ there is no point preferable to $x(\theta', \beta')$ to type (θ', β') then we can just set $t(\theta', \beta') = t(\theta, \beta)$. On the other hand, if there is such a point in the budget set associated with $t(\theta, \beta)$, then we have to choose a different tax function for type (θ', β'). From Lemma

4 there exists $\tilde{t}(\theta', \beta')$ such that in the budget set associated the optimal choice for a type (θ', β') agent is $x(\theta, \beta)$. We could use this tax function for type (θ', β'), but then the local behavior of taxes around $x(\theta, \beta)$ might be affected even though the incentive constraint is strictly slack. But given that the incentive constraint is strictly slack, we can reduce $\tilde{t}_0(\theta', \beta')$ up to a point in which the associated budget set has a point that makes agent (θ', β') indifferent between their own allocation and the best option in this budget set. Denote by $t(\theta', \beta')$ this respective tax function.

Notice that the budget sets induced by $t(\theta', \beta')$ and $t(\theta, \beta)$ are generally different, but all of them contain $x(\theta, \beta)$. Therefore their intersection contains $x(\theta, \beta)$ and there is no choice in their intersection that is better than the allocations designed for types (θ, β) or (θ', β'). But the tax function defining this intersection is given by

$$T_1(y, a) = \max_{(\theta'', \beta'')} \left\{ t_0(\theta'', \beta'') + t_y(\theta'', \beta'') y + t_a(\theta'', \beta'') a \right\}$$

which concludes the proof. □

Lemma 6. *Assume $\beta_1 > 0$. Then for any incentive compatible allocation $\mathcal{A}$, there exists a piecewise linear tax function of the form*

$$T_1(y, a) = \min_i \max_j \left\{ t_{i,j} + t^y_{i,j} y + t^a_{i,j} a \right\}$$

that implements this allocation as a competitive equilibrium with $b(y) = 0$ and $T_2(Ra) = 0$.

Proof. From Lemma 5 it follows that for each type (θ, β) there exists a function

$$\hat{T}_1^{(\theta,\beta)}(y, a) = \max_j \left\{ t_j + t^y_j y + t^a_j a \right\}$$

such that $x(\theta, \beta)$ belongs to the budget set associated with this tax function and no other type would strictly prefer to deviate to an allocation in this budget set. Therefore, if we take the union of these budget sets, then there is no other point in this union that agents would strictly prefer to the points determined by $\mathcal{A}$. But this union is defined by the tax function

$$T_1(y, a) = \min_{(\theta,\beta)} \hat{T}_1^{(\theta,\beta)}(y, a)$$

which concludes the proof. □

Together, Lemmata 4–6 prove the statement that forms part 1 of the proposition.

2. Recall from part 1 of Lemma 1 that $y(\theta_n, \beta_1)$ is weakly increasing in n. Define a strictly increasing sequence y_k for $k = 1, \ldots, K$ of the K unique values that $y(\theta_n, \beta_1)$ takes for $n = 1, \ldots, N$. Then for each k define

$$c_{2,k} = \min_n \left\{ c_2(\theta_n, \beta_1) \,\middle|\, y(\theta_n, \beta_1) = y_k \right\}$$

and

$$b_k(y) = c_{2,k-1} + \left(\frac{c_{2,k} - c_{2,k-1}}{y_k - y_{k-1}} \right) (y - y_{k-1})$$

so that $b_k(y_k) = c_{2,k}$ and $b_k(y_{k-1}) = c_{2,k-1}$.[47] With that definition, we can finally define $b(y)$ as

$$b(y) = \begin{cases} b_k(y) & y \in (y_{k-1}, y_k], k = 2, \ldots, K \\ b_1(y) & y \in [0, y_1] \\ b_K(y_K) & y > y_K \end{cases}$$

Recall that in our equilibrium definition we have $a_j \geq 0$, therefore we need to check that $c_2(\theta, \beta) \geq b(y(\theta, \beta))$ for all (θ, β). We can always add nodes to $b(y)$ to make sure this is true. In fact notice that if $c_2(\theta', \beta') < b(y(\theta', \beta'))$, then we can set $y(\theta', \beta')$ as another node j' with

$$c_{2,j'} = \min_{\theta,\beta} \left\{ c_2(\theta, \beta) \,\middle|\, y(\theta, \beta) = y_{j'} \right\}$$

Once we add this node we obtain $c_2(\theta', \beta') \geq b(y(\theta', \beta'))$. Since the number of types is finite, we can add nodes to $b(y)$ until $c_2(\theta, \beta) \geq b(y(\theta, \beta))$ holds for all (θ, β). From part (1), there exists $\tilde{T}_1(y, a)$ that implements the desired allocation as a competitive equilibrium with $b(y) = 0$. Define

$$T_1(y, a) = \tilde{T}_1 \left(y, a + \frac{b(y)}{R} \right) + \frac{b(y)}{R} \tag{9}$$

We need to show that the budget set under (T_1, b) is equivalent to the budget set under $\tilde{T}_1$.

[47] We set $c_{2,0} = 0$ and $y_0 = 0$ for the definition of $b_1(y)$.

A vector (c_1, c_2, y) is in the new budget set if only if there exists $a \geq 0$ such that

$$c_1 + a = y - T_1(y, a)$$
$$c_2 = b(y) + Ra$$

But since we can let $\tilde{a} = a + \frac{b(y)}{R}$, the above holds if and only if

$$c_1 + \tilde{a} = y - \tilde{T}_1(y, \tilde{a})$$
$$c_2 = R\tilde{a}$$

Notice that since $a \geq 0$ and $b(y) \geq 0$, we have $\tilde{a} \geq 0$. This concludes the proof of part (2).

□

A.5.2 Proof of Proposition 2

Proof. We build upon the implementation in part 2 of Proposition 8 to obtain a policy that implements the original allocation with the desired properties. Indeed, from part 2, there exists (T_1, b) that implements this allocation. Let us define

$$\tilde{a}(\theta_n, \beta_m) = \frac{c_2(\theta_n, \beta_m) - b(y(\theta_n, \beta_m))}{R}$$

This is how much an agent of type (θ_n, β_m) saves in this decentralization. Let us consider the decision wedge

$$t_{n,m} = 1 - \frac{u'(c_1(\theta_n, \beta_m))}{R\beta_m\delta u'(c_2(\theta_n, \beta_m))}$$

then let us define the intermediary object

$$\tau_{n,m}(y) = t_{n-1,m} + \left(\frac{t_{n,m} - t_{n-1,m}}{y_{n,m} - y_{n-1,m}}\right)(y - y_{n-1,m})$$

where $y_{n,m} = y(\theta_n, \beta_m)$ and $\tau_{1,m}(y) = t_{1,m}y/y_{1,m}$. Therefore, in a system with as many savings accounts as there are β-types, labeled $m = 1, \ldots, M$, we can write the marginal tax on withdrawals

from account m as

$$\tau_m(y) = \begin{cases} \tau_{1,m}(y) & y \leq y_{1,m} \\ \tau_{n,m}(y) & y \in (y_{n-1,m}, y_{n,m}], n = 2, \ldots, N \\ \tau_{N,m}(y_{N,m}) & y > y_{N,m} \end{cases}$$

Given this marginal tax on account m we can then construct the contribution limits recursively. For $m = 1$, we have second period consumption of type (θ_n, β_1) given by

$$c_2(\theta_n, \beta_1) = b(y(\theta_n, \beta_1)) + (1 - t_{n,1}) R\bar{a}_1(y_{n,1})$$

We can then define recursively, starting with $m = 1$:

$$\alpha_{n,1} = \frac{1}{R(1 - t_{n,1})} [c_2(\theta_n, \beta_1) - b(y_{n,1})]$$

Since by construction we have $c_2(\theta_n, \beta_1) \geq b(y_{n,1})$, then $\alpha_{n,1} \geq 0$. So that for $y \in (y_{n-1,1}, y_{n,1}]$ we have the linear piece

$$\bar{a}_{n,1}(y) = \alpha_{n-1,1} + \left(\frac{\alpha_{n,1} - \alpha_{n-1,1}}{y_{n,1} - y_{n-1,1}} \right) (y - y_{n-1,1})$$

where $\bar{a}_{1,1}(y) = \alpha_{1,1} y / y_{1,1}$. We can then combine these into a piecewise linear function giving account contribution limits:This completes the description of savings limits for account $m = 1$. Now assume that the account limits are defined for all accounts lower than index m. For this m, we have second period consumption of type (θ_n, β_m) given by

$$c_2(\theta_n, \beta_m) = b(y(\theta_n, \beta_m)) + (1 - t_{n,m}) R\tilde{\alpha}_{n,m} + \sum_{k=1}^{m-1} (1 - \tau_k(y_{n,m})) R\bar{a}_k(y_{n,m})$$

Then for account m, we can define

$$\tilde{\alpha}_{n,m} = \frac{c_2(\theta_n, \beta_m) - b(y(\theta_n, \beta_m)) - \sum_{k=1}^{m-1} (1 - \tau_k(y_{n,m})) R\bar{a}_k(y_{n,m})}{(1 - t_{n,m}) R}$$

If type (θ_n, β_m) does not save additional funds in account m at the solution to problem, then $\tilde{\alpha}_{n,m} \leq 0$ and the account limit will be set to zero, and to a positive level otherwise: $\alpha_{n,m} = \max\{\tilde{\alpha}_{n,m}, 0\} \geq$

0. Then we can define the segments of the account limits as

$$\bar{a}_{n,m}(y) = \alpha_{n-1,m} + \left(\frac{\alpha_{n,m} - \alpha_{n-1,m}}{y_{n,m} - y_{n-1,m}}\right)(y - y_{n-1,m})$$

We end up with an account contribution limit function that is piecewise linear in income:

$$\bar{a}_m(y) = \begin{cases} \bar{a}_{1,m}(y) & y \leq y_{1,m} \\ \bar{a}_{n,m}(y) & y \in (y_{n-1,m}, y_{n,m}], n = 1, \ldots, N \\ \bar{a}_{N,m}(y_{N,m}) & y > y_{N,m} \end{cases}$$

This completes the construction of savings account caps and taxes. We now need to construct the tax function during working life that makes the budget set with savings accounts and marginal distribution taxes equivalent to the budget set available in part 2. Note that in part 2 we have

$$c_2 = b(y) + R\tilde{a}$$

Therefore for the same c_2 and $b(y)$ we must have

$$\tilde{a} = \sum_{m=1}^{M(y,\tilde{a})} (1 - \tau_m(y))\, a_m$$

where

$$M(y, \tilde{a}) = \min\left\{K : \sum_{m=1}^{M(y,\tilde{a})} (1 - \tau_m(y))\, \bar{a}_m(y) \geq \tilde{a}\right\},$$

imposing that agents use accounts $m' = 1, \ldots, m-1$ up to their respective caps whenever they save positive amounts in account m. Given contributions have to satisfy the account order, we can solve for a_m as a function of $\tilde{a}$ and y

$$a_m(y, \tilde{a}) = \begin{cases} \bar{a}_m(y) & \text{if } m < M(y, \tilde{a}) \\ \frac{1}{1-\tau_m(y)}\left[\tilde{a} - \sum_{k=1}^{M(y,\tilde{a})-1} (1 - \tau_k(y))\, \bar{a}_k(y)\right] & \text{if } m = M(y, \tilde{a}) \\ 0 & \text{if } m > M(y, \tilde{a}) \end{cases}$$

The essential thing to show is that this $a_m(y, \tilde{a})$ yields a bijective relation between consumption in periods 1 and 2 in the decentralized economy compared to that in the planner's solution. We

show that a period 1 tax schedule $T_1\left(y,(a_m)_{m=1}^M\right)$ exists that achieves this. By construction we have $a_m(y,\tilde{a}) \geq 0$. From part 2 we also have

$$c_1 + \tilde{a} = y - \tilde{T}_1(y,\tilde{a})$$

where $\tilde{T}_1(y,\tilde{a})$ is the period 1 tax as a function of income and savings from equation 9 in part 2 of Proposition 8. Since $\tilde{a} = \sum_{m=1}^M (1-\tau_m(y))\,a_m$ from above, we have

$$c_1 + \sum_{m=1}^M (1-\tau_m(y))\,a_m = y - \tilde{T}_1\left(y, \sum_{m=1}^M (1-\tau_m(y))\,a_m\right)$$

So that

$$c_1 + \sum_{m=1}^M a_m = y - T_1\left(y,(a_m)_{m=1}^M\right)$$

where

$$T_1\left(y,(a_m)_{m=1}^M\right) = \tilde{T}_1\left(y, \sum_{m=1}^M (1-\tau_m(y))\,a_m\right) - \sum_{m=1}^M \tau_m(y)\,a_m \tag{10}$$

Finally, we define T_2 by

$$T_2\left(y,\{a_m\}_{m=1}^M\right) = \sum_{m=1}^M \tau_m(y)\,Ra_m \tag{11}$$

Then with period 1 tax function given by equation (10) and period 2 savings taxes aggregated across accounts given by equation (11), we obtain the identical budget set as in part 2 of Proposition 8. Therefore, this set of policies implements the optimal allocation $\mathcal{A}^*$. □

A.5.3 Proof of Corollary 2

Proof. That a system of retirement savings policies decentralizes the optimal allocation follows directly from Proposition 2. Under the stated conditions, it follows that θ_1-types are bunched, while θ_n-types are separated across β-levels (Theorem 1). Since bunched θ_1-types all have the same income y, they receive the same old-age benefits $b(y)$. Therefore, their allocation is independent of β and there is no need or desire for them to use any of the voluntary savings accounts. Similarly, since separated θ_n-types have differentially distorted savings margins (Theorem 2), then the old-age benefits schedule $b(y)$ by itself is insufficient to implement the optimum. Therefore, agents that share the same ability level θ_n must use different voluntary retirement savings accounts. □

A.6 Multi-period life-cycle model

A.6.1 Model setup

In this section, we present a multi-period life-cycle model with stochastic earnings ability and self-control shocks. We characterize the efficient dynamic provision of insurance and commitment in this environment. Extending our results from the 2-period model, we show that a trade-off between providing insurance and providing commitment arises for agents who experience high income shocks, but not for agents with low income shocks. As a result, commitment is optimally provided only at low income levels.

In the following setup, we assume hyperbolic preferences shocks over the life cycle. While related work uses off-equilibrium path allocations to separate different degrees of time-inconsistency (Esteban and Miyagawa, 2004; Galperti, 2015; Yu, 2016), we effectively sidestep these intricacies by introducing stochastic time inconsistency levels. While studying a model with constant present bias is of great theoretical interest, our setup simplifies the analysis significantly and has two further advantages. First, our setup allows for changes in individuals' present bias over the life cycle, such as myopia that decreases with age. Second, our setup is robust to small stochastic perturbations in the hyperbolic discount factor, which the other setup abstracts from in order to generate perfectly persistent private information.

The economy is composed of a measure one of agents whose life cycle consists of $T \geq 3$ periods, divided into T_w periods of working life and $T - T_w$ periods of retirement.[48] At each $t = 1, \ldots, T_w$, agents face an earnings ability shock $\theta_t \in \Theta = \{\theta_1, \ldots, \theta_N\}$, where $\theta_1 < \ldots < \theta_N$, with transition probabilities $\rho_{t+1}(\theta_{t+1}|\theta_t)$. We allow transition probabilities to vary over the life-cycle and and assume full support over Θ at all t and for all $\theta_t \in \Theta$. We also assume that ρ_{t+1} is stochastically ordered so that higher levels of θ_t imply a distribution that first order stochastically dominates a distribution for lower levels of θ_t. With a slight abuse of notation, we denote by $\rho_1(\theta_1)$ the probability distribution over the initial earnings ability θ_1 and assume that it also has full support.

