

Retirement Financing: An Optimal Reform Approach^{*}

Roozbeh Hosseini

UGA & FRB Atlanta

roozbeh@uga.edu

Ali Shourideh

Carnegie Mellon University

ashourid@andrew.cmu.edu

January 13, 2019

Abstract

We study Pareto optimal policy reforms aimed at overhauling retirement financing as an integral part of the tax and transfer system. Our framework for policy analysis is a heterogeneous-agent overlapping-generations model that performs well in matching the aggregate and distributional features of the U.S. economy. We present a test of Pareto optimality that identifies the main source of inefficiency in the status quo policies. Our test suggests that lack of asset subsidies late in life is the main source of inefficiency when annuity markets are incomplete. We solve for Pareto optimal policy reforms and show that progressive asset subsidies provide a powerful tool for Pareto optimal reforms. On the other hand, earnings tax reforms do not always yield efficiency gains. We implement our Pareto optimal policy reform in an economy that features demographic change. The reform reduces the present discounted value of net resources consumed by each generation by about 7 to 11 percent in the steady state. These gains amount to a one-time lump-sum transfer to the initial generation equal to 10.5 percent of GDP.

Keywords: retirement, optimal taxation, social security

JEL classification: H21, H55, E62

^{*}We would like to thank three anonymous referees, Laurence Ales, Tony Braun, V. V. Chari, Mariacristina De Nardi, Berthold Herrendorf, Karen Kopecky, Dirk Krueger, Ellen McGrattan, Chris Sleet, Chris Telmer, Venky Venkateswaran, Gianluca Violante (the editor), Shu Lin Wee, Sevin Yeltekin, Ariel Zetlin-Jones, and participants at various conferences and seminars for helpful comments and suggestions. We specially thank Mariacristina De Nardi, Eric French, and John Bailey Jones for sharing their data and codes. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Supplementary materials are available at https://people.terry.uga.edu/roozbeh/supp/retire_reform_supp.html.

1 Introduction

The government in the United States, and in many other developed countries, plays a crucial role in the provision of old-age consumption. In the United States, for example, a major fraction of the older population relies heavily on their social security income. Old-age benefits provided by the social security program are 40 percent of all income of older people. Moreover, these benefits are the main source of income for half of the older population.¹ On the other hand, these programs are a major source of cost for governments. In the United States, social security payouts are 30 percent of total government outlays. The severity of these costs together with an aging population has made reforms in the retirement system a necessity.

Various reforms have been proposed to reduce the cost of these programs or raise revenue to fund them. Typically, these proposals only target reform of the payroll tax and old-age benefits. Moreover, with a few exceptions, they focus on gains to future generations and often ignore the impact of reforms on current generations (see our discussion of related literature in section 1.1). While such reforms have their merit, they require interpersonal comparison of utilities and are not necessarily robust to the variety of the political arrangements through which these reforms are determined. Alternatively, one can consider Pareto improving reforms: reforms that improve everyone's welfare. It is thus important to know under what conditions Pareto improving policy reforms are feasible. Moreover, what policy instruments are essential in achieving such reforms, and how large are the efficiency gains arising from these reforms?

In this paper, we propose a theoretical and quantitative analysis of Pareto improving policy reforms which view payroll taxes, old-age benefits, etc. as part of a comprehensive fiscal policy. On the theory side, we expand on [Werning \(2007\)](#) and provide a test of Pareto optimality of a tax and transfer schedule in an overlapping-generation economy with many tax instruments (i.e., taxes on earnings and savings). We then use the theory to investigate the possibility of Pareto optimal reforms in a quantitative model consistent with aggregate and distributional features of the U.S. economy. Our main result is that earnings tax reforms are not always a major source of efficiency gains in a Pareto optimal reform, but asset subsidies play an essential role in producing efficiency gains.

We use an overlapping-generation framework in which individuals of each cohort are heterogeneous in their earning ability, mortality and discount factor. We assume those with higher earning ability have lower mortality. This assumption is motivated by the empirical research that documents a negative correlation between lifetime income and mortality (see, for example, [Cristia \(2009\)](#); [Waldron \(2013\)](#)). We also assume higher-ability individuals are more patient. The moti-

¹Social security benefits are more than 83 percent of the income for half of the older population (see Table 6 in [Poterba \(2014\)](#)).

vation for this assumption is the observed heterogeneity in savings rates across income groups (see, for example, [Dynan et al. \(2004\)](#)). This feature also allows us to match the distribution of wealth in our calibration. Finally, annuity markets are incomplete.²

Our goal is to characterize the set of Pareto optimal fiscal policies, that is, non-linear earnings tax and transfers during working age, asset taxes and social security benefits. The evaluation of fiscal policies is based on the allocations that they induce in a competitive equilibrium where economic agents face these policies. In particular, a sequence of fiscal policies is Pareto optimal if one cannot find another sequence of policies whose induced allocations deliver at least the same welfare to each type of individual in each generation at a lower resource cost.

In this environment, the key question is whether a Pareto optimal reform (henceforth “Pareto reform”) is feasible. We show that, absent dynamic inefficiencies, a Pareto reform is only possible when there are inefficiencies within each generation. In other words, determining whether a sequence of policies can be improved upon comes down to checking the same property within each generation. An important implication of this result is that Pareto improvements cannot be achieved by simply replacing distortionary tax policies. This is because in an economy with heterogeneity, distortionary taxes may be efficient, as they serve a purpose: they balance redistributive motives in a society with incentives. It is well known that the set of Pareto optimal non-linear income taxes are potentially large.³ In other words, judgment about the Pareto optimality of a tax system is not possible by simply examining the tax rates.

In order to examine the optimality of a given tax and transfer system, we extend the analysis of [Werning \(2007\)](#) to our overlapping-generations economy and derive the criteria for optimality for each generation. A tax system is optimal if it satisfies two criteria, an inequality constraint for the earnings tax schedule and a tax-smoothing relationship between various taxes (between contemporaneous earnings and savings taxes and between savings taxes over time). The inequality test of earnings taxes is standard from [Werning \(2007\)](#), and it is equivalent to the existence of non-negative Pareto weights on different individuals that rationalize the observed tax function. The novel prediction of our analysis is the tax smoothing relationship between various taxes. Together, these conditions can be tested for any tax schedule, as we do in our quantitative exercise.

Our tests imply that optimality of the asset tax schedule is tied to the incompleteness in the annuity markets and to earnings taxes. In other words, if redistributive motives inherent in observed policies are captured in earnings taxes, then the tax-smoothing relationship ties the optimal level of asset taxes to these redistributive motives (earnings taxes). This condition implies

²The private annuity market in the United States is small and plays a minor role in financing retirement. See [Poterba \(2001\)](#) and [Benartzi et al. \(2011\)](#) for surveys and our discussion in section 3.

³See [Mirrlees \(1971\)](#) and [Werning \(2007\)](#) for static examples.

that optimal asset taxes must have two components. First, they must have a subsidy component that captures the inefficiencies arising from incompleteness in annuity markets. More specifically, with incomplete annuity markets, a subsidy to savings can index asset returns to individual mortality rates and therefore complete the market. Second, optimal asset taxes must have a tax component that stems from the increasing demand for savings from more productive individuals above and beyond usual consumption-smoothing reasons. In effect, since more productive individuals have a higher valuation for consumption in the future (due to their lower mortality and higher discount factor), taxation of future consumption can relax redistributive motives by the government, which in turn leads to lower taxes on earnings. The nature and magnitude of optimal asset taxes is determined by the balance of these two effects.

With this theoretical characterization as a guide, we turn to a quantitative version of our model. Specifically, we calibrate our model economy to the status quo policies in the United States (income taxes, payroll taxes and old-age transfers), aggregate measures of hours worked and capital stock, and the distribution of earnings and wealth. Our model can successfully match the key features of the U.S. data, particularly the cross-sectional distribution of earnings and wealth.

Using this quantitative model, we first apply our Pareto optimality test to assess the optimality of the status quo policies. Our tests show that these policies fail the efficiency test described above. While the earnings tax inequality is violated, this violation only occurs at the income levels close to the social security maximum earnings cap. In fact, since marginal tax rates fall around this cap, the tax is regressive and thus fails the inequality criterion. Beside this violation, earnings taxes pass our inequality test for all other earning levels, and their deviation from optimality tests is small. On the other hand, our results show that the asset tax schedule violates our equality test at almost all ages and for all income levels. This suggests that savings tax (or subsidy) reforms—as opposed to earnings tax reforms—are a source of gains.

Next, we solve the problem of minimizing the cost of delivering the status quo welfare to each individual in each generation (i.e., the welfare associated with allocations induced by the status quo policies). The cost savings associated with this problem capture the potential efficiency gains in optimal reforms and identify the main elements of a Pareto optimal reform. This exercise confirms the results of the test: earnings taxes barely change compared to the status quo, while asset taxes are negative and progressive; that is, assets must be subsidized and asset-poor individuals must face a higher subsidy rate than asset-rich individuals.

That assets must be subsidized shows that the incompleteness in the annuity markets is the primary source of welfare gains. In addition, it shows that heterogeneity in mortality and discount rates play a secondary role in determining asset taxes. Furthermore, since in our model, poorer individuals have a higher mortality rate, they must face a higher subsidy in order for the return

on their savings to be indexed to their mortality. This effect leads to progressive subsidies.

We conduct our quantitative exercises in two forms. First, we consider the steady state of an economy with currently observed U.S. demographics. This exercise shows that asset subsidies could be significant. In particular, the average subsidy rate post-retirement is 5 percent. Overall, implementing optimal policies reduces the present value of net resources used by each cohort by 11 percent. This is equivalent to a 0.82 percent reduction in the status quo consumption of all individuals, keeping their welfare unchanged.⁴

Second, we consider an aging economy that experiences a fall in population growth and mortality (as projected by the U.S. Census Bureau). In this economy, and along the demographic transition, we solve for Pareto optimal reform policies that do not lower the welfare of any individual in any birth cohort relative to the continuation of status quo. Our numerical results concerning the transition economy confirm our main findings: assets subsidies are significant and crucial in generating efficiency gains. However, the gains for each birth cohort are smaller relative to the previous exercise. The present discount value of net resources used by each cohort in the new steady state falls by about 7 percent. We distribute all the gains along the transition path to the initial generations in a lump-sum fashion. This amounts to a one-time lump-sum transfer of about 10.5 percent of current U.S. GDP.

In order to highlight the importance of asset subsidies, we conduct another quantitative exercise in which we restrict reforms to policies that do not include asset subsidies and old-age transfers. In a sense, this is the best that can be achieved by phasing out retirement benefits and reforming payroll taxes. We find that these policies do not improve efficiency. In other words, they deliver the status quo welfare at a higher resource cost than the status quo policies. Finally, we also check the robustness of our results to the inclusion of other saving motives, namely, presence of out-of-pocket medical expenditure late in life (as emphasized by the seminal work of [De Nardi et al. \(2010\)](#)) and warm-glow bequests. Our quantitative exercises illustrate that our main findings are robust to these changes.

Asset subsidies are central to our proposed optimal policy. These subsidies resemble some of the features of the U.S. tax code and retirement system. Tax breaks for home ownership, retirement accounts (eligible IRAs, 401(k), 403(b), etc.), and subsidies for small business development are a few examples of such programs, whose estimated cost was \$367 billion in 2005 (about 2.8 percent of GDP). Moreover, these programs mostly benefit higher-income individuals.⁵ One view of our proposed optimal policy is to extend and expand such policies to include broader asset categories and, more importantly, continue during the retirement period. Our result also highlights

⁴In the steady state analysis, we do not take a stand on how these gains are distributed. For the economy in transition, gains are distributed to initial generations.

⁵See [Woo and Buchholz \(2006\)](#).

the need for progressivity in these subsidies, contrary to the current observed outcome. An important feature of the U.S. tax code is that it penalizes the accumulation of assets in tax-deferred accounts beyond the age of 70 and a half. Our analysis implies that these features are at odds with the optimal policy prescribed by our model and their removal can potentially yield significant efficiency gains.

1.1 Related Literature

Our paper contributes to various strands in the literature on policy reform. We contribute to the large and growing literature on retirement financing, most of which studies the implications of a specific set of policy proposals. For example, [Nishiyama and Smetters \(2007\)](#) study the effect of privatization of social security. [Kitao \(2014\)](#) compares different combinations of tax increase and benefit cuts within the current social security system. [McGrattan and Prescott \(2017\)](#) propose phasing out social security and Medicare benefits and removing payroll taxes. [Blandin \(2018\)](#) studies the effect of eliminating the social security maximum earnings cap.

We depart from the existing literature in two important aspects. First, we do not restrict the set of policies at the outset. Therefore, our results can inform us about which policy instrument is an essential part of a reform. As a result, we find that changing the marginal tax rates on labor earnings is not a major contributor to an optimal policy reform. Second, we focus explicitly on Pareto optimal policies and derive the condition that can inform us about the feasibility of Pareto improving policy reforms. In that regard, our paper is close to [Conesa and Garriga \(2008\)](#), who characterize a Pareto optimal reform in an economy without heterogeneity within each cohort and find Pareto optimal linear taxes (a Ramsey exercise).

Our paper is also related to a large literature on optimal policy design. The common approach in this literature is to take a stand on specific social welfare criteria and find optimal policies that maximize social welfare. For example, [Conesa and Krueger \(2006\)](#) and [Heathcote et al. \(2014\)](#) study the optimal progressivity of a tax formula for a parametric set of tax functions, while [Fukushima \(2011\)](#), [Huggett and Parra \(2010\)](#) and [Heathcote and Tsujiyama \(2015\)](#) do the same using a Mirrleesian approach that does not impose a parametric restriction on policy instruments (similar to our paper). One drawback of this approach is that it relies on the choice of the social welfare function. Consequently, the resulting policy proposals can improve efficiency while at the same time provide redistribution across individuals.⁶ Moreover, the resulting policies are conditional on a particular welfare function which might or might not be conforming to the political institutions that are determinants of government policies in a certain country. The benefit of our approach is that it does not rely on an arbitrary welfare function by providing

⁶See [Benabou \(2002\)](#) and [Floden \(2001\)](#) for the decomposition of the gains into redistribution, efficiency, and social insurance.

non-negative gains to all individuals. To the best of our knowledge, this is the first paper that proposes this approach to optimal policy reform in a dynamic quantitative setting.⁷

Our paper also contributes to the literature on dynamic optimal taxation over the life cycle. Similar to [Weinzierl \(2011\)](#), [Golosov et al. \(2016\)](#) and [Farhi and Werning \(2013b\)](#), we provide analytical expressions for distortions and summarize insights from those expressions. However, unlike these cited works, which focus on labor distortions over the life cycle, we focus on intertemporal distortions. Furthermore, we emphasize the role of policy during the retirement period, thus relating our work to [Golosov and Tsyvinski \(2006\)](#), who study the optimal design of the disability insurance system, and [Shourideh and Troshkin \(2017\)](#) and [Ndiaye \(2017\)](#), who focus on an optimal tax system that provides incentive for an efficient retirement age.

Another strand of literature our paper is related to studies the role of social security in providing longevity insurance. [Hubbard and Judd \(1987\)](#), [İmrohoroglu et al. \(1995\)](#), [Hong and Ríos-Rull \(2007\)](#) and [Hosseini \(2015\)](#) (among many others) have examined the welfare-enhancing role of providing an annuity income through social security when the private annuity insurance market has imperfections. [Caliendo et al. \(2014\)](#) point out that the welfare-enhancing role of social security in providing annuitization is limited because social security does not affect individuals' intertemporal trade-offs. In this paper, we pinpoint to the optimal distortions and policies that address this shortcoming in the system by emphasizing that any optimal retirement system (whether public, private or mixed) must include features that affect individuals' intertemporal decisions on the margin. In our proposed implementation, those features take the form of a non-linear subsidy on assets.

Finally, our paper is related to the literature on the observed lack of annuitization in the United States. [Friedman and Warshawsky \(1990\)](#) show that if one is to consider the high fees (what they referred to as “load factor”) on annuities provided in the market together with adverse selection, the standard model without bequest motives can go a long way in explaining the lack of annuitization. [Diamond \(2004\)](#) and [Mitchell et al. \(1999\)](#) point to taxes on insurance companies as well as high overhead costs (marketing and administrative costs as well as other corporate overhead) behind the high transaction costs. In particular, observing that the government cost of handling social security is much lower, [Diamond \(2004\)](#) suggests government-provided annuities—a task that our saving subsidies achieve. Our paper can be thought of as a quantitative evaluation of this idea in reforming the retirement benefit system in the United States.

The rest of the paper is organized as follows: section 2 lays out a two-period OLG framework where we provide intuition for our results; in section 3, we describe the benchmark model used in our quantitative exercise; in section 4, we calibrate the model; in section 5, we discuss our quantitative results in steady state; in section 6, we discuss reforms in an aging economy; in

⁷See [Werning \(2007\)](#) for a theoretical analysis in a static framework.

section 7, we study various robustness exercises; and in section 8 we present our conclusions.

2 Pareto Optimal Policy Reforms: A Basic Framework

In this section, we use a basic framework to provide a theoretical analysis of Pareto optimal policy reforms. In particular, we extend the static analysis in [Werning \(2007\)](#) to a dynamic OLG economy in order to characterize the determinants of a Pareto optimal policy reform.

To do so, we consider an OLG economy where the population in each cohort is heterogeneous with respect to their preferences over consumption and leisure. In particular, suppose time is discrete and indexed by $t = 0, 1, \dots$. There is a continuum of individuals born in each period. Each individual lives for at most two periods. Upon birth, each individual draws a type $\theta \in \Theta = [\underline{\theta}, \bar{\theta}]$ from a continuous distribution $H(\theta)$ that has density $h(\theta)$. This type determines various characteristics of the individual such as labor productivity, mortality risk and discount rate. We assume that an individual's preferences are represented by the following utility function over bundles of consumption and hours worked, y/θ

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

where $\beta(\theta)$ is the discount factor, $P(\theta)$ is the survival probability, θ is labor productivity, $u(\cdot)$ is strictly concave and $v(\cdot)$ is strictly convex. For simplicity, we assume that $v(\ell) = \psi \ell^{1+1/\varepsilon} / (1 + 1/\varepsilon)$ where ℓ is hours worked.

Production is done using labor and capital, with the production function given by $F(K, L)$, where K is capital and L is total effective labor; for ease of notation, $F(K, L)$ here is taken to be NDP (net domestic product). In addition, population grows at rate n , and N_t is total population at t .

Government policy is given by taxes and transfers paid during each period. Taxes and transfers in the first period depend on earnings, while in the second period, they depend on asset holdings and earnings in the first period. Thus, the individual maximization problem is

$$\max U\left(c_1, c_2, \frac{y}{\theta}\right)$$

s.t.

$$\begin{aligned} c_1 + a &= w_t y - T_y(w_t y) \\ c_2 &= (1 + r_{t+1}) a - T_a((1 + r_{t+1}) a, w_t y), \end{aligned}$$

where $r_t = F_{K,t}(K_t, L_t)$ is the net return on investment after depreciation, while $w_t = F_L(K_t, L_t)$

is the average wage rate in the economy. Note that in the above equations, we have allowed the second period taxes, $T_a(\cdot, \cdot)$, to depend on wealth and earnings, which can potentially capture a redistributive and history-dependent social security benefit formula together with taxes on assets. In addition, we have imposed incomplete annuity markets. In particular, the price of assets purchased when individuals are young is the same for all individuals and normalized to 1, even though individuals could be heterogeneous in their survival probability. This assumption is consistent with the observation that private annuity markets in the United States are very small.⁸ Finally, we assume that upon the death of an individual, his or her non-annuitized asset is collected by the government.

Given these tax functions and market structure, an allocation is a sequence of consumption, assets and effective hours distributions, and aggregate capital over time represented by $\{c_{1,t}(\theta), c_{2,t}(\theta), y_t(\theta), a_t(\theta)\}_{\theta \in \Theta}$ together with K_t and L_t , where subscript t represents the period in which the individual is born, total K_t is capital in period t and N_t is total effective hours. Such allocation is feasible if it satisfies the usual market clearing conditions:

$$N_t \int c_{1,t}(\theta) dH(\theta) + N_{t-1} \int P(\theta) c_{2,t-1}(\theta) dH(\theta) + K_{t+1} = F\left(K_t, N_t \int y_t(\theta) dH(\theta)\right) + K_t$$

$$N_t \int q_t a_t(\theta) dH(\theta) = K_t.$$

For any allocation, we refer to the utility of an individual of type θ born at t as $W_t(\theta)$. For a given set of taxes and initial stock of physical capital, we refer to the profile of utilities that arise in equilibrium as *induced* by policies T_y, T_a .

In this context, for a given policy $T_{y,t}(\cdot), T_{a,t}(\cdot, \cdot)$ and its induced welfare profile, $W_t(\theta)$, a *Pareto reform* is a sequence of policies $\hat{T}_{y,t}(\cdot), \hat{T}_{a,t}(\cdot, \cdot)$ whose induced welfare, $\hat{W}_t(\theta)$, satisfies $\hat{W}_t(\theta) \geq W_t(\theta)$ with strict inequality for a positive measure of θ 's and some t . Notice that in our definition of Pareto reforms, we allowed for policies to be time-dependent in order to have flexibility in the reforms. A pair of policies is thus said to be *Pareto optimal* if a Pareto reform does not exist.

The following proposition shows our first result about the existence of Pareto optimal reforms:

Proposition 1. (Diamond) *Consider an allocation $\{\{\hat{c}_{1,t}(\theta), \hat{c}_{2,t}(\theta), \hat{y}_t(\theta), \hat{a}_t(\theta)\}_{\theta \in \Theta}, K_t, L_t\}$ induced by a pair of policies $\hat{T}_{a,t}, \hat{T}_{y,t}$. Suppose that $r_t = F_K(K_t, L_t) - n > \gamma$ for some positive γ ; then the pair $\hat{T}_{a,t}$ and $\hat{T}_{y,t}$ is Pareto optimal if and only if, for all $t = 0, 1, \dots$*

$$\{\hat{c}_{1,t}(\theta), \hat{c}_{2,t}(\theta), \hat{y}_t(\theta)\}_{\theta \in \Theta} \in \arg \max_{c_1(\theta), c_2(\theta), y(\theta)} \int \left[y(\theta) - c_1(\theta) - \frac{P(\theta)}{1 + r_{t+1}} c_2(\theta) \right] dH(\theta) \quad (\text{P})$$

⁸In section 3, we provide a detailed discussion of the reasons behind this market incompleteness.

subject to

$$\theta \in \arg \max_{\hat{\theta}} U \left(c_1(\hat{\theta}), c_2(\hat{\theta}), \frac{y(\hat{\theta})}{\theta} \right) \quad (1)$$

$$U \left(c_1(\theta), c_2(\theta), \frac{y(\theta)}{\theta} \right) \geq W_t(\theta). \quad (2)$$

The proof can be found in the appendix.

The above proposition is an extension of the results in [Diamond \(1965\)](#) to an environment with heterogeneity and second best policies. It states that when the economy is dynamically efficient, $F_{K,t} > n$, then the possibility of a Pareto optimal reform depends on whether tax and transfer schemes exhibit inefficiencies within some generation. To the extent that dynamic efficiency seems to be the case in the data, the only possible Pareto optimal reforms can come from within-generation inefficiencies.⁹ In other words, the Pareto reform problem can be separated across generations and comes down to finding inefficiencies of policies within each generation. Note that a usual asymmetric information assumption is imposed on allocations, to reflect that not all tax policies are feasible. In particular, tax policies that directly depend on individuals' characteristics (e.g., ability types and mortality) are not available. As is well-known from the public finance literature, the set of Pareto efficient tax functions is potentially large.¹⁰ This implies that distortionary taxes (payroll, earnings, etc.) cannot necessarily be removed, since they could satisfy the condition in [Proposition 1](#).

[Proposition 1](#) and the above discussion highlight the main task at hand in finding Pareto optimal reforms: we have to characterize tax schedules, T_a and T_y , that solve problem (P). This is similar to the standard Pareto optimal tax problem as studied by [Werning \(2007\)](#) for a static economy. The difference compared to Werning's model is that the government has access to multiple instruments (i.e., tax on earnings and assets). As we establish, the fact that the government has access to multiple instruments introduces new restrictions on optimal taxes. The key implication is that Pareto optimal taxes must satisfy the following property: distortions along different margins adjusted by elasticities must be equated for all individuals of the same type. This result is akin to smoothing of distortions along different margins. The following proposition presents this result:

Proposition 2. *Consider a pair of policies \tilde{T}_y and \tilde{T}_a and suppose that it induces an allocation without bunching, i.e., $c_1(\theta)$ and $y(\theta)$ are one-to-one functions of θ . Then the pair \tilde{T}_y, \tilde{T}_a is Pareto*

⁹See [Abel et al. \(1989\)](#) for assessment of dynamic efficiency in U.S. data.

¹⁰See [Mirrlees \(1971\)](#) and [Werning \(2007\)](#).

optimal only if it satisfies:

$$[P(\theta) \tilde{\tau}_{a,t}(\theta) + 1 - P(\theta)] \frac{1}{\frac{\beta'(\theta)}{\beta(\theta)} + \frac{P'(\theta)}{P(\theta)}} = \frac{\theta}{(1 + \frac{1}{\varepsilon})} \frac{\tilde{\tau}_{l,t}(\theta)}{1 - \tilde{\tau}_{l,t}(\theta)}, \quad (3)$$

where $\tilde{\tau}_l(\theta)$ and $\tilde{\tau}_a(\theta)$ are the wedges induced by the tax schedule; and

$$\tilde{\tau}_{l,t}(\theta) = 1 - \frac{v'(y_t/\theta)}{w_t \theta u'(c_{1,t}(\theta))}, \tilde{\tau}_{a,t}(\theta) = 1 - \frac{q_t u'(c_{1,t}(\theta))}{(1 + r_{t+1}) \beta(\theta) P(\theta) u'(c_{2,t}(\theta))},$$

where the allocations are those induced by the policies.

The proof can be found in the appendix.

Equation (3) is the main dynamic implication of the test of Pareto optimality. It states that distortions to labor and assets margin must comove, holding other things constant. In other words, given any profile of labor taxes, which is determined by the profile of Pareto weights, the asset tax profile is determined by (3). Note that in (3), $P(\theta) \tilde{\tau}_{a,t}(\theta) + 1 - P(\theta)$ is the increase in government's revenue per person from a unit increase in assets of workers of type θ ,¹¹ while $\tilde{\tau}_{l,t}$ is the same thing except for earnings. As we describe below, equation (3) states that the behavioral increase¹² in government's revenue from a small increase in asset taxes for individuals of type θ must be equal to that of earnings taxes. In this sense, this result states that with two non-linear taxes, the distortions adjusted by behavioral responses must be equated across the two schedules.

To see the intuition behind (3), consider a slightly simpler model where the preferences of individuals are given by $c_1 + \beta(\theta) P(\theta) u(c_2) - v(y/\theta)$. In this formulation, there is no income effect and, therefore, the calculation of individual responses to tax perturbations is simpler. In the Appendix B.2, we show how this analysis works in a model with income effect. Starting with the tax function $T_y(y)$ and $T_a(a)$,¹³ consider the following perturbation of any tax schedule:

$$\tilde{T}_y(y) = \begin{cases} T_y(y) & y \leq y(\theta) \\ T_y(y) + d\tau(y - y(\theta)) & y \in [y(\theta), y(\theta) + \delta] \\ T_y(y) + d\tau\delta & y \geq y(\theta) + \delta \end{cases}$$

¹¹The government collects $\tilde{\tau}_{a,t}$ when the individual survives to the second period, and all of the assets when the individual dies in the second period.

¹²By behavioral increase, we mean the increase in government revenue resulting from behavioral response of individuals to a tax change. See Saez (2001) for the precise definition.

¹³For simplicity, we assume that all taxes are paid in the first period.

$$\tilde{T}_a(a) = \begin{cases} T_a(a) & a \leq a(\theta) \\ T_a(a) - d\tau(a - a(\theta)) & a \in [a(\theta), a(\theta) + \delta] \\ T_a(a) - d\tau\delta & a \geq a(\theta) + \delta \end{cases}$$

In the above perturbation, the marginal earnings tax rate for the bracket $[y(\theta), y(\theta) + \delta]$ increases by $d\tau > 0$, while the marginal asset tax rate for the bracket $[a(\theta), a(\theta) + \delta]$ decreases by $d\tau$, where $d\tau$ and δ are two small positive numbers. Note that for all types with assets higher than $a(\theta) + \delta$ and earnings higher than $y(\theta) + \delta$, this perturbation leaves their welfare, income and marginal taxes unchanged. This is because for these types, the change in tax on earnings cancels out that of the tax on assets. As for types close to θ , since only their marginal tax changes (taxes paid on their last earned unit of earnings and assets), their welfare change is second order. By the envelope theorem, the change in welfare for them is proportional to the size of the tax change, and the measure of people affected is also small. This implies that the above tax perturbation is feasible, up to a possible second order violation of the participation constraint (2); the utility of individuals close to θ changes by a small amount which leads to a second order change in welfare. Therefore, at the optimum, it should not raise government revenue. Note that the same holds for the reverse of this perturbation and, as a result, at the optimum the perturbation should keep government revenue unchanged.

Similar to Saez (2001), this perturbation can have a mechanical effect (the increase in revenue coming from the change in taxes, holding individual responses fixed), and a behavioral effect (the increase in revenue coming from the behavioral response of individuals) on government revenue. Since this tax perturbation only affects a small measure of individuals, its mechanical effect is zero. Therefore, we must have

$$\tau_l(\theta) dy(\theta) g_y(y(\theta)) = [1 - P(\theta) + \tau_a(\theta) P(\theta)] da(\theta) g_a(a(\theta)),$$

where $dy(\theta)$ is the behavioral response of earnings to an earnings tax increase of magnitude $d\tau$, and $da(\theta)$ is the response of assets to an increase in asset tax of magnitude $d\tau$. Moreover, $g_y(y(\theta))$ is the measure of individuals whose marginal earnings taxes increase, while $g_a(a(\theta))$ is the measure of individuals whose marginal asset taxes decrease.¹⁴ Some algebra, deferred to the Appendix B, shows that the above equation becomes (3).

This discussion highlights the key implication of Pareto optimality in dynamic environments where the government can impose multiple non-linear taxes along different margins. As we have argued, small offsetting perturbations of non-linear taxes preserve Pareto optimality, up to a second order effect on people whose marginal taxes are perturbed. Since these perturbations

¹⁴A crucial assumption made here is that there is no bunching of types; that is, a positive measure of types does not choose the same level of output or assets.

have offsetting mechanical effects, it must be that their behavioral effect on government's revenue must be equated. This equalization of the behavioral response across different instruments can be thought of as sort of a tax smoothing. As we show in section 3.1, in an extended version of this model the same results hold. Moreover, as our quantitative analysis establishes, the failure of this test of Pareto optimality is significant for status quo U.S. policies and leads to the main source of efficiency gains in Pareto optimal reforms.

A rewriting of (3) clarifies the main roles it plays in this model:

$$\tilde{\tau}_{a,t}(\theta) = 1 - \frac{1}{P(\theta)} + \frac{1}{P(\theta)} \left(\frac{\beta'(\theta)}{\beta(\theta)} + \frac{P'(\theta)}{P(\theta)} \right) \frac{1}{\theta \left(1 + \frac{1}{\varepsilon}\right)} \frac{\tilde{\tau}_{l,t}(\theta)}{1 - \tilde{\tau}_{l,t}(\theta)}, \quad (4)$$

The first component of the right hand side of the above formula, $1 - 1/P(\theta)$, captures the inefficiencies arising from the incompleteness of annuity markets. This reflects the fact that in the absence of annuities, a subsidy to savings can provide annuity returns and thus complete the market. We should note that even absent any heterogeneity, the market incompleteness assumption implies that $\tilde{\tau}_{a,t}$ is non-zero and equal to $1 - 1/P$ where P is the probability of survival.

The second component is more subtle and stems from the increasing demand for savings from more productive individuals above and beyond usual consumption-smoothing reasons. In effect, since more productive individuals have a higher valuation for consumption in the second period (they have a higher discount factor and a higher survival rate), taxation of second-period consumption can relax redistributive motives by the government, which in turn leads to lower taxes on earnings.¹⁵ Note that when $\beta'(\theta) = 0$ and $P(\theta) = 1$, our model becomes the model studied by [Atkinson and Stiglitz \(1976\)](#) and, as a result, the above formula becomes $\tilde{\tau}_{a,t}(\theta) = 0$; that is, savings taxes should be zero.

We should note a subtle point about forces towards progressivity of savings tax or subsidies in our setup. When income and mortality are positively correlated, i.e., $P'(\theta) > 0$, the market incompleteness component, $1 - 1/P(\theta)$, is negative and increases with θ . In other words, workers with lower productivity face a higher subsidy. This can be interpreted as a progressive subsidy on savings. This force towards “progressivity” in the subsidy on savings is independent of government's redistributive motive and purely comes from efficiency reasons. As an example, suppose that there is no government expenditure and government does not care about redistribution at all. In this case, the optimal labor income taxes are zero; $\tilde{\tau}_{l,t} = 0$, yet saving subsidies are progressive.

In addition to the above, a Pareto optimal tax system must also satisfy another condition that

¹⁵The literature on optimal taxation has typically used such an argument for positive (or non-zero) taxes on savings. However, the implied magnitudes vary across different papers. See for example [Golosov et al. \(2013\)](#), [Piketty and Saez \(2013\)](#), [Farhi and Werning \(2013a\)](#) and [Bellofatto \(2015\)](#), among many others.

is equivalent to the existence of Pareto weights. That is, for any Pareto optimal tax schedule, non-negative Pareto weights on individuals must exist so that the tax functions maximize the value of a weighted average of the utility of individuals. As shown by [Werning \(2007\)](#), the existence of such Pareto weights is equivalent to inequalities in terms of taxes, distribution of productivities and labor supply elasticities. This inequality must also be satisfied in our model:

Proposition 3. *A pair of policies \tilde{T}_y and \tilde{T}_a is efficient only if it satisfies the following relationships:*

$$1 \geq -\theta \frac{\varepsilon}{1 + \varepsilon} \frac{\tilde{\tau}_{l,t}(\theta)}{1 - \tilde{\tau}_{l,t}(\theta)} \left[\frac{h'(\theta)}{h(\theta)} + \frac{1}{\theta} + \frac{\tilde{\tau}'_{l,t}(\theta)}{\tilde{\tau}_{l,t}(\theta)(1 - \tilde{\tau}_{l,t}(\theta))} + \frac{-u''(c_{1,t}) c_{1,t}(\theta) c'_{1,t}(\theta)}{u'(c_{1,t}(\theta)) c_{1,t}(\theta)} \right] \quad (5)$$

In addition, if optimal allocations under the tax functions are fully characterized by an individual's first-order conditions, then (3) and (5) are sufficient for efficiency.

The proof is relegated to the Appendix.

The above formula implies that a tax schedule is more likely to be negative (1) the higher is the rate of change in the skill distribution, (2) the higher is the slope of the marginal tax rate, (3) the stronger is the income effect and (4) the lower is the Frisch elasticity of labor supply. These forces can be identified in (5). An important observation is that when taxes become regressive, i.e., $\tau'_l < 0$, a Pareto improving reform is more likely.¹⁶

Our analysis here points toward the key properties that can, in principle, provide sources of gain for Pareto optimal reforms. Note that given the generality of our result, our analysis will apply whether transitional issues in policies are considered or not. In other words, either taxes are inefficient, in which case one can always find a rearrangement of resources across generations and find a possible Pareto improvement, or taxes are efficient, in which case it is impossible to find such an improvement.

In what follows, we develop a quantitative model that does fairly well in matching basic moments of consumption, earnings and wealth distribution. We will use this model to test for potential inefficiencies and compute the magnitude of cost savings that Pareto optimal reforms can provide.

3 The Model

In this section, we develop a heterogeneous-agent overlapping-generations model that extends the ideas discussed in section 2 and is suitable for our quantitative policy analysis. Our description of the policy instruments is general and includes the current U.S. status quo policies as a special

¹⁶As we will see in section 5.1, the main source of inefficiency in the earnings tax schedule, albeit small, comes from the sudden drop of marginal tax rate around the social security maximum taxable earnings cap.

case. The model is rich enough and is calibrated in section 4 to match U.S. aggregate data and cross-sectional observations on earnings and asset distribution. In section 3.1, we show how this model can be used to derive Pareto optimal policies.

Demographics, Preferences and Technology Time is discrete, and the economy is populated by $J + 1$ overlapping generations. A cohort of individuals is born in each period $t = 0, 1, 2, \dots$. The number of newborns grows at rate n_t . Upon birth, each individual draws a type $\theta \in \Theta = [\underline{\theta}, \bar{\theta}]$ from a continuous distribution $H(\theta)$ that has density $h(\theta)$. This parameter determines three main characteristics of an individual: life-cycle labor productivity profile, survival rate profile, and discount factor. In particular, an individual of type θ has a labor productivity of $\varphi_j(\theta)$ at age j . We assume that $\varphi'_j(\theta) > 0$ and thus refer to individuals with a higher value of θ as more productive. Everyone retires at age R , and $\varphi_j(\theta) = 0$ for $j > R$. Moreover, an individual of type θ and of age j who is born in period t has a survival rate $p_{j+1,t}(\theta)$ (this is the probability of being alive at age $j + 1$, conditional on being alive at age j).¹⁷ Nobody survives beyond age J (with $p_{J+1,t}(\theta) = 0$ for all θ and t). As a result, the survival probability at age j for those who are born in period t is

$$P_{j,t}(\theta) = \prod_{i=0}^j p_{i,t}(\theta).$$

Additionally, an individual of type θ has a discount factor given by $\beta(\theta)$. Thus, that individual's preferences over streams of consumption and hours worked are given by

$$\sum_{j=0}^J \beta(\theta)^j P_{j,t}(\theta) [u(c_{j,t}) - v(l_{j,t})]. \quad (6)$$

Here, $c_{j,t}(\theta)$ and $l_{j,t}(\theta)$ are consumption and hours worked for an individual of θ at j who is born in period t .

We assume that the economy-wide production function uses capital and labor and is given by $F(K_t, L_t)$. In this formulation, K_t is aggregate per capita stock of capital, and L_t is the aggregate effective units of labor per capita. Effective labor is defined as labor productivity, $\varphi_j(\theta)$, multiplied by hours, $l_j(\theta)$. Its aggregate value is the sum of the units of effective labor across all

¹⁷Arguably, the assumption that mortality risk and lifetime productivity are perfectly correlated (i.e., they are controlled by the same random variable θ) is unrealistic. However, it helps us in characterizing optimal policies, especially since solving mechanism design problems with multiple sources of heterogeneity is known to be a very difficult problem.

individuals alive in each period. In other words,

$$L_t = \int \sum_{j=0}^J \mu_t(\theta, j) \varphi_j(\theta) l_{j,t}(\theta) dH(\theta),$$

where $\mu_t(\theta, j)$ is the share of type θ of age j in the population in period t . Finally, capital depreciates at rate δ . Therefore, the return on capital net of depreciation is $F_K(K_t, L_t) - \delta$.

Markets and Government We assume that individuals supply labor in the labor market and earn a wage w_t per unit of effective labor. In addition, individuals have access to a risk-free asset and cannot borrow. The assets of the deceased in each period t convert to bequests and are distributed equally among the living population in period t .¹⁸ Our main assumption here is that annuity markets do not exist. As discussed in section 2, this assumption is in line with the observed low volume of trade in annuity markets in the United States and other countries.¹⁹

The government uses non-linear taxes on earnings from supplying labor, including the social security tax, while we assume that there is a linear tax on capital income and consumption. The revenue from taxation is then used to finance transfers to workers and social security payments to retirees. While transfers are assumed to be equal for all individuals, social security benefits are not and depend on individuals' lifetime income.

Given the above market structure and government policies, each individual born in period t faces a sequence of budget constraints of the following form:²⁰

$$\begin{aligned} (1 + \tau_c) c_j + a_{j+1} &= (w_{t+j} \varphi_j l_j - T_{y,j,t+j}(w_{t+j} \varphi_j l_j) + Tr_{j,t+j}) \mathbf{1}[j < R] \\ &\quad + (1 + r_{t+j}) a_j - T_{a,j,t+j}((1 + r) a_j) + S_{j,t+j}(\mathcal{E}_t) \mathbf{1}[j \geq R] + B_{t+j}, \\ a_{j+1} &\geq 0. \end{aligned} \quad (7)$$

Here, r_{t+j} is the rate of return on assets a_{j+1} ; $T_{y,j,t}(\cdot)$, and $T_{a,j,t}(\cdot)$ are the earnings tax and asset tax functions, respectively; $Tr_{j,t}$ are transfers to working individuals; $S_{j,t}(\cdot)$ is the retirement benefit from the government; and B_{t+j} is the income earned from bequests. The dependence of

¹⁸An alternative and equivalent specification is one where the government collects all assets upon the death of individuals. Given the availability of lump-sum taxes and transfers, the way in which the assets of the deceased are allocated among the living agents does not change our results.

¹⁹See, for example, Benartzi et al. (2011), James and Vittas (2000) and Poterba (2001), among many others.

²⁰To avoid clutter, we drop the explicit dependence of individual allocations on birth year, t , whenever there is no risk of confusion.

retirement benefits on lifetime earnings is captured by \mathcal{E} , which is given by

$$\mathcal{E}_t = \frac{1}{R+1} \sum_{j=0}^R w_{t+j} \varphi_j l_j.$$

All tax functions and transfers can potentially depend on age and birth cohort (e.g., along a demographic transition).

There is a corporate tax rate τ_K paid by producers. Therefore, the return on assets, r_t , is equal to $(1 - \tau_K) (F_K(K_t, L_t) - \delta)$.²¹ We assume that the government taxes households' holding of government debt at an equal rate and, therefore, the interest paid on government debt is also r_t .

Given the above assumptions, the government budget constraint is given by

$$\begin{aligned} & \int \sum_{j=0}^J \mu_t(\theta, j) Tr_{j,t} dH(\theta) + \int \sum_{j=R+1}^J \mu_t(\theta, j) S_{j,t}(\mathcal{E}_{t-j}(\theta)) dH(\theta) + G_t + (1 + r_t) D_t = \\ & \tau_C \int \sum_{j=0}^J \mu_t(\theta, j) c_{j,t-j}(\theta) dH(\theta) + \int \sum_{j=0}^J \mu_t(\theta, j) T_{y,j,t}(w_t \varphi_j(\theta) l_{j,t-j}(\theta)) dH(\theta) + \\ & \int \sum_{j=0}^J \mu_t(\theta, j) T_{a,j,t}((1 + r_t) a_{j,t-j}(\theta)) dH(\theta) + \tau_K (F_K(K_t, L_t) - \delta) + (1 + \hat{n}_{t+1}) D_{t+1}, \end{aligned} \quad (8)$$

where G_t is per capita government purchases, D_t is per capita government debt, and \hat{n}_t is population growth rate at t , which can be calculated as a function of mortality rates and n_t . Finally, goods and asset market clearing implies

$$\int \sum_{j=0}^J \mu_t(\theta, j) c_{j,t-j}(\theta) dH(\theta) + G_t + (1 + n_{t+1}) K_{t+1} = F(K_t, L_t) + (1 - \delta) K_t, \quad (9)$$

$$\int \sum_{j=0}^J \mu_t(\theta, j) p_{j+1,t-j}(\theta) a_{j+1,t-j}(\theta) dH(\theta) = (1 + \hat{n}_{t+1}) (K_{t+1} + D_{t+1}), \quad (10)$$

$$\int \sum_{j=0}^J \mu_t(\theta, j) (1 - p_{j+1,t-j}(\theta)) a_{j+1,t-j}(\theta) dH(\theta) = (1 + \hat{n}_{t+1}) B_{t+1}. \quad (11)$$

Equilibrium The equilibrium of this economy is defined as allocations where individuals maximize (6) subject to (7), while the government budget constraint (8), market clearings (9), (10) and (11) must hold. The equilibrium is stationary (or in steady state) when all policy functions, de-

²¹We interpret the tax rate τ_K as the effective marginal corporate tax rate on capital gains that captures all the distortions caused by the corporate income tax code and capital gain taxes. Our optimal reform exercise does not contain an overhaul of the capital tax schedule. As a result, in our economy, we take as a given the after-tax interest rate earned on all types of assets.

mographics parameters, allocations and prices are independent of calendar period t .

This sums up our description of the economy. In the next section, we describe our approach to analyzing an optimal reform within the framework specified above. Note that we have not specified any details of the status quo policies yet. We will do that in section 4 where we impose detailed parametric specifications of the U.S. tax and social security policies and calibrate this model to the U.S. data. We can then apply our optimal reform approach to the calibrated model and conduct our optimal reform exercise.

When the tax function and social security benefits are calibrated to those for the United States, we refer to the resulting equilibrium allocations and welfare as *status quo* allocations and welfare. We refer to the status quo welfare of an individual of type θ who is born in period t by $W_t^{sq}(\theta)$.

Remark on Annuity Markets Throughout the analysis in this paper, we assume that there are no markets for annuities. This is in line with the observed lack of annuitization in the United States. As [Poterba \(2001\)](#), [Benartzi et al. \(2011\)](#), and many others have mentioned, the annuity market in the United States is very small. According to [Hosseini \(2015\)](#)’s calculation based on HRS, only 5 percent of the elderly hold private annuities in their portfolio.²² Moreover, the offered annuities have very high transaction costs and low yields (see [Friedman and Warshawsky \(1990\)](#) and [Mitchell et al. \(1999\)](#)), and are not effectively used by individuals (see [Brown and Poterba \(2006\)](#)).²³

Various reasons have been proposed as leading to lack of annuitization in the United States: the presence of social security as an imperfect substitute, adverse selection in the annuity market, low yields on offered annuities due to overhead and other costs, bequest motives and complexity of choice faced by individuals (see [Benartzi et al. \(2011\)](#) and [Diamond \(2004\)](#)). All of these reasons warrant government intervention in annuity markets. In our paper, we have focused on the extreme case where the government fully takes over the annuity market. This role for the government is also discussed in detail by [Diamond \(2004\)](#). It would be interesting to study the case where annuity markets are present and government intervention crowds out the private market. This, however, is beyond the scope of our paper.

3.1 Optimal Policy Reform in the Quantitative Framework

Our optimal policy-reform exercise is very similar to the one in the two-period model provided in section 2. It builds on the positive description of the economy in section 3. In particular, we

²²Furthermore, private annuities make up only 0.5 percent of the portfolio of people over 65 years of age.

²³As discussed by [Brown and Poterba \(2006\)](#), the asset class called “variable annuities” has the option of conversion to life annuities during retirement. In practice, most individuals do not convert. As a result, they do not provide insurance against longevity risk.

use the distribution of welfare implied by the model in section 3 and consider a planning problem that chooses policies in order to minimize the cost of delivering this distribution of welfare, the status quo utility profile $\{W^{sq}(\theta)\}_{\theta \in \Theta}$, to a particular representative cohort of individuals. We show how the efficiency tests discussed in section 2 extend to the dynamic environment. For simplicity, we assume steady state and do not consider the changes in prices resulting from the reforms. Later, in our quantitative exercise, we allow for both transitions and changes in prices.

3.1.1 A Planning Problem

The set of policies that we allow for in our optimal reform are very similar to those described in section 3. In particular, we allow for non-linear and age-dependent taxation of assets. Moreover, we allow for non-linear and age-dependent taxation of earnings together with flat social security benefits (i.e., social security benefits are independent of lifetime earnings). Therefore, given any tax and benefit structure, each individual maximizes utility (6) subject to the budget constraints (7).

The planning problem associated with the optimal reform finds the policies described above to maximize the net revenue for the government (i.e., present value of receipts net of expenses). In this maximization, the government is constrained by the optimizing behavior by individuals—as described above, the feasibility of allocations and the requirement that each individual’s utility must be above $W^{sq}(\theta)$. We also focus on the steady state problem for the government and ignore issues related to transition.

Using standard techniques, in the Appendix we show that the problem of finding Pareto optimal reforms for each generation²⁴ can be written as a planning problem and in terms of allocations. This planning problem maximizes the revenue from delivering an allocation of consumption and labor supply over the life of a generation subject to an implementability constraint and a minimum utility requirement given by

$$\max \int \sum_{j=0}^J \frac{P_j(\theta)}{(1+r)^j} [\varphi_j(\theta) l_j(\theta) - c_j(\theta)] dH(\theta) \quad (\text{P1})$$

²⁴Note that Proposition 1 from section 2 applies here, and we only need to consider the Pareto reform within each generation.

subject to

$$U(\theta) = \sum_{j=0}^J \beta(\theta)^j P_j(\theta) [u(c_j(\theta)) - v(l_j(\theta))] \quad (12)$$

$$U'(\theta) = \sum_{j=0}^J \beta(\theta)^j P_j(\theta) \frac{\varphi'_j(\theta) l_j(\theta)}{\varphi_j(\theta)} v'(l_j(\theta)) \quad (13)$$

$$+ \sum_{j=0}^J \left(\frac{j\beta'(\theta)}{\beta(\theta)} + \frac{P'_j(\theta)}{P_j(\theta)} \right) \beta^j P_j(\theta) [u(c_j(\theta)) - v(l_j(\theta))] \\ U(\theta) \geq W^{sq}(\theta). \quad (14)$$

The objective in the above optimization problem is equal to the present discounted value of government tax receipts net of outlays from a given cohort of individuals.²⁵

3.1.2 Test of Pareto Optimality

Given our environment and the optimal reform problem described above, we can provide tests of Pareto optimality. Note that as before, the labor and savings wedge is defined as

$$\tau_{l,j}(\theta) = 1 - \frac{w\varphi_j(\theta) u'(c_j(\theta))}{v'(l_j(\theta))}, \tau_{a,j+1}(\theta) = 1 - \frac{u'(c_j(\theta))}{(1+r)\beta(\theta)p_j(\theta)u'(c_{j+1}(\theta))}, \quad (15)$$

where w and r are steady state wage and interest rate, respectively.

The following proposition presents tests of Pareto optimality for the quantitative framework developed above:

Proposition 4. *The wedges induced by any Pareto optimal tax schedule must satisfy the following equality constraints*

$$\frac{1 - p_{j+1}(\theta)(1 - \tau_{a,j+1}(\theta))}{\frac{p'_{j+1}(\theta)}{p_{j+1}(\theta)} + \frac{\beta'(\theta)}{\beta(\theta)}} = \frac{\tau_{l,j}(\theta)}{1 - \tau_{l,j}(\theta)} \frac{1}{\frac{\varphi'_j(\theta)}{\varphi_j(\theta)} \left(1 + \frac{1}{\varepsilon}\right)}, \quad (16)$$

$$\frac{1 - p_{j+1}(\theta)(1 - \tau_{a,j+1}(\theta))}{\frac{p'_{j+1}(\theta)}{p_{j+1}(\theta)} + \frac{\beta'(\theta)}{\beta(\theta)}} = \frac{1 - p_j(\theta)(1 - \tau_{a,j}(\theta))}{\frac{p'_j(\theta)}{p_j(\theta)} + \frac{\beta'(\theta)}{\beta(\theta)}} \frac{1}{p_j(\theta)(1 - \tau_{a,j}(\theta))}, \quad (17)$$

²⁵Our planning problem is related to the one solved by [Huggett and Parra \(2010\)](#). There, the authors take the present discounted value of tax and transfers to a generation in the status quo economy as a given and find an allocation that maximizes the utilitarian social welfare function that costs no more than the status quo allocation (in terms of present discounted value of net transfers to a generation). Our planning problem, instead, takes the distribution of welfare in the status quo economy as a given and finds the least costly way of delivering that welfare.

as well as the following inequality

$$1 \geq -\frac{\tau_{l,0}}{1 - \tau_{l,0}} \frac{\varepsilon}{1 + \varepsilon} \frac{\varphi_0}{\varphi'_0} \left(\frac{1}{\sigma_0} \frac{c'_0}{c_0} + \frac{\tau'_{l,0}}{\tau_{l,0} (1 - \tau_{l,0})} + \frac{\varphi'_0}{\varphi_0} - \frac{\varphi''_0}{\varphi'_0} + \frac{h'}{h} \right) \quad (18)$$

where σ_0 is the intertemporal elasticity of substitution at age 0.