Furthermore, At each period $t = 1, \ldots, T-1$ each agent faces a hyperbolic self-control shock $\beta_t \in B = \{\beta_1, \ldots, \beta_M\}$, where $\beta_1 < \ldots < \beta_M$, which we assume to be independently distributed both over time and from earnings ability shocks. We allow the probability distribution of self-

[48] We implicitly assume that retirement lasts for at least one period.

control shocks at period t, denoted $\gamma_t(\beta_t)$, to vary over the life-cycle as long as there is full support.

We denote an agent's joint type by $h_t = (\beta_t, \theta_t) \in H_t$ and its distribution at time t by π_t. We mark by superscript t the history of types realized until period t, so that $h^t = (h_1, \ldots, h_t) \in H^t$. We let π_t denote the probability distribution over H^t.

The period payoff during working life periods $t = 1, \ldots, T_w$ over consumption and obtained earnings is given by $u_W(c_t, y_t; h_t) = u(c_t) - v(y_t)/\theta_t$, where we assume $u' > 0$, $u'' < 0$, $v' > 0$, $v'(0) = 0$, $v'' > 0$, and $v(0) = 0$. During retirement periods $t = T_w + 1, \ldots, T$, the agent is retired and consumes without working ($y_t = 0$), with period payoff given by $u_R(c_t) = u(c_t)$. The generalized period payoff function is then

$$u_t(c_t, y_t; h_t) = \begin{cases} u_W(c_t, y_t; h_t) & \text{for } t \leq T_w \\ u_R(c_t) & \text{for } t > T_w \end{cases}$$

A planner cannot directly observe agents' types but designs an incentive compatible and feasible mechanism that maximizes social welfare. As previously, we apply the Revelation Principle to characterize implementable allocations in this environment.[49] In this environment, an allocation can be written as a sequence of functions $(c_t, y_t) : H^t \to \mathbb{R}^2_+$ for each t. We define an *allocation* as $\mathcal{A} = (c, y)$, where c and y denote the entire set of history-dependent consumption and labor allocations. The planner evaluates welfare according to the period 0 preferences, or *experienced utility*, of agents in the economy:

$$W_t(c, y) = \sum_{s=1}^{T} \delta^{s-1} \sum_{h^s} \pi_s(h^s) u_s(c_s(h^s), y_s(h^s); h_s) \tag{12}$$

Following a large strand in the behavioral public finance literature, we interpret this as the problem of an agent at period 0 seeking the optimal level of insurance for earnings ability shocks and a commitment device for self-control shocks over the life-cycle. Therefore, the efficient allocation could be implemented either by the government or by competitive private insurance companies, as long as both are able to enforce the contract.

Agents, once they reach the decision stage, have a present-biased evaluation of life-time utility,

[49] We show below that it suffices to consider mechanisms in which at each period agent report their current type instead of their whole history of types. This result follows from our assumption that hyperbolic preference shocks are independent over time, and differs from the approach taken in Galperti (2015) and Yu (2016).

$U_t(c, y; h_\tau)$, according to:

$$u_t\left(c_t\left(h^t\right), y_t\left(h^t\right); \theta_t\right) + \beta_\tau \sum_{s=t+1}^{T} \delta^{s-t} \sum_{h^s} \pi_s\left(h^s\right) u_s\left(c_s\left(h^s\right), y_s\left(h^s\right); \theta_s\right)$$

where we assume agents to be sophisticated in that they expect their future selves to be subject to some degree of present bias. Hence, there is dynamic disagreement between different period selves of the same (β, θ)-type as in Laibson (1997). A contract satisfies IC at time t if

$$h_t = \arg\max_{h'_t} U_t\left(c, y; h'_t\right) \tag{13}$$

A contract is *feasible* at time t if

$$\sum_{s=t}^{T} \frac{1}{R^{s-1}} \sum_{h^s} \pi\left(h^s\right)\left[y_s\left(h^s\right) - c_s\left(h^s\right)\right] \geq 0 \tag{14}$$

An allocation is *implementable* if $\forall t$ it satisfies IC (13) and feasibility (14).

A.6.2 General results

We now characterize the planner's problem solution, which provides efficient insurance against earnings ability shocks and self-control shocks.

Bunching and separation. Our first result shows that in the dynamic economy full commitment is provided only to parts of the population.

Theorem 4. *Fix* $\{\theta_2, \ldots, \theta_{N-1}\}$ *and* $\{\beta_2, \ldots, \beta_M\}$. *Then there exist* $\underline{\theta} > 0$, $\overline{\theta} < +\infty$, *and* $\underline{\beta} > 0$ *such that at the solution to the planner's problem:*

1. *If* $\theta_1 < \underline{\theta}$, *then for any* $t = 1, \ldots, T-1$ *and history* h^{t-1} *agents with types* $\left\{\left(h^{t-1}, (\theta_1, \beta)\right) : \beta \in B\right\}$ *are all assigned the same level of consumption and earnings in period* t *and are assigned the same*

continuation allocation for all future periods:

$$\begin{aligned} c_t\left(h^{t-1},(\theta_1,\beta)\right) &= c_t\left(h^{t-1},(\theta_1,\beta')\right) \\ y_t\left(h^{t-1},(\theta_1,\beta)\right) &= y_t\left(h^{t-1},(\theta_1,\beta')\right) \\ c_{t+s}\left(h^{t-1},(\theta_1,\beta),(h_{t+1},\ldots,h_{t+s})\right) &= c_{t+s}\left(h^{t-1},(\theta_1,\beta'),(h_{t+1},\ldots,h_{t+s})\right) \\ y_{t+s}\left(h^{t-1},(\theta_1,\beta),(h_{t+1},\ldots,h_{t+s})\right) &= y_{t+s}\left(h^{t-1},(\theta_1,\beta'),(h_{t+1},\ldots,h_{t+s})\right) \end{aligned}$$

for all $\beta,\beta' \in B$ and for all $s \geq 1$;

2. *If $\theta_N > \overline{\theta}$ and $\beta_1 \leq \underline{\beta}$, then for any $t = 1,\ldots,T-1$ and history h^{t-1} not all agents with types $\left\{\left(h^{t-1},(\theta_N,\beta)\right) : \beta \in B\right\}$ are assigned the same current allocation and continuation allocations.*

Proof. See Appendix A.6.5. □

The planner values insurance against both earnings ability shocks and self-control shocks. Theorem 4 shows that it is efficient to provide perfect commitment at low earnings but not at high earnings. This result is due to the interaction between the planner's two motives. Agents value flexibility after the realization of a self-control shock, demanding more immediate gratification than their prior selves' plans. Without an insurance motive, the planner would provide no such flexibility and instead provide commitment to all agents.[50] However, the planner also pursues the motive of consumption insurance, which in the presence of asymmetric information will be imperfectly provided. Therefore, the planner can charge high-ability agents for flexibility and use the proceeds to improve insurance against labor earnings shocks accrued to lower-ability agents. At low earnings, such a trade is feasible but not optimal since low-ability agents are unable to compensate the planner for the welfare loss associated with flexibility.

Optimal savings distortions. Our second result characterizes the distortions of time-inconsistent agents in this dynamic environment. A natural measure of distortions is the wedge relative to a path of time-consistent intertemporal consumption decisions. Without self-control shocks ($\beta = 1$),

[50] For example, this is the case when $\theta_t = \theta_0$ for all agents in the economy at all histories.

efficient insurance implies that intertemporal choices satisfy an *inverse Euler equation*:[51]

$$\sum_{\theta_{t+1}\in\Theta_{t+1}} \rho_{t+1}\left(\theta_{t+1}|\theta_t\right) \frac{u'\left(c_t\left(\theta^t\right)\right)}{\delta R u'\left(c_{t+1}\left(\theta^t,\theta_{t+1}\right)\right)} = 1$$

Whenever this intertemporal condition holds, the detrimental effects of time-inconsistency have been completely dealt with. We can define the *time inconsistency wedge* in our economy for agents with history h^t as

$$\tau\left(h^t\right) = \sum_{h_{t+1}\in H_{t+1}} \pi_{t+1}\left(h_{t+1}|h^t\right) \frac{u'\left(c_t\left(h^t\right)\right)}{\delta R u'\left(c_{t+1}\left(h^t,h_{t+1}\right)\right)} - 1$$

If agents face self-control problems when left on their own absent commitment devices, this would be represented as a negative time consistency wedge. Our second main result extends our characterization of savings wedges to our dynamic economy.

Theorem 5. *Fix* $\{\theta_2,\ldots,\theta_{N-1}\}$ *and* $\{\beta_2,\ldots,\beta_M\}$. *Then there exist scalars* $\underline{\theta} > 0$, $\overline{\theta} < +\infty$, *and* $\underline{\beta} > 0$ *such that at the solution to the planner's problem:*

1. *If* $\theta_1 \leq \underline{\theta}$, *then for any* $t = 1,\ldots,T-1$ *and history* h^{t-1} : $\tau\left(h^{t-1},(\theta_1,\beta)\right) \geq 0$ *for all* $\beta \in B$;

2. *If* $\theta_N > \overline{\theta}$ *and* $\beta_1 \leq \underline{\beta}$, *then for any* $t = 1,\ldots,T-1$ *and history* h^{t-1}:

   - $\tau\left(h^{t-1},(\theta_N,\beta_M)\right) \geq 0$;
   - $\tau\left(h^{t-1},(\theta_N,\beta_1)\right) < 0$.

Proof. See Appendix A.6.6. □

Savings distortions optimally vary throughout the income distribution. For low enough productivity types, the planner fully undoes low-ability types' self-control problem, and at times may induce savings above the first-best rate ($\tau\left(h^t\right) = 0$) as a screening device. On the other hand, high-ability types are differentially distorted, with the most patient agents ($\beta_M = 1$) weakly oversaving, but the lowest ability types strictly under-saving relative to the efficient level. Thus, not only does the planner provide imperfect commitment at higher ability levels, but it is also optimal to offer greater choice in savings for this part of the population.

[51] For applications in the context of optimal taxation see Rogerson (1985), Golosov et al. (2003), Golosov and Tsyvinski (2006), Farhi and Werning (2012), Farhi and Werning (2013b), Stantcheva (2015), and Golosov et al. (2016).

A.6.3 Proofs of general results in multi-period life-cycle model

Problem reformulation. Types are unobservable and we rely on the Revelation Principle to characterize implementable allocations. To this end, we define an allocation as a pair of functions $(c_t, y_t) : H^1 \times \cdots \times H^{t-1} \times H^t \to \mathbb{R}^2_+$ for each period t that assigns a consumption level and an earnings level for any reported history $h^t \in H^t$ at period t and any past reported history $\hat{r}^{t-1} = \left(h^1, \ldots, h^{t-1}\right) \in H^1 \times \cdots \times H^{t-1}$. A strategy for an agent is a sequence of reporting strategies $\sigma_t : H^1 \times \cdots \times H^{t-1} \times H^t \to H^t$. The overall payoff after history h^t, previous reports $\hat{r}^{t-1} = \left(r^1, \ldots, r^{t-1}\right) \in H^1 \times \cdots \times H^{t-1}$ and following a strategy $(\sigma_s)_{s=t}^T$ from period t on is $U_t\left(\hat{r}^{t-1}, h^t, (\sigma_s)_{s=t}^T\right)$, given by:

$$u\left(c_t\left(\hat{r}^{t-1}, \sigma_t\left(\hat{r}^{t-1}, h^t\right)\right)\right) - \frac{v\left(y_t\left(\hat{r}^{t-1}, \sigma_t\left(\hat{r}^{t-1}, h^t\right)\right)\right)}{\theta_t}$$
$$+ \beta_t \sum_{s=t+1}^{T} \delta^{s-t} \sum_{h^s \succ h^t} \pi_s\left(h^s | \theta_t\right) \left[u\left(c_s\left(\sigma_s\left(\hat{r}^{s-1}, h^s\right)\right)\right) - \frac{v\left(y_s\left(\sigma_s\left(\hat{r}^{s-1}, h^s\right)\right)\right)}{\theta_s}\right]$$

Note that preferences are hyperbolic with quasi-geometric discount factor β_t in period t.[52]

We assume that agents are sophisticated in that they take into account their present bias problems in the future. Define the truth-telling strategy as $\sigma_t^{Truth}\left(\hat{r}^{t-1}, h^t\right) = h^t$. An allocation satisfies IC if truth-telling is a sub-game perfect equilibrium of the game played between the selves in different periods, so that after any history of reports $\hat{r}^{t-1} \in H^1 \times \cdots \times H^{t-1}$ and any realized type h^{t-1} truth-telling is the optimal one-shot deviation:

$$\sigma_t^{Truth} \in \arg\max_{\sigma_t'} U_t\left(\hat{r}^{t-1}, h^t, \left(\sigma_t', \left(\sigma_s^{Truth}\right)_{s=t+1}^T\right)\right)$$

Taking into account that future selves will consider it optimal to report the truth, reporting the truth in period t after history h^t is optimal given any reports history $\hat{r}^{t-1}$. Since this is a Bayesian game with positive probabilities at all nodes of the game, the Revelation Principle guarantees that the outcome of any mechanism can be obtained using the allocations defined above.

Our assumptions of full support over types, the Markovian nature of the stochastic process over types and the planner's objective allow us to further simplify IC constraints in this environment. The Markovian structure implies that, conditional on $\hat{r}^{t-1}$, the preferences after any history

[52] We denote by $h^s \succ h^t$ the continuation histories at times $s > t$ that are consistent with h^t.

$\tilde{h}^t \in H^t$ with $h_t = \tilde{h}_t$ have the same ordering as the preferences after history h^t. As we will show below, the planner's objective function is strictly concave, which implies that the optimal allocation in period t treats agents of type $\tilde{h}^t$ and h^t identically. Hence we can write

$$c_{t+s}\left(\hat{r}^{t-1}, h^t, \ldots, h^s\right) = c_{t+s}\left(\hat{r}^{t-1}, h_t, \ldots, h^s\right)$$
$$y_{t+s}\left(\hat{r}^{t-1}, h^t, \ldots, h^s\right) = y_{t+s}\left(\hat{r}^{t-1}, h_t, \ldots, h^s\right)$$

Using this argument recursively for all periods $s > t$ we obtain

$$c_{t+s}\left(\hat{r}^{t-1}, h^t, \ldots, h^s\right) = c_{t+s}\left(\hat{r}_1, \ldots, \hat{r}_{t-1}, h_t, \ldots, h_s\right)$$
$$y_{t+s}\left(\hat{r}^{t-1}, h^t, \ldots, h^s\right) = y_{t+s}\left(\hat{r}_1, \ldots, \hat{r}_{t-1}, h_t, \ldots, h_s\right)$$

where we used that $\hat{r}_1, \ldots, \hat{r}_t$ are optimal reports for an agent with that history of types. Therefore it is without loss of generality that the mechanism requires only reporting of the current period type and not of the full history of types.[53]

From here onward, we denote by $\left(u_t\left(h^t\right), v_t\left(h^t\right)\right)$ the intra-period allocation in utility space.

A.6.4 Precursory results

The following result provides a useful bound on deviation utilities in the dynamic economy.

Lemma 7. *Given $\{\theta_2, \ldots, \theta_N\}$, there is $\underline{\theta} > 0$ such that if $\theta_1 < \underline{\theta}$ then at the solution to the planner's problem we have*

$$\underbrace{U_t\left(h^{t-1}, (\beta, \theta_1)\right)}_{\text{truthful report}} \geq \underbrace{U_t\left((\beta', \theta_1)\,|h^{t-1}, (\beta, \theta_1)\right)}_{\text{deviation in } \beta} > \underbrace{U_t\left((\beta', \theta')\,|h^{t-1}, (\beta, \theta_1)\right)}_{\text{deviation in } \theta}$$

and

$$v_t\left(h^{t-1}, (\beta', \theta')\right) > v_t\left(h^{t-1}, (\beta, \theta_1)\right)$$

for all $\theta' > \theta_1$, for all $\beta' \neq \beta$, for all h^{t-1}, and for all $t = 1, \ldots, T$.