Moreover, when the first order conditions are sufficient for describing the behavior of consumers, the above conditions are also sufficient for any Pareto optimal tax schedule.²⁶

The above formulas are the equivalents of the optimality formulas in section 2. In particular, as it might be apparent, equations (16) and (17) are the equivalents of the tax smoothing relationship (3).

Equation (16) is identical to (3) and has a similar intuition. It states that at the optimum the behavioral response of the government revenue to a perturbation of marginal tax rate of earnings should be equal to that of asset taxes at every age and for every type. Note that in the dynamic model, a perturbation of any tax rate affects all margins—assets and earnings over the life cycle—through an income effect. As a result, writing the precise equation that describes the behavioral increase in government revenue in response to a tax increase is rather cumbersome. Nevertheless, (16) is identical to its equivalent in the two-period model and therefore, the intuition behind it is the same. Similar to the discussion in section 2, (16) can be written as

$$\tau_{a,j+1}(\theta) = 1 - \frac{1}{p_{j+1}(\theta)} + \frac{\tau_{l,j}(\theta)}{1 - \tau_{l,j}(\theta)} \frac{1}{\frac{\varphi'_j(\theta)}{\varphi_j(\theta)} \left(1 + \frac{1}{\varepsilon}\right)} \left(\frac{p'_{j+1}(\theta)}{p_{j+1}(\theta)} + \frac{\beta'(\theta)}{\beta(\theta)} \right)$$

which again identifies the main forces that lead to distortion of savings, i.e., market incompleteness and higher saving demand by more productive individuals.

Similarly, equation (17) states that the behavioral increase in government revenue must be equated for an increase in taxes at age j and age $j + 1$. Note that since these tax perturbations are done in two different ages, in order for them to be welfare neutral, the magnitude of the perturbations must be adjusted. The last term in (17) performs this age adjustment; by definition of the savings wedge, $(p_j (1 - \tau_{a,j}))^{-1} = \beta (1 + r) u_{c,j} / u_{c,j-1}$. In other words, the age adjustment is equal to the ratio of marginal utilities of consumption adjusted by the interest rate. Finally, as before, inequality (18) is equivalent to the non-negativity of the implied Pareto weights.

Equation (17) is informative about the behavior of savings wedges—marginal tax rates—over time. Specifically, it highlights the role of the changes in the gradient of the survival over the life cycle, i.e., $p'_j(\theta) / p_j(\theta)$. An increase in this gradient with age leads to a decline in subsidies

²⁶The necessary conditions above also hold for general disutility of labor; see proofs in Appendix A for general results.

or increase in taxes on savings. This goes back to a mechanism that we have already discussed in section 2: a higher gradient of survival leads to a higher value of consumption late in life and higher demand for saving. Taxation of savings is then a way to prevent productive individuals from earning less and saving less and, as a result, reducing the deadweight loss of taxation of earnings. From an alternative perspective, when the gradient of survival is high, saving increases quickly with productivity, which in turn reduces the density of number of workers at a certain asset level. Hence, the tax smoothing intuition in section 2 would imply that taxes (subsidies) must be higher (lower). As we show in our quantitative exercise, the gradient of survival is positive and increases with age. This would imply that subsidies become more progressive as individuals age.

Another way of getting an understanding about Pareto optimality tests is to use (16) and rewrite (17) as

$$\frac{\tau_{l,j+1}(\theta)}{1 - \tau_{l,j+1}(\theta)} \frac{1}{\frac{\varphi'_{j+1}(\theta)}{\varphi_{j+1}(\theta)}} = \frac{\tau_{l,j}(\theta)}{1 - \tau_{l,j}(\theta)} \frac{1}{\frac{\varphi'_j(\theta)}{\varphi_j(\theta)}} \frac{1}{p_{j+1}(\theta) (1 - \tau_{a,j+1}(\theta))} \quad (19)$$

This relationship can be thought of as the tax-smoothing relationship between earnings taxes across two ages. In our quantitative exercise, we use this relationship to shed light on Pareto optimality of earnings taxes.

Finally, note that whether the above tests are satisfied determines the Pareto optimality of a tax system. Away from the optimum, i.e., when these conditions are violated, in general it is difficult to determine what margins must be adjusted. In our numerical simulations, however, often the magnitude of the violation is indicative of the significance of the reform. In other words, when the above conditions are violated significantly, the gains from a reform in the associated policies are significant. In what follows, we show that the main source of violation is the equations associated with savings wedges and, thus, they are the main source of gains.

3.1.3 Optimal Taxes

So far, we have mainly focused on optimal allocations and wedges. It is possible to construct taxes whose marginals coincide with the wedges described above. In the supplementary appendix, we provide a monotonicity condition which if satisfied implies the existence of tax functions that implement the efficient allocation. This monotonicity condition is a condition on allocations that result from the planning problem. While we have no way of theoretically checking that the monotonicity conditions are satisfied, our numerical simulations always involve a check that ensures that they are. Needless to say, in all our simulations the monotonicity constraints are satisfied.

Furthermore, while in most of our analysis we focus on taxes on savings and earnings, it is possible to think about alternative implementations of efficient allocations. For example, another implementation of efficient allocations is via earnings taxes and social security benefit formulas that are indexed to income as well as assets. If the goal is to provide a progressive asset subsidy, this indexation must occur so that an increase in households' saving increases their retirement benefit, while this indexation must be progressive, i.e., it must be higher for workers with lower income and assets. For an asset tax, the implication is similar.²⁷ An alternative way of implementing this policy is to use a personalized non-linear consumption tax that varies with age. A declining (increasing) consumption tax can then replicate the saving subsidies (taxes).

4 Calibration

In this section, we calibrate the model described in section 3 by choosing parametric specifications and parameter values. We will estimate some of the parameters independently (e.g., wage/productivity profiles or mortality profiles), and we choose the rest of the parameters (e.g., discount factor) so that the model matches targets from the U.S. data.

Earning Ability Profiles We assume that individual productivity $\varphi_j(\theta)$ at age j can be written as

$$\varphi_j(\theta) = \theta \tilde{\varphi}_j,$$

where θ is an individual fixed effect, while $\tilde{\varphi}_j$ is an age dependent productivity shifter given by

$$\log \tilde{\varphi}_t = \xi_0 + \xi_1 \cdot j + \xi_2 \cdot j^2 + \xi_3 \cdot j^3.$$

To estimate the productivity parameters, we follow a large part of the literature (e.g., [Altig et al. \(2001\)](#), [Nishiyama and Smetters \(2007\)](#) and [Shourideh and Troshkin \(2017\)](#)) and use the effective reported labor earnings per hour in Panel Study of Income Dynamics as a proxy for $\varphi_j(\theta)$. We calculate this as the ratio of all reported labor earnings to total reported hours. For labor earnings, we use the sum over a list of variables on salaries and wages, separate bonuses, the labor portion of business income, overtime pay, tips, commissions, professional practice or trade payments and other miscellaneous labor income converted to constant 2000 dollars. In order to avoid well-known issues in the raw data, we use [Heathcote et al. \(2010\)](#)'s version of the PSID data. The resulting estimated parameters are $\xi_0 = 0.879$, $\xi_1 = 0.1198$, $\xi_2 = -0.00171$ and $\xi_3 = 7.26 \times 10^{-6}$.

²⁷In our model, since there is no aggregate shock and everyone has access to the same type of assets, indexation or subsidizing saving is very simple. Its implication for the real world application, however, is not necessarily obvious. The easiest way of implementing it is to use the saving in retirement accounts, 401(k)'s and IRA's.

Moreover, we assume the type-dependent fixed effect θ has a Pareto-lognormal distribution with parameters $(\mu_\theta, \sigma_\theta, a_\theta)$. This distribution approximates a lognormal distribution with parameters μ_θ and σ_θ at low incomes and a Pareto distribution with parameter a_θ at high values. It therefore allows for a heavy right tail at the top of the ability and earnings distribution. For this reason, it is commonly used in the literature (see [Golosov et al. \(2016\)](#), [Badel and Huggett \(2014\)](#) and [Heathcote and Tsujiyama \(2015\)](#)).²⁸ We choose the tail parameter and variance parameter to be $a_\theta = 3$ and $\sigma_\theta = 0.6$, respectively. The location parameter is set to $\mu_\theta = -1/a_\theta$ so that $\log \theta$ has mean 0. With these parameters, the cross-section variance of log hourly wages in the model is 0.36. Also, the ratio of median hourly wages to the bottom decile of hourly wages is 2.3. These statistics are consistent with the reported facts on the cross-section distribution of hourly wages in [Heathcote et al. \(2010\)](#).

Demographics and Mortality Profiles Population growth n_t is constant and is equal to 1 percent. The model period is 1 year. Individuals start earning income at age 25, they all retire at age 65, and nobody survives beyond 100 years of age. Each individual has a Gompertz force of mortality

$$M_j(\theta) = \frac{\eta_0}{\theta^{\eta_1}} \left(\frac{\exp(\eta_2 j)}{\eta_2} - 1 \right). \quad (20)$$

The Gompertz distribution is widely used in the actuarial literature (see, for example, [Horiuchi and Coale \(1982\)](#)) and economics (see, for example, [Einav et al. \(2010\)](#)). The second term in equation (20) determines the changes in mortality by age and is common across all types. The first term is decreasing in θ and determines the gradient of mortality in the cross section. Therefore, a higher-ability person has a lower mortality at all ages. The key parameter is η_1 , which determines how mortality varies with ability. To choose this parameter, we use data on the male mortality rate across lifetime earnings deciles reported in [Waldron \(2013\)](#). She uses Social Security Administration data to estimate mortality differentials at ages 67–71 by lifetime earnings deciles. Table 1 shows the estimated annual mortality rates for 67- to 71-year-old males born in 1940.²⁹ This piece of evidence points to large differences in death rates across different income groups, with the poorest deciles almost 4 times more likely to die than the richest decile. We use this data to calibrate parameter η_1 .

Parameter η_2 is chosen to match the average survival probability from cohort life tables for the Social Security area by year of birth and sex for males of the 1940 birth cohort (table 7 in [Bell and](#)

²⁸See [Reed and Jorgensen \(2004\)](#) for more details on Pareto-lognormal distribution, its properties and relation to other better-known distributions.

²⁹[Waldron \(2013\)](#) estimates the death rates by lifetime earnings deciles for a sample of fully insured individuals. To make sure the total death rates add up to the population death rates, we adjust the reported number by the population mortality rate for the 1940 birth cohort.

Table 1: Death rates by lifetime earnings deciles for males age 67–71

	Lifetime Earnings Deciles ^a									
	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Deaths (per 10000)	369	307	286	205	204	211	204	167	142	97

^asource: Table A-1 in Waldron (2013) (adjusted by average mortality rate of 1940 birth cohort).

Miller (2005)). Finally, η_0 is chosen so that mortality at age 25 is 0. The parameters that give the best fit to the mortality data in Table 1 and average mortality data are $\eta_0 = 0.0006$, $\eta_1 = 0.5545$ and $\eta_2 = 0.0855$. Figure 1 shows the fit of the model in terms of matching mortality across the lifetime earnings deciles in Waldron (2013). Once we have the mortality hazard $M_j(\theta)$, we can find the survival probability $P_j(\theta) = \exp(-M_j(\theta))$.

Using this parametrization, we find there are 4 workers per each retiree in the steady state. This is consistent with U.S. Census Bureau estimates.³⁰

Preferences and Technology We assume a constant relative risk aversion over consumption, $u(c) = \frac{c^{1-\sigma}-1}{1-\sigma}$, and constant Frisch elasticity for disutility over hours worked, $v(l) = \psi \frac{l^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}$. The risk aversion parameter is $\sigma = 1$, and the elasticity of labor supply is $\varepsilon = 0.5$. The weight of leisure in utility ψ is chosen so that, in the model, the average number of annual hours worked is 2000.

To capture the heterogeneity in the discount factor across different-ability types, we assume

$$\beta(\theta) = \beta_0 \cdot \theta^{\beta_1}.$$

We choose β_0 to match a capital to output ratio of 4.³¹ The other parameter, β_1 , determines the degree of heterogeneity in the discount factor. The larger β_1 , the larger the dispersion in the discount factor across ability types. We choose this parameter to match the wealth Gini index of 0.78 based on the 2007 Survey of Consumer Finances (SCF).

The aggregate production function is Cobb-Douglas with a capital share parameter $\alpha = 0.435$, and the depreciation rate is $\delta = 0.048$. These are chosen to match the average ratio of capital income and investment relative to GDP in the U.S. over the period 2000–2010.

³⁰These estimates can be found here: <http://www.census.gov/population/projections/data/national/2014.html>

³¹Our measure of capital includes fixed private and government assets, consumer durables, inventories and land. The average value of capital relative to GDP over the 2000–2010 period is 4.07. On the other hand, the average value of total non-financial assets (household and non-profits, non-financial corporates, non-financial non-corporates and government) relative to GDP over the same period is 3.97. A more detailed discussion is provided in the supplementary appendix.

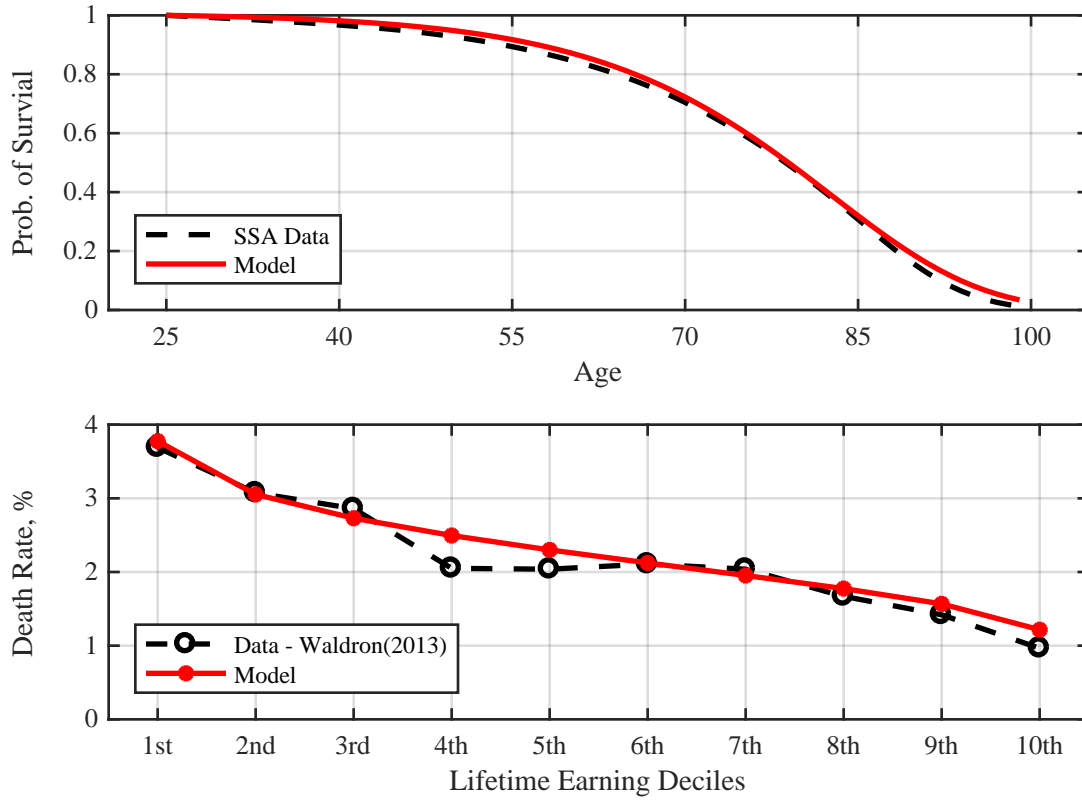


Figure 1: Fit of the mortality model. The top panel shows the average survival probability in the model vs. social security data. The bottom panel shows death rates at age 67 in the model vs. those reported in [Waldron \(2013\)](#).

Social Security Social security taxes are levied on labor earnings, up to a maximum taxable, as in the actual U.S. system. Benefits are paid as a nonlinear function of the average taxable earnings over lifetime.³² Let e be labor earnings and e_{max} be maximum taxable earnings. We set e_{max} equal to 2.47 times the average earnings in the economy, \bar{E} . The social security tax rate is $\tau_{ss} = 0.124$.³³ There is also a Medicare tax rate, $\tau_m = 0.029$, which applies to the entire earnings.

Each individual's benefits are a function of that individual's average lifetime earnings (up to e_{max}). We use the same benefit formula that the U.S. Social Security Administration uses to determine the primary insurance amount (PIA) for retirees:

³²The Social Security Administration uses only the highest 35 years of earnings to calculate the average lifetime earnings. We use the entire earnings history, for easier computation.

³³We account for disability insurance tax and benefits by aggregating them with social security.

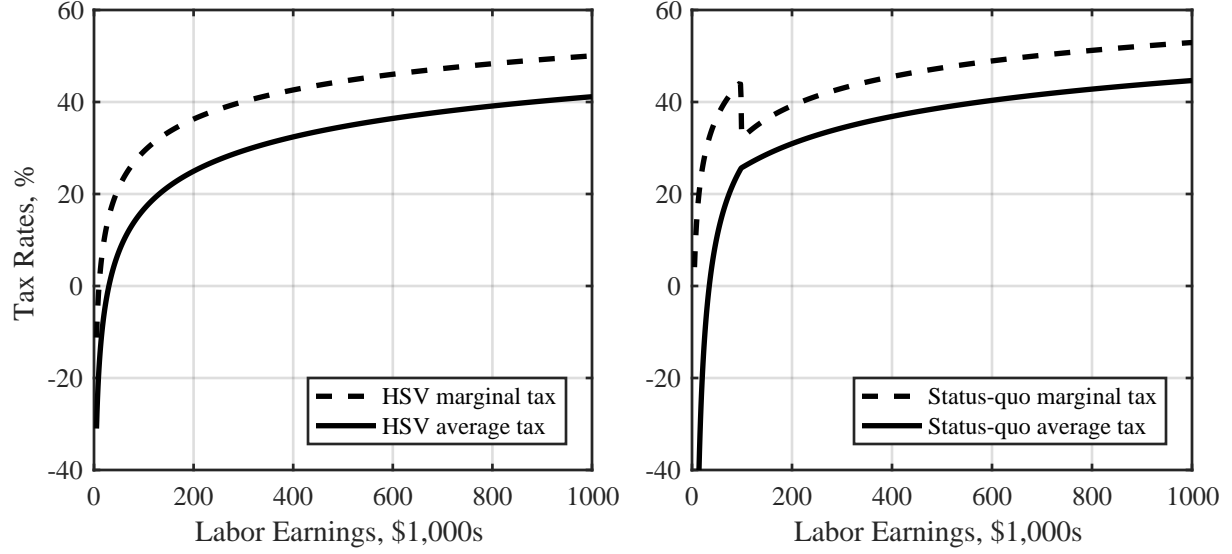


Figure 2: Tax functions. The left panel is the calibrated HSV tax function, $\mathcal{T}_{HSV}(\cdot)$. The right panel is the effective tax function (including HSV tax, payroll tax and transfers). The discontinuity is due to the social security cap on taxable earnings.

$$S^{sq}(\mathcal{E}) = \begin{cases} 0.9 \times \mathcal{E} & \mathcal{E} \leq 0.2\bar{E} \\ 0.18\bar{E} + 0.33 \times (\mathcal{E} - 0.2\bar{E}) & 0.2\bar{E} < \mathcal{E} \leq 1.24\bar{E} \\ 0.5243\bar{E} + 0.15 \times (\mathcal{E} - 1.24\bar{E}) & \mathcal{E} > 1.24\bar{E} \end{cases}$$

To account for Medicare benefits, we assume each individual in retirement will receive an additional transfer independent of that individual's earnings history. We choose this value so that the aggregate Medicare benefits are 3 percent of GDP.³⁴

Income Taxes and Government Purchases In addition to social security, the government has an exogenous spending G , which we assume to be 8 percent of GDP.³⁵ For the income tax function, we use

$$\mathcal{T}_{HSV}(y) = y - \lambda y^{1-\tau},$$

where y is taxable income. During the working age, the taxable income for each individual is $w\varphi_j(\theta)l_j(\theta) - 0.5T_{ss}$, in which $w\varphi_j(\theta)l_j(\theta)$ is labor earnings and T_{ss} is the social security and

³⁴Our analysis abstracts from the health expenditure risks that this program helps to insure. In this regard, it is similar to [Huggett and Ventura \(1999\)](#). Our approach can be applied to a more detailed model that includes these risks as well as a more detailed model of Medicare benefits. We leave this for future research. However, in section 7.3 we consider a model with exogenous out-of-pocket medical expenditures.

³⁵This is the sum of all government consumption expenditures on national defense, general public service, public order and safety, and economic affairs in the NIPA Table 3.16. We use the average over the period 2000 to 2010. A more detailed discussion is provided in the supplementary appendix.

Table 2: Parameters chosen outside the model

Parameter	Description	Values/source
Demographics		
J	maximum age	75 (100 years old)
R	retirement age	40 (65 years old)
n	population growth rate	0.01
$M_j(\theta)$	mortality hazard	see text
Preferences		
σ	risk aversion parameter	1
ε	elasticity of labor supply	0.5
Labor productivity		
σ_θ	Pareto-lognormal variance parameter	0.6
a_θ	Pareto-lognormal tail parameter	3
μ_θ	Pareto-lognormal location parameter	-0.33
Technology		
α	capital share	0.435
δ	depreciation rate	0.048
Government policies		
τ_{SS}, τ_m	social security and Medicare tax rates	0.124, 0.029
S^{sq}	social security benefit formula	see text
τ_c	consumption tax	0.055
τ, λ	parameters of income tax function	0.151, 4.74
G	government purchases	8% of GDP
D	government debt	47% of GDP

Note: For details on the calculation of capital share, government expenditure and government debt, see supplementary appendix.

Medicare payroll taxes that the worker pays. The second term reflects the effective tax credit individuals get for the portion of social security tax paid by their employers. We assume retirement benefits are not taxed.

The tax function of this form is extensively used to approximate the effective income taxes in the United States. The parameter τ determines the progressivity of the tax function, while λ determines the level (the lower λ is, the higher are the total tax revenues for a given τ). [Heathcote et al. \(2014\)](#) estimate a value of 0.151 for τ , based on PSID income data and income tax calculations using NBER's TAXSIM program. We use their estimated value for τ and choose λ . We refer to this tax function as HSV tax function. The left panel in Figure 2 illustrates the resulting marginal and average taxes as functions of annual earnings in constant 2000 dollars.

Finally, we assume the government debt is 47 percent of GDP.³⁶ The transfers Tr are chosen

³⁶This is the sum of the state and local municipal securities and federal treasury securities. We use the average over the period from 2000 to 2010. A more detailed discussion is provided in the supplementary appendix.

Table 3: Parameters calibrated using the model

Parameters	Description	Values	
β_0	discount factor: level	0.976	
β_1	discount factor: elasticity w.r.t θ	0.014	
ψ	weight on leisure	0.675	
Targeted Moments		Data	Model
Capital-output ratio		4.00	4.00
Wealth Gini (SCF (2007))		0.78	0.78
Average annual hours		2000	2000

such that the government budget constraint (equation (8)) is satisfied in stationary equilibrium.

To summarize, individuals pay three different types of taxes on their earnings: HSV nonlinear tax, social security payroll tax (subject to a maximum taxable cap), and Medicare tax. In addition, they receive the transfer Tr prior to retirement. The right panel in Figure 2 shows the resulting marginal and average tax on the sum of all these taxes and transfers. The discontinuity in the marginal tax is due to social security’s maximum taxable earnings cap. In addition to earnings taxes, we assume that there is a proportional tax on consumption. This tax allows us to match the government’s balance sheet. In particular, part of the government’s revenue comes from consumption tax, which is not captured by the earnings tax and transfers, as estimated by [Heathcote et al. \(2014\)](#). In our steady-state analysis, the value of this consumption tax, represented by τ_c , is fixed and is set to 5.5 percent, as calculated in [Mendoza et al. \(1994\)](#). In our analysis of the economy under transition, we assume that this consumption tax is increased to finance the increase in the retirement benefits paid out by the government.

Finally, we assume that there is a (corporate) capital income tax of 33 percent, which is paid by the firms. We assume that this rate is fixed and remains unchanged under the reform. As a result, the implied after tax return on all assets is $r = 0.0405$ and is the same for everyone.³⁷ This is also the interest rate that the government pays on its debt. We assume that households do not pay any tax on their savings—beyond the corporate income tax. In general, measurement of savings taxes in the cross-section is very difficult.³⁸ This is because of the vast differences in the tax code in the treatment of different types of savings. In reality, a significant fraction of savings are held in tax-deferred retirement accounts³⁹ (which are tax deductible and are treated as income during retirement), whose tax treatment is possibly progressive. On the other hand, richer individuals

³⁷This is consistent with the average real return of stocks and long-term bonds over the period 1946–2001, as reported in [Siegel and Coxe \(2002\)](#), Tables 1-1 and 1-2.

³⁸A series of papers in the ’80s and ’90s tried to measure savings taxes faced by households—see the survey by [Sørensen \(2004\)](#). Unfortunately, there are large differences between estimates depending on methodology, and it is hard to find agreement on the sign of these taxes and on their progressivity.

³⁹See [McGrattan and Prescott \(2017\)](#).

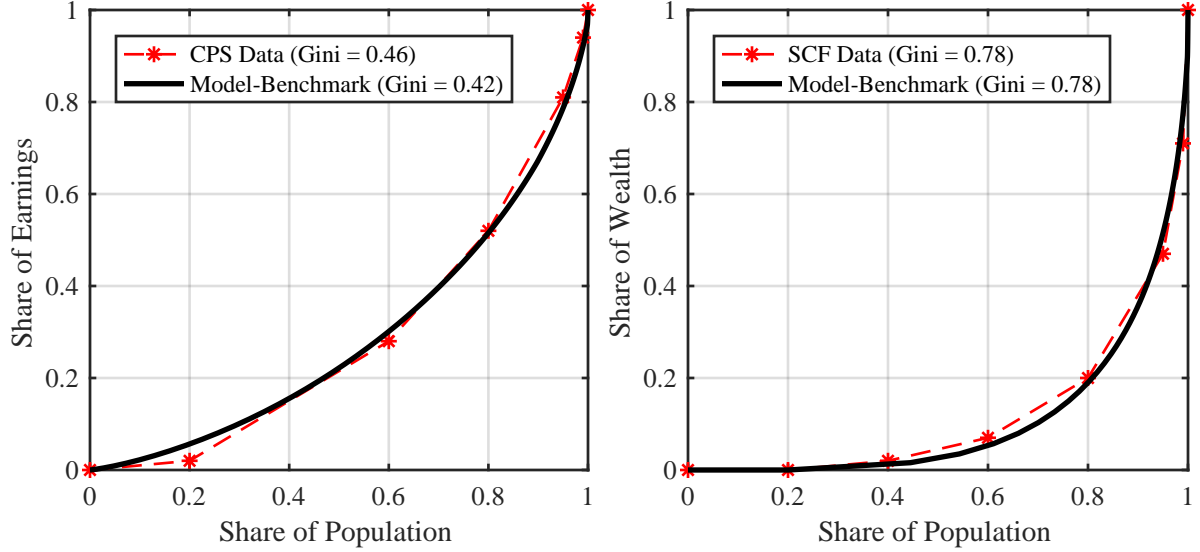


Figure 3: Fit of the distribution of earnings (left panel) and wealth (right panel).

who hold stocks and bonds can have more sophisticated strategies to minimize their tax burden. These facts motivate us to use no savings taxes in our benchmark calibration. In order to check the robustness of this assumption, we provide two robustness exercises in the supplementary appendix: an exercise with a flat and positive savings tax and one with a progressive savings tax.⁴⁰

Calibration Results Table 2 lists the parameters that are either taken from other studies, or estimated or calculated independent of the model structure. Their sources and estimation or calculation procedures are outlined in the previous paragraphs. Table 3 lists the parameters that are calibrated using the model by matching some moments in the U.S. data. The top panel lists the parameter values. The bottom panel shows the targeted moments in the data and the resulting values in the model.

As a check of the model's ability to capture the extent of inequality in the data, we compute the concentration of earnings and wealth in the model and compare them with the data. The results are presented in Figure 3. The left panel shows the concentration of earnings. The dashed line indicates the commutative share of earnings at each commutative population share for individuals age 25 to 60 in the CPS (1994). The solid line shows the same measure in the model. Overall, the model does a good job at capturing the extent of earnings inequality in the data. The Gini index of earnings is 0.43 in the model and 0.46 in the data. Moreover, the model is able to capture the concentration of earnings at the top. The share of earnings of the top 1 percent is 8 percent in

⁴⁰As we show in the supplementary appendix, our results on optimal policies are unchanged by these alternative calibrations, while the cost saving measures are magnified.

the model and 6 percent in the data. This is achieved through the use of a Pareto-lognormal distribution for ability distribution (even though we did not directly target this moment).

Finally, the right panel in Figure 3 shows the concentration of wealth. The dashed line is the cumulative share of wealth owned by each cumulative population share in the SCF (2007). The model matches the Gini index of wealth by construction (see Table 3). Heterogeneity in the discount factor allows us to generate a high concentration of wealth in the model. The share of wealth owned by the top 1 percent is 23 percent in the model and 29 percent in the data. In the supplementary appendix, we plot consumption and earnings profiles in the model and discuss their relationship to the data.

5 Quantitative Results: Steady State

In this section we apply the tools developed in section 3.1 to our calibrated economy described in section 4. We first make a case for policy reforms by demonstrating that status quo policies fail the Pareto optimality tests derived in Proposition 4. We then use the procedure outlined in the supplementary appendix to solve for optimal policies that implement efficient distortions in the economy. Finally, we report the effect that an optimal reform has on individual choices, macro aggregates and government budget. Note that our optimal policies minimize the present value of consumption net of labor income for each generation. We report the reduction in this cost as a measure of efficiency gains from optimal reform policies.

Two points are worth emphasizing about our exercise: First, the efficiency gains from our Pareto optimal policy reforms can be redistributed across individuals in various ways. In this section, we do not specify how the gains are distributed. In the next section, we provide one way to distribute these gains to a subset of the population. Second, since it is important to disentangle the partial and general equilibrium effects of the reform, in section 5.1-5.5, we assume that prices—interest rates and wages—are fixed at the status quo level. We also assume the same demographics as the current U.S. economy. In sections 5.4 and 6, we report the results with endogenous factor prices, future demographics and transition.

5.1 Test of Pareto Optimality

We start our analysis by testing the Pareto optimality of the status quo allocations. We do this by computing the intertemporal and intratemporal distortions for the status quo allocations and checking how much the formulas (16), (17) and (18) are violated.

In Figures 4 and 5, we plot the implications of Pareto optimality tests for the status quo econ-

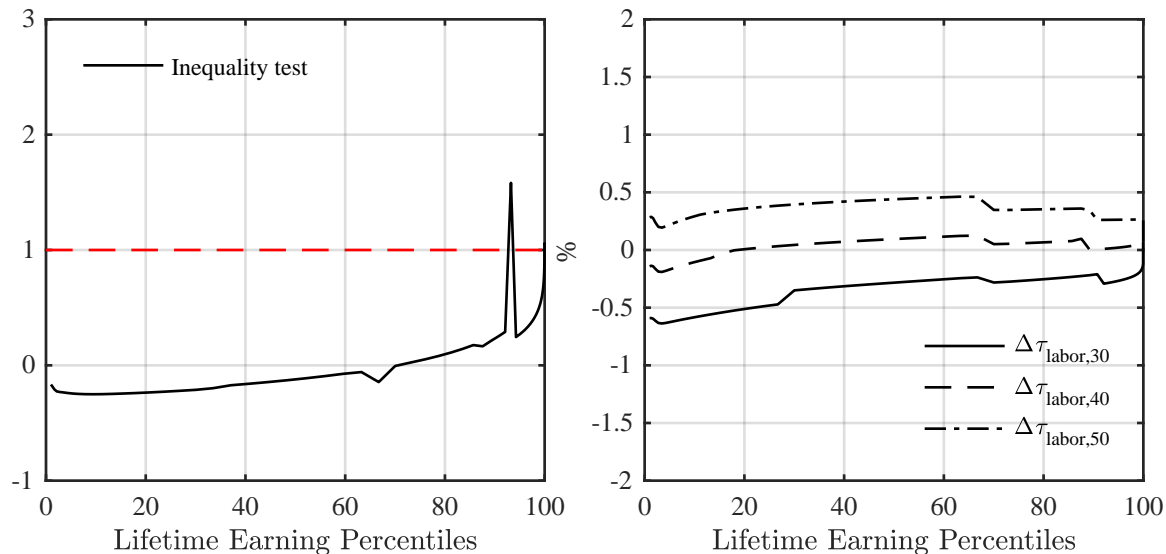


Figure 4: Test of Pareto optimality for status quo policies. The left panel plots the two sides of the inequality (18). The right panel depicts the change in earnings wedges required for (19) to hold.

omy. Figure 4 plots the performance of the tests for labor wedges.⁴¹ The left panel depicts the inequality (18); the dashed red line is the left hand side, and the black solid line is the right hand side. As it illustrates, the inequality only fails to hold over a small range of earnings. This is where effective earnings taxes are regressive, due to the social security maximum taxable earnings cap (see Figure 2). In this range, the term $\tau'_{\text{labor},0}(\theta)$ is a large negative number. This pushes the left hand side of inequality (18) up. The right panel depicts the change required in the labor wedge so that the tax smoothing relationship in (19) holds. As we see, the percentage change in the labor wedges to restore (19) is around ± 0.5 percent. In other words, it suggests that given the earnings taxes at $j = 0$, i.e., age 25, and the redistributive motives they represent, earnings taxes are roughly optimal and should not change by much. This observation is suggestive of one of our main findings: that earnings taxes are not a large source of inefficiency of the tax code.

Moving to the efficiency properties of savings taxes, Figure 5 shows the efficiency properties of savings wedges. In the left panel, we focus on working life and examine (16). The figure depicts the change required in savings taxes so that (16) holds at ages 30, 40 and 50. Interestingly, for young people, savings wedges must increase. This is because these young individuals face a borrowing constraint (they cannot borrow) and thus face a negative wedge on their intertemporal savings margin.⁴² For individuals with mid-values of lifetime earnings, the required change is

⁴¹Note that under status quo, many policies and institutional features distort earnings and savings decisions: consumption tax, earnings tax, the social security benefit formula, and borrowing constraints. Because of this, we focus on wedges, which are defined in (15).

⁴²Under a binding borrowing constraint, the current marginal utility of consumption is high relative to that in the future. This is equivalent to a negative savings wedge.

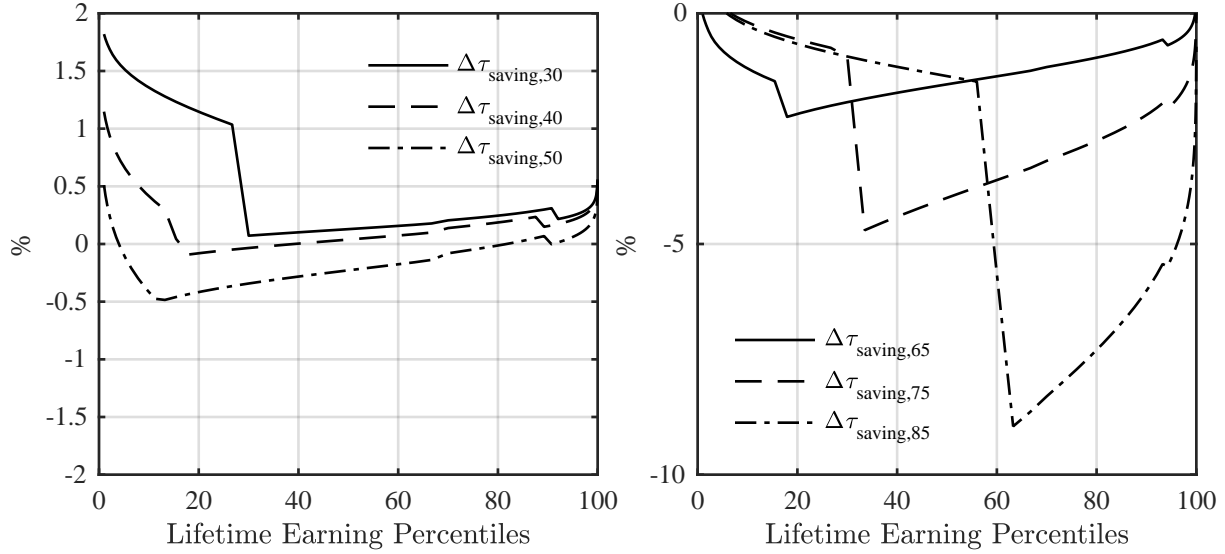


Figure 5: Test of Pareto optimality for status quo policies. The left panel depicts the required change in the savings wedge so that (16) holds at ages 30, 40 and 50. The right panel depicts the required change in the savings wedge so that (17) holds at ages 65, 75 and 85.

minimal, close to 0. This is mainly because their mortality risk is small and not very sensitive to their lifetime income. Finally, for working-age rich individuals, the savings wedge must increase. This is because of the discount rate differentials for productive workers. As mentioned in section 2, since more productive individuals value consumption more in the future, a tax on savings of an individual incentivizes everyone with a higher productivity to work harder and save more. Nevertheless, this figure illustrates that a reform must significantly change the tax treatment of savings.

The right panel of Figure 5 depicts the change required in savings taxes so that (17) holds with equality—holding the values on the RHS fixed. As it can be seen, savings wedges must decline significantly for the majority of individuals older than 65. This is mainly capturing the fact that markets are incomplete, and a subsidy to savings completes the market, i.e., provides annuity insurance to workers. Note that the required change in savings subsidies is small at the extremes. At the bottom of the distribution, individuals face a binding borrowing constraint and thus face a negative wedge. Therefore, the required decline in their savings wedge is not high. For individuals at the top of the income distribution, the mortality risk is very small and, as a result, the required decline in their savings wedge is small.

In summary, the results of our tests suggest the following: First, earnings taxes pass the Pareto optimality tests to a great extent, except around the social security earnings cap. Second, savings taxes strongly fail the Pareto optimality tests. This result suggests that in a reform, the focus must be on asset taxes as opposed to earnings taxes. Our numerical results below confirm this

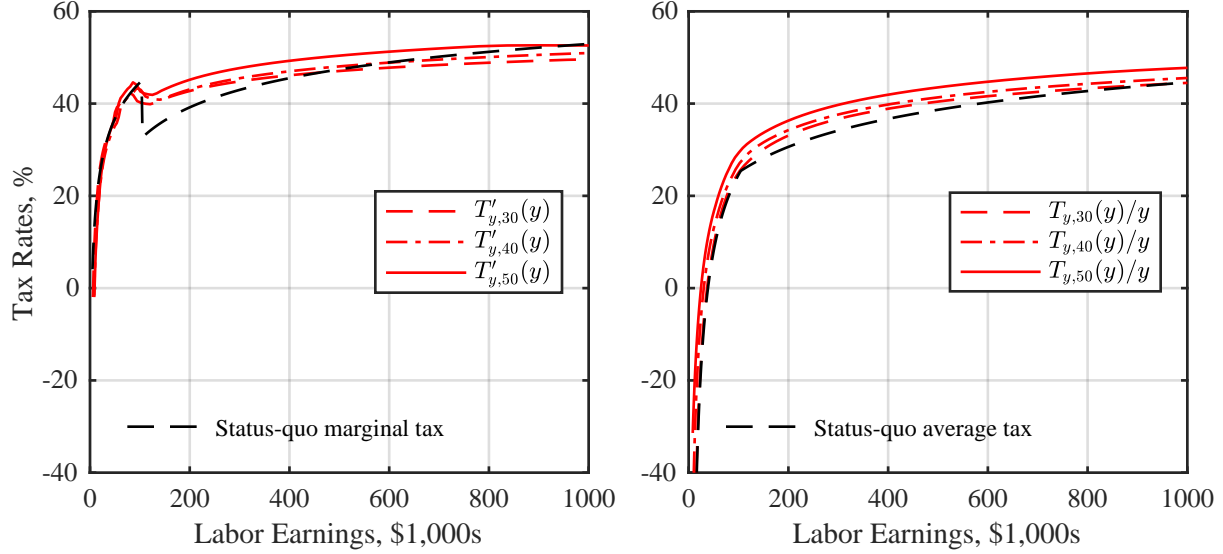


Figure 6: Optimal labor income tax functions. The left panel shows marginal taxes, and the right panel shows average taxes. The black dashed line is the effective status quo tax schedule.

intuition.

5.2 Optimal Policies

We solve for optimal policies using the planning problem (P1) outlined in section 3.1. These are (1) non-linear, age-dependent taxes on assets upon survival, $T_{a,j}((1+r)a_j)$; (2) non-linear, age-dependent taxes on labor income, $T_{y,j}(y_j)$; (3) transfers to workers before retirement, Tr_j ; and (4) transfers to workers after retirement, S_j . Note that transfers are independent of individual choices, but they do depend on age. Note also that the level of transfers and assets of households is not uniquely determined, due to the presence of lump-sum transfers. As a result, we choose transfers such that the lowest-ability type opts not to hold any asset. Moreover, we assume that individuals face linear consumption taxes. We fix the consumption tax rate at the calibrated level for the status quo economy. This assumption eases the comparison of labor income taxes across economies.⁴³ Finally, we fix the corporate income tax rate at the calibrated status quo level. This implies that the pre-tax return on assets is the same in the status quo economy and in the optimal reform.

Figure 6 shows the optimal marginal and average labor income tax functions for ages $j = 25, 35, 45$ (solid lines). We also plot the status quo tax functions for comparison (dashed lines). Notice that except for the region where there is a sharp drop in the status quo tax rates (due to

⁴³ As in any optimal tax exercise, we can uniquely determine the intratemporal (labor) wedges. Consumption taxes and labor income taxes are not separately pinned down.

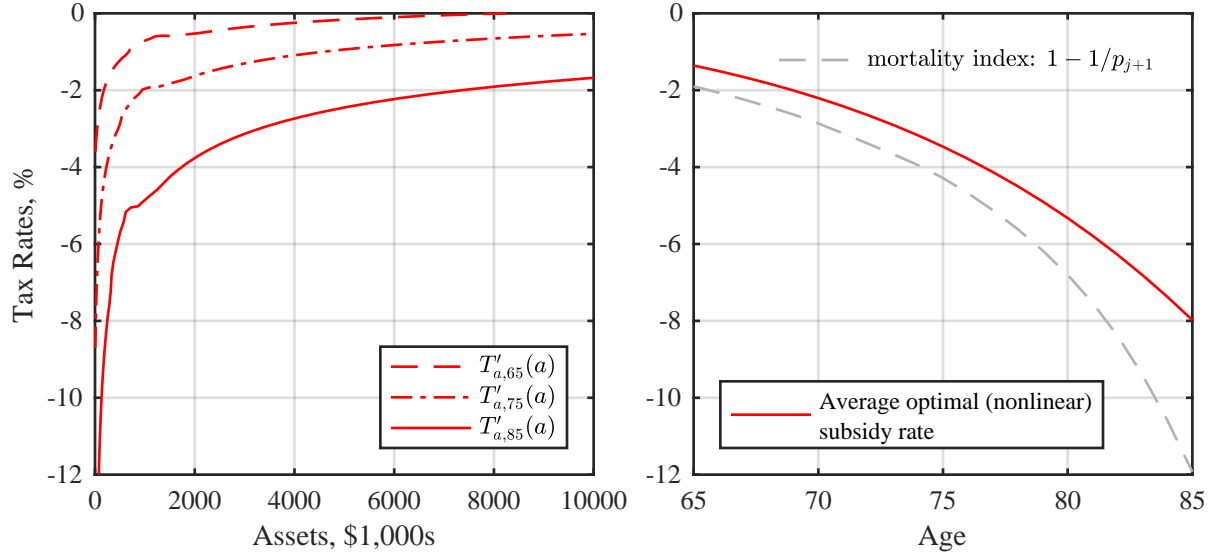


Figure 7: Optimal asset tax functions. The left panel shows the marginal taxes over all asset levels at ages 65, 75 and 85, while the right panel shows the average marginal rates at each age from 65 to 85. The dashed line is the population mortality index.

social security maximum taxable earnings), the optimal taxes are very close to those in the status quo. Furthermore, there is little age dependence in the optimal labor income taxes.⁴⁴ These results imply that there is little room for improvement in efficiency by reforming labor income taxes. In essence, our exercise confirms the insight from the Pareto optimality tests performed in section 5.1 regarding earnings taxes.

The left panel of Figure 7 shows the optimal marginal taxes (subsidies) on assets for ages $j = 65, 75, 85$. Since mortality is larger for asset-poor individuals, the rates are larger for these individuals at all ages. In contrast, asset-rich individuals have higher ability, and hence lower mortality. The inefficiency due to the absence of an annuity market is smaller for these individuals; therefore, asset subsidies are smaller (taxes are higher). In this sense, optimal asset taxes (subsidies) are progressive. Figure 7 also illustrates that subsidies are large, around 5 percent, and thus can play an important role in the provision of retirement benefits by the government.

The right panel shows the average marginal rates at each age from 65 to 85 years in comparison to the average mortality of the population. The difference between the two implies the following: first, progressivity of the subsidies is significant and cannot be ignored; second, policies are above and beyond completing the annuity market, as would be the case in a world where mortality were observed by the government (or mortality were uniform in the population).

As before, the implied magnitudes of asset subsidies and their progressivity confirms the

⁴⁴The result that earnings taxes are independent of age is because there are no shocks to productivity and labor productivity profiles are parallel.

Table 4: Sources of retirement income				
Income quartiles	Share of public transfers in retirement income (%)			
	Data ^a	Status quo	Optimal reform	
			(incl.asset subsidies)	(excl.asset subsidies)
1st	95	100	80	47
2nd	90	94	70	27
3rd	67	79	63	17
4th	34	33	48	6

^aSource: Table 6 in [Poterba \(2014\)](#).

results of our optimality tests. In other words, asset subsidies are an important part of our Pareto optimal reform.

5.3 Sources of Retirement Income

It is useful to compare the sources of retirement income in the status quo economy and that of the optimal reform. This comparison would shed light on the burden of the reform for the government and on changes in individual budgets.

Table 4 compares the share of government transfers out of the total income for retired individuals (asset income plus government transfers). In our calculation for the status quo economy, the government transfers consist of social security (and Medicare) benefits. For comparison, we also include the share of government transfers in retirement income as measured in the CPS data (reported in [Poterba \(2014\)](#)).⁴⁵

The numbers in our status quo economy are close to the CPS data, particularly for the lower half of the income distribution.

An important feature of the reform economy is the significant reduction in the share of government transfers in retirement income for all income groups except the top quartile. This is mainly a result of the presence of asset subsidies. In particular, asset subsidies imply that individuals will save more. As a result, asset income constitutes a higher fraction of retirement income and, therefore, the share of government transfers in income declines.

5.4 Aggregate Effects of Reforms

Table 5 shows the summary statistics of the aggregate variables for our economy. In the first column, we report the aggregate quantities in the calibrated benchmark with the status quo U.S. policies. The second column shows the aggregate variables under Pareto optimal reform policies,

⁴⁵To make the CPS statistics comparable to our model, we exclude labor earnings (we calculate the share of government transfers out of all incomes excluding labor earnings).

holding factor prices fixed. In this case, the stock of capital in the economy is 12.29 percent higher relative to the status quo. This is due to higher incentives to save provided by optimal asset subsidies. As a result, GDP is higher by 4.33 percent and consumption by 1.66 percent relative to the status quo. However, consumption as share of GDP falls slightly from 0.69 to 0.67. This is, again, due to a higher desire for savings under optimal reform policies. Overall, the present discounted value of consumption, net of labor income, for each cohort falls by 11.08 percent in the optimal reform relative to the status quo. In terms of flow of consumption, this is equivalent to a 0.82 percent fall in consumption for all types in all ages. That is the amount of decline in status quo consumption needed to equalize the present discount value of consumption, net of labor income, equalized across status quo and optimal reform allocations.

The third column in Table 5 shows aggregate quantities under Pareto optimal policies with endogenous factor prices (but with benchmark demographics, i.e., current U.S. demographics). In this case, the capital stock is higher by only 5.14 percent. This is due to the general equilibrium effect of the lower real return (3.81 percent relative to 4.05 percent). GDP is higher by 1.72 percent and consumption by 0.58 percent relative to the status quo. The cost savings in this case are significantly larger relative to the case with fixed factor prices. In other words, the present discounted value of consumption, net of labor income, for each generation falls by 29.43 percent (this is equivalent to a 2.18 percent fall in the flow of consumption for all types at all ages). This large difference in cost savings can be accounted for entirely by the fall in the interest rate.⁴⁶

5.5 Distributional and Budgetary Effects of the Reform

While our exercise keeps the distribution of welfare the same, an optimal reform can affect the allocation of resources across individuals. In this section, we describe the effect of our optimal reform exercise on the distribution of allocations.

Figure 8 shows the Lorenz curve for earnings and wealth distribution for status quo and efficient allocations. As we see, the optimal reform policies do not have a significant effect on the distribution of earnings, which is in line with the fact that earnings taxes exhibit very little change. On the other hand, the efficient distribution of assets is less concentrated than in the status quo. In particular, the wealth Gini under reform policies is 0.64, which is significantly lower than the wealth Gini of 0.78 under the status quo. This is mainly because the consumption of low-productivity individuals increases late in life, due to subsidies on assets and, as a result, the asset distribution becomes less skewed.

Table 5 shows how the optimal reform affects the government's tax revenue and transfers.

⁴⁶As we show in section 6, when general equilibrium analysis includes demographic changes, factor prices are not very different between the status quo and reform economies. Hence, the general equilibrium effects are smaller in the presence of a demographic change.

Table 5: Aggregate effects of reform for current U.S. demographics

	Current U.S. (1)	Optimal Reform	
		(2)	(3)
Factor prices			
Interest rate (%)	4.05	4.05	3.81
Wage	1.00	1.00	1.03
Values relative to GDP			
Consumption	0.69	0.67	0.68
Capital	4.00	4.31	4.13
Tax revenue (total)	0.26	0.27	0.27
Earnings tax	0.14	0.14	0.15
Consumption tax	0.04	0.04	0.04
Capital (corporate) tax	0.08	0.09	0.08
Transfers	0.16	0.15	0.15
To retirees	0.08	0.02	0.02
To workers	0.08	0.05	0.05
Asset subsidies	0.00	0.08	0.08
Change (%) (relative to current U.S.)			
GDP	–	4.33	1.72
Consumption	–	1.66	0.58
Capital	–	12.29	5.14
Labor input	–	-1.80	-0.83
PDV of net resources	–	-11.08	-29.43
Consumption equivalence		0.82	2.18

Note: Column (1) is the benchmark calibration to the current U.S. economy. Column (2) is the optimal reform policies with prices and demographics fixed at the current U.S. values. Column (3) is the optimal reform policies with equilibrium prices but fixed demographics (at current U.S. levels).

There is little difference in total tax revenue and total transfers as a fraction of GDP. However, the nature of transfers changes significantly in an optimal reform. Pure transfers before and after retirement fall as a percentage of GDP; instead, asset subsidies, which amount to 8 percent of GDP, are introduced. Optimal reform policies can achieve the same welfare as status quo policies by collecting more taxes and transferring less resources. This is possible because optimal reform policies remove inefficiencies due to a lack of annuitization and inefficiencies in the status quo income tax.