[53] This characterization implies that only equilibrium path allocations are important for IC (Fernandes and Phelan, 2000; Kapička, 2013). This argument can break down in problems with perfectly correlated types, demonstrated by an example in Battaglini and Lamba (2018). The assumption of full support of β_t for all t and histories is crucial for this characterization to be valid. If there is no full support in β_t, then it is possible to design a mechanism in which off-equilibrium path allocations relax incentive constraints on the equilibrium path, as in Galperti (2015) and Yu (2016).

Proof. The first, weak inequality always holds by IC. The following argument justifies the second, strict inequality. If $\theta_1 = 0$, then $y_t\left(h^{t-1}, (\beta, \theta_1)\right) = 0$ for all $\beta \in B$ and all h^{t-1}. For any $\theta' > 0$, by zero marginal disutility around zero labor supply we have $y_t\left(h^{t-1}, (\beta', \theta')\right) > 0$, therefore

$$U_t\left((\beta', \theta') \,|h^{t-1}, (\beta, \theta_1)\right) = -\infty$$

which proves the second, strict inequality. By an application of the maximum theorem (part 3 of Lemma 2), following an argument essentially identical to that in part 3 of Lemma 3, for fixed $\{\theta_2, \theta_3, \ldots, \theta_N\}$ there exists $\underline{\theta} > 0$ such that for all $\theta_1 \leq \underline{\theta}$ the desired inequality holds. □

Lemma 8. *Given $\{\theta_1, \ldots, \theta_{N-1}\}$, there exists $\overline{\theta} < +\infty$ such that if $\theta_N > \overline{\theta}$ then at the solution to the planner's problem we have*

$$U_t\left(h^{t-1}, (\beta', \theta')\right) > U_t\left((\beta, \theta_N) \,|h^{t-1}, (\beta', \theta')\right)$$

and

$$v_t\left(h^{t-1}, (\beta, \theta_N)\right) > v_t\left(h^{t-1}, (\beta', \theta')\right)$$

for all h^{t-1} , for all $\theta' \in \{\theta_1, \ldots, \theta_{N-1}\}$, and for all $\beta, \beta' \in B$.

Proof. The proof is analogous to that of Lemma 7 but for the highest-productivity level relative to any lower productivity type. Again, the proof relies on an application of the maximum theorem (part 3 of Lemma 2), by which the planner's solution is continuous in $\theta > 0$. □

Lemma 9. *Given $\{\theta_1, \ldots, \theta_{N-1}\}$ and $\{\beta_2, \ldots, \beta_M\}$, there exists $\overline{\theta} < +\infty$ and $\underline{\beta} > 0$ such that if $\theta_N > \overline{\theta}$ and $\beta_1 < \underline{\beta}$ then at the solution to the planner's problem we have $u_t\left(h^{t-1}, (\beta_1, \theta_N)\right) > u_t\left(h^{t-1}, (\beta, \theta)\right)$ for all h^{t-1}, for all $\theta \in \{\theta_1, \ldots, \theta_{N-1}\}$, for all $\beta \in B$ and for all $t = 1, \ldots, T-1$.*

Proof. From Lemma 8, we know that $v_t\left(h^{t-1}, (\beta_1, \theta_N)\right) > v_t\left(h^{t-1}, (\beta, \theta)\right)$. Note that if $\beta_1 = 0$, IC requires $u_t\left(h^{t-1}, (\beta_1, \theta_N)\right) > u_t\left(h^{t-1}, (\beta, \theta)\right)$. By an application of the maximum theorem (part 3 of Lemma 2), the solution to this problem is continuous in $\beta \in [0, 1]$, so there exists $\underline{\beta}_t > 0$ such that this inequality remains strict for all $\beta_1 < \underline{\beta}_t$. Since $T < \infty$ we can pick a uniform level of $\underline{\beta} = \min_t\left\{\underline{\beta}_t\right\} > 0$ that satisfies the desired property. □

A.6.5 Proof of Theorem 4

Part 1.

Proof. Consider the problem in terms of utility levels from consumption and disutility levels from working. Assume by way of contradiction that for a fixed $t < T$ and fixed history $h^{t-1} \in H^{t-1}$ and for $\beta, \beta' \in B_t$ the solution to the planner's problem features

$$u_t\left(h^{t-1}, (\beta, \theta_1)\right) > u_t\left(h^{t-1}, (\beta', \theta_1)\right)$$

Consider a new allocation that is a convex combination between between $\left(h^{t-1}, (\beta^*, \theta_1)\right)$-types' allocations for all $\beta^* \in B$ and that is offered after history h^{t-1}:

$$\tilde{u}_t\left(h^{t-1}, (\beta^*, \theta_1), h_{t+1}, \ldots, h_T\right) = \sum_{b \in B} \frac{\pi_t\left((b, \theta_1) \,|h^{t-1}\right)}{\sum_{b' \in B} \pi_t\left((b', \theta_1) \,|h^{t-1}\right)} u_t(\cdot)$$
$$\tilde{v}_t\left(h^{t-1}, (\beta^*, \theta_1), h_{t+1}, \ldots, h_T\right) = \sum_{b \in B} \frac{\pi_t\left((b, \theta_1) \,|h^{t-1}\right)}{\sum_{b' \in B} \pi_t\left((b', \theta_1) \,|h^{t-1}\right)} v_t(\cdot)$$

By Lemma (7), there exists $\underline{\theta} > 0$ such that for $\theta_1 < \underline{\theta}$ we have

$$U_t\left(h^{t-1}, (\beta, \theta_1)\right) \geq U_t\left((b, \theta_1) \,|h^{t-1}, (\beta, \theta_1)\right) > U_t\left((\beta'', \theta') \,|h^{t-1}, (\beta, \theta_1)\right)$$

for all $\theta' > \theta_1$ and for all $\beta'' \in B$. Therefore, IC constraints at nodes $\left(h^{t-1}, (b, \theta_1)\right)$ are satisfied for all $b \in B$. From linearity of the objective function, IC constraints of $\left(h^{t-1}, (b, \theta)\right)$-types are also satisfied for all $\theta > \theta_1$ and all $b \in B$. Therefore this perturbation preserves IC at period t. Furthermore, from the planner's point of view the continuation utility at h^{t-1} remains unchanged under this perturbation. Therefore, welfare is unchanged, while for agents with hyperbolic preferences all IC constraints for period $s \leq t-1$ are satisfied. For histories $h^s \succ h^{t-1}$ for $s > t$, taking a convex combination leaves incentives unchanged because the objective is linear. But $C(u) = u^{-1}(u)$ is strictly convex, so for $u_t\left(h^{t-1}, (\beta, \theta_1)\right) > u_t\left(h^{t-1}, (\beta', \theta_1)\right)$ and $\pi_t(\cdot)$ having full support this perturbation saves a strictly positive amount of resources—a contradiction. Hence, agents with types $\left\{\left(h^{t-1}, (\theta_1, \beta)\right) : \beta \in B\right\}$ are bunched. □

Part 2.

Proof. Assume by way of contradiction that for some h^{t-1} we have that $\left(h^{t-1}, (\beta_t, \theta_N)\right)$-types for

all $\beta_t \in B_t$ share the same allocation. Then

$$\mathbb{E}_t\left[\frac{C'\left(u_{t+1}\left(h^{t-1},\left(\beta_t,\theta_N\right),\left(\beta_{t+1},\theta_{t+1}\right)\right)\right)}{\delta R_t C'\left(u_t\left(h^{t-1},\left(\beta_t,\theta_N\right)\right)\right)}|\theta_N\right]=\kappa$$

for some constant $\kappa > 0$. Recalling that $\beta_M = 1$, consider the following perturbation for the allocation of type (β_M, θ_N):

$$\begin{aligned}\tilde{u}_t\left(h^{t-1},\left(\beta_M,\theta_N\right)\right)&=u_t\left(h^{t-1},\left(\beta_M,\theta_N\right)\right)-\varepsilon\\ \tilde{u}_{t+1}\left(h^{t-1},\left(\beta_m,\theta_n\right),\left(\beta_{t+1},\theta_{t+1}\right)\right)&=u_{t+1}\left(h^{t-1},\left(\beta_M,\theta_N\right),\left(\beta_{t+1},\theta_{t+1}\right)\right)+\frac{1}{\delta}\varepsilon\end{aligned}$$

for $\varepsilon > 0$. Welfare of type $\left(h^{t-1},\left(\beta_M,\theta_N\right)\right)$ is kept constant by such a change. Types $\left(h^{t-1},\left(\beta_j,\theta_N\right)\right)$ for $\beta_j < 1$ dislike this perturbation, so it preserves IC. The marginal resource cost dE is

$$\begin{aligned}&-C'\left(u_t\left(h^{t-1},\left(\beta_M,\theta_N\right)\right)\right)\varepsilon+\frac{\mathbb{E}_t\left[C'\left(u_{t+1}\left(h^{t-1},\left(\beta_M,\theta_N\right),\left(\beta_{t+1},\theta_{t+1}\right)\right)\right)|\theta_N\right]}{R_t\delta}\varepsilon\\ &=\left(\kappa-1\right)C'\left(u_t\left(h^{t-1},\left(\beta_M,\theta_N\right)\right)\right)\varepsilon\end{aligned}$$

For the original allocation to be optimal, we require $dE \geq 0$ and thus $\kappa \geq 1$.

Suppose further that $\kappa > 1$. Recall that $\beta_t \leq 1$ for all $\beta_t \in B_t$ and consider the following perturbation to all types $\left(h^{t-1},\left(\beta_t,\theta_N\right)\right)$:

$$\begin{aligned}\tilde{u}_t\left(h^{t-1},\left(\beta_t,\theta_N\right)\right)&=u_t\left(h^{t-1},\left(\beta_t,\theta_N\right)\right)+\varepsilon\\ \tilde{u}_{t+1}\left(h^{t-1},\left(\beta_t,\theta_N\right),\left(\beta_{t+1},\theta_{t+1}\right)\right)&=u_{t+1}\left(h^{t-1},\left(\beta_t,\theta_N\right),\left(\beta_{t+1},\theta_{t+1}\right)\right)-\frac{\varepsilon}{\delta}\end{aligned}$$

Type $\left(h^{t-1},\left(\beta_M,\theta_N\right)\right)$ is indifferent between the original allocation and the new one, while types $\left(h^{t-1},\left(\beta_t,\theta_N\right)\right)$ with $\beta_t < 1$ strictly prefer the new allocation for $\varepsilon > 0$. Since no IC constraint for types $\{\theta_1, \theta_2, \ldots, \theta_{N-1}\}$ are binding with respect to θ_N-types, then the perturbation preserves IC for $\varepsilon > 0$ small enough. The associated resource cost is

$$dE=\left(1-\kappa\right)C'\left(u_t\left(h^{t-1},\left(\beta_t,\theta_N\right)\right)\right)\varepsilon$$

But $\kappa > 1$, leading to a resource gain—a contradiction. This leaves us with the case when $\kappa = 1$.

Suppose that $\kappa = 1$ so that for agents bunched at θ_N the inverse Euler equation holds. By

Lemma 9 and θ_N-type agents are bunched, we have $u_t\left(h^{t-1},(\beta_t,\theta_N)\right) > u_t\left(h^{t-1},(\beta_t,\theta_j)\right)$ for all $j < N$. Then consider the following perturbation:

$$\tilde{u}_t\left(h^{t-1},(\beta_1,\theta_N)\right) = u_t\left(h^{t-1},(\beta_1,\theta_N)\right) + \varepsilon - \nu$$
$$\tilde{u}_{t+1}\left(h^{t-1},(\beta_1,\theta_N),(\beta_{t+1},\theta_{t+1})\right) = u_{t+1}\left(h^{t-1},(\beta_1,\theta_N),(\beta_{t+1},\theta_{t+1})\right) - \frac{\varepsilon}{\delta\beta_2}$$

From the point of view of β_1-types the payoff change is

$$dU(\beta_1) = \left(1-\frac{\beta_1}{\beta_2}\right)\varepsilon - \nu$$

Since $\beta_1 < \beta_2$, we can choose $\varepsilon > 0$ and $\nu > 0$ such that $(1-\beta_1/\beta_2)\,\varepsilon = \nu$. All $\left(h^{t-1},(\beta_1,\theta_N)\right)$-types are left indifferent by this perturbation. Furthermore, agents with type $\beta > \beta_1$ dislike this perturbation. Therefore, since other IC constraints with respect to type $\left(h^{t-1},(\beta_1,\theta_N)\right)$ are slack, the perturbation preserves IC. However, the associated resource cost, dE, is

$$\begin{aligned}
&C'\left(u_t\left(h^{t-1},(\beta_1,\theta_N)\right)\right)(\varepsilon-\nu) - \frac{\mathbb{E}_t\left[C'\left(u_{t+1}\left(h^{t-1},(\beta_M,\theta_N),(\beta_{t+1},\theta_{t+1})\right)\right)|\theta_N\right]\varepsilon}{R_t\beta_2\delta}\\
=&C'\left(u_t\left(h^{t-1},(\beta_1,\theta_N)\right)\right)\left(\frac{\beta_1}{\beta_2}-\frac{\kappa}{\beta_2}\right)\varepsilon
\end{aligned}$$

Since $\beta_2 \leq 1$, then for $\varepsilon > 0$ and $\nu > 0$ we get $dE < 0$, so that the planner saves resources. Since $u_t\left(h^{t-1},(\beta_t,\theta_N)\right) > u_t\left(h^{t-1},(\beta_t,\theta_j)\right)$ for all $j < N$, there exists $\varepsilon > 0$ small enough such that redistributing these extra resources improves welfare—a contradiction.. □

A.6.6 Proof of Theorem 5

Part 1.

Proof. Fix a period t and a history h^{t-1}. From Theorem 4, we know that there exists $\underline{\theta} > 0$ such that all agents with a history in $\left\{\left(h^{t-1},(\beta,\theta_1)\right) : \beta \in B\right\}$ for $\theta_1 < \underline{\theta}$ are bunched at the same continuation allocation. In particular, those agents face the same inverse Euler equation distortion

$$\sum_{(\beta_{t+1},\theta_{t+1})\in B\times\Theta} \gamma_{t+1}(\beta_{t+1})\,\rho_{t+1}(\theta_{t+1}|\theta_1)\left[\frac{C'\left(u_{t+1}\left(h^{t-1},(\beta,\theta_1),(\beta_{t+1},\theta_{t+1})\right)\right)}{\delta R_t C'\left(u_t\left(h^{t-1},(\beta,\theta_1)\right)\right)}|\theta_N\right] = \kappa$$

for all $\beta \in B$ and for some constant $\kappa > 0$. The desired result holds if and only if $\kappa \geq 1$. Assume

by way of contradiction that $\kappa < 1$. Then consider the following perturbation:

$$\tilde{u}_t\left(h^{t-1},(\beta,\theta_1)\right) = u_t\left(h^{t-1},(\beta,\theta_1)\right) - \delta\varepsilon$$
$$\tilde{u}_{t+1}\left(h^{t-1},(\beta,\theta_1),(\beta_{t+1},\theta_{t+1})\right) = u_{t+1}\left(h^{t-1},(\beta,\theta_1),(\beta_{t+1},\theta_{t+1})\right) + \varepsilon$$

for all $(\beta_{t+1},\theta_{t+1}) \in B \times \Theta$. This perturbation keeps welfare constant, hence does not affect IC at period $s < t$, when due to quasi-geometric discounting agents and the planner agree about the intertemporal trade-off between periods t and $t+1$. From Lemma 7, there exists $\underline{\theta} > 0$ such that for $\theta_1 < \underline{\theta}$ agents with histories in $\left\{\left(h^{t-1},(\beta,\theta_1)\right) : \beta \in B\right\}$ have strictly slack IC constraints with respect to any other agent not in this group. For $\varepsilon > 0$, agents with $\beta \leq 1$ find themselves weakly worse off under this perturbation, so the perturbation preserves incentive compatible. The marginal resource cost is

$$dE = (\kappa - 1)\,\delta C'\left(u_t\left(h^{t-1},(\beta,\theta_1)\right)\right)\varepsilon$$

For $\varepsilon > 0$ and $\kappa < 1$ we get $dE < 0$, so the perturbation saves resources—a contradiction. □

Part 2.