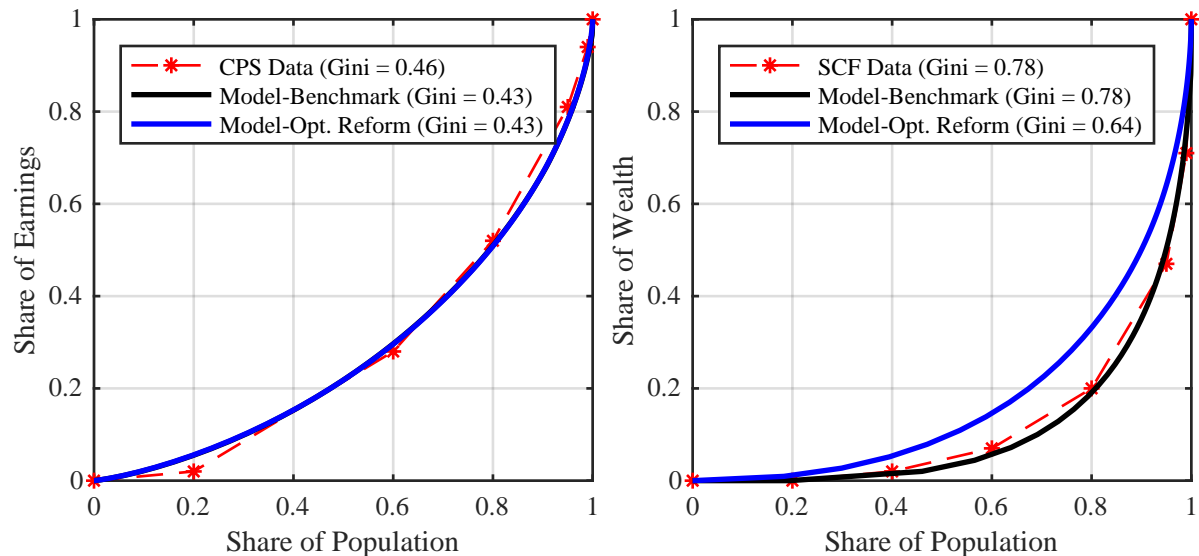


Figure 8: Distribution of earnings and wealth: status quo vs. optimal reform. The black line shows the results in the calibrated economy with current U.S. status quo policies. The blue line shows the results under Pareto optimal policies (for current U.S. demographic parameters and holding factor prices fixed).

6 Quantitative Results: Transition

The above analysis points toward the key reforms that are relevant for an overhaul of the fiscal policies including social security in the steady state. While the results are informative, the analysis assumes that there is no demographic change and, therefore, downplays the role of a policy reform. In this section, we repeat our quantitative exercise in an aging society with a declining population growth and mortality rate. Our quantitative results confirm the importance of asset tax reforms and the lack of importance of earnings tax reforms.

An Aging Economy. We assume that the status quo economy is initially in a steady state determined by the calibrated parameters, as described in Section 4. The economy then experiences a demographic transition which starts at $t = 0$ and ends in 50 years. At the conclusion of the demographic transition, the population growth is 0.5 percent (down from 1 percent), consistent with U.S. Census Bureau’s projections (see Colby and Ortman (2015)). In addition, the new population mortality rates match the mortality rates of 2040 birth cohort males (Table 7 in Bell and Miller (2005)). We calibrate equation (20) to match the differences in mortality rates among lifetime earnings deciles reported in Waldron (2013), as well as the new population mortality rates.⁴⁷ All parameters change gradually according to a linear trend over the 50-year transition period. These assumptions imply that the ratio of workers to retirees falls from 4 (its current value) to

⁴⁷We assume that the ratios of mortality among lifetime earnings deciles do not change.

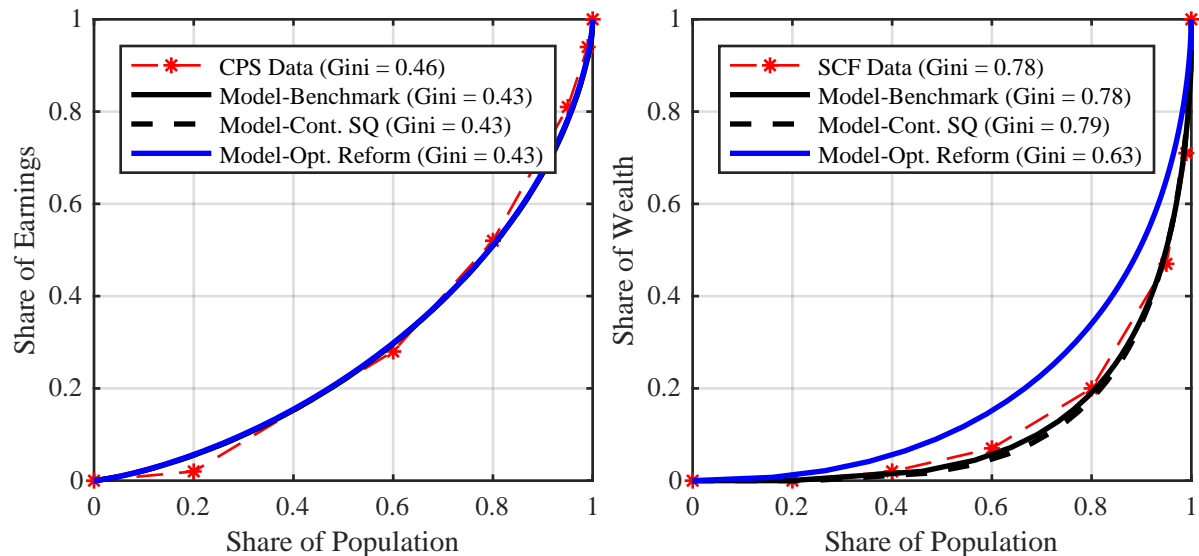


Figure 9: Distribution of earnings and wealth: status quo vs. optimal reform. The black solid line shows the results in the calibrated economy with current U.S. status quo policies. The black dashed line shows the steady state results with projected demographic parameters and continuation of status quo policies. The blue line shows the steady state results under Pareto optimal policies for projected demographic parameters. Factor prices are endogenous.

2.4 (its projected value). This is consistent with U.S. Census Bureau’s projections (see [Colby and Ortman \(2015\)](#)).

Transition in the Status Quo Economy. In order to solve for optimal policies, we need to know the distribution of lifetime welfare for each birth cohort along the transition path for the status quo economy. Since under the status quo and in an aging economy, the social security program is not sustainable, we have to take a stance on what status quo policies will be implemented in order to make the social security system sustainable. To this end, we make the following assumptions: First, we assume that the income tax schedules and social security benefit formula do not change. Second, the debt to GDP ratio is held constant at its initial calibrated value of 47 percent. Third, and most importantly, we assume that the consumption tax adjusts in each period to balance the government budget constraint and hence finance the transition. It is important to note that, due to political uncertainties, it is impossible to know how status quo policies evolve in response to demographic changes. Here, we use the simplest benchmark to conduct our analysis. However, our methodology can be applied to any alternative assumption for the future path of status quo policies.

The second column in Table 6 shows how the demographic change and continuation of status quo policies affect the aggregates. Since mortality is lower, individuals live longer and, therefore, have a higher demand for savings. This, in turn, increases the stock of capital by 7.96 percent.

However, due to the lower number of workers as share of population, the labor input falls by 9.26 percent, resulting in a 2.13 percent decline in GDP.

While continuation of the status quo policies does not change the tax revenue as percentage of GDP, there is a significant increase in old-age transfers as percentage of GDP. This is because there are more retirees in the economy. On the other hand, to offset the effect of a rise in old-age transfers on government budget, the consumption tax rate must rise to 10 percent (from the original value of 5.5 percent). This increase in the consumption tax rate increases the share of government revenue from consumption tax and contributes to a decline in inequality. As a result, inequality overall does not change very much. The cross-sectional distribution of earnings and wealth in the new steady state are depicted in Figure 9. There is no change in the distribution of earnings, while the distribution of wealth becomes slightly more unequal (the wealth Gini index rises from 0.78 to 0.79).

Reform Exercise. Using the time path of the distribution of welfare for each generation, we solve the problem of minimizing the resource cost of delivering the status quo welfare to each individual in each birth cohort. We do this while keeping the corporate income tax rate and consumption tax rate at its status quo level.⁴⁸

A complication that arises when performing an optimal policy reform in an economy in transition is the treatment of existing generations: generations that are alive at the time of the reform. The complication arises from an information problem. At the time of the reform, households who have worked and saved previously have revealed their types. Thus, if the government has a flexible enough tax function (e.g., generation-specific taxes on their assets at the time of the reform), it can achieve first best and fully bypass the incentive problem. We think this ability of the government to completely bypass the incentive problem is unrealistic. It also creates a discontinuity on allocations for people who are alive at the time of the reform relative to future generations, which makes it harder to accept it as a reasonable reform.

In order to solve this problem, we make the following assumptions: any person who is alive at the beginning of the reform ($t = 0$) will face the status quo policies together with an additional one-time lump-sum transfer. All other individuals will face optimal reform policies. Note that this means that the generations that are alive at the start of the reform receive all the gains from the reform.

Optimal Reforms. Our quantitative exercise for the transition mainly confirms our previous findings in our steady state analysis: asset subsidies play a key role in the reform, while earnings taxes do not change it by much. Figure 10 shows the changes in the earnings taxes over time. Since in the course of transition to the new steady state, inequality remains somewhat constant,

⁴⁸Note that, during transition, the status quo consumption tax rate changes. We take the time path of this consumption tax rate as given when we solve for optimal reform policies.

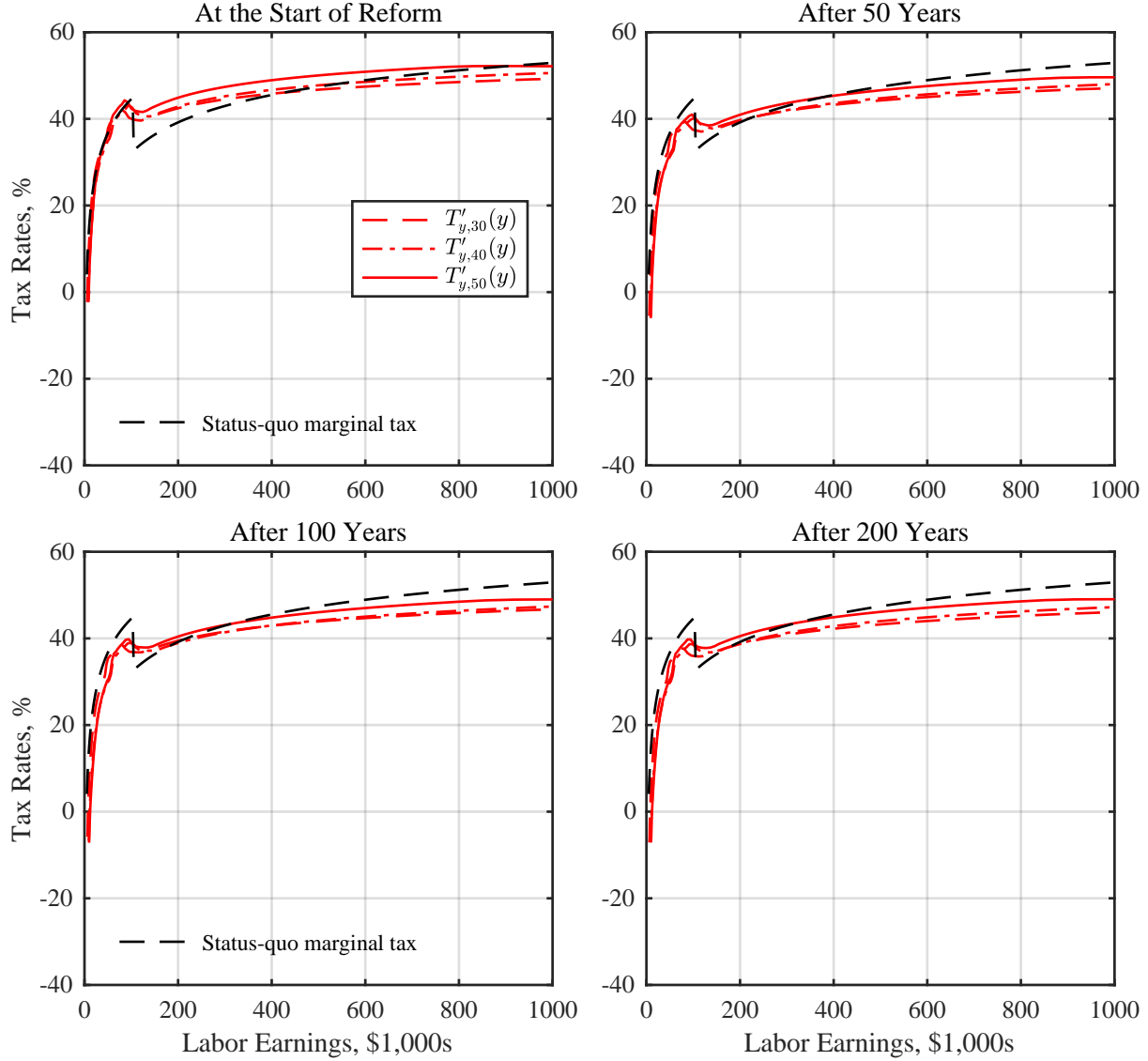


Figure 10: Evolution of optimal marginal labor income tax functions over the transition

earnings taxes should not change by much. Furthermore, asset subsidies are still significant, although slightly lower, due to the decline in the mortality rate (Figure 11).

The last column of Table 6 shows the impact of these policies on aggregate allocations and on government budget. Capital stock rises more relative to the status quo economy. This leads to a smaller decline in GDP and aggregate consumption.⁴⁹ Figure 12 shows the path of the aggregate variables over the transition. The jump in the primary surplus as share of GDP is due to the initial lump-sum distribution.

Importantly, reform policies reduce the cost of delivering the status quo welfare to each birth

⁴⁹The decline is primarily driven by a fall in the labor supply, caused by a decline in the number of workers.

Table 6: Aggregate effects of demographic transition and policy change

	Current U.S. (1)	Continue (2)	Optimal reform		
			(3)	(4)	(5)
Factor prices					
Interest rate (%)	4.05	3.37	4.05	3.81	3.31
Wage	1	1.08	1	1.03	1.09
Values relative to GDP					
Consumption	0.69	0.69	0.67	0.68	0.68
Capital	4.00	4.41	4.31	4.13	4.45
Tax revenue (total)	0.26	0.29	0.27	0.27	0.27
Earnings tax	0.14	0.14	0.14	0.15	0.13
Consumption tax	0.04	0.07	0.04	0.04	0.07
Capital (corporate) tax	0.08	0.07	0.09	0.08	0.07
Transfers	0.16	0.19	0.15	0.15	0.13
To retirees	0.08	0.12	0.02	0.02	0.02
To workers	0.08	0.07	0.05	0.05	0.04
Asset subsidies	0.00	0.00	0.08	0.08	0.07
Change (%) (relative to status quo)					
GDP	–	-2.13	4.33	1.72	-1.44
Consumption	–	-2.38	1.66	0.58	-2.00
Capital	–	7.96	12.29	5.14	9.73
Labor input	–	-9.26	-1.80	-0.83	-9.26
PDV of net resources	–	–	-11.08	-29.43	-4.24
Consumption equivalence	–	–	0.82	2.18	0.98

Note: Column (1) is the benchmark calibration to the current U.S. economy. Column (2) is the continuation of U.S. status quo policies (with consumption tax adjusted to balance government' budget constraint). Column (3) is the optimal reform policies with prices and demographics fixed at the current U.S. values. Column (4) is the optimal reform policies with equilibrium prices but fixed demographics (at current U.S. levels). Column (5) is the optimal reform policies with equilibrium prices and future demographics. In column (3) and (4), the percentage change in the PDV is calculated relative to column (1). In column (5) the percentage change in the PDV is calculated relative to column (2).

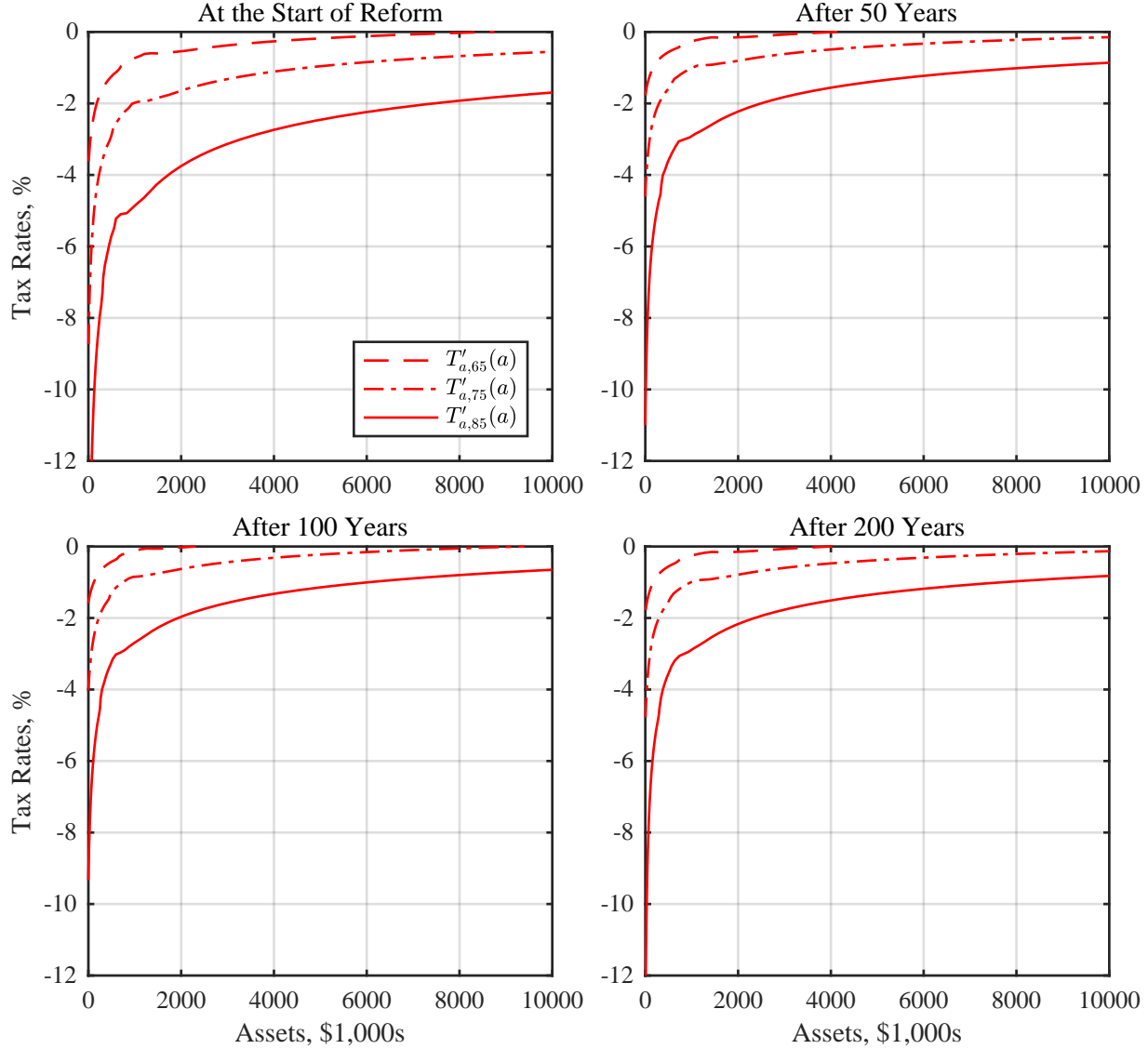


Figure 11: Evolution of optimal marginal asset tax functions over the transition

cohort. Under optimal reform policies, the present discounted value of consumption net of labor income for a newborn is 4.24 percent lower relative to what it would be under the continuation of the status quo policies in the steady state (this is equivalent to 0.98 percent lower consumption for all types and all ages). As we discuss above, we distribute these resources to those who are alive at the start of the reform in a lump-sum fashion. This transfer is equivalent to 10.5 percent of GDP in the initial steady state.

Overall, we view the results of our quantitative exercises, one for the aging economy and one in the steady state, as pointing toward the importance of asset subsidies to all individuals as an integral part of any fiscal policy reform. This is in contrast with much of the discussion in policy

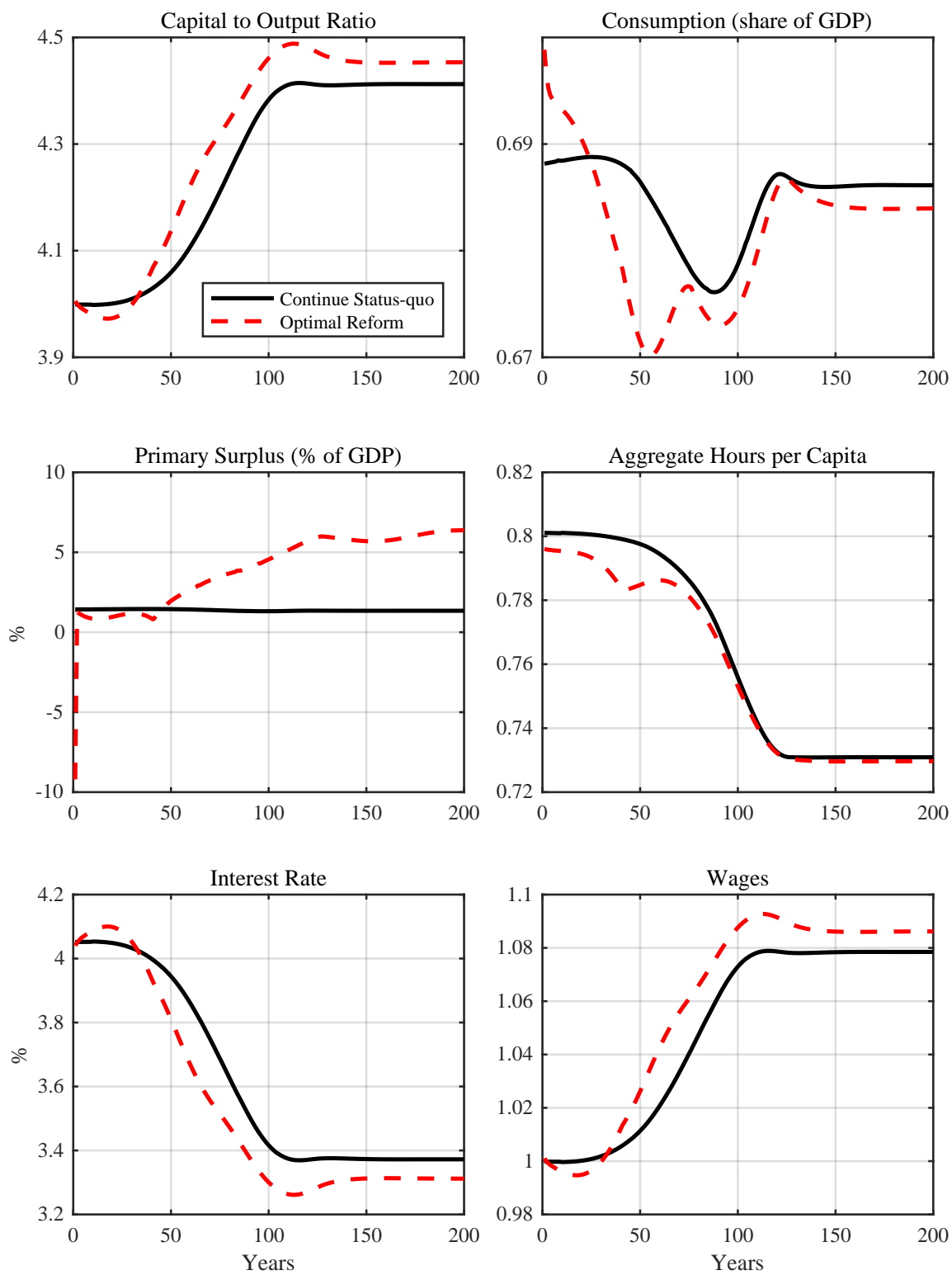


Figure 12: Evolution of aggregates along the transition

circles on earnings tax reform (reform of the payroll taxes, etc.).

7 Extensions and Robustness

In this section we investigate the importance of our results relative to other commonly considered reforms. Furthermore, we discuss the robustness of our results to alternative calibrations of status quo policies as well as alternative motives for saving.

7.1 Optimal Privatization Reform

As we discussed in Section 5, savings subsidies play an important role in our Pareto optimal reforms. In particular, the optimality tests and the optimal reform exercise indicate a reform of the earnings taxes does not seem to play an important role. One might, however, think that this is due to the generality and flexibility of the asset taxes. Here, we briefly describe an exercise that further highlight the role of asset subsidies and their progressivity. The details of the analysis are in the appendix.

In this exercise, we assume that there are no asset taxes or subsidies. This exercise is similar to a particular proposal that has received considerable attention in the literature: *privatization* of retirement financing. More precisely, this is the proposal to eliminate social security retirement benefits and reduce payroll taxes and move towards a save-for-retirement system.⁵⁰ These *privatization* policies differ from our optimal reform policy in two very important ways. First, our optimal reform policy does not involve a major adjustment of labor income taxes. Second, our optimal reform policy relies crucially on asset subsidies.

We solve for the best reform policies that feature no old-age transfers and no asset taxes or subsidies. In this regard, the efficiency gains from these policies can be viewed as an upper bound on what can be gained through privatization policies. The additional constraint that we impose on the planning problem relative to that in section 5 is that the savings wedge as defined in (15) must be 0. This implies that earnings taxes are chosen without constraint, and savings subsidies are 0 for everyone.

Figure 13 (left panel) shows the optimal marginal taxes under privatization policies. Note that marginal rates are lower than the status quo, especially at the lower income levels. Moreover, the drop in marginal taxes matches the level of payroll taxes. In this regard, our optimal policies mimic a key feature of the privatization proposals. However, there is also a crucial difference that our optimal labor tax rates are negative for the poorest individuals. The no-subsidy restriction tilts the optimal profiles of consumption towards younger ages. To accommodate this

⁵⁰See Nishiyama and Smetters (2007) and McGrattan and Prescott (2017) for example.