Proof. For the first part, let

$$\sum_{\beta_{t+1},\theta_{t+1}} \gamma_{t+1}(\beta_{t+1})\,\rho_{t+1}(\theta_{t+1}|\theta_N)\left[\frac{C'\left(u_{t+1}\left(h^{t-1},(\beta_M,\theta_N),(\beta_{t+1},\theta_{t+1})\right)\right)}{\delta R_t C'\left(u_t\left(h^{t-1},(\beta_M,\theta_N)\right)\right)}|\theta_N\right] = \kappa_H$$

for some $\kappa_H > 0$. Assume by way of contradiction that $\kappa_H < 1$ and let

$$\tilde{u}_t\left(h^{t-1},(\beta_M,\theta_N)\right) = u_t\left(h^{t-1},(\beta_M,\theta_N)\right) - \delta\varepsilon$$
$$\tilde{u}_{t+1}\left(h^{t-1},(\beta_M,\theta_N),(\beta_{t+1},\theta_{t+1})\right) = u_{t+1}\left(h^{t-1},(\beta_M,\theta_N),(\beta_{t+1},\theta_{t+1})\right) + \varepsilon$$

for all $(\beta_{t+1},\theta_{t+1}) \in B \times \Theta$. Since $\beta_M = 1$, this perturbation preserves IC for $\varepsilon > 0$. The marginal resource cost is

$$dE = (\kappa_H - 1)\,\delta C'\left(u_t\left(h^{t-1},(\beta_M,\theta_N)\right)\right)\varepsilon$$

so for $\kappa_H < 1$ we have that $dE < 0$ whenever $\varepsilon > 0$, which saves a strictly positive amount of resources—a contradiction.

For the second part, note that from Lemma 9 and from Theorem 4 we have that there exists

$\overline{\theta} < +\infty$ and $\underline{\beta} > 0$ such that for $\theta_N > \overline{\theta}$ and $\beta_1 < \underline{\beta}$ we have that agents with histories in $\left\{\left(h^{t-1}, (\beta, \theta)\right) : \beta \in B, \theta < \theta_N\right\}$ strictly prefer their own allocation to the allocation of any agent with history $\left\{\left(h^{t-1}, (\beta, \theta_N)\right) : \beta \in B\right\}$. Furthermore, we have $u_t\left(h^{t-1}, (\beta_1, \theta_N)\right) > u_t\left(h^{t-1}, (\beta, \theta)\right)$ for all $\theta < \theta_N$ and all $\beta \in B$ by Lemma 9. Suppose now that

$$\sum_{\beta_{t+1}, \theta_{t+1}} \gamma_{t+1}(\beta_{t+1}) \rho_{t+1}(\theta_{t+1}|\theta_N) \left[\frac{C'\left(u_{t+1}\left(h^{t-1}, (\beta_1, \theta_N), (\beta_{t+1}, \theta_{t+1})\right)\right)}{\delta R_t C'\left(u_t\left(h^{t-1}, (\beta_1, \theta_N)\right)\right)} | \theta_N\right] = \tilde{\kappa}_H$$

for some $\tilde{\kappa}_H > 0$. Assume by way of contradiction that $\tilde{\kappa}_H \geq 1$. Let

$$\begin{aligned}
\tilde{u}_t\left(h^{t-1}, (\beta_1, \theta_N)\right) &= u_t\left(h^{t-1}, (\beta_1, \theta_N)\right) + \beta_1 \delta \varepsilon + \nu \\
\tilde{u}_{t+1}\left(h^{t-1}, (\beta_1, \theta_N), (\beta_{t+1}, \theta_{t+1})\right) &= u_{t+1}\left(h^{t-1}, (\beta_1, \theta_N), (\beta_{t+1}, \theta_{t+1})\right) - \varepsilon \\
\tilde{u}_t\left(h^{t-1}, (\beta, \theta)\right) &= \tilde{u}_t\left(h^{t-1}, (\beta, \theta)\right) + \nu
\end{aligned}$$

for all $(\beta_{t+1}, \theta_{t+1}) \in B \times \Theta$ and all $(\beta, \theta) \neq (\beta_1, \theta_N)$. For $\varepsilon > 0$ and $\nu > 0$, this perturbation is incentive compatible since agents with $\beta \geq \beta_1$ find it (weakly) less attractive. To keep welfare unchanged, we require

$$dW = \pi\left(\beta_1, \theta_N | h^{t-1}\right)(\beta_1 - 1)\delta\varepsilon + \nu = 0$$

so that $\nu = \pi\left(\beta_1, \theta_N | h^{t-1}\right)(1 - \beta_1)\delta\varepsilon$. The marginal resource cost dE is

$$\begin{aligned}
&\pi\left(\beta_1, \theta_N | h^{t-1}\right) C'(\cdot)\delta(\beta_1 - \tilde{\kappa}_H)\varepsilon + \nu \sum_{\beta, \theta} \pi\left(\beta, \theta | h^{t-1}\right) C'\left(u_t\left(h^{t-1}, (\beta, \theta)\right)\right) \\
=&\pi\left(\beta_1, \theta_N | h^{t-1}\right) \delta C'(\cdot)(1 - \beta_1)\left[\left(\frac{\beta_1 - \tilde{\kappa}_H}{1 - \beta_1}\right) + \sum_{\beta, \theta} \pi\left(\beta, \theta | h^{t-1}\right) \frac{C'(\cdot)}{C'\left(u_t\left(h^{t-1}, (\beta_1, \theta_N)\right)\right)}\right]\varepsilon
\end{aligned}$$

Since $u_t\left(h^{t-1}, (\beta_1, \theta_N)\right) > u_t\left(h^{t-1}, (\beta, \theta)\right)$ for $\theta < \theta_N$ by Lemma 9, and clearly $u_t\left(h^{t-1}, (\beta_1, \theta_N)\right) \geq u_t\left(h^{t-1}, (\beta, \theta_N)\right)$ for all $\beta \in B$, then

$$\sum_{\beta, \theta} \pi\left(\beta, \theta | h^{t-1}\right) \frac{C'\left(u_t\left(h^{t-1}, (\beta, \theta)\right)\right)}{C'\left(u_t\left(h^{t-1}, (\beta_1, \theta_N)\right)\right)} < 1$$

Since $\beta_1 < 1 \leq \tilde{\kappa}_H$ we conclude that $dE < 0$ for $\varepsilon > 0$, meaning that this perturbation saves a strictly positive amount of resources—a contradiction. □

B Numerical Solution Algorithm

B.1 Proof of Proposition 3

Proof. The proof relies on convexity of the planner's problem (part 1 of Lemma 2). First, note that the algorithm converges in iteration $k \geq 0$ if and only if it has found a point $x^*_{\mathcal{W}_k}$ that satisfies

$$x^*_{\mathcal{W}_k} = \min_{x \in \mathcal{F}_{\mathcal{W}_k}} f(x) \quad \text{s.t.} \quad c_i(x) \geq 0 \text{ for } i \in \mathcal{W}_k,$$

so that $\left|\mathcal{V}_{\mathcal{W}_k}\left(x^*_{\mathcal{W}_k}\right)\right| = 0$. For this to be the case, $x^*_{\mathcal{W}_k}$ must satisfy IC and feasibility of the planner's problem, so $f\left(x^*_{\mathcal{W}_k}\right) = f^* \geq f(x^*_{\mathcal{C}})$ for some constant f^*. We already know that at least one such f^* exists for our planner's problem (part 2 of Lemma 2). Since the cardinality of the constraint set, $|\mathcal{C}| = I$, is finite and since the algorithm avoids loops with probability one, it follows that such an exit criterion is found in finite time. Formally, there exists $f^* \geq f(x^*_{\mathcal{C}})$ such that for any $\varepsilon > 0$, we have:

$$\lim_{k \to \infty} \mathbb{P}\left(\left|f\left(x^*_{\mathcal{W}_k}\right) - f^*\right| \geq \varepsilon\right) = 0$$

Therefore, $\operatorname{plim}_{k\to\infty} f\left(x^*_{\mathcal{W}_k}\right) = f^*$. For general nonlinear optimization problems, f^* need not coincide with the global solution, $f(x^*_{\mathcal{C}})$. But since the solution to the planner's problem is unique (part 2 of Lemma 2), it follows that $\operatorname{plim}_{k\to\infty} f\left(x^*_{\mathcal{W}_k}\right) = f(x^*_{\mathcal{C}})$ and hence $\operatorname{plim}_{k\to\infty} x^*_{\mathcal{W}_k} = x^*_{\mathcal{C}}$. □

B.2 Semi-matrix representation of the transformed planner's problem

Let x denote the vector of stacked elements of the transformed allocation, $\tilde{\mathcal{A}}$. Our goal is to write the planner's problem as a minimization problem of the following semi-matrix form:[54]

$$\min_{x \in F} Ax \quad \text{s.t.} \quad \begin{cases} Bx \geq 0, & \text{representing the set of linear IC constraints} \\ g(x) \geq 0, & \text{representing the nonlinear feasibility constraint,} \end{cases}$$

where x is the choice vector, $F \subseteq \mathbb{R}^n$ for some $n \in \mathbb{N}$ is the domain, A is the objective comatrix, B is the IC constraint set comatrix, and $g(\cdot)$ is the feasibility constraint function.

[54] We choose the label "semi-matrix" for this representation because the program—although nonlinear—would be linear if it were not for the one nonlinear feasibility constraint.

Notation. We index agents' types by $i = 1, \ldots, N$, where $N = N_\theta \times N_\beta$ is total number of types. We order type labels by first fixing a θ-type and looping through β-types, then moving to the next θ-type and repeating the procedure until we have exhausted all (θ, β)-combinations:

$$\begin{aligned}
(\theta_1, \beta_1) &\mapsto i = 1 \\
(\theta_1, \beta_2) &\mapsto i = 2 \\
&\vdots \\
(\theta_1, \beta_M) &\mapsto i = M \\
(\theta_2, \beta_1) &\mapsto i = M + 1 \\
(\theta_2, \beta_2) &\mapsto i = M + 2 \\
&\vdots \\
(\theta_N, \beta_{M-1}) &\mapsto i = N \cdot M - 1 \\
(\theta_N, \beta_M) &\mapsto i = N \cdot M = N
\end{aligned}$$

Henceforth, we index all objects corresponding to type $i = 1, \ldots, N$ by superscript i. For example, the allocation for type i is denoted $\left(u_1^i, u_2^i, \tilde{v}^i\right)$. We write the type vector of dimension $N \times 1$, with types sorted in ascending order from the first (top) entry to the last (bottom) entry, as:

$$\begin{bmatrix} \text{type } 1 \\ \text{type } 2 \\ \vdots \\ \text{type } N_\beta \\ \text{type } N_\beta + 1 \\ \text{type } N_\beta + 2 \\ \vdots \\ \text{type } N_\theta \times N_\beta - 1 \\ \text{type } N_\theta \times N_\beta \end{bmatrix}_{N \times 1} = \begin{bmatrix} \left(\theta^1, \beta^1\right) \\ \left(\theta^2, \beta^2\right) \\ \vdots \\ \left(\theta^M, \beta^M\right) \\ \left(\theta^{M+1}, \beta^{M+1}\right) \\ \left(\theta^{M+2}, \beta^{M+2}\right) \\ \vdots \\ \left(\theta^{N \cdot M - 1}, \beta^{N \cdot M - 1}\right) \\ \left(\theta^{N \cdot M}, \beta^{N \cdot M}\right) \end{bmatrix}_{N \times 1} = \begin{bmatrix} (\theta_1, \beta_1) \\ (\theta_1, \beta_2) \\ \vdots \\ (\theta_1, \beta_M) \\ (\theta_2, \beta_1) \\ (\theta_2, \beta_2) \\ \vdots \\ (\theta_N, \beta_{M-1}) \\ (\theta_N, \beta_M) \end{bmatrix}_{N \times 1}$$

Choice variable vector. The choice variable vector is:

$$x = \begin{bmatrix} u_1^1 \\ u_2^1 \\ \tilde{v}^1 \\ \vdots \\ u_1^N \\ u_2^N \\ \tilde{v}^N \end{bmatrix}_{3N\times 1}$$

with lower bound

$$L = \begin{bmatrix} \underline{u} \\ \underline{u} \\ \underline{\tilde{v}} \\ \vdots \\ \underline{u} \\ \underline{u} \\ \underline{\tilde{v}} \end{bmatrix}_{3N\times 1}$$

and upper bound

$$U = \begin{bmatrix} \overline{u} \\ \overline{u} \\ \overline{\tilde{v}} \\ \vdots \\ \overline{u} \\ \overline{u} \\ \overline{\tilde{v}} \end{bmatrix}_{3N\times 1}$$

Let $F = \left\{ x \in \mathbb{R}^{3N} \middle| \, L \leq x \leq U \right\}$ denote the choice vector's domain.

Objective covector. The objective covector is:

$$A = -\begin{bmatrix} -\pi^1\lambda^1 & -\pi^1\lambda^1\delta & \pi^1\lambda^1 D\left(\theta^1\right) & \ldots & -\pi^N\lambda^N & -\pi^N\lambda^N\delta & \pi^N\lambda^N D\left(\theta^N\right) \end{bmatrix}_{1\times 3N}$$

IC constraint set comatrix. We order the IC constraints as follows:

$$\begin{bmatrix} \text{IC from type } 1 \to \text{type } 2 \\ \text{IC from type } 1 \to \text{type } 3 \\ \vdots \\ \text{IC from type } 1 \to \text{type } N \\ \text{IC from type } 2 \to \text{type } 1 \\ \text{IC from type } 2 \to \text{type } 3 \\ \vdots \\ \text{IC from type } N \to \text{type } N-2 \\ \text{IC from type } N \to \text{type } N-1 \end{bmatrix}_{N(N-1)\times 1}$$

Hence, we can write the IC constraint set comatrix as:

$$B = \begin{bmatrix} 1 & \beta^1\delta & -D\left(\theta^1\right) & -1 & -\beta^1\delta & D\left(\theta^1\right) & 0 & \dots & & & & 0 \\ 1 & \beta^1\delta & -D\left(\theta^1\right) & 0 & 0 & 0 & -1 & -\beta^1\delta & D\left(\theta^1\right) & 0 & \dots & 0 \\ \vdots & & & & & & & & & & & \vdots \\ 1 & \beta^1\delta & -D\left(\theta^1\right) & 0 & \dots & & & & 0 & -1 & -\beta^1\delta & D\left(\theta^1\right) \\ -1 & -\beta^2\delta & D\left(\theta^2\right) & 1 & \beta^2\delta & -D\left(\theta^2\right) & 0 & \dots & & & & 0 \\ 0 & 0 & 0 & 1 & \beta^2\delta & -D\left(\theta^2\right) & -1 & -\beta^2\delta & D\left(\theta^2\right) & 0 & \dots & 0 \\ \vdots & & & & & & & & & & & \vdots \\ 0 & \dots & 0 & -1 & -\beta^N\delta & D\left(\theta^N\right) & 0 & 0 & 0 & 1 & \beta^N\delta & -D\left(\theta^N\right) \\ 0 & \dots & & & & 0 & -1 & -\beta^N\delta & D\left(\theta^N\right) & 1 & \beta^N\delta & -D\left(\theta^N\right) \end{bmatrix}_{N(N-1)\times 3N}$$

with lower bound

$$-\infty_{N(N-1)\times 1}$$

and upper bound

$$\mathbf{0}_{N(N-1)\times 1}$$

Nonlinear feasibility constraint. The feasibility constraint function is:

$$g\left(\tilde{\mathcal{A}}\right) = \sum_{i=1}^{N} \pi^i \left[Y\left(\tilde{v}^i\right) - C\left(u_1^i\right) - \frac{C\left(u_2^i\right)}{R} \right],$$

with scalar bounds $-\infty$ from below and 0 from above. The Jacobian (first derivative) matrix with respect to the choice variable vector is:

$$\left[\begin{array}{cccccc} -\pi^1 C'\left(u_1^1\right) & -\frac{\pi^1}{R} C'\left(u_2^1\right) & \pi_1 Y'\left(\tilde{v}^1\right) & \ldots & -\pi^N C'\left(u_1^N\right) & -\frac{\pi^N}{R} C'\left(u_2^N\right) \quad \pi^N Y'\left(\tilde{v}^N\right) \end{array} \right]_{1\times 3N}$$

The Hessian (second derivative) matrix with respect to the choice variable vector is:

$$\left[\begin{array}{ccccccccc} -\pi^1 C''\left(u_1^1\right) & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 \\ 0 & -\frac{\pi^1}{R} C''\left(u_1^1\right) & 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \vdots \\ \vdots & 0 & \pi^1 Y''\left(\tilde{v}^1\right) & 0 & \ldots & \ldots & \ldots & \ldots & \vdots \\ \vdots & \vdots & 0 & -\pi^2 C''\left(u_{12}\right) & 0 & \ldots & \ldots & \ldots & \vdots \\ \vdots & \vdots & \vdots & 0 & \vdots & \ddots & \ldots & \ldots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & -\pi^N C''\left(u_1^N\right) & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & -\frac{\pi^N}{R} C''\left(u_1^N\right) & 0 \\ 0 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & 0 & \pi^N Y''\left(\tilde{v}^N\right) \end{array} \right]_{3N\times 3N}$$

In summary, we can write the transformed planner's problem in the semi-matrix representation:

$$\min_{x\in F} Ax \quad \text{s.t.} \quad \begin{cases} Bx \geq 0, & \text{representing the set of linear IC constraints} \\ g\left(x\right) \geq 0, & \text{representing the nonlinear feasibility constraint} \end{cases}$$

B.3 Algorithm initialization

Our point of departure is the transformed planner's problem in consumption utility-labor disutility space introduced in the proof of Lemma 2. Recall that the transformed allocation is

$$\tilde{\mathcal{A}} = \left\{u_1\left(\theta,\beta\right), u_2\left(\theta,\beta\right), \tilde{v}\left(\theta,\beta\right)\right\}_{(\theta,\beta)}.$$

For the remainder of this section, we assume the power utility formulation that we later use in our quantitative application:

$$u\left(c\right) = \begin{cases} \frac{c^{1-1/\sigma}-1}{1-1/\sigma} & \text{for } \sigma \neq 1 \\ \ln\left(c\right) & \text{for } \sigma = 1 \end{cases}, \quad v\left(\ell\right) = \kappa \frac{\ell^{1+1/\gamma}}{1+1/\gamma}$$

Uniform allocation across types. One possible initialization is to the economy with everyone consuming and producing the same amount. To this end, consider $\overline{y} > 0$ and $\overline{c} > 0$ such that:

$$u_1(\theta, \beta) = u(\overline{c}) \tag{15}$$

$$u_2(\theta, \beta) = u(\overline{c}) \tag{16}$$

$$\tilde{v}(\theta, \beta) = v(\overline{y}) \tag{17}$$

$$\overline{c} + \frac{\overline{c}}{R} = \overline{y}$$

Consumption is equalized across periods and, like labor supply, across agents. Consequently, individual output equals individual net present value of lifetime consumption. Normalizing the intratemporal Euler equation to be undistorted for types with $\theta = 1$, we get:

$$\begin{aligned} v'(\overline{y}) &= u'\left(\frac{R\overline{y}}{1+R}\right) \\ \implies \overline{y} &= \left(\frac{1+R}{\kappa^{\eta} R}\right)^{\frac{\gamma}{\eta-\gamma}} \end{aligned} \tag{18}$$

Plugging equation (18) back into equations (15)–(17) yields the desired initialization:

$$\overline{u}_1 = u(\overline{c})$$

$$\overline{u}_2 = u(\overline{c})$$

$$\overline{\tilde{v}} = v(\overline{y})$$

Laissez-faire allocation. Another initialization has every agent consume and produce at their respective laissez-faire levels. To this end, $c(\theta, \beta)$ and $y(\theta, \beta)$ must satisfy agents' individual Euler equations and budget constraint:

$$\begin{aligned} u'\left(c_1^{LF}(\theta, \beta)\right) &= \beta\delta R u'\left(c_2^{LF}(\theta, \beta)\right) \\ \implies c_2^{LF}(\theta, \beta) &= (\beta\delta R)^{\eta} c_1^{LF}(\theta, \beta) \end{aligned} \tag{19}$$

and

$$\begin{aligned} u'\left(c_1^{LF}(\theta,\beta)\right) &= \tilde{v}'\left(y^{LF}(\theta,\beta)/\theta\right) \\ \Longrightarrow y^{LF}(\theta,\beta) &= \frac{\theta}{\kappa}\left[c_1^{LF}(\theta,\beta)\right]^{-\frac{\gamma}{\eta}} \end{aligned} \tag{20}$$

and

$$c_1^{LF}(\theta,\beta) + \frac{c_2^{LF}(\theta,\beta)}{R} = y^{LF}(\theta,\beta) \tag{21}$$

Solving the system of three equations (19)–(21) in three unknowns yields:

$$\begin{aligned} c_1^{LF}(\theta,\beta) + \frac{(\beta\delta R)^{\eta} c_1^{LF}(\theta,\beta)}{R} &= \frac{\theta}{\kappa}\left[c_1^{LF}(\theta,\beta)\right]^{-\frac{\gamma}{\eta}} \\ \Longrightarrow c_1^{LF}(\theta,\beta) &= \left\{\frac{\kappa}{\theta}\left[1+\frac{(\beta\delta R)^{\eta}}{R}\right]\right\}^{\frac{\eta}{-\gamma-\eta}} \end{aligned} \tag{22}$$

Plugging equation (22) back into equations (19)–(20) yields the desired initialization:

$$\begin{aligned} u_1^{LF}(\theta,\beta) &= u\left(c_1^{LF}(\theta,\beta)\right) \\ u_2^{LF}(\theta,\beta) &= u\left(c_2^{LF}(\theta,\beta)\right) \\ \tilde{v}^{LF}(\theta,\beta) &= v\left(y^{LF}(\theta,\beta)\right) \end{aligned}$$

B.4 Algorithm benchmarking

Performance. To benchmark the performance of our active-set algorithm vis-à-vis a more direct approach that solves the global problem directly, we solve a sequence of problems of increasing size using both methods. To make things simple, we use a type space with a constant number of present bias levels, $N_\beta = 6$, a growing number of ability levels, N_θ—both distributed between 0.1 and 1.0—a uniform distribution, $\pi(\theta,\beta) = 1/\left(N_\theta N_\beta\right)$, and constant Pareto weights across all types, $\lambda(\theta) = 1$. All computations are performed on a 2017 iMac Pro with 3.0GHz 10-core Intel Xeon W processor and 128GB 2666MHz DDR4 memory. Table 7 shows the results of this performance benchmark.

Table 7. Comparison of active-set algorithm vs. global solution method

Number of types, $N = N_\theta \times N_\beta$	Time (seconds), active-set	Time (seconds), global
60 = 10 × 6	0.3	0.6
180 = 30 × 6	1.4	5.5
600 = 100 × 6	6.6	185.8
1,800 = 300 × 6	101.0	6,653.9
6,000 = 1,000 × 6	1,315.7	114,707.0

The advantage of our algorithm is apparent throughout all problem sizes. For the smallest instance of the problem with 10×6 types, our algorithm is approximately twice as fast as the global solution method. With 30×6 types, it is almost four times as fast as the global solution method. This factor increases to 28 for a problem with 100×6 types, and to 66 for a problem with 300×6 types. For the largest version of our problem with 1,000×6 types, which we use in our quantitative analysis in Section 7, our algorithm finds a solution in 22 minutes, while the global method takes almost 32 hours to complete. We conclude that our active-set algorithm offers large efficiency gains, particularly when solving large-scale programs.

Next, we compare the structure of IC constraints under the global (or, equivalently, active-set) solution method vis-à-vis a a local solution that ignores any nonlocal IC constraints. Table 8 compares the pattern of binding IC constraints across different problem sizes.

Table 8. Structure of IC constraints

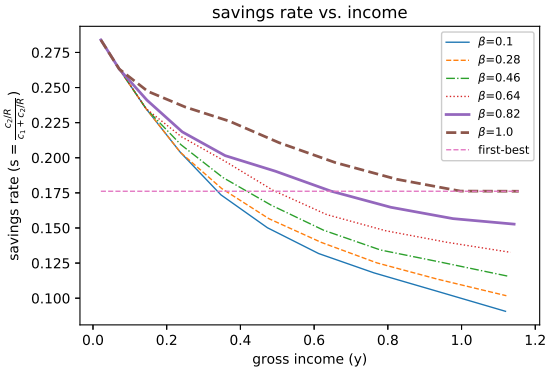
Number of types, $N = N_\theta \times N_\beta$	Number of choice variables	Number of global IC constraints	Number of local IC constraints	Number of binding IC constraints at optimum	Number of binding nonlocal IC constraints at optimum
60 = 10 × 6	180	3,540	388	257	93
180 = 30 × 6	540	32,220	1,228	678	201
600 = 100 × 6	1,800	359,400	4,168	1,759	330
1,800 = 300 × 6	5,400	3,238,200	12,568	6,228	2,269
6,000 = 1,000 × 6	18,000	35,994,000	41,968	59,158	42,808

While the number of global IC constraints grows approximately with the square of the number of types, the number of local constraints grows roughly linearly. The number of binding constraints at the optimum grows faster than linearly, but the share of binding constraints as a fraction of all global constraints vanishes. Nonlocal IC constraints always bind and make up a signficant share of all binding constraints across all problem instances. This suggests that the first-order approach—a local solution method—does not work well for our problem.

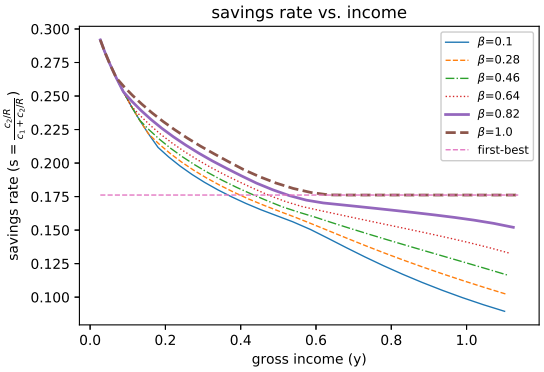
Comparison of results for different problem sizes. We illustrate our theoretical results from Section 3—bunching at the bottom versus separation at the top of the income (or ability) distribution—for different problem sizes in computations. To this end, we use the same model parameters as above. Figure 12 shows the results of this exercise.

Figure 12. Optimal savings rates across income and present bias levels for different problem sizes

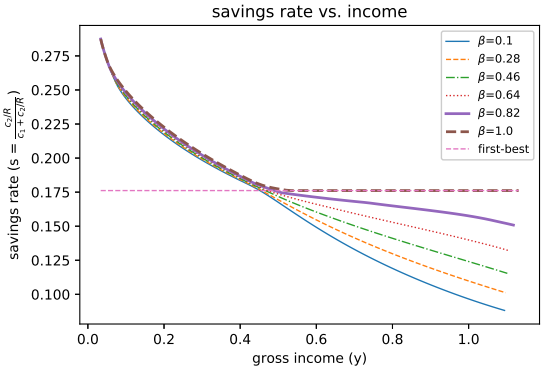
(a) $N_\theta = 10, N_\beta = 6$



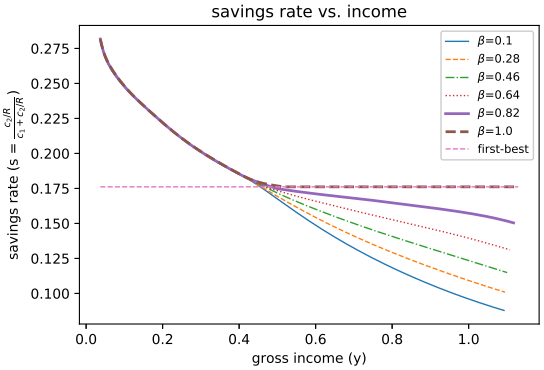
(b) $N_\theta = 30, N_\beta = 6$



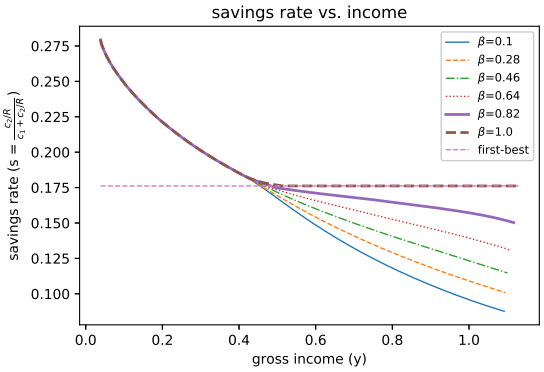
(c) $N_\theta = 100, N_\beta = 6$



(d) $N_\theta = 300, N_\beta = 6$



(e) $N_\theta = 1000, N_\beta = 6$



The distribution of savings rates across types for different problem sizes always shows bunching at the bottom and separation at the top. At the same time, the optimal allocation features bunching over a greater range of income (or ability) levels when the type space is more dense. This observation is symptomatic of the underlying fact that the optimal allocation differs markedly across type spaces. Going from $N_\theta = 10$ to $N_\theta = 300$ leads to visibly different savings rates throughout the income distribution, holding fixed $N_\beta = 6$. Meanwhile, going from $N_\theta = 300$ to $N_\theta = 1000$, is associated with a much smaller—although quantitatively noticable—change in the optimal allocation. We conclude that a precise type grid is desirable when trying to accurately represent a dense population of agents for numerical solutions of optimal policies.