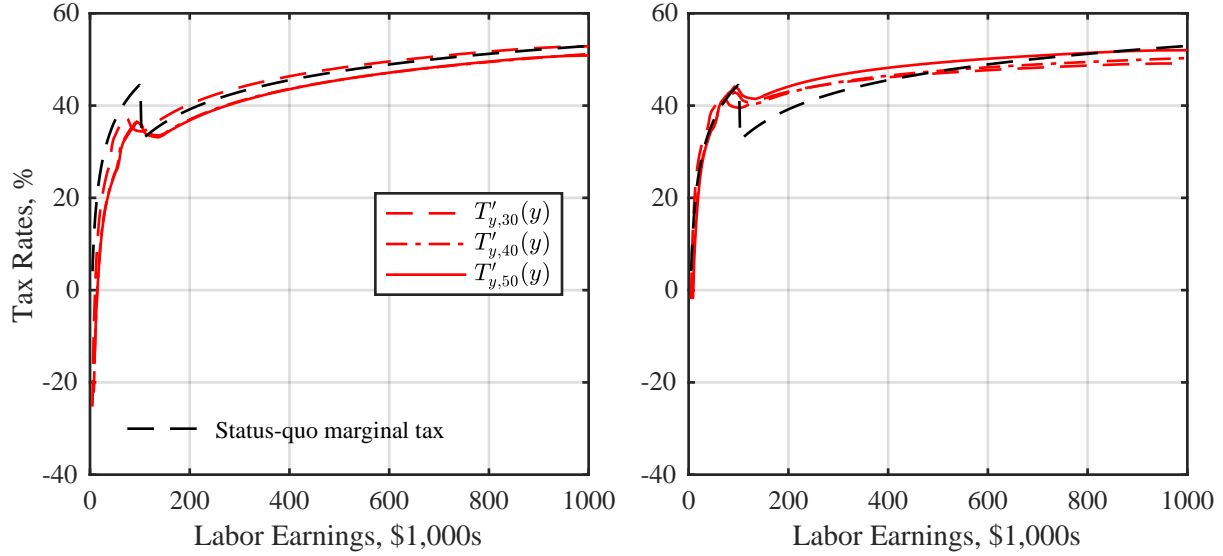


Figure 13: Optimal labor income tax functions with *privatization* (no old-age transfers and no asset subsidies). The left panel is optimal marginal taxes under *privatization*, while the right panel shows the same for the benchmark calibration. The black dashed line is the effective status quo tax schedule.

higher consumption, low-income individuals must work more. The negative marginal income tax provides the incentive needed for these low-ability individuals to increase their work effort.

Under privatization policies the present discounted value of consumption, net of labor income, rises relative to the status quo under all scenarios regarding prices and demographics. In other words, imposing 0 taxes on savings—as opposed to subsidizing them—is stringent enough on allocations that it raises the costs of delivering utilities in the steady state. This highlights the importance of subsidies in any reform. Details of the calculations and aggregate effects of privatization policies are provided in the supplementary appendix.⁵¹

While this exercise highlights the importance of savings subsidies, one can also question how important the progressivity of the subsidy system is in a reform. In supplementary appendix, we perform an optimal reform exercise while imposing that savings taxes must be linear. Our calculations establish that two-thirds of the gains can be achieved by linear subsidies and optimal earnings taxes. Thus, progressivity of subsidies is an integral part of an optimal reform.

⁵¹In supplementary appendix, we show that privatization can indeed lead to cost savings when labor supply elasticity is as large as 2.5. However, the gains from privatization are significantly smaller than those of optimal reforms: around a quarter to one-third.

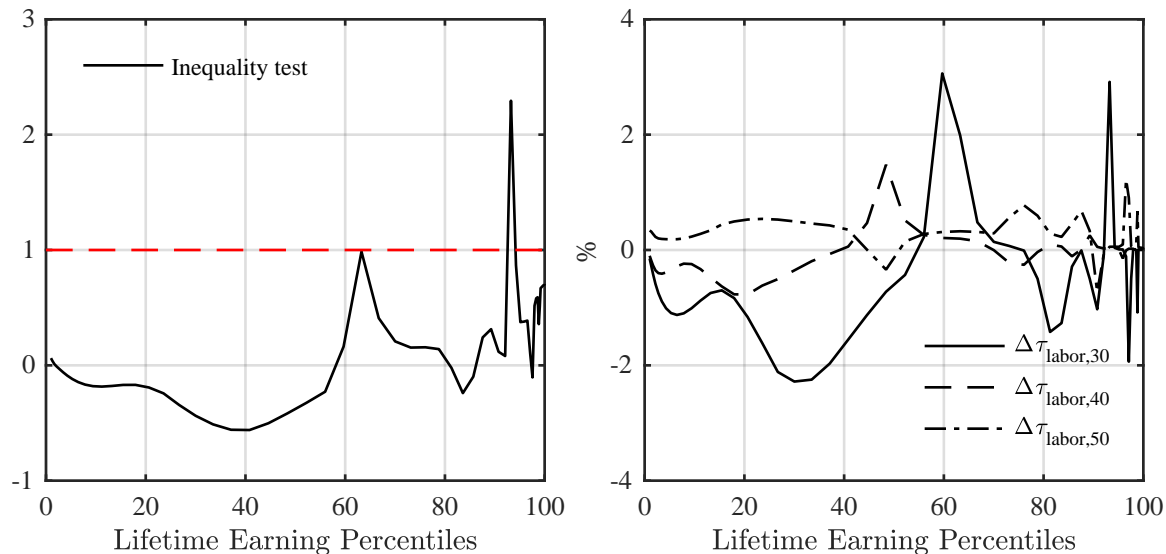


Figure 14: Test of Pareto optimality for status quo policies with a CBO effective tax rates. The left panel plots the two sides of the inequality (18). The right panel depicts the change in earnings wedges required for equation (19) to hold.

7.2 Alternative Status Quo Tax Function

One of our key findings in section 5 is that the status quo earnings tax function in the U.S. fails the Pareto efficiency test only at the maximum social security taxable income. This is partly because for the most part we used a smooth function to approximate the U.S. earnings taxes. A concern is that the actual earnings tax in the U.S. contains many thresholds which lead to a non-smooth tax function and could potentially lead to inefficiencies. To address this concern, we repeat our exercise using the calculations of effective marginal tax rates on labor income provided by the Congressional Budget Office; see Harris (2005). In particular, for the status quo earnings taxes, we use the effective marginal federal (and state and local) tax rate for a head of household with one child in 2005.⁵² This tax approximation includes many intricate features of the tax code including EITC phase in and phase out, AMT, CTC and itemized deductions.

Figure 14 depicts the earnings tax test. Despite a non-smooth status quo tax function, the earnings tax function fails the inequality test only at the social security maximum taxable income. However, it comes close to being violated at a lower income level as well. This level of earnings is the large drop in the effective marginal tax rate due to transition from the EITC phase-out (which implies an effective marginal rate of 31 percent) to the 15 percent bracket. Furthermore,

⁵²Harris (2005) calculates these effective tax rates only for hypothetical families. They do indeed vary by family details and state of residence. This is one of the reasons that approximations such as Heathcote et al. (2014), which are done using actual tax and income data, are advantageous. However, we use this example as a test of whether our results change if we move to a non-smooth tax function.

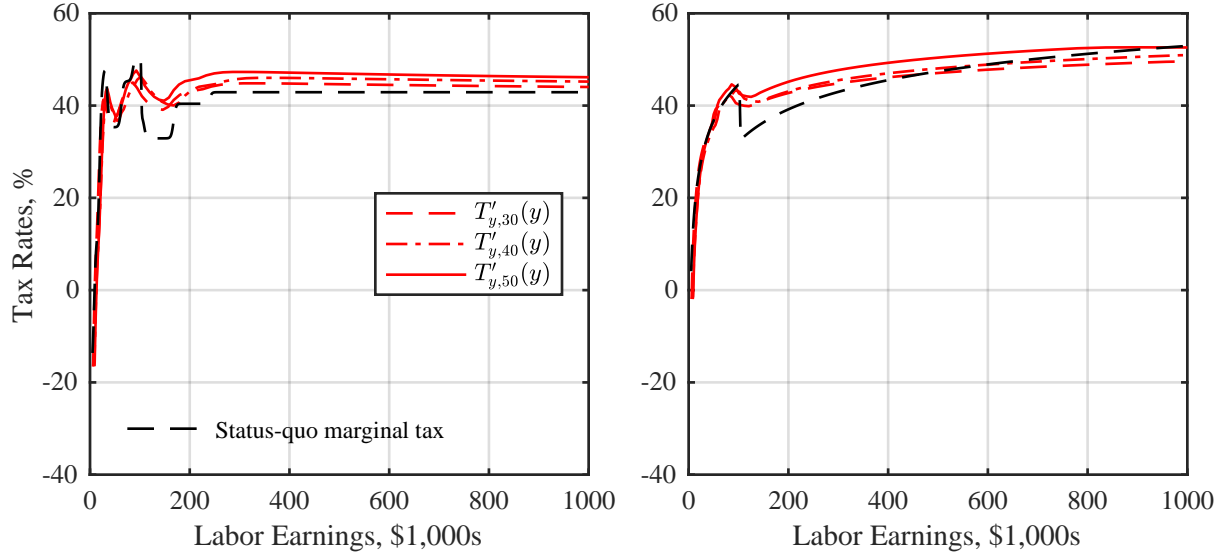


Figure 15: Optimal labor income tax functions for the alternative status quo tax policy. The left panel shows optimal marginal taxes and compares them with CBO effective tax rates. The right panel shows the same for the benchmark economy (with HSV tax function).

as the right panel in Figure 14 depicts, the deviations from the tax smoothing equation (19) are higher than before and of a magnitude of up to 3 percent. Nevertheless, as depicted in Figure 15, the optimal labor income taxes are not very far from optimal. Intuitively, despite having many ups and downs, there is not much variation in the marginal income taxes relative to a smooth approximation to this schedule. As a result, the earnings taxes do not vary by much in an optimal reform.⁵³

7.3 Additional Motives for Saving

In our analysis so far, we have assumed that the drop in income during retirement and demand for insurance against mortality risk are the only motives for saving. In other words, a large source of inefficiency comes from households' desire to finance old-age consumption and self-insure against outliving their assets. Absent any other motive for saving, the model may over-emphasize the role of lifecycle saving and, hence, exaggerates inefficiencies caused by the annuity market incompleteness. To check the robustness of our findings, in this section we consider other motives for saving commonly considered in the literature: out-of-pocket medical expenditures and bequest motives.

⁵³Further quantitative results related to this exercise are provided in the supplementary appendix.

7.3.1 Out-of-pocket Medical Expenditures

As [De Nardi et al. \(2010\)](#) document, out-of-pocket medical expenditures rise rapidly with age and income. As people get older, this increase in their medical needs provides a strong motive to save. This motive is even stronger for those with a higher lifetime income. To examine how this additional saving motive affects our results, we introduce exogenous out-of-pocket medical expenditure to the model.⁵⁴ We assume these medical expenditures increase with age and ability type θ . More specifically, let $m_j(\theta)$ be the medical expenditure of a person of type θ at age j . In other words, we assume $m'_j(\theta) > 0$ for all θ . Moreover, to focus on saving in old age we assume there are no out-of-pocket medical expenditures prior to retirement, i.e., $m_j(\theta) = 0$ for all $j \leq R$. Finally, we assume monotonicity with respect to age. For $j, j' > R$ we assume $m_j(\theta) > m_{j'}(\theta)$ if $j > j'$ for all θ .

Let $c_j(\theta)$ be the total consumption expenditure. Individual preferences are given by

$$\sum_{t=0}^{\infty} \beta(\theta)^j P_j(\theta) [u(c_j(\theta) - m_j(\theta)) - v(l_j(\theta))].$$

In other words, individuals have preferences over non-medical expenditure and hours worked.

⁵⁵ The rest of the model is identical to what we described in Section 3. The formulation of the optimal policy problem is also very similar. Since medical expenditures vary with type, the implementability constraint is different from (13) and is given by

$$\begin{aligned} U'(\theta) = & \sum_{j=0}^J \beta(\theta)^j P_j(\theta) \frac{\varphi'_j(\theta) l_j(\theta)}{\varphi_j(\theta)} v'(l_j(\theta)) \\ & + \sum_{j=0}^J \beta(\theta)^j P_j(\theta) \left(j \frac{\beta'(\theta)}{\beta(\theta)} + \frac{P'_j(\theta)}{P_j(\theta)} \right) (u(c_j(\theta) - m_j(\theta)) - v(l_j(\theta))) \\ & - \sum_{j=0}^J \beta(\theta)^j P_j(\theta) m'_j(\theta) u'(c_j(\theta) - m_j(\theta)). \end{aligned}$$

The above constraint captures the fact that due to the presence of type-dependent medical expenditure, individuals with different types value consumption differently, and this can be used to change the incentive to work. In particular, the presence of medical expenditure that increases with types leads to forces towards the optimality of positive taxes on savings. Due to higher

⁵⁴[De Nardi et al. \(2010\)](#) find that shocks to out-of-pocket medical expenditures are not very important in accounting for the saving behavior of the elderly.

⁵⁵Here, we have assumed that health expenditure or, more generally, health status does not enter utility directly. This is in line with the results of [De Nardi et al. \(2010\)](#), who show that health-dependent utility does not explain the saving behavior of retirees.

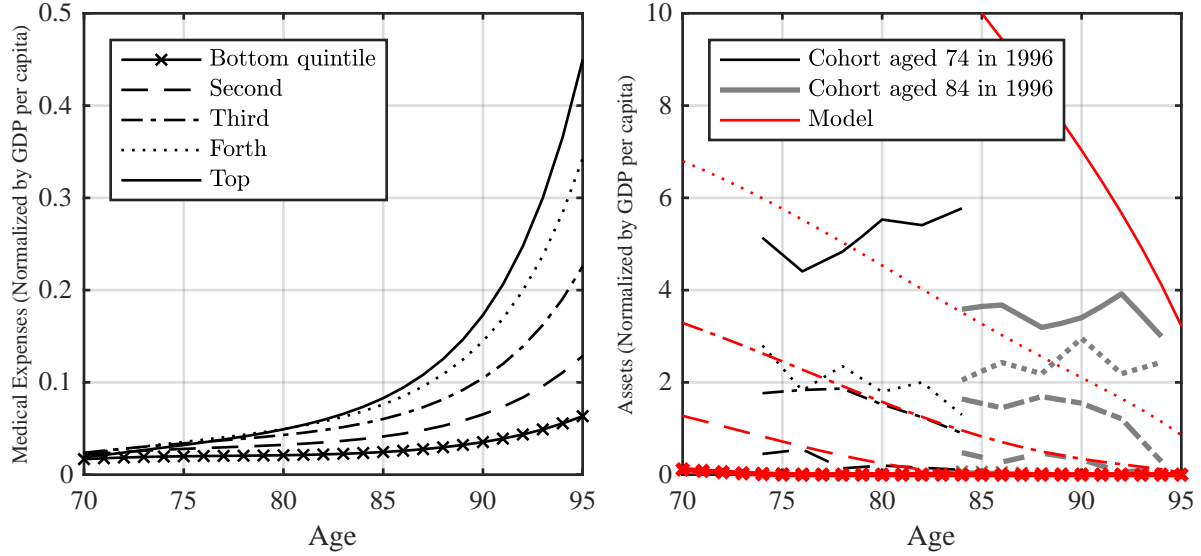


Figure 16: The left panel shows the average out-of-pocket medical expenditures by permanent income. The right panel shows the median assets by permanent income quintile in the model (solid line) and data (dashed line). Data source: [De Nardi et al. \(2010\)](#) calculations for AHEAD cohorts who were 74 and 84 years old in 1996. Note that the lowest quintile has 0 assets in the model and in the data.

medical expenditure, productive individuals have a stronger incentive to save, which creates an inelastic source of income for the government and can be taxed in order to lower the deadweight loss of earnings taxation.

In order to calibrate the out-of-pocket medical expenditure profiles, we closely follow [De Nardi et al. \(2010\)](#) and use data from the AHEAD survey between 1996 and 2006. We allow medical expenditure to depend on age and permanent income ranking (the individual's average income quintile, which can be thought of as associated with θ). This is depicted in the left panel of Figure (16). Furthermore, in order to better match the patterns of asset decumulation, we assume that σ , the coefficient of absolute risk-aversion, takes a value of 2. As before, we calibrate the average discount factor in order to match the capital output ratio. We leave the variation in the discount factor, represented by parameter β_1 , the same as in the benchmark model.

To show how well the model captures the pattern of dissaving in retirement, we plot the median assets by permanent income quintile in the model as well as the medial assets by permanent income quintile in the AHEAD data in Figure (16) (right panel). The data are based on [De Nardi et al. \(2010\)](#) calculations for AHEAD cohorts who were 74 and 84 years old in 1996. As we see, the model (solid line) captures the pattern of dissaving very well except for the assets of the top income quintile.

Using the calibrated model we compute the optimal earnings tax and asset subsidies. These

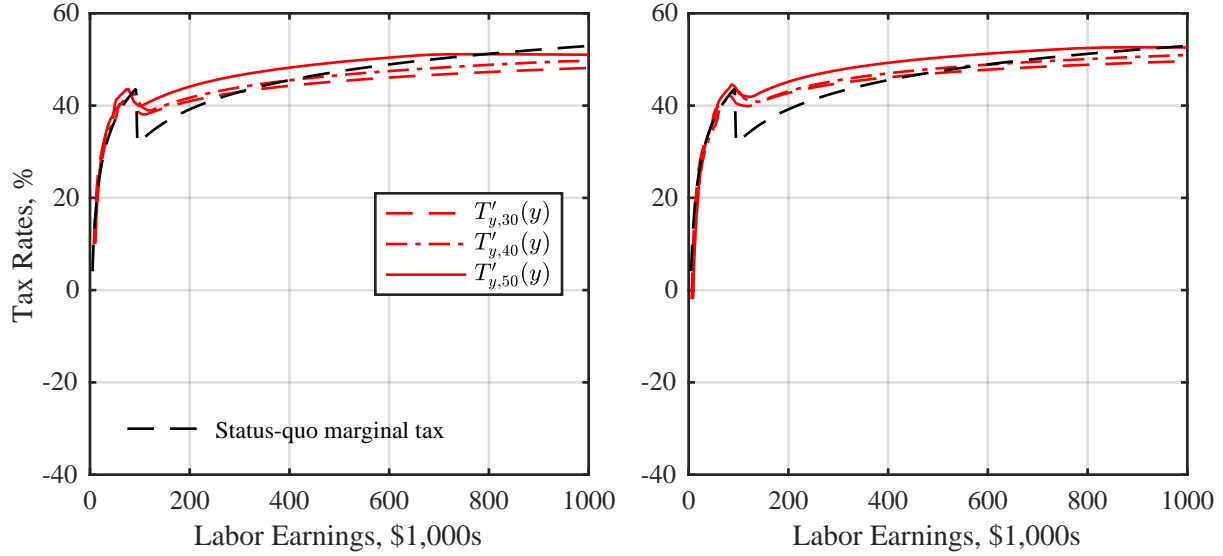


Figure 17: Optimal labor income tax functions with out-of-pocket medical expenditures. The left panel is optimal marginal taxes with out-of-pocket medical expenditures. The right panel shows the same for the benchmark model. The black dashed line is the effective status quo tax schedule.

are presented in Figure 17 and 18. As these figures demonstrate, there are no significant differences in optimal policies derived from the model with medical expenditures relative to the ones presented in the previous sections. In other words, the presence of medical expenditure does not change the prescription of our model about policy reforms. As we have mentioned before, the presence of medical expenditure that increases with earnings leads to forces towards taxation of savings. Yet, our analysis shows that such forces are not strong enough to overcome the forces to subsidize savings. This is mainly because the gradient of medical expenditure is not large enough to generate a strong motive for savings taxes.

Finally, table 7 shows the effect of optimal policies on aggregate quantities. The last row presents the efficiency gains, measured in decline in the present discounted value of lifetime consumption, next to labor income for each cohort. The magnitude of the cost savings is not very different from the ones in the main exercise. This is mainly because of the way optimal reforms affect consumption profiles over the lifetime. In particular, due to annuitization, consumption does not fall as people age and, thus, the same level of utilities can be delivered with a lower level of consumption. As a result, when we calculate the present values, the drop in consumption early in life is more pronounced because of discounting. Because of this, the cost saving measures do not drop significantly.

In summary, the inclusion of out-of-pocket-medical expenditure results in a richer model that is able to capture more details in the patterns of asset accumulation or decumulation. However, the model's implication for an optimal policy does not change. Moreover, the efficiency gains

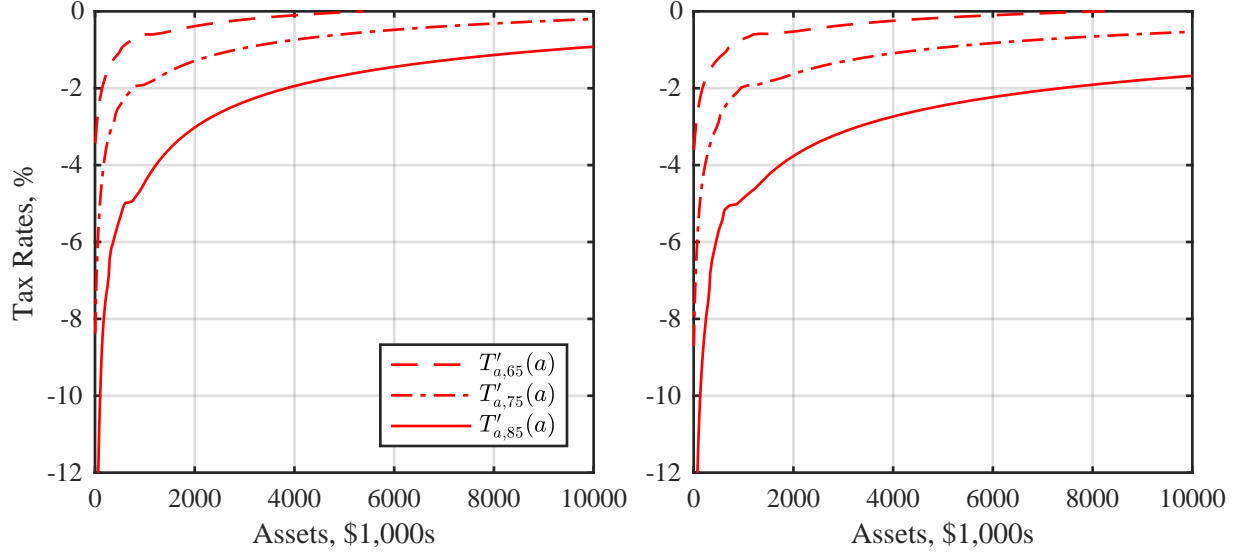


Figure 18: Optimal asset tax functions with out-of-pocket medical expenditures. The left panel shows optimal marginal taxes over all asset levels at ages 65, 75 and 85 for an economy with out-of-pocket medical expenditure. The right panel shows the same for the benchmark model.

from implementing optimal policies, although lower, are still significant and imply that the reform is effective even in the presence of out-of-pocket medical expenditure.

7.3.2 Bequest Motive

Another potentially important reason for saving is the bequest motive. When individuals want to leave assets behind, either for their descendants (altruism) or the society (joy-of-giving), they save more. In order to investigate the robustness of our results to the addition of this motive, we extend the model in (7.3.1) to allow for bequest motives. We assume that individuals have joy-of-giving bequest motives (see [De Nardi et al. \(2010\)](#), [Lockwood \(2012\)](#), and [De Nardi and Yang \(2016\)](#), among many others) given by

$$\sum_{j=0}^J \beta(\theta)^j P_j(\theta) [u(c_j(\theta) - m_j(\theta)) - v(l_j(\theta)) + \beta(1 - p_{j+1}(\theta)) w(b_{j+1}(\theta))], \quad (21)$$

where $b_{j+1}(\theta)$ is the amount of bequest left by type θ if they do not survive to age $j + 1$. We use a utility from bequest $w(\cdot)$ that implies that bequests are luxury goods. Moreover, we allow the government to differentially tax savings conditional on survival and bequests. In order to calibrate this model, we match the moments related to the distribution of bequests in the U.S.: the fraction of the deceased individuals that leave no bequests and the ratio of bequests to wealth in the aggregate.

Table 7: Aggregate effects of optimal policies with out-of-pocket medical expenditures

	Current U.S. (1)	Continue (2)	Optimal reform		
			(3)	(4)	(5)
Factor prices					
Interest rate (%)	4.05	3.14	4.05	3.84	3.06
Wage	1	1.11	1	1.02	1.12
Change (%) (relative to status quo)					
GDP	–	1.70	2.18	0.9	1.92
Consumption	–	0.19	0.34	-0.31	-0.06
Capital	–	16.21	7.63	3.95	17.97
Labor input	–	-8.24	-2.02	-1.39	-8.93
PDV of net resources	–	–	-9.67	-28.76	-7.94
Consumption equivalence			0.66	1.97	0.99

Note: Column (1) is the benchmark calibration to the current U.S. economy. Column (2) is the continuation of the U.S. status quo policies (with the consumption tax adjusted to balance government' budget constraint). Column (3) is the optimal reform policies with prices and demographics fixed at the current U.S. values. Column (4) is the optimal reform policies with equilibrium prices but fixed demographics (at the current U.S. levels). Column (5) is the optimal reform policies with equilibrium prices and future demographics. In column (3) and (4), the percentage change in the PDV is calculated relative to column (1). In column (5), the percentage change in the PDV is calculated relative to column (2).

Using this calibrated model, we perform an optimal policy reform exercise on the calibrated model that includes the medical expenditure profiles estimated in section 7.3.1. Figure 19 depicts the optimal asset subsidies compared to those in the benchmark model. Especially for lower values of assets, optimal subsidies are as large as those in the benchmark models. This is mainly because it is optimal for these individuals not to leave bequests. For higher values of assets, subsidies fall relative to the benchmark model due to the demand for bequests by richer individuals. Nevertheless, our benchmark implications for optimal policies remain roughly unchanged. An integral part of this policy is bequest taxation. In particular, for many individuals, bequests must be fully taxed away in order to solve the market incompleteness problem faced by these households. Interestingly, in an optimal policy reform most of the cost savings come from a reduction in bequests. This is because the only way for individuals to save is a risk-free asset and, as a result, bequests are too high for the status quo economy. The detailed theoretical and quantitative analysis of this model is in supplementary appendix.