B.5 Example of a loop in a deterministic active-set algorithm

In designing our active-set algorithm, we introduced a stochastic element when adding and dropping constraints across iterations. To demonstrate that this design element matters, we demonstrate below that infinite loops indeed exist when running a deterministic version of the active-set algorithm, i.e. one in which the adding probabilities p_n and the dropping probabilities q_n are either 0 or 1. To this end, we tried to solve a large-scale version of the problem with $N_\theta \times N_\beta = 1,000 \times 6$ types and came across the following loop (from step 6 onwards, aborted at step 10):

```
*** start active-set algorithm
* step 1
   --> total number of IC constraints = 65964
   --> problem successfully solved in 115.218s
   --> number of violated IC constraints = 35337
* step 2
   --> number of IC constraints dropped = 55941
   --> number of IC constraints added = 35337
   --> total number of IC constraints = 45360
   --> problem successfully solved in 109.517s
   --> number of violated IC constraints = 27
* step 3
   --> number of IC constraints dropped = 35035
   --> number of IC constraints added = 27
   --> total number of IC constraints = 10352
   --> problem successfully solved in 213.991s
   --> number of violated IC constraints = 9
```

```
* step 4
   --> number of IC constraints dropped = 7
   --> number of IC constraints added = 9
   --> total number of IC constraints = 10354
   --> problem successfully solved in 130.748s
   --> number of violated IC constraints = 1
* step 5
   --> number of IC constraints dropped = 7
   --> number of IC constraints added = 1
   --> total number of IC constraints = 10348
   --> problem successfully solved in 205.145s
   --> number of violated IC constraints = 4
* step 6
   --> number of IC constraints dropped = 1
   --> number of IC constraints added = 4
   --> total number of IC constraints = 10351
   --> problem successfully solved in 217.941s
   --> number of violated IC constraints = 1
* step 7
   --> number of IC constraints dropped = 4
   --> number of IC constraints added = 1
   --> total number of IC constraints = 10348
   --> problem successfully solved in 191.119s
   --> number of violated IC constraints = 4
* step 8
   --> number of IC constraints dropped = 1
   --> number of IC constraints added = 4
   --> total number of IC constraints = 10351
   --> problem successfully solved in 204.047s
   --> number of violated IC constraints = 1
* step 9
   --> number of IC constraints dropped = 4
   --> number of IC constraints added = 1
   --> total number of IC constraints = 10348
   --> problem successfully solved in 190.416s
   --> number of violated IC constraints = 4
# etc.
```

Note that at times loops may appear when the working set is very "close" to the set of relevant global constraints—in the sense of containing almost all the relevant global constraints—but that at other times the algorithm may get stuck in a loop very "far" from the relevant constraint set. Furthermore, the example above is a loop of the second order, i.e. once we are in the loop, then

every second term is a repetition of itself. At times, we have also encountered higher-order loops.

C Calibration

C.1 Description of datasets

Health and Retirement Study (HRS). Our analysis is largely based on the HRS microdata, which follow households before and after retirement and provide rich information on demographics, family structure, health status and expenditures, income, financial wealth, Social Security, and retirement savings plans use.[55] These data consist of a panel of 37,495 households that are followed across twelve waves between 1992 and 2014, with five entry cohorts between 1992 and 2010.

We also use tax information from the HRS,[56] which contain separate information about federal, state, and FICA taxes for respondents in the 2000–2014 HRS surveys. We use this additional file to compute individuals' gross and net income based on the NBER-Internet TAXSIM calculator, Version 9.3 (Feenberg and Coutts, 1993).

Panel Study of Income Dynamics (PSID). To supplement our analysis with life-cycle income dynamics for a broader age range, we use the family and individual-merged files of the PSID.[57] These data have been used by several researchers in the past and we refer to Meghir and Pistaferri (2004) for a detailed description of the data. The dataset follows approximately 4,500 households over time and reports a wide range of socioeconomic characteristics, including age, industry, occupation, labor income, and income from self-employment.

U.S. Life Tables. The U.S. Life Tables provide mortality statistics for the population as a whole and for various subpopulations by gender, race, and ethnicity (Arias, 2010). We deliberately use

[55] Health and Retirement Study, RAND HRS Longitudinal File 2014 (V2) public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740) and the Social Security Administration. Ann Arbor, MI, 2018. The RAND HRS Longitudinal File 2014 (V2) is available online from https://www.rand.org/labor/aging/dataprod/hrs-data.html.

[56] Health and Retirement Study, RAND HRS Tax Calculations 2014 (V2) public use dataset. Produced and distributed by the University of Michigan with funding from the National Institute on Aging (grant number NIA U01AG009740) and the Social Security Administration. Ann Arbor, MI, 2018. The RAND HRS Tax Calculations 2014 (V2) is available online from https://www.rand.org/labor/aging/dataprod/tax-calculations.html.

[57] Panel Study of Income Dynamics, public use dataset. Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, 2017. The PSID microdata are available online from https://psidonline.isr.umich.edu/.

2006 as an older vintage of the U.S. Life Tables to address the fact that mortality rates are changing over time.[58]

Health Inequality Project. The Health Inequality Project (Chetty et al., 2016) publishes statistics on mortality rates by income groups based on administrative data, including 1.4 billion tax returns and death records from the Social Security Administration.[59] Statistics are reported separately by gender, age, year, and income quantile.

Other data. To deflate prices over time in our data, we use the "Consumer Price Index for All Urban Consumers: All Items [CPIAUCSL]" (in short, CPI) retrieved from FRED (U.S. Bureau of Labor Statistics, 2018).

C.2 Data cleaning and sample selection

We use all twelve waves of the HRS data, where possible, although certain variables are available only for some waves.[60] We focus all our analysis on the household-level by identifying the Financial Respondent in couple households. While we collapse the data on individual household members into one observation per household, our analysis uses information on both the respondent and their spouse, where available. For parts of our analysis, we focus on the reference person's characteristics, for example when classifying a single household as working, retired, or deceased. Finally, we omit from our analysis households that appear in only one survey wave.

Next, we fill in missing observations retirement status, private and public wealth, expected bequests, and medical expenditures between survey waves when this information is missing. For a small number of variables—including education and year of birth of the reference person and their spouse, and year of death of the reference person—we adopt their modal value across waves. We drop any households who report mising values in their education or income throughout all years. We also drop households whose reference person is retired or deceased for all waves, and those who never retire during the coverage period of the HRS survey. Finally, we restrict our analysis to households whose reference person is between 25 and 99 years old. Finally, for the

[58] The U.S. Life Tables, including the 2006 vintage and other years of data, are available online from https://www.cdc.gov/nchs/products/life_tables.htm.

[59] All statistics from the Health Inequality Project, together with replication codes and data, are available online from https://healthinequality.org/.

[60] For example, the value of a secondary home owned by the household members is not reported in wave 3, and predicted Social Security Wealth of pre-retirees is available only in waves 1, 4, 7, and 10.

selected sample, we create an annual balanced panel dataset by imputing key variables through interpolation between years.

We combine the US Life Tables data on mortality rates by age and the Health Inequality Project data on mortality differentials across income groups. To this end, we augment the average mortality rate profile from the US Life Tables with the income decile-specific mortality rate intercept from the Health Inequality Project.

C.3 Variable construction

Annual income. Using the HRS data, we construct annual gross income as the sum of individual earnings and other household income, which consist of wage and salary income in the primary and secondary job, bonuses and overtime pay, commissions, and tips, military reserve earnings, professional practice and trade income, alimony, insurance payment receipts, pensions, and inheritance.[61] We construct an aggregate income measure at the household-level, summing earnings of the spouse and the reference person within each survey year. We deflate all nominal variables by the CPI

To impute a complete panel of income over the life-cycle, we record for each household the reference person's education level, which we use together with age to connect the part of the lifetime earnings profile covered by the HRS data with that in the PSID data. To this end, we first estimate the following regression on a pooled sample of each household member i across each survey year t:

$$y_{it} = X_{it}\beta + \sum_{\tau} \mathbf{1}\left[t = \tau\right]\gamma_{\tau} + \varepsilon_{it} \tag{23}$$

where y_{it} is log income, X_{it} is a matrix of household covariates, including educational attainment dummies interacted with a cubic polynomial in age, β is a coefficient vector for these characteristics, γ_{τ} is a year effect, and ε_{it} is a residual term. We estimate equation (23) using ordinary least squares on the HRS data and the PSID data separately. From the PSID data, which covers a wider age range, we store the estimated coefficients on the polyomial terms in age, $\hat{\beta}^{PSID}$. We then

[61] We find that annuities are held by 2.14% (between 0.46% and 2.65% across waves) of all HRS survey respondents. On average, they make up 1.97% (between 0.30% and 3.17% across waves) of total household income among respondents across all waves of the HRS data. The phenomenon that few households convert their savings into steady income streams is sometimes referred to as the "Annuity Puzzle" (Mitchell et al., 1999). We have included annuity income, as reported in the HRS microdata, in our retirement wealth calculations as the discounted net present value at the time of retirement given the individual's observed (when covered by the HRS wave) or else the expected (applying the estimated mortality profiles by income group from the U.S. Life Tables and Health Inequality Project data) retirement length without any noticable changes to our estimates.

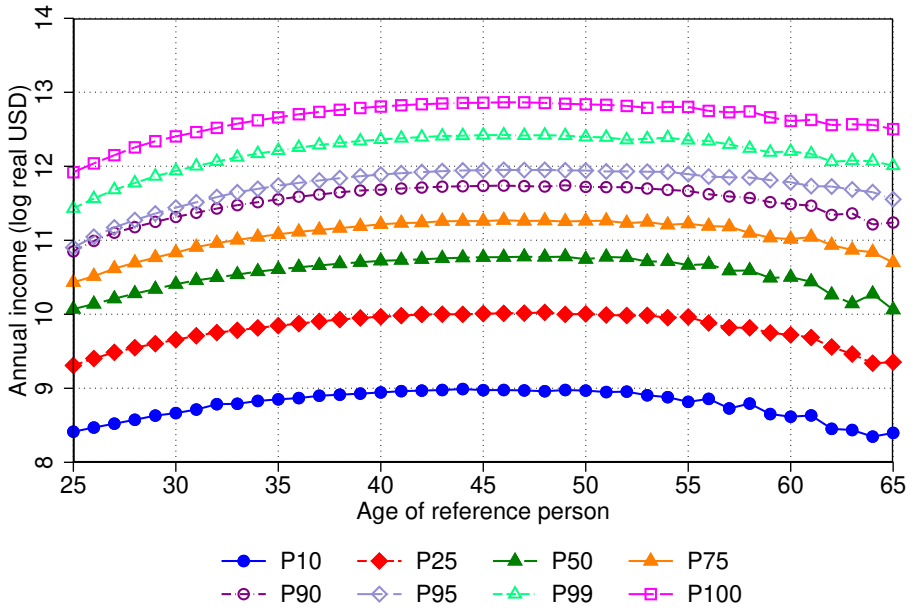
predict in the HRS data for each household member their entire lifecycle of earnings:

$$\hat{y}_{it}^{HRS-PSID} = X_{it}^{HRS}\hat{\beta}^{PSID} + \sum_{\tau} \mathbf{1}\left[t = \tau\right]\hat{\gamma}_{\tau}^{HRS} + \hat{\varepsilon}_{it}^{HRS}$$

where $\hat{y}_{it}^{HRS-PSID}$ is the imputer income variable of interest and variables with hats denote the estimate from the respective sample. In doing this, we fill in income information for each household years that were not covered in the HRS either if they fall between survey waves or else if they fall outside of the age window that a given household is covered between the first and the last interview.

Lifetime income. We use deflated annual income from above and discount values over time to calculate the net present value of their lifetime income using an annual interest rate of 3.44% (Gourinchas and Parker, 2002). Figure 13 shows the estimated annual income profiles by lifetime earnings percentiles applying the estimated PSID income growth rates to the HRS income data.

Figure 13. Mean log income profiles by lifetime income percentile, conditional on working



Notes: Figure shows imputed age profiles of income conditional on being in lifetime income percentile, e.g. P10 is 10th lifetime income percentile. Source: Authors' computations based on HRS and PSID.

Gross versus net income. Using the RAND HRS Tax Calculations 2014 (V2) supplement, we construct gross income at the household-level as the sum of wage and salary income of the respondent and their spouse, dividends, property income, taxable pensions and IRA distributions, gross social security benefits, and other nontaxable transfer income. We separately sum tax all reported tax deductions related to student loan interest, tuition fees, and qualifying investment losses. We form net income by subtracting all deductable income and tax payments from the household's gross income.

Post-retirement income. Some individuals in the HRS data report post-retirement income other than Social Security benefits and withdrawals from private savings plans, such as IRA or Keogh accounts. For these individuals, we and compute the net present discounted value of all income streams over their retirement life and add this to our measure of lifetime income.

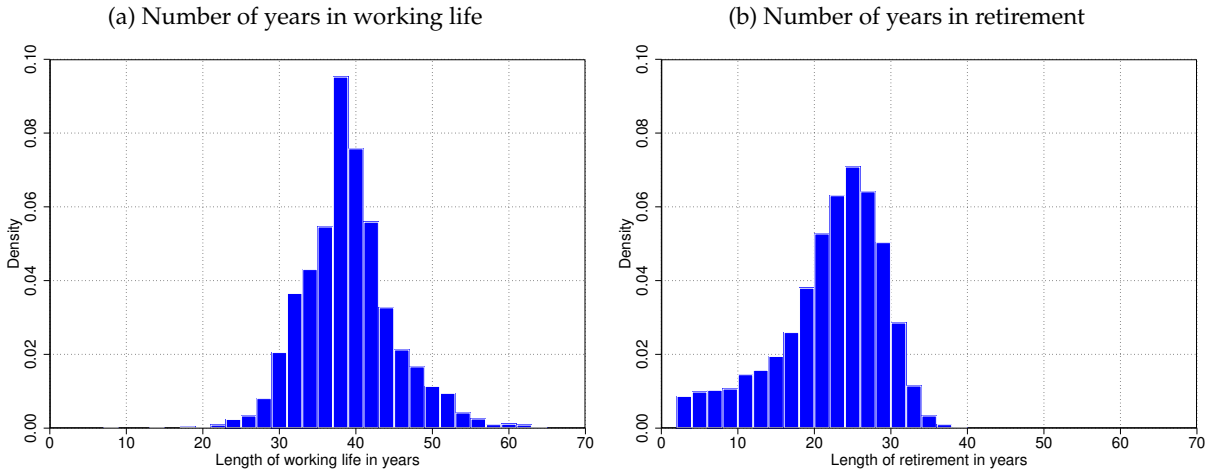
Medical expenditures. We follow a similar imputation strategy as we used for income to impute medical expenditures over the lifecycle. To this end, we use information on medical out-of-pocket expenditures reported in the HRS data. Such out-of-pocket expenditures include hospital costs, nursing home costs, doctor visits costs, dentist costs, outpatient surgery costs, average monthly prescription drug costs, home health care costs, and special facilities costs, to the extent that they are not covered by medical insurance. We do not include in our medical expenditures measure any expenses covered by private insurance plans or through public health care provision such as Medicare or Medicaid. We calculate out-of-pocket medical expenditures separately before and after retirement. For pre-retirement expenditures, we simply compute the net present value of all expenses. For post-retirement expenditures, we group households by lifetime income decile and compute the empirical distribution of medical expenditures, approximated by a lognormal whose mean and standard deviation we estimate separately for each group.

Bequests. From survey wave 6 onwards, we observe individuals' subjective bequest probabilities over three bequest brackets: the probability of leaving a bequest of more than USD 10,000, more than USD 100,000, and more than USD 500,000. We use these probabilities to calculate expected bequests by choosing "midpoints" for each bracket as USD 50,000, USD 250,000, and the maximum of USD 1,000,000 and the average bequest in multiples of income by parental productivity quartile from De Nardi (2004), capped at USD 10,000,000.

Working life, retirement life, and death. We define retirement as the first time that a household self-reports as retired. For most households, this is a binary choice, meaning that individuals report either being (full-time or part-time) employed, or else retired. A small number of respondents being retired and also working, disabled, or unemployed at the same time. We classify those individuals as retired and treat them identically to those who only report being retired.

We assume that household members have been working since age 25. Since we restricted attention to households entering retirement during one of the HRS survey waves, we see a distribution of retirement ages for every household in our sample. For households whose reference person dies during one of the survey waves, we also know the length of their retirement life. For households who retire but do not die during one of the survey years, we compute their expected retirement life span based on the conditional survival probabilities from merging the US Life Tables with the Health Inequality Project data. Figure 14 shows the empirical distribution of the length of working lifes (before retirement) and the length of retirement lifes (before death) in the HRS data.

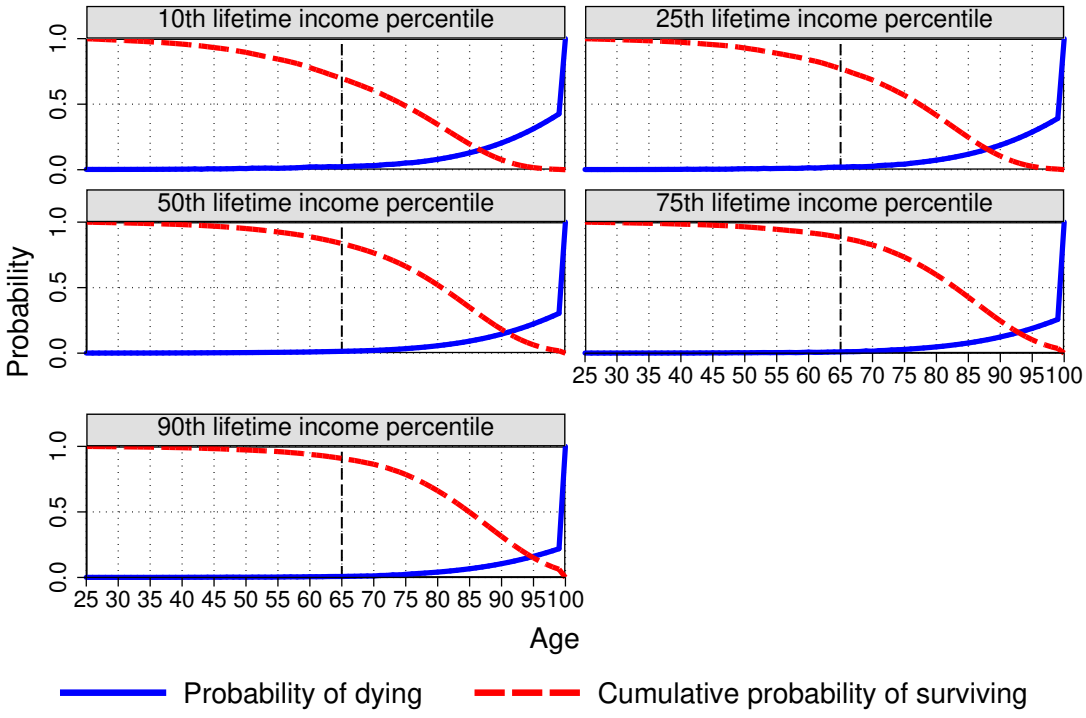
Figure 14. Length of working life and length of retirement



Notes: Histograms use a bin width of 2 years. Source: Authors' computations based on HRS.

We combine the U.S. Life Tables with the Health Inequality Project data to estimate survival probabilities over the lifecycle. Figure 15 shows the estimated probability of dying and cumulative probability of surviving for different lifetime income percentiles. Using these estimates, we compute for each individual their probability of surviving until their observed retirement age.

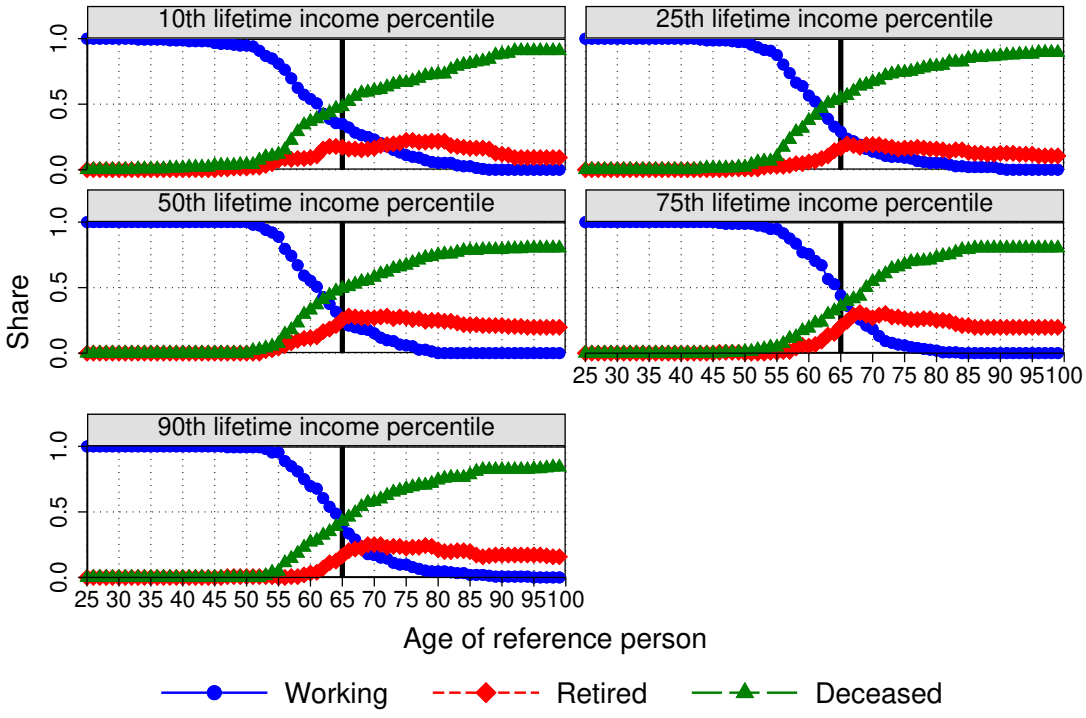
Figure 15. Probability of dying and cumulative probability of surviving



Notes: Figure assumes that the probability of dying conditional on age 100 is 1. Source: Authors' computations based on U.S. Life Tables and the Health Inequality Project.

To summarize, Figure 16 shows the empirical population shares working, retired, and deceased by age from the HRS data.

Figure 16. Population share working, retired, and deceased by age



Notes: Shares computed separately by age. Source: Authors' computations based on HRS.

Retirement wealth. Using the HRS data, we construct a measure of household wealth at retirement that consists of public wealth (predicted Social Security wealth), and private wealth (net value of real estate, vehicles, businesses, IRA and Keogh accounts, stocks, mutual funds, investment trusts, checking, savings, and other money market accounts, bonds, and other savings, net of any outstanding debt obligations). Predicted Social Security wealth is reported directly in the HRS and is based on the retirement insurance benefit calculated based on individual earnings records and assuming normal retirement claiming age. We sometimes add up public and private savings and refer to the sum as total retirement wealth, or lifetime savings.

Lifetime savings rates. Finally, we construct lifetime savings rates as the share of retirement wealth out of lifetime income, both in net present value terms.

C.4 Calibration details

C.4.1 Positive calibration

Model setup. As it will be convenient to write the problem in annual terms, we think of each period consisting of a number of years, where agents of type β have aggregate discount factor $\psi = \beta\delta$. Conditional on surviving and the old-age medical shock realization, an old agent solves the following problem:

$$\begin{aligned} U_2\left(y_2;\phi_1,T_R\right) &= \max_{c_2,b^{late}}\left\{\sum_{t=1}^{T_R}\psi^{t-1}u\left(c_2\right)+\Phi\left(b^{late};\phi_1\right)\right\} \\ \text{s.t.}\quad \sum_{t=1}^{T_R}R^{-(t-1)}c_2+R^{-T_R}\tilde{b} &= \max\left\{y_2,\sum_{t=1}^{T_R}R^{-(t-1)}\underline{c}_2\right\} \\ b^{late} &= \tilde{b}-T_b\left(\tilde{b}\right), \end{aligned}$$

Taking as given the problem of the old agent across possible future states, a young agent solves the following problem:

$$\begin{aligned} U_1\left(\theta,\psi;h,\phi_1,T_W,T_R,P\left[death\right]\right) &= \max_{c_1,s^{401k},s^{IRA},s^{taxable},y_1}\left\{\begin{array}{c}\sum_{t=1}^{T_W}\psi^{t-1}\left[u\left(c_1\right)-v\left(\frac{y_1}{\theta}\right)\right] \\ +\psi\left[\begin{array}{c}\left(1-P\left[death\right]\right)\times\mathbb{E}_{m_2}U_2\left(y_2\left(m_2\right);\phi_1,T_R\right) \\ +P\left[death\right]\times\Phi\left(b^{early};\phi_1\right)\end{array}\right]\end{array}\right\} \\ \text{s.t.}\quad \sum_{t=1}^{T_W}R^{-(t-1)}\left[c_1+s^{401k}+s^{IRA}+s^{taxable}\right] &= \sum_{t=1}^{T_W}R^{-(t-1)}\max\left\{y_1-T_y\left(\max\left\{y_1-s^{401k}-s^{IRA}-m_1,0\right\}\right)-m_1,\underline{c}_1\right\} \\ \tilde{y}_2^{pre-tax} &= 0.85\times SS\left(y_1\right)+R\times match\left(s^{401k}\right)+Rs^{IRA}+\left(R-1\right)s^{taxable} \\ \tilde{y}_2^{post-tax} &= 0.15\times SS\left(y_1\right)+s^{taxable} \\ y_2\left(m_2\right) &= \sum_{t=1}^{T_W}R^{(T_w-t+1)}\left[\tilde{y}_2^{pre-tax}-T_y\left(\max\left\{\tilde{y}_2^{pre-tax}-m_2,0\right\}\right)-m_2+\tilde{y}_2^{post-tax}\right] \\ \tilde{\tilde{b}} &= R\times match\left(s^{401k}\right)+Rs^{IRA} \\ \tilde{b} &= \tilde{\tilde{b}}-T_y\left(\tilde{\tilde{b}}\right)+Rs^{taxable} \\ b^{early} &= \tilde{b}-T_b\left(\tilde{b}\right) \\ m_2 &\sim H\left(\cdot;h\right) \\ c_1,s^{401k},s^{IRA},s^{taxable},y_1 &\geq 0 \\ s^{401k} &\leq \bar{s}^{401k}\left(y_1\right) \\ s^{IRA} &\leq \bar{s}^{IRA}\left(y_1\right) \end{aligned}$$

We denote annualized variables by a tilde, which is particularly convenient since it allows us

to write tax schedules, Social Security benefits, and subsidy rates and caps at an annual level, corresponding to the real-world system.

Calibration targets and parameters. In the model and in the data, we compute the target vector

$$\left(ln\left(y_1\right), s^{private}, \ln\left(\mathbb{E}b^{late}\right)\right), \tag{24}$$

where $s^{private} = s^{401k} + s^{IRA} + s^{taxable}$ is the private savings rate, and the expectation of old-age bequests, $\mathbb{E}b^{late}$, is taken with respect to the agent's information at the end of the first period in the model, and it is reported directly in the HRS data. Individual-by-individual, we pick a model parameter vector $(\theta, \psi, \phi_1) \in P = \mathbb{R}_+ \times [0,1] \times \mathbb{R}$ to minimize the L^2-norm between the model target vector (24) and its empirical analogue associated with the individual:[62]

$$(\theta, \psi, \phi_1) = \underset{(\theta', \psi', \phi_1') \in P}{\arg\min} \left\{ \begin{array}{c} \left[\ln\left(\left(y_1\right)_{model}\right)(\theta', \psi', \phi_1') - \ln\left(\left(y_1\right)_{data}\right)\right]^2 \\ + w_s\left[\left(s^{private}\right)_{model}(\theta', \psi', \phi_1') - \left(s^{private}\right)_{data}\right]^2 \\ + w_b\left[\left(\ln\left(\mathbb{E}b^{late}\right)\right)_{model}(\theta', \psi', \phi_1') - \left(\ln\left(\mathbb{E}b^{late}\right)\right)_{data}\right]^2 \end{array} \right\},$$

where w_s and w_b are the relative weights on the savings rate and expected bequest, respectively. We find that putting relatively more weight on the savings rate, $w_s = 30$ and $w_b = 1$, compensates for the difference in units between moments and helps to identify compound discount factors that closely match empirical savings decisions.

After our point-by-point calibration of (θ, ψ, ϕ_1), we drop observations with a loss function above 10% for which our model cannot provide a good fit. We summarize the joint distribution of θ and β by first approximating θ as a log-normal distribution with mean μ_θ and standard deviation σ_θ over 1000 grid points. Then, for each percentile of the approximated θ-distribution we fit a flexible beta distribution over 6 grid points to approximate the distribution of calibrated values of

[62] In our baseline calibration, we restrict the compound discount factor to $\psi \in [0, 1/R]$ because we later impose $\delta = 1/R$. As a result, $\beta \in [0,1]$. In a previous version of the paper, we allowed for the possibility that $\beta > 1$, so agents could be more patient than the planner, which is a common tenet in the political economy literature (Aguiar and Amador, 2011; Halac and Yared, 2014, 2018). In our current calibration with many more "rational" motives to save, we find that the restriction to $\beta \leq 1$ is less binding compared to a model where other savings motives are ignored.

the annualized discount factor ψ_{annual}:[63]

$$\psi_{annual}|\,\theta \sim beta\left(a_\psi\left(\theta\right), b_\psi\left(\theta\right)\right),$$

where we let the shape parameters $a\left(\theta\right)$ and $b\left(\theta\right)$ vary linearly across θ-ranks. That is, if we let $\rho\left(\theta\right) \in [0,1]$ denote the rank of θ in the population, with $\rho = 0$ representing the lowest and $\rho = 1$ representing the highest value of θ, then we estimate the shape parameters as

$$a_\psi\left(\theta\right) = a_{\psi,0} + a_{\psi,1} \times \rho\left(\theta\right)$$
$$b_\psi\left(\theta\right) = b_{\psi,0} + b_{\psi,1} \times \rho\left(\theta\right)$$

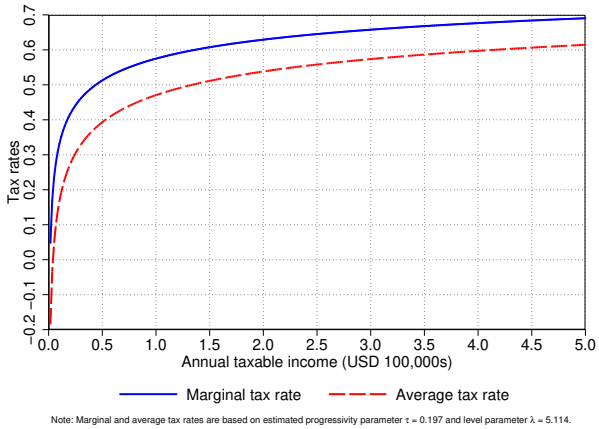
This flexible specification allows for a wide range of shapes of the ψ-distribution conditional on a given ability level, and for a different shape—including the mean and variance—of this distribution distribution across θ-levels, while retaining tractability.

Income tax and transfer function. The implied marginal tax and average tax schedules are shown in Figure 17(a). The estimated tax function provides a good fit to the relation between pre- and post-tax income in the HRS tax data, particularly for the lower 90 percent of the income distribution. Panel (b) of the figure shows the empirical fit to disposable income and taxable income.

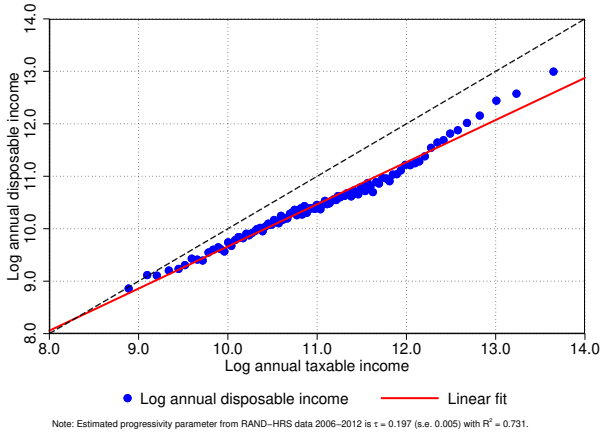
[63] We convert the discount factor ψ in to an annualized discount factor ψ_{annual} for an individual with T_W years of working life and T_R years of retirement by solving for the root of the equation $\psi = \psi_{annual}^{T_W}(1-\psi_{annual}^{T_R+1})/(1-\psi_{annual}^{T_W+1})$.