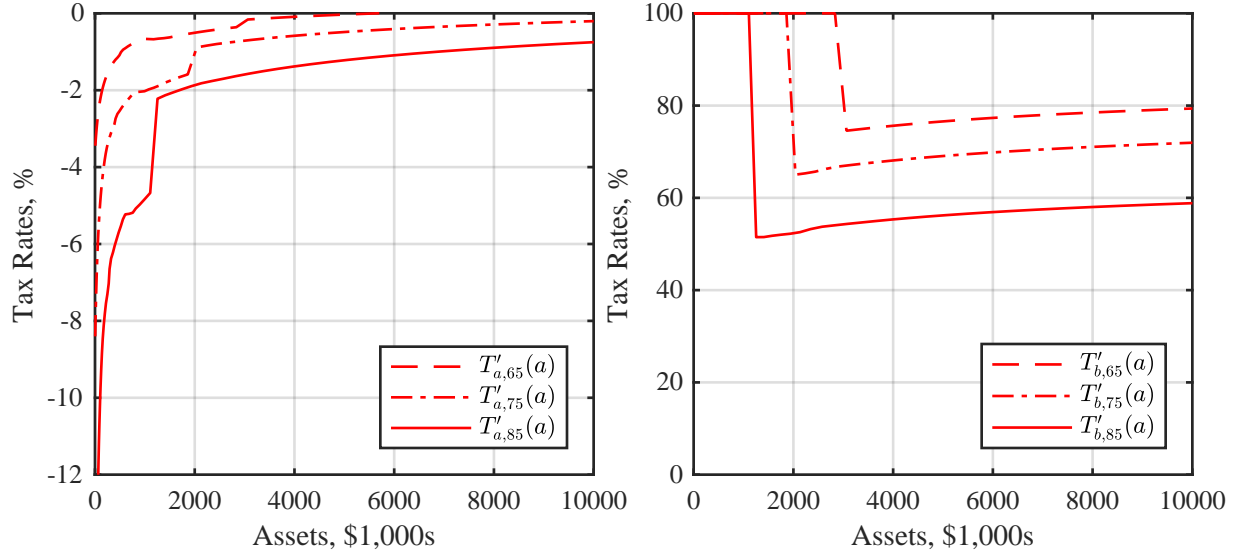


Figure 19: Optimal asset tax functions with out-of-pocket medical expenditures and bequest motives. The left panel shows the optimal marginal asset taxes over all asset levels for surviving individuals at ages 65, 75 and 85. The left panel shows the marginal bequest taxes for the same ages.

8 Conclusion

In this paper, we have provided a theoretical and quantitative analysis of Pareto optimal policy reforms aimed at financing retirement. These are reforms that intend to separate the efficiency of such schemes from their distributional consequences. Our optimal reform approach points toward the importance of subsidization of asset holdings late in life. At the same time, our analysis shows that reforms aimed at earnings taxes (such as a decline in payroll taxes or an extension of social security maximum earnings cap) are not integral to Pareto optimal reforms.

To keep our analysis tractable, we have focused on permanent ability types and abstracted from idiosyncratic shocks that are the focus of most of the optimal dynamic tax literature. Inclusion of these shocks introduces additional reasons for taxing capital (as in [Goloso et al. \(2003\)](#) and [Goloso et al. \(2016\)](#)) in the pre-retirement period. As shown by others, such shocks induce very little reason to tax capital income (see [Farhi and Werning \(2012\)](#)), compared to the magnitude of our savings distortions. Hence, we have good reasons to believe that including shocks to earnings does not alter our results.

A key feature of our model is the correlation between earning ability and mortality. In choosing this assumption, we were guided by the large body of evidence that points to a strong correlation between socioeconomic factors (such as income or education) and mortality rates. We take an extreme view and assume that this correlation is exogenously given and individuals' choice

has no effect on their mortality. In reality, many individuals affect their mortality through the decisions they make over their lifetime. We choose to ignore these effects due to two reasons. First, as Ales et al. (2014) show, when individuals differ in their earning ability, and mortality is endogenous, efficiency implies more investment in the survival of the higher-ability individuals. Hence, it is never efficient to eliminate the correlation between ability and mortality. Second, in any model in which the length of life is endogenous, the level of utility flow becomes important in marginal decisions by individuals. This makes analysis of such models very complicated and intractable. It is important, however, to know how inclusion of endogenous mortality affects our analysis of optimal policy. We leave this for future research.

References

- ABEL, A. B., N. G. MANKIW, L. H. SUMMERS, AND R. J. ZECKHAUSER (1989): “Assessing Dynamic Efficiency: Theory and Evidence,” *Review of Economic Studies*, 56, 1–19.
- ALES, L., R. HOSSEINI, AND L. E. JONES (2014): “Is There Too Much Inequality in Health Spending Across Income Groups?” *Unpublished*.
- ALTIG, D., A. J. AUERBACH, L. J. KOTLIKOFF, K. A. SMETTERS, AND J. WALLISER (2001): “Simulating fundamental tax reform in the United States,” *American Economic Review*, 91, 574–595.
- ATKINSON, A. B. AND J. E. STIGLITZ (1976): “The design of tax structure: direct versus indirect taxation,” *Journal of public Economics*, 6, 55–75.
- BADEL, A. AND M. HUGGETT (2014): “Taxing top earners: a human capital perspective,” *Federal Reserve Bank of St. Louis Working Paper Series*.
- BELL, F. C. AND M. L. MILLER (2005): “Life Tables for the United States Social Security Area 1900–2100,” Actuarial Study 120, Social Security Administration.
- BELLOFATTO, A. A. (2015): “Wealth Taxation and Life Expectancy,” *Unpublished*.
- BENABOU, R. (2002): “Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?” *Econometrica*, 70, 481–517.
- BENARTZI, S., A. PREVITERO, AND R. H. THALER (2011): “Annuitization puzzles,” *The Journal of Economic Perspectives*, 143–164.
- BLANDIN, A. (2018): “Human Capital and the Social Security Tax Cap,” *Unpublished*.

- BROWN, J. R. AND J. M. POTERBA (2006): "Household ownership of variable annuities," *Tax Policy and the Economy*, 20, 163–191.
- CALIENDO, F. N., N. L. GUO, AND R. HOSSEINI (2014): "Social security is NOT a substitute for annuity markets," *Review of Economic Dynamics*, 17, 739–755.
- COLBY, S. L. AND J. M. ORTMAN (2015): "Projections of the Size and Composition of the U.S. Population: 2014 to 2060," *Current Population Reports*.
- CONESA, J. C. AND C. GARRIGA (2008): "Optimal Fiscal Policy In The Design Of Social Security Reforms," *International Economic Review*, 49, 291–318.
- CONESA, J. C. AND D. KRUEGER (2006): "On the optimal progressivity of the income tax code," *Journal of Monetary Economics*, 53, 1425–1450.
- CRISTIA, J. P. (2009): "Rising mortality and life expectancy differentials by lifetime earnings in the United States," *Journal of Health Economics*, 28, 984–995.
- DE NARDI, M., E. FRENCH, AND J. B. JONES (2010): "Why do the elderly save? The role of medical expenses," *Journal of Political Economy*, 118, 39–75.
- DE NARDI, M. AND F. YANG (2016): "Wealth Inequality, Family Background, and Estate Taxation," *Journal of Monetary Economics*, 77.
- DIAMOND, P. (2004): "Social Security," *American Economic Review*, 94, 1–24.
- DIAMOND, P. A. (1965): "National debt in a neoclassical growth model," *The American Economic Review*, 55, 1126–1150.
- DYNAN, K. E., J. SKINNER, AND S. P. ZELDES (2004): "Do the Rich Save More?" *Journal of Political Economy*, 112.
- EINAV, L., A. FINKELSTEIN, AND P. SCHRIMPF (2010): "Optimal mandates and the welfare cost of asymmetric information: Evidence from the uk annuity market," *Econometrica*, 78, 1031–1092.
- FARHI, E. AND I. WERNING (2012): "Capital taxation: Quantitative explorations of the inverse Euler equation," *Journal of Political Economy*, 120, 398–445.
- (2013a): "Estate Taxation with Altruism Heterogeneity," *American Economic Review*, 103, 489–95.
- (2013b): "Insurance and Taxation over the Life Cycle," *The Review of Economic Studies*, 80, 596–635.

- FLODEN, M. (2001): "The Effectiveness of Government Debt and Transfers as Insurance," *Journal of Monetary Economics*, 48, 81–108.
- FRIEDMAN, B. AND M. WARSHAWSKY (1990): "The Cost of Annuities: Implications for Saving Behavior and Bequests," *Quarterly Journal of Economics*, 105, 135–54.
- FUKUSHIMA, K. (2011): "Quantifying the welfare gains from flexible dynamic income tax systems," *manuscript*.
- GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70, 569–587.
- GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2016): "Redistribution and social insurance," *American Economic Review*, 106, 359–86.
- GOLOSOV, M., M. TROSHKIN, A. TSYVINSKI, AND M. WEINZIERL (2013): "Preference heterogeneity and optimal capital income taxation," *Journal of Public Economics*, 97, 160–175.
- GOLOSOV, M. AND A. TSYVINSKI (2006): "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy*, 114, 257–269.
- HARRIS, E. (2005): "Effective Marginal Tax Rates on Labor Income," Congress of the United States, Congressional Budget Office.
- HEATHCOTE, J., F. PERRI, AND G. L. VIOLANTE (2010): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic dynamics*, 13, 15–51.
- HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2014): "Optimal tax progressivity: An analytical framework," Tech. rep., National Bureau of Economic Research.
- HEATHCOTE, J. AND H. TSUJIYAMA (2015): "Optimal Income Taxation: Mirrlees Meets Ramsey," *Research Department Staff Report 507*.
- HONG, J. AND J.-V. RÍOS-RULL (2007): "Social Security, Life Insurance and Annuities for Families," *Journal of Monetary Economics*, 54, 118–40.
- HORIUCHI, S. AND A. J. COALE (1982): "A simple equation for estimating the expectation of life at old ages," *Population Studies*, 36, 317–326.
- HOSSEINI, R. (2015): "Adverse Selection in the Annuity Market and the Role for Social Security," *Journal of Political Economy*, 123, 941–984.

- HUBBARD, R. G. AND K. L. JUDD (1987): "Social Security and Individual Welfare: Precautionary Saving, Borrowing Constraints, and the Payroll Tax," *American Economic Review*, 77, 630–46.
- HUGGETT, M. AND J. C. PARRA (2010): "How Well Does the U.S. Social Insurance System Provide Social Insurance?" *Journal of Political Economy*, 118, 76–112.
- HUGGETT, M. AND G. VENTURA (1999): "On the distributional effects of social security reform," *Review of Economic Dynamics*, 2, 498–531.
- İMROHOROĞLU, A., S. İMROHOROĞLU, AND D. H. JOINES (1995): "A Life Cycle Analysis of Social Security," *Economic Theory*, 6, 83–114.
- JAMES, E. AND D. VITTAS (2000): *Annuity markets in comparative perspective: Do consumers get their money's worth?*, vol. 2493, World Bank Publications.
- KITAO, S. (2014): "Sustainable Social Security: Four Options," *Review of Economic Dynamics*, 17, 756–779.
- LOCKWOOD, L. (2012): "Bequest Motives and the Annuity Puzzle," *Review of Economic Dynamics*, 15, 226–243.
- LUENBERGER, D. G. (1997): *Optimization by vector space methods*, John Wiley & Sons.
- MCGRATTAN, E. R. AND E. C. PRESCOTT (2017): "On financing retirement with an aging population," *Quantitative Economics*, 8, 75–115.
- MENDOZA, E. G., A. RAZIN, AND L. L. TESAR (1994): "Effective Tax Rates in Macroeconomics: Cross-Country Estimates of Tax Rates on Factor Incomes and Consumption," *Journal of Monetary Economics*, 34, 297–323.
- MIRRLEES, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208.
- MITCHELL, O., J. POTERBA, M. WARSHAWSKY, AND J. BROWN (1999): "New Evidence on the Money's Worth of Individual Annuities," *American Economic Review*, 89, 1299–1318.
- NDIAYE, A. (2017): "Flexible Retirement and Optimal Taxation," .
- NISHIYAMA, S. AND K. SMETTERS (2007): "Does Social Security Privatization Produce Efficiency Gains?" *Quarterly Journal of Economics*, 122, 1677–1719.
- PIKETTY, T. AND E. SAEZ (2013): "A theory of optimal inheritance taxation," *Econometrica*, 81, 1851–1886.

- POTERBA, J. (2001): “The History of Annuities in the United States,” in *The Role of Annuity Markets in Financing Retirement*, ed. by J. Brown, O. Mitchell, J. Poterba, and M. Warshawsky, Cambridge: MIT Press, 23–56.
- POTERBA, J. M. (2014): “Retirement security in an aging population,” *The American Economic Review*, 104, 1–30.
- REED, W. J. AND M. JORGENSEN (2004): “The double Pareto-lognormal distribution—a new parametric model for size distributions,” *Communications in Statistics-Theory and Methods*, 33, 1733–1753.
- SAEZ, E. (2001): “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68, 205–229.
- SHOURIDEH, A. AND M. TROSHKIN (2017): “Providing Efficient Incentives to Work: Retirement Ages and the Pension System,” *Available at SSRN 2062355*.
- SIEGEL, J. J. AND D. G. COXE (2002): *Stocks for the long run*, vol. 3, McGraw-Hill New York.
- SØRENSEN, P. B. (2004): “Measuring Taxes on Capital and Labor: An Overview of Methods and Issues,” .
- WALDRON, H. (2013): “Mortality Differentials by Lifetime Earnings Decile: Implications for Evaluations of Proposed Social Security Law Changes,” *Social Security Bulletin*, 73.
- WEINZIERL, M. (2011): “The Surprising Power of Age-Dependent Taxes,” *Review of Economic Studies*, 78, 1490–1518.
- WERNING, I. (2007): “Pareto Efficient Income Taxation,” *Massachusetts Institute of Technology, Manuscript*.
- WOO, L. AND D. BUCHHOLZ (2006): “Subsidies for Assets: A New Look at the Federal Budget,” in *Assets Learning Conference – A Lifetime of Assets, sponsored by the Federal Reserve and CFED*.

Appendix

A Proofs

A.1 Proof of Proposition 1

We first show the following lemma:

Lemma 5. *A feasible allocation $\{c_{1,t}(\theta), c_{2,t}(\theta), y_t(\theta)\}_{\theta \in \Theta, t \geq 0}$ together with capital allocation K_t is induced by some sequence of tax functions $T_{y,t}(\cdot), T_{a,t}(\cdot, \cdot)$ if and only if*

$$\theta \in \arg \max_{\hat{\theta}} U \left(c_{1,t}(\hat{\theta}), c_{2,t}(\hat{\theta}), \frac{y(\hat{\theta})}{\theta} \right). \quad (22)$$

Proof. Suppose that an allocation is induced by a sequence of tax functions and suppose that for some types θ and θ'

$$U \left(c_{1,t}(\theta'), c_{2,t}(\theta'), \frac{y(\theta')}{\theta} \right) > U \left(c_{1,t}(\theta), c_{2,t}(\theta), \frac{y(\theta)}{\theta} \right)$$

Then facing the tax functions, an agent of type θ at t can choose $c_{1,t}(\theta'), y_t(\theta'), c_{2,t}(\theta')$ - this is a feasible choice since budget constraints are independent of agent's types. This implies that the original allocations cannot be induced by the tax functions as the allocations are not optimal under the tax codes.

Now consider a feasible allocation that satisfies the condition in the statement of Lemma. Let $a_t(\theta)$ be defined by

$$a_t(\theta) = \frac{w_t y_t(\theta) - c_{1,t}(\theta)}{q_t}$$

where $w_t = F_L(K_t, N_t \int y_t(\theta) dH)$. Then let $\hat{T}_{a,t+1}$ be defined by

$$\hat{T}_{a,t+1}(w_t y_t(\theta), (1 + r_{t+1}) a_{t+1}(\theta)) = c_{2,t}(\theta) - (1 + r_{t+1}) a_{t+1}(\theta)$$

where $r_t = F_K(K_t, N_t \int y_t dH(\theta))$. Note that this tax function is well-defined as if $y_t(\theta)$ and $a_{t+1}(\theta)$ are the same for two types, then the incentive compatibility constraint implies that $c_{2,t}(\theta)$ must also be the same and therefore so is the value of \hat{T}_a . Furthermore, for a value $(w_t y, (1 + r_{t+1}) a)$ with $(y, a) \neq (y_t(\theta), a_{t+1}(\theta))$, we choose a value for $\hat{T}_{a,t+1}$ so that these points are not chosen by any type θ - this is easily done by considering the value for the highest type that benefits from such a point and choosing it high enough so that such type does not want to choose this point. If under this construction, $q_t \int a_{t+1} dH \neq K_{t+1}$, then we can adjust the tax

function by a constant in order to make this equality be satisfied.

By the incentive compatibility and the construction of $\hat{T}_{a,t+1}$ and $\hat{T}_y := 0$, it is optimal for an individual of type θ to choose the desired allocation. Since this allocation is feasible, it must be induced by the constructed tax functions. \square

Now we prove Proposition 1

Proof. Given the above Lemma, we can focus on allocations. In particular, among the set of feasible and incentive compatible allocations (those satisfying (22)), those induced by Pareto optimal tax functions must be Pareto optimal themselves. In what follows, we characterize the set of Pareto optimal allocation. A useful property that helps us in our analysis is that under our assumption of the utility function, the set of incentive compatible allocations is linear in the utility space. This property allows us to use standard separating hyperplane arguments to show that an allocation is Pareto optimal if and only if a positive continuous function $\alpha(t, \theta)$ exists so that this allocation is the solution to the following planning problem

$$\max \sum_{t=0}^{\infty} \int \alpha(t, \theta) U \left(c_{1,t}(\theta), c_{2,t}(\theta), \frac{y_t(\theta)}{\theta} \right) dH(\theta)$$

subject to

$$\begin{aligned} \theta &\in \arg \max_{\hat{\theta} \in \Theta} U \left(c_{1,t}(\hat{\theta}), c_{2,t}(\hat{\theta}), \frac{y_t(\hat{\theta})}{\theta} \right) \\ K_t + F \left(K_t, N_t \int y_t dH \right) &\geq N_t \int c_{1,t} dH + N_{t-1} \int c_{2,t-1} dH + K_{t+1} \end{aligned}$$

Since if we rewrite the constraint set in terms of utilities, it is a convex set, we can write the above in its dual form

$$\max \sum_{t=0}^{\infty} \lambda_t \left[F \left(K_t, N_t \int y_t dH \right) + K_t - K_{t+1} - N_t \int c_{1,t} dH - N_{t-1} \int c_{2,t-1} dH \right] \quad (\text{P1})$$

subject to

$$\begin{aligned} \theta &\in \arg \max_{\hat{\theta} \in \Theta} U \left(c_{1,t}(\hat{\theta}), c_{2,t}(\hat{\theta}), \frac{y_t(\hat{\theta})}{\theta} \right) \\ U \left(c_{1,t}(\theta), c_{2,t}(\theta), \frac{y_t(\theta)}{\theta} \right) &\geq W_t(\theta) \end{aligned}$$

where $W_t(\theta)$ is the utility of each individual at date t under the specified allocation. Note that since the objective is strictly concave - if we rewrite things in terms of utilities - and the constraint set is convex, the solution to this planning problem is unique.

Now consider the solution to the above problem for a sequence of λ_t 's. Then the First Order Conditions with respect to K_t satisfy

$$\lambda_t F_{K,t} + \lambda_t = \lambda_{t-1}$$

Now, if we let

$$w_t = F_L \left(K_t, N_t \int y_t dH \right),$$

then the solution to the above optimization problem is also a solution to

$$\max \sum_{t=0}^{\infty} \lambda_t \left[w_t N_t \int y_t dH - N_t \int c_{1,t} dH - N_{t-1} \int c_{2,t-1} dH \right]$$

subject to

$$\begin{aligned} \theta &\in \arg \max_{\hat{\theta} \in \Theta} U \left(c_{1,t}(\hat{\theta}), c_{2,t}(\hat{\theta}), \frac{y_t(\hat{\theta})}{\theta} \right) \\ U \left(c_{1,t}(\theta), c_{2,t}(\theta), \frac{y_t(\theta)}{\theta} \right) &\geq W_t(\theta) \end{aligned}$$

given $\{w_t\}_{t \geq 0}$. We can rewrite the above optimization as

$$\max \sum_{t=0}^{\infty} \lambda_t N_t \left[w_t \int y_t dH - \int c_{1,t} dH - \frac{\lambda_{t+1}}{\lambda_t} \int c_{2,t} dH \right]$$

subject to

$$\begin{aligned} \theta &\in \arg \max_{\hat{\theta} \in \Theta} U \left(c_{1,t}(\hat{\theta}), c_{2,t}(\hat{\theta}), \frac{y_t(\hat{\theta})}{\theta} \right) \\ U \left(c_{1,t}(\theta), c_{2,t}(\theta), \frac{y_t(\theta)}{\theta} \right) &\geq W_t(\theta) \end{aligned}$$

If we define $1 + r_{t+1} = \lambda_t / \lambda_{t+1}$, then since each generation's contribution to the objective is additively separable, the solution to the above must also solve the optimization (P). Now, if an allocation solves optimization (P), then it must be the solution of the above problem where $\lambda_t =$

$\prod_{s=0}^t (1 + F_{K,s})^{-1}$. By assumption $\gamma < F_{K,t} - n$ as a result, $N_t \lambda_t \rightarrow 0$ and the objective in the above is well-defined. Now since given these values of λ_t , the solution to the above satisfies the FOC's associated with (P1) and the solution to (P1) is unique, the allocation must be Pareto optimal. This concludes the proof. \square

A.2 Proof of Proposition 2

Proof. For the class of preferences considered, any Pareto optimal allocation induced by some tax function must solve planning problem (P). By the no-bunching assumption, we can replace the incentive compatibility constraint with its associated first order condition

$$U'(\theta) = \left(\frac{\beta'(\theta)}{\beta(\theta)} + \frac{P'(\theta)}{P(\theta)} \right) \beta(\theta) P(\theta) u(c_2(\theta)) + \frac{y(\theta)}{\theta^2} v' \left(\frac{y(\theta)}{\theta} \right)$$

where $U(\theta)$ is the utility of individual θ . The first order conditions associated with this problem are given by

$$\begin{aligned} -1 + \eta u'(c_{1,t}) &= 0 \\ w_t - \eta \frac{1}{\theta} v' - \mu \frac{1}{\theta^2} \left(v' + \frac{y_t}{\theta} v'' \right) &= 0 \\ -\frac{P}{1 + r_{t+1}} + \eta \beta P u'(c_{2,t}) - \mu \left(\frac{\beta'}{\beta} + \frac{P'}{P} \right) \beta P u'(c_{2,t}) &= 0 \\ -\eta - \frac{1}{h} (\mu h)' + \gamma &= 0 \\ \mu(\bar{\theta}) h(\bar{\theta}) + \gamma(\bar{\theta}) &= 0 \\ \mu(\bar{\theta}) h(\bar{\theta}) - \gamma(\bar{\theta}) &= 0 \end{aligned}$$

Pareto optimality of the allocation implies that $\gamma \geq 0$ for all values of θ . By definition of the labor and saving wedges, we have

$$\begin{aligned} 1 - \tau_{l,t} &= \frac{v'(y_t/\theta)}{w_t u'(c_{1,t}) \theta} \\ 1 - \tau_{a,t} &= \frac{u'(c_{1,t})}{\beta P (1 + r_{t+1}) u'(c_{2,t})} \end{aligned}$$

The above implies that

$$\begin{aligned}
\mu &= \frac{w_t - \eta \frac{1}{\theta} v'}{\frac{1}{\theta^2} (v' + \frac{y}{\theta} v'')} = \theta^2 \frac{w_t - \frac{1}{u'(c_1)} \frac{1}{\theta} v'}{v' (1 + \frac{1}{\varepsilon})} \\
&= w_t \theta^2 \frac{\tau_{l,t}}{v'} \frac{\varepsilon}{1 + \varepsilon} \\
&= \theta \frac{\tau_{l,t}}{1 - \tau_{l,t}} \frac{\varepsilon}{1 + \varepsilon} \frac{1}{u'(c_{1,t})}
\end{aligned}$$

Note that the FOC's also imply that

$$\begin{aligned}
-\frac{u'(c_{1,t})}{(1 + r_{t+1}) \beta u'(c_{2,t})} + 1 &= u'(c_{1,t}) \mu \left(\frac{\beta'}{\beta} + \frac{P'}{P} \right) \\
-\frac{P}{q_t} (1 - \tau_{a,t}) + 1 &= \theta \frac{\tau_{l,t}}{1 - \tau_{l,t}} \frac{\varepsilon}{1 + \varepsilon} \left(\frac{\beta'}{\beta} + \frac{P'}{P} \right) \\
\tau_{a,t} &= 1 - \frac{1}{P} + \frac{1}{P} \theta \frac{\tau_{l,t}}{1 - \tau_{l,t}} \frac{\varepsilon}{1 + \varepsilon} \left(\frac{\beta'}{\beta} + \frac{P'}{P} \right).
\end{aligned}$$

By rearranging the terms, we can rewrite the above as

$$P \tau_{a,t} + 1 - P = \theta \frac{\tau_{l,t}}{1 - \tau_{l,t}} \frac{1}{\frac{1}{\varepsilon} + 1} \left(\frac{\beta'}{\beta} + \frac{P'}{P} \right)$$

which results in the condition stated in the proposition. \square

A.3 Proof of Proposition 3

Proof. Consider the first order conditions derived in section A.2. Then Pareto optimality of the allocation implies that for $\theta, \gamma \geq 0$. Therefore, we must have

$$\begin{aligned}
\eta &\geq -\frac{1}{h} (\mu h)' \\
\frac{1}{u'(c_1)} &\geq -\frac{1}{h} (\mu h)' = \mu \left(\frac{h'}{h} + \frac{\mu'}{\mu} \right) \\
\frac{1}{u'(c_{1,t})} &\geq -\theta \frac{\tau_{l,t}}{1 - \tau_{l,t}} \frac{\varepsilon}{1 + \varepsilon} \frac{1}{u'(c_{1,t})} \left[\frac{h'}{h} + \frac{1}{\theta} + \frac{\tau'_{l,t}}{\tau_{l,t} (1 - \tau_{l,t})} + \frac{-u''(c_{1,t}) c_{1,t} c'_{1,t}}{u'(c_{1,t}) c_{1,t}} \right] \\
1 &\geq -\theta \frac{\tau_{l,t}}{1 - \tau_{l,t}} \frac{\varepsilon}{1 + \varepsilon} \left[\frac{h'}{h} + \frac{1}{\theta} + \frac{\tau'_{l,t}}{(1 - \tau_{l,t}) \tau_{l,t}} + \frac{-u''(c_{1,t}) c_{1,t} c'_{1,t}}{u'(c_{1,t}) c_{1,t}} \right].
\end{aligned}$$

This complete the proof of the first part.