Figure 17. Estimated progressivity and empirical fit of the income tax schedule

(a) Marginal and average tax rates



(b) Empirical fit: disposable vs. taxable income

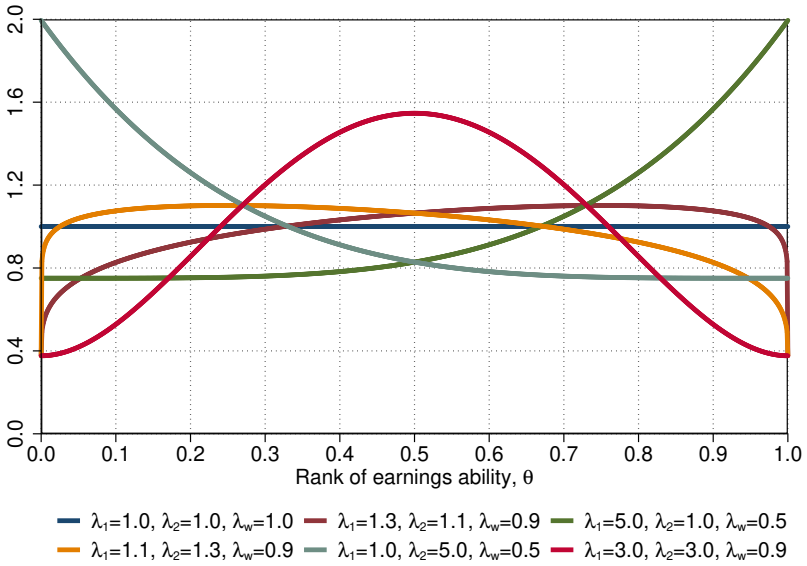


Notes: Tax function estimation follows Heathcote et al. (2017). Source: Authors' computations based on HRS.

C.4.2 Normative calibration

Flexible Pareto weights specification. Figure 18 shows the flexible Pareto weights specification we employ for our normative calibration in Section 6.2.

Figure 18. Illustration: Possible Pareto weight distributions



Notes: Figure shows examples of possible distributions of Pareto weights $\lambda(\theta) = \lambda_w \times beta(\rho(\theta), \lambda_1, \lambda_2) + (1 - \lambda_w)$, where λ_w is the relative weight on a beta distribution versus a uniform weight of 1 on all agents, $beta(\cdot, \lambda_1, \lambda_2)$ is a beta distribution with shape parameters λ_1 and λ_2, and $\rho(\theta) \in [0, 1]$ is the rank of earnings ability, with $\rho = 1$ representing the highest ability and $\rho = 0$ representing the lowest ability. Source: Authors' calculations.

Calibration targets and parameters. To pin down a distribution of Pareto weights, one must pick the three parameters $(\lambda_1, \lambda_2, \lambda_w) \in L = \mathbb{R}_{++} \times \mathbb{R}_{++} \times [0,1]$. We first obtain an updated positive model allocation by taking as given the pair (θ, β) from the positive model calibration but shutting down all other margins from the extended positive model, which are not part of our normative framework. Specifically, we shut down survival risk, longevity heterogeneity, bequest motives, and medical expenditure shocks, and minimum consumption floors. We also set the number of years of working life and the number of years of retirement to their respective median population values.

For each calibrated (θ, β)-pair, we then record the resulting allocation $(c_1(\theta,\beta), c_2(\theta,\beta), y(\theta,\beta))$ from the stripped down positive model. Finally, we compare the resulting allocation from the stripped down positive model, $\{(c_1(\theta,\beta), c_2(\theta,\beta), y(\theta,\beta))\}_{\theta,\beta}$, with the optimal allocation from the normative model under the same parameterization and social welfare weights given by $(\lambda_1', \lambda_2', \lambda_w')$, $\{(c_1^*(\theta,\beta), c_2^*(\theta,\beta), y^*(\theta,\beta))\}_{\theta,\beta}(\lambda_1', \lambda_2', \lambda_w')$.

We are now ready to calibrate social preferences. In our baseline calibration of the normative model, to reduce the number of parameters and for ease of presentation, we fix $\delta = 1/R = 0.214$, where R is the compounded gross interest rate between periods that corresponds to an annual net interest rate of 3.44% (Gourinchas and Parker, 2002). We use a grid with 1000 θ-types and 6 β-types to approximate the positive calibration results. We then calibrate $(\lambda_1, \lambda_2, \lambda_w)$ to minimize the L^2-norm between the allocations in the normative and positive versions of our model according to the following criterion:

$$(\lambda_1, \lambda_2, \lambda_w) = \underset{(\lambda_1', \lambda_2', \lambda_w') \in L}{\arg\min} \left\{ \begin{array}{c} [\ln(c_1(\theta,\beta)) - \ln(c_1^*(\theta,\beta)(\lambda_1', \lambda_2', \lambda_w'))]^2 \\ + [\ln(c_2(\theta,\beta)) - \ln(c_2^*(\theta,\beta)(\lambda_1', \lambda_2', \lambda_w'))]^2 \\ + [\ln(y(\theta,\beta)) - \ln(y^*(\theta,\beta)(\lambda_1', \lambda_2', \lambda_w'))]^2 \end{array} \right\}$$

C.5 Comparison of empirical findings to most related results in the literature

Naturally, we are not the first to study the sources of heterogeneity in wealth and savings behavior. Venti and Wise (1998) link large differences in retirement wealth back to different savings choices. Browning and Lusardi (1996) discuss several potential explanations, including life-cycle

consumption smoothing, bequest motives, and preferences over consumption streams.[64] They also consider the role of "nonstandard" (i.e., behavioral) models in explaining observed savings outcomes, suggesting that a model with multiple dimensions of heterogeneity is needed to assess the relative importance of rational or behavioral factors behind savings decisions.

Many previous works have highlighted the importance of time preference heterogeneity in explaining observed savings and wealth patterns, see for example Falk et al. (2018), Hendricks (2007), Krusell and Smith, Jr. (1998), and Carroll et al. (2017). Recent studies suggest that present bias due to hyperbolic discounting can rationalize empirical regularities in savings behavior.[65] Laibson (1997) shows that such present bias can explain realistic income-consumption patterns and portfolio allocations. Bernheim et al. (2001) evaluate competing explanations for savings heterogeneity, distinguishing between rational and behavioral savings motives. They find that consumption growth rates near retirement do not vary systematically with wealth at retirement, leading them to reject a model with differences in exponential discount factors in favor of models with present bias heterogeneity. Aguiar and Hurst (2005) show that substitution of work-related expenses for home production explains part of the observed consumption drop upon retirement. However, Bernheim and Taubinsky (2018) argue that the observed relationship between the decline in consumption and income replacement rates, and more recent evidence on personal financial choices around retirement by Olafsson and Pagel (2018) are not easily explained by the rational benchmark model.[66] Angeletos et al. (2001) calibrate a dynamic savings model with a partially liquid asset, concluding that the hyperbolic discounting model fits the data better. Laibson et al. (2017) estimate discount functions in a richer life-cycle model and reject the restriction to exponential discounting across almost all specifications.

Our estimates also fit well with a range of empirical lab and field estimates. See Augenblick et al. (2015) and Beshears et al. (2015) for laboratory experiments; Ashraf et al. (2006), Tanaka et al. (2010), Jones and Mahajan (2015), and Kaur et al. (2015) for field experiments; and Laibson et al. (1998, 2017) for natural field data estimates of present bias. The positive gradient of estimated

[64] In their words, "the improvement motive," "the avarice motive," and "intertemporal substitution motive" refer to preferences over present and future consumption (see p.1797 of Browning and Lusardi, 1996). In earlier work, Friedman (1953) concluded that "one cannot rule out the possibility that a large part of the existing inequality of wealth can be regarded as produced by men to satisfy their tastes and preference."

[65] Strotz (1955) formalizes time-inconsistent decision making and the demand for commitment devices. Phelps and Pollak (1968) develops the hyperbolic discount function that has since become popular.

[66] Note that a level shift in desired retirement wealth is not per se a problem for our analysis as our focus lies on heterogeneity. Potentially more problematic would be dispersion in the reliance on unmeasured home production, which could lead us to overestimate the variation in present bias conditional on earnings ability.

discount factors in income we find is in line with empirical correlations between lifetime income and savings rates (Dynan et al., 2004) as well as with present bias estimates in Paserman (2008), Meier and Sprenger (2015), and Lockwood (2016). Analyzing data on identical and fraternal twins, Cronqvist and Siegel (2015) find that the genetic component of savings behavior is correlated with smoking and obesity, suggesting that it may be related to behavioral issues of self-control. See also De Nardi and Fella (2017) for a comprehensive overview of dynamic models linking wealth heterogeneity to preference heterogeneity.

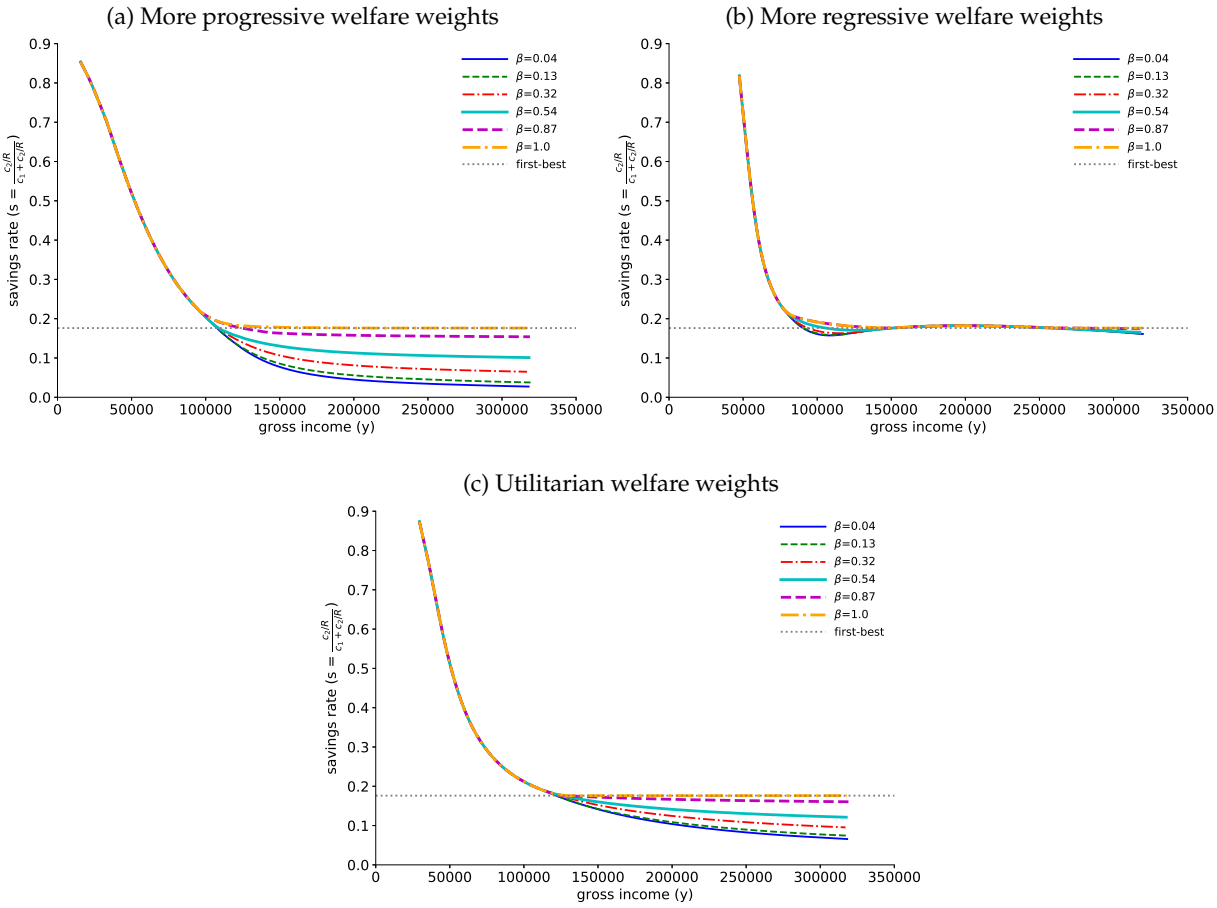
D Optimal Policies

D.1 Further results on the optimal allocation and optimal policies

D.1.1 More on the Role of Redistributive Preferences

We now compare optimal savings rates under different sets of Pareto weights. Figure 19(a) shows that under more progressive welfare weights $(\lambda_1 = 1.000, \lambda_2 = 3.000, \lambda_w = 0.499)$, optimal savings rates are even a little more spread out, particularly in the income region of USD 100,000 to USD 200,000. As before, offering more savings choice facilitates redistribution. Figure 19(b) shows that under more regressive welfare weights $(\lambda_1 = 3.000, \lambda_2 = 1.000, \lambda_w = 0.499)$, savings rate dispersion significantly shrinks, as now there is less of a redistributive motive. Figure 19(c) shows that the utilitarian benchmark lies somewhere in between these extremes.

Figure 19. Optimal savings rates under various redistributive preferences

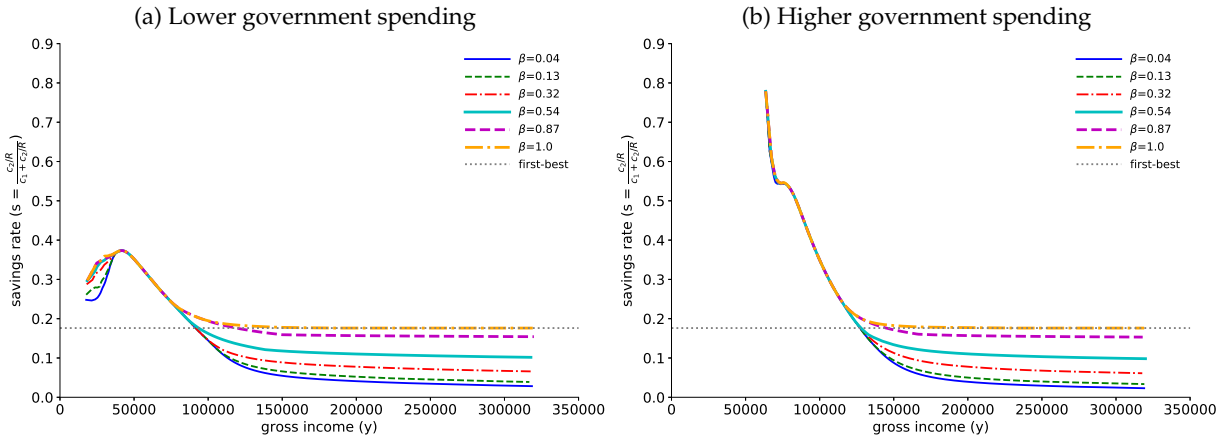


Notes: Savings rate is defined as $s = (c_2/R) \,/\, (c_1 + c_2/R)$. Source: Authors' calculations based on model.

D.1.2 The Role of the Revenue Generation Motive

We now explore optimal savings rates computed under different exogenous levels of government spending. Recall that in our benchmark calibration we set exogenous government spending to match net tax payments under the current system. Figure 20(a) shows optimal savings rates when there are no government expenditures to finance. Relative to our benchmark results, savings rates are significantly lower for incomes below USD 100,000. In contrast, in an economy with twice the benchmark amount of exogenous government spending, savings rates among low-income individuals are significantly higher than in the benchmark, as shown in Figure 20(b).

Figure 20. Optimal savings rates under various levels of exogenous government spending



Notes: Savings rate is defined as $s = (c_2/R) \,/\, (c_1 + c_2/R)$. Benchmark exogenous government spending is set to USD 40,058 per capita. Panel (b) sets exogenous government spending equal to zero, while panel (b) doubles exogenous government spending relative to the benchmark. Source: Authors' calculations based on model.