Note that under the assumption that first order conditions fully characterize optimal allocation, the local incentive constraint is sufficient for global incentive compatibility. As a result, the

planning problem for each generation is given by

$$\max_{c_1(\theta), c_2(\theta), y(\theta)} \int \left[y(\theta) - c_1(\theta) - \frac{P(\theta)}{1 + r_{t+1}} c_2(\theta) \right] dH(\theta)$$

subject to

$$\begin{aligned} U'(\theta) &= \left(\frac{\beta'(\theta)}{\beta(\theta)} + \frac{P'(\theta)}{P(\theta)} \right) \beta(\theta) P(\theta) u(c_2(\theta)) + \frac{y(\theta)}{\theta^2} v' \left(\frac{y(\theta)}{\theta} \right) \\ U(\theta) &= u(c_1(\theta)) + \beta(\theta) u(c_2(\theta)) - v(y(\theta)/\theta) \\ U(\theta) &\geq W_t(\theta). \end{aligned}$$

Under the assumption that $v(\ell) = \psi \ell^\gamma / \gamma$, we have

$$\ell v'(\ell) = \gamma v(\ell)$$

Now, if we define the variables $v(\theta) \equiv v(\ell(\theta))$, $u_1(\theta) \equiv u(c_1(\theta))$, and $u_2(\theta) \equiv u(c_2(\theta))$ and let $C(u)$ and $L(v)$ be the inverse of $u(\cdot)$ and $v(\cdot)$ respectively, the above optimization problem can be written as

$$\max_{u_1(\theta), u_2(\theta), v(\theta)} \int \left[\theta L(v(\theta)) - C(u_1(\theta)) - \frac{P(\theta)}{1 + r_{t+1}} C(u_2(\theta)) \right] dH(\theta)$$

subject to

$$\begin{aligned} U'(\theta) &= \left(\frac{\beta'(\theta)}{\beta(\theta)} + \frac{P'(\theta)}{P(\theta)} \right) \beta(\theta) P(\theta) u_2(\theta) + \gamma \frac{1}{\theta} v(\theta) \\ U(\theta) &= u_1(\theta) + \beta(\theta) u_2(\theta) - v(\theta) \\ U(\theta) &\geq W_t(\theta). \end{aligned}$$

As it can be seen, the constraint set of the above optimization problem is linear in $u_1(\theta)$, $u_2(\theta)$, $v(\theta)$. Since its objective is strictly concave in these variables – $C(\cdot)$ and $L(\cdot)$ are strictly convex and concave respectively – by Theorem 1, Chapter 8 of [Luenberger \(1997\)](#), the first order conditions are sufficient. This implies that (3) and (5) are sufficient for Pareto optimality of a tax schedule. \square

A.4 Proof of Proposition 4

The planning problem (P1), replaces a global implementability as in (1) with its local equivalent (13). We start by deriving this local implementability constraint for the planning problem:

A.4.1 Derivation of Local Implementability Constraint (13)

Consider the individual maximization problem for type θ where hours l_j are replaced by $y_j = \varphi_j(\theta) l_j$:

$$U(\theta) = \max_{c_j, y_j, a_{j+1}} \sum_{j=0}^J \beta(\theta)^j P_j(\theta) \left(u(c_j(\theta)) - v\left(\frac{y_j(\theta)}{\varphi_j(\theta)}\right) \right)$$

subject to (7). Note that θ does not appear in the budget constraint.

Now take envelope condition with respect to θ

$$\begin{aligned} U'(\theta) &= \sum_{j=0}^J \left(\frac{j\beta'(\theta)}{\beta(\theta)} + \frac{P'_j(\theta)}{P_j(\theta)} \right) \beta(\theta)^j P_j(\theta) \left[u(c_j) - v\left(\frac{y_j}{\varphi_j(\theta)}\right) \right] \\ &\quad + \sum_{j=0}^J \beta(\theta)^j P_j(\theta) \left[\frac{\varphi'_j(\theta) y_j}{\varphi_j(\theta)^2} v'\left(\frac{y_j}{\varphi_j(\theta)}\right) \right]. \end{aligned}$$

Now replace l_j back and evaluate at the solution $\{c_j(\theta), l_j(\theta)\}$

$$\begin{aligned} U'(\theta) &= \sum_{j=0}^J \beta(\theta)^j P_j(\theta) \left[\frac{\varphi'_j(\theta) l_j(\theta)}{\varphi_j(\theta)} v'(l_j(\theta)) \right] \\ &\quad + \sum_{j=0}^J \left(\frac{j\beta'(\theta)}{\beta(\theta)} + \frac{P'_j(\theta)}{P_j(\theta)} \right) \beta(\theta)^j P_j(\theta) [u(c_j(\theta)) - v(l_j(\theta))]. \end{aligned}$$

We now turn to the proof of Proposition 4. To avoid clutter, assume $u_{c,j}(\theta) \equiv u'(c_j(\theta))$ and $v_{l,j}(\theta) \equiv v'(l_j(\theta))$. Also we will drop dependence on θ whenever possible.

Proof. Let $H'(\theta) = h(\theta)$ where $h(\theta)$ is the density function. Let $\eta(\theta)h(\theta)$, $\mu(\theta)h(\theta)$ and $\gamma(\theta)h(\theta)$ be multipliers on equations (12), (13) and (14) respectively. The first order conditions for the problem (P1) are

$$\left(\eta + \mu \left(\frac{P'_j}{P_j} + j \frac{\beta'}{\beta} \right) \right) u'(c_j) = \frac{1}{\beta^j (1+r)^j} \quad (23)$$

$$\left(\eta - \mu \frac{\varphi'_j}{\varphi_j} \left(1 + l_j \frac{v''(l_j)}{v'(l_j)} \right) + \mu \left(\frac{P'_j}{P_j} + j \frac{\beta'}{\beta} \right) \right) v'(l_j) = \frac{\varphi_j}{\beta^j (1+r)^j} \quad (24)$$

$$\eta - \mu \frac{h'}{h} - \gamma = \mu', \quad (25)$$

and the boundary conditions

$$\mu(\underline{\theta}) = \mu(\bar{\theta}) = 0.$$

Recall that

$$1 - \tau_{l,j} = \frac{v_{l,j}}{\varphi_j u_{c,j}}$$

$$p_{j+1} (1 - \tau_{a,j+1}) = \frac{u_{c,j}}{\beta(1+r)u_{c,j+1}}$$

Evaluate these equations at $j = 0$, we get

$$\frac{1}{u_{c,0}} = \eta$$

$$\frac{\varphi}{v'_{l,0}} = \eta - \mu \frac{\varphi_{0,\theta}}{\varphi_0} \left(1 + \frac{1}{\varepsilon}\right)$$

$$\frac{\tau_{l,0}}{1 - \tau_{l,0}} \frac{1}{u_{c,0}} = -\mu \frac{\varphi_{0,\theta}}{\varphi_0} \left(1 + \frac{1}{\varepsilon}\right)$$

$$\mu = -\frac{\tau_{l,0}}{1 - \tau_{l,0}} \frac{1}{u_{c,0}} \frac{\varphi_0}{\varphi_{0,\theta}} \frac{\varepsilon}{1 + \varepsilon}.$$

Also

$$\eta - \mu \frac{h'}{h} - \gamma = \mu'.$$

As in Proposition 2, the allocation is Pareto optimal if $\gamma \geq 0$

$$\eta - \mu \frac{h'}{h} \leq \mu'.$$

Replacing all the terms gives the inequality at $j = 0$

$$1 \geq -\frac{\tau_{l,0}}{1 - \tau_{l,0}} \frac{\varphi_0}{\varphi_{0,\theta}} \frac{\varepsilon}{1 + \varepsilon} \left(\frac{h'}{h} + \frac{\tau'_{l,0}}{(1 - \tau_{l,0}) \tau_{l,0}} + \frac{\varepsilon'_{F,0}}{\varepsilon_{F,0} (1 + \varepsilon_{F,0})} + \frac{\varphi'_0}{\varphi_0} - \frac{\varphi''_0}{\varphi'_0} + \sigma \frac{c'_0}{c_0} \right), \quad (26)$$

where $\sigma = -\frac{u_{cc,0}c_0}{u_{c,0}}$ and $\varepsilon_{F,0}$ is the Frisch elasticity of labor supply, $\frac{v'(\ell)}{\ell v''(\ell)}$, at $j = 0$.

Note, also that combining (23) and (24) we get

$$-(\beta(1+r))^j \mu \frac{\varphi_{j,\theta}}{\varphi_j} \left(1 + \frac{1}{\varepsilon}\right) = \frac{\varphi_j}{v_{l,j}} - \frac{1}{u_{c,j}}$$

$$= \frac{\tau_{l,j}}{1 - \tau_{l,j}} \frac{1}{u_{c,j}}.$$

Additionally, we can combine (23) for two consecutive ages to arrive at

$$\begin{aligned}
(\beta(1+r))^j \left\{ \eta + \mu \left[\frac{P'_{j+1}}{P_{j+1}} + (j+1) \frac{\beta'}{\beta} \right] - \left(\eta + \mu \left[\frac{P'_j}{P_j} + j \frac{\beta'}{\beta} \right] \right) \right\} &= \frac{1}{\beta(1+r)u_{c,j+1}} - \frac{1}{u_{c,j}} \\
(\beta(1+r))^j \mu \left[\frac{p'_{j+1}}{P_{j+1}} + \frac{\beta'}{\beta} \right] &= \frac{1}{\beta(1+r)u_{c,j+1}} - \frac{1}{u_{c,j}} \\
&= \frac{1}{u_{c,j}} \left[\frac{u_{c,j}}{\beta(1+r)u_{c,j+1}} - 1 \right] \\
(\beta(1+r))^j \mu \left[\frac{p'_{j+1}}{P_{j+1}} + \frac{\beta'}{\beta} \right] &= \frac{1}{u_{c,j}} [(1 - \tau_{a,j+1})p_{j+1} - 1]
\end{aligned} \tag{27}$$

Therefore

$$[1 - (1 - \tau_{a,j+1})p_{j+1}] \frac{1}{\frac{p'_{j+1}}{P_{j+1}} + \frac{\beta'}{\beta}} = \frac{\tau_{l,j}}{1 - \tau_{l,j}} \frac{1}{\frac{\varphi_{j,\theta}}{\varphi_j} (1 + \frac{1}{\varepsilon})}$$

which establishes (16). Equation (17) can be simply derived by rewriting (27) for two consecutive ages.

The sufficiency of these conditions can be shown using an argument which is very similar to the sufficiency in Proposition 3; it uses the linearity of the incentive constraint in utility space. It is thus omitted to avoid repetition. \square

B Intuitive Derivation of Tax Smoothing Formulas

In this section, we describe how the tax smoothing formula can be derived from a perturbation of the earnings and savings tax schedules. To do this, first observe that by a duality argument, our optimal policy problem is equivalent to maximizing a weighted average of utilities of the individuals subject to incentive compatibility

B.1 No Income Effect

We start our analysis with preferences that have no income effect; they are given by:

$$c_1 + \beta(\theta) P(\theta) u(c_1) - v(y/\theta).$$

Consider a tax schedule $T_y(y), T_a(a)$ and suppose that both taxes are paid in the first period⁵⁶ and suppose that it induces choices $\{y(\theta), a(\theta)\}$ without bunching, i.e., both $y(\theta)$ and $a(\theta)$ are one-to-one functions. This implies that both of these functions must be increasing.

Now consider a perturbation of the tax functions given by $T_y(y) + \delta T_y(y), T_a(a) + \delta T_a(a)$ where

$$\delta T_y(y) = \begin{cases} 0 & y \leq y(\theta) \\ d\tau_y(y - y(\theta)) & y \in [y(\theta), y(\theta) + dy] \\ d\tau_y dy & y \geq y(\theta) + dy \end{cases}$$

$$\delta T_a(a) = \begin{cases} 0 & a \leq a(\theta) \\ d\tau_a(a - a(\theta)) & a \in [a(\theta), a(\theta) + da] \\ d\tau_a da & a \geq a(\theta) + da \end{cases}$$

where $d\tau_y, dy, da, d\tau_a$ are sufficiently small real numbers and θ is any type. As is standard (see Saez (2001)), this perturbation has three effect: a mechanical effect, a behavioral effect and a welfare effect.

The mechanical effect of this perturbation is given by

$$d\tau_y dy \int_{\theta}^{\bar{\theta}} dF(\hat{\theta}) + d\tau_a da \int_{\theta}^{\bar{\theta}} dF(\hat{\theta})$$

⁵⁶Here, for simplicity, we assume that the tax functions are separable, that they are independent of each other, and that both taxes are paid in the first period. This significantly simplifies the analysis. Relaxing these assumptions significantly complicates the analysis without much benefit.

while the welfare effect is given by

$$-d\tau_y dy \int_{\theta}^{\bar{\theta}} \mathcal{W}(\hat{\theta}) dF(\hat{\theta}) - d\tau_a da \int_{\theta}^{\bar{\theta}} \mathcal{W}(\hat{\theta}) dF(\hat{\theta})$$

where $\mathcal{W}(\theta)$ is the social marginal value of giving one unit of income to an individual of type θ – this is evaluated according to the social welfare function associated with the Pareto optimal tax schedule. Intuitively, this perturbation simply decreases the after tax income of workers with type $\hat{\theta} \geq \max \left\{ \theta + \frac{dy}{y'(\theta)}, \theta + \frac{da}{a'(\theta)} \right\}$. Moreover, From an envelope condition, the change in utility for these individuals are simply $d\tau_y dy + d\tau_a da$. As da and dy become small and converge to zero, the welfare and mechanical effects converge to the above integrals. Note that for types in the interval $\hat{\theta} \in \left[\theta, \max \left\{ \theta + \frac{dy}{y'(\theta)}, \theta + \frac{da}{a'(\theta)} \right\} \right]$, only the marginal taxes change and therefore, the change in their utilities is of higher order of magnitude relative to $d\tau_y dy$ and $d\tau_a da$.

Note that we can always choose the perturbation so that $d\tau_a = -d\tau_y$ and $da = dy$. This implies that the welfare and mechanical effects are both zero. Therefore, this perturbation only has a behavioral effect on the savings and earnings of individuals in a small interval above θ . Note that since there is no income effect, the earnings tax perturbation, $\delta T_y(\cdot)$, only affects earnings while savings tax perturbation, $\delta T_a(\cdot)$, only affects saving behavior.

Since $dy > 0$ is small enough, we can say that the set of types that change their earnings is given by $\theta \in \left[\theta, \theta + \frac{dy}{y'(\theta)} \right]$ and their change in earnings must satisfy

$$\forall \hat{\theta} \in [\theta, \theta + d\theta_y], 1 - T'_y(y(\hat{\theta}) + \delta y) - d\tau_y = \frac{1}{\hat{\theta}} v' \left(\frac{y(\hat{\theta}) + \delta y}{\hat{\theta}} \right)$$

where $y'(\theta) d\theta_y = dy$. We can write the above approximately as

$$\begin{aligned} -T''_y(y(\hat{\theta})) \delta y - d\tau_y &= \frac{1}{\hat{\theta}^2} v''(y(\hat{\theta})/\hat{\theta}) \delta y \\ -T''_y(y(\hat{\theta})) \delta y - d\tau_y &= \frac{1}{\hat{\theta}^2} v'(y(\hat{\theta})/\hat{\theta}) \frac{1}{\varepsilon(\hat{\theta})} \frac{\hat{\theta}}{y(\hat{\theta})} \delta y \\ -T''_y(y(\hat{\theta})) \delta y - d\tau_y &= \left(1 - T'_y(y(\hat{\theta})) \right) \frac{1}{\varepsilon(\hat{\theta})} \frac{1}{y(\hat{\theta})} \delta y \\ \frac{\delta y}{y(\hat{\theta})} \frac{1 - T'_y}{d\tau_y} &= - \frac{1}{\frac{y T''_y}{1 - T'_y} + \frac{1}{\varepsilon(\hat{\theta})}} \end{aligned}$$

This implies that the gain in government budget from this behavioral response is approximately

given by

$$T'_y(y(\theta)) \delta y d\theta_y h(\theta) = -\frac{T'_y}{1-T'_y} y(\theta) \frac{1}{\frac{yT''_y}{1-T'_y} + \frac{1}{\varepsilon(\theta)}} \frac{1}{y'(\theta)} h(\theta) d\tau_y dy$$

In the above, $d\theta_y h(\theta)$ is the size of the bracket that is affected by this perturbation while $T'_y \delta y$ is the behavioral change in government revenue; simply marginal tax rate multiplied by the behavioral change in earning. Similarly, the set of types that change their savings is given by $\theta \in \left[\theta, \theta + \frac{da}{a'(\theta)}\right]$ and their change in savings must satisfy

$$\forall \hat{\theta} \in [\theta, \theta + d\theta_a], 1 + T'_a(a(\hat{\theta}) + \delta a) + d\tau_a = \beta R P u'(Ra(\hat{\theta}) + \delta a)$$

and $a'(\theta) d\theta_a = da$. We can write the above approximately as

$$\begin{aligned} T''_a(a(\hat{\theta})) \delta a + d\tau_a &= \beta R P u''(Ra(\hat{\theta})) R \delta a \\ T''_a(a(\hat{\theta})) \delta a + d\tau_a &= -\beta R P u'(Ra(\hat{\theta})) \frac{-u''(Ra(\hat{\theta})) Ra(\hat{\theta})}{u'(Ra(\hat{\theta}))} \frac{\delta a}{a(\hat{\theta})} \\ T''_a(a(\hat{\theta})) \delta a + d\tau_a &= -\left[1 + T'_a(a(\hat{\theta}))\right] \frac{-u''(Ra(\hat{\theta})) Ra(\hat{\theta})}{u'(Ra(\hat{\theta}))} \frac{\delta a}{a(\hat{\theta})} \\ \frac{\delta a}{a(\theta)} \frac{1 + T'_a}{d\tau_a} &= -\frac{1}{\frac{aT''_a}{1+T'_a} + \frac{1}{\sigma(\theta)}} \end{aligned}$$

where $1/\sigma(\theta)$ is the intertemporal elasticity of substitution. This implies that the gain in government budget from this behavioral response is approximately given by

$$[T'_a(a(\theta)) + (1 - P)] \delta a d\theta_a h(\theta) = -\frac{T'_a + (1 - P)}{1 + T'_a} a(\theta) \frac{1}{\frac{aT''_a}{1+T'_a} + \frac{1}{\sigma(\theta)}} \frac{1}{a'(\theta)} h(\theta) d\tau_a da$$

Note that in the above, we capture the fact that upon death the government confiscates the assets. Since $d\tau_a da = -d\tau_y dy$, at the optimum,

$$\left[1 - \frac{P}{1 + T'_a}\right] a(\theta) \frac{1}{\frac{aT''_a}{1+T'_a} + \frac{1}{\sigma(\theta)}} \frac{1}{a'(\theta)} h(\theta) = \frac{T'_y}{1 - T'_y} y(\theta) \frac{1}{\frac{yT''_y}{1-T'_y} + \frac{1}{\varepsilon(\theta)}} \frac{1}{y'(\theta)} h(\theta) \quad (28)$$

Note that if we take the FOCs associated with a and y and take a derivative with respect to θ , we

have

$$1 - T'_y = \frac{1}{\theta} v' \left(\frac{y(\theta)}{\theta} \right) \Rightarrow -T''_y y' = \frac{1 - T'_y}{\varepsilon} \frac{y'}{y} - \frac{1}{\theta} \left(1 + \frac{1}{\varepsilon} \right) (1 - T'_y)$$

$$\frac{y'}{y} \left[\frac{T''_y y}{1 - T'_y} + \frac{1}{\varepsilon} \right] = \frac{1}{\theta} \left(1 + \frac{1}{\varepsilon} \right)$$

Moreover,

$$1 + T'_a = \beta P R u' (Ra) \Rightarrow T''_a a' = -\frac{1 + T'_a}{\sigma} \frac{a'}{a} + \left(\frac{\beta'}{\beta} + \frac{P'}{P} \right) (1 + T'_a)$$

$$\frac{a'}{a} \left[\frac{a T''_a}{1 + T'_a} + \frac{1}{\sigma} \right] = \left(\frac{\beta'}{\beta} + \frac{P'}{P} \right)$$

Replacing the above in (28), we have

$$\left[1 - \frac{P}{1 + T'_a} \right] \frac{1}{\frac{\beta'}{\beta} + \frac{P'}{P}} = \frac{T'_y}{1 - T'_y} \frac{1}{\frac{1}{\theta} \left(1 + \frac{1}{\varepsilon} \right)}$$

Note that $1 + T'_a = \frac{1}{1 - \tau_a}$ and thus the above can be written as

$$[1 - P(1 - \tau_a)] \frac{1}{\frac{\beta'}{\beta} + \frac{P'}{P}} = \frac{\tau_l}{1 - \tau_l} \frac{1}{\frac{1}{\theta} \left(1 + \frac{1}{\varepsilon} \right)}$$

which is the same as (3).

B.2 Income Effect

With income effect, cross-elasticities matter as well; earnings and savings tax perturbations affect both earnings and savings.

Consider the same tax perturbation as above. Note that using the same argument, the welfare effect and mechanical effects cancel each other. We thus need to understand the behavioral effects.

Let $\delta y, \delta a$ be the response of type θ to a marginal tax perturbation $(d\tau_y, d\tau_a)$. Then we must have that

$$\begin{pmatrix} \delta y \\ \delta a \end{pmatrix} = \mathbf{\Gamma}^{-1} \begin{pmatrix} d\tau_y \\ d\tau_a \end{pmatrix}$$

where in the above $\mathbf{\Gamma}^{-1}$ is the compensated elasticity matrix and is given by

$$\begin{aligned}\mathbf{\Gamma} &= \begin{pmatrix} \gamma_{yy} & -\gamma_{ay} \\ -\gamma_{ay} & \gamma_{aa} \end{pmatrix} \\ \gamma_{yy} &= \frac{u_{cc1}}{u_{c1}} (1 - T'_y)^2 - T''_y - \frac{1}{\theta^2} v'' \\ \gamma_{ay} &= \frac{u_{cc1}}{u_{c1}} (1 - T'_y) (1 + T'_a) \\ \gamma_{aa} &= \frac{\beta P R^2 u_{cc2}}{u_{c1}} - T''_a + \frac{u_{cc1}}{u_{c1}} (1 + T'_a)\end{aligned}$$

This can simply be derived from the individuals optimality conditions given by

$$\begin{aligned}u_{c1} (1 + T'_a) &= \beta P R u_{c2} \\ u_{c1} (1 - T'_y) &= \frac{1}{\theta} v' \left(\frac{y}{\theta} \right)\end{aligned}$$

We can therefore write

$$\begin{pmatrix} \frac{\partial y}{\partial \tau_y} & \frac{\partial y}{\partial \tau_a} \\ \frac{\partial a}{\partial \tau_y} & \frac{\partial a}{\partial \tau_a} \end{pmatrix} = \mathbf{\Gamma}^{-1} = \frac{1}{\Delta} \begin{bmatrix} \gamma_{aa} & \gamma_{ay} \\ \gamma_{ay} & \gamma_{yy} \end{bmatrix}$$

At the optimum, the behavioral change in government revenue must be zero. This behavioral change is given by

$$\frac{h(\theta)}{y'(\theta)} \left(T'_y \frac{\partial y}{\partial \tau_y} d\tau_y + (1 - P + P T'_a) \frac{\partial a}{\partial \tau_y} d\tau_y \right) + \frac{h(\theta)}{a'(\theta)} \left(T'_y \frac{\partial y}{\partial \tau_a} d\tau_a + (1 - P + P T'_a) \frac{\partial a}{\partial \tau_a} d\tau_a \right) = 0$$

Since $d\tau_y = -d\tau_a$, we can write this as

$$(1 - P + P T'_a) \left[a' \frac{\partial a}{\partial \tau_y} - y' \frac{\partial a}{\partial \tau_a} \right] + T'_y \left[a' \frac{\partial y}{\partial \tau_y} - y' \frac{\partial y}{\partial \tau_a} \right] = 0$$

or

$$-(1 - P + P T'_a) [-a' \gamma_{ay} + y' \gamma_{yy}] + T'_y [a' \gamma_{aa} - y' \gamma_{ay}] = 0$$

and we can write the above in matrix form

$$\begin{pmatrix} -(1 - P + P T'_a) & T'_y \end{pmatrix} \mathbf{\Gamma} \begin{pmatrix} y' \\ a' \end{pmatrix} = 0$$

When, we take derivative of the first order conditions with respect to θ , we have

$$\Gamma \begin{pmatrix} y' \\ a' \end{pmatrix} = - \begin{pmatrix} \frac{1}{\theta} \left(1 + \frac{1}{\varepsilon}\right) (1 - T_y') \\ \left(\frac{\beta'}{\beta} + \frac{P'}{P}\right) (1 + T_a') \end{pmatrix}$$

and therefore

$$\begin{pmatrix} -(1 - P + PT_a') & T_y' \end{pmatrix} \begin{pmatrix} \frac{1}{\theta} \left(1 + \frac{1}{\varepsilon}\right) (1 - T_y') \\ \left(\frac{\beta'}{\beta} + \frac{P'}{P}\right) (1 + T_a') \end{pmatrix} = 0$$

which becomes

$$-(1 - P + PT_a') \frac{1}{\theta} \left(1 + \frac{1}{\varepsilon}\right) (1 - T_y') + T_y' \left(\frac{\beta'}{\beta} + \frac{P'}{P}\right) (1 + T_a') = 0$$

which is the same as the desired equation.