Retirement Financing: An Optimal Reform Approach*

Roozbeh Hosseini
University of Georgia
Federal Reserve Bank of Atlanta
roozbeh@uga.edu

Ali Shourideh
Carnegie Mellon University
ashourid@andrew.cmu.edu

August 13, 2018

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 Laurence Ales, Tony Braun, V. V. Chari, Mariacristina De Nardi, Berthold Herrendorf, Karen Kopeky, Dirk Krueger, Ellen McGrattan, Chris Sleet, Chris Telmer, Venky Venkateswaran, Shu Lin Wee, Sevin Yeltekin, Ariel Zetlin-Jones, and various seminars and conference participants for their comments and suggestions. We specially thank Mariacristina De Nardi, Eric French, and John Jones for sharing their data and codes. We are grateful to Gianluca Violante (editor) and three anonymous referees for their feedback. 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.
Please visit http://www.roozbeh-hosseini.com/ for updates.

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-

[1]Social 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

[2]The 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.

[3]See 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

[4] In 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.

[5] See 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. As a result, it is hard to separate the redistribution aspects of the optimal policy from efficiency gains. The benefit of our approach is that it does not rely on a welfare function and it takes the distribution of welfare in the economy as a given. To the best of our knowledge, this is the first paper that proposes this approach to optimal policy reform in a dynamic quantitative setting.[6]

[6] See Werning (2007) for a theoretical analysis in a static framework.

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), İmrohoroğlu 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 nonlinear 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 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 $t = 0, 1, \cdots$. 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 $\theta \in \Theta = \left[\underline{\theta}, \overline{\theta}\right]$ 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.[7] 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 $\theta's$ 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* $\left\{\{\hat{c}_{1,t}(\theta), \hat{c}_{2,t}(\theta), \hat{y}_t(\theta), \hat{a}_t(\theta)\}_{\theta\in\Theta}, K_t, L_t\right\}$ *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* $t = 0, 1, \cdots$

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

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

subject to

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

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

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.[8] 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.[9] 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*

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

[9]See Mirrlees (1971) and Werning (2007).

optimal only if it satisfies:

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

where $\tilde{\tau}_{l}\left(\theta\right)$ and $\tilde{\tau}_{a}\left(\theta\right)$ are the wedges induced by the tax schedule; and

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

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), $P\left(\theta\right)\tilde{\tau}_{a,t}\left(\theta\right)+1-P\left(\theta\right)$ is the increase in government's revenue per person from a unit increase in assets of workers of type θ, [10] while $\tilde{\tau}_{l,t}$ is the same thing except for earnings. As we describe below, equation (3) states that the behavioral increase [11] 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 $c_1+\beta\left(\theta\right)P\left(\theta\right)u\left(c_2\right)-v\left(y/\theta\right)$. In this formulation, there is no income effect and, therefore, the calculation of individual responses to tax perturbations is simpler. In the Appendix C.2, we show how this analysis works in a model with income effect. Starting with the tax function $T_y\left(y\right)$ and $T_a\left(a\right)$,[12] consider the following perturbation of any tax schedule:

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

[10] The 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.

[11] By 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.

[12] For 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.[13] Some algebra, deferred to the Appendix C, 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

[13] A 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:

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

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. 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.[14] 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 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

[14] The 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.

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:*

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

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.[15]

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 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.

[15] As 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.

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 $t = 0, 1, 2, \ldots$. The number of newborns grows at rate n_t. Upon birth, each individual draws a type $\theta \in \Theta = \left[\underline{\theta}, \overline{\theta}\right]$ 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).[16] 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

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

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

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

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 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 $F_K(K_t, L_t) - \delta$.

[16] Arguably, 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.

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.[17] 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.[18]

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:[19]

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

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 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 $(1-\tau_K)\left(F_K(K_t, L_t) - \delta\right)$.[20] We assume that the government taxes households' holding of

[17] An 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.

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

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

[20] We 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

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

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

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

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

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

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

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, demographics 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 $W_t^{sq}(\theta)$.

rate earned on all types of assets.

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.[21] 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)).[22]

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 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

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

[22] As 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.

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 generation [23] 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

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

subject to

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

$$\begin{aligned} U'(\theta) = {} & \sum_{j=0}^{J} \beta(\theta)^j P_j(\theta) \frac{\varphi_j'(\theta) l_j(\theta)}{\varphi_j(\theta)} v'(l_j(\theta)) \\ & + \sum_{j=0}^{J} \left(\frac{j\beta'(\theta)}{\beta(\theta)} + \frac{P_j'(\theta)}{P_j(\theta)} \right) \beta^j P_j(\theta) \left[u(c_j(\theta)) - v(l_j(\theta))\right] \end{aligned} \tag{13}$$

$$U(\theta) \geq W^{sq}(\theta). \tag{14}$$

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.[24]

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

[24] Our 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.

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

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

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*

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

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

as well as the following inequality

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

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.[25]

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 intution beind it is

[25] The necessary conditions above also hold for general disutility of labor; see proofs for the general results.

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 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

$$\frac{\tau_{l,j+1}(\theta)}{1-\tau_{l,j+1}(\theta)} \frac{1}{\frac{\varphi_{j+1}'(\theta)}{\varphi_{j+1}(\theta)}} = \frac{\tau_{l,j}(\theta)}{1-\tau_{l,j}(\theta)} \frac{1}{\frac{\varphi_j'(\theta)}{\varphi_j(\theta)}} \frac{1}{p_{j+1}(\theta)\left(1-\tau_{a,j+1}(\theta)\right)} \tag{19}$$

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 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.[26] 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.

[26] In 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.

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 $\xi_3 = 7.26 \times 10^{-6}$.

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)).[27] 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

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

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

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

[a]source: Table A-1 in Waldron (2013) (adjusted by average mortality rate of 1940 birth cohort).

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.[28] 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 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.[29]

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.

[28]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.

[29]These estimates can be found here: http://www.census.gov/population/projections/data/national/2014.html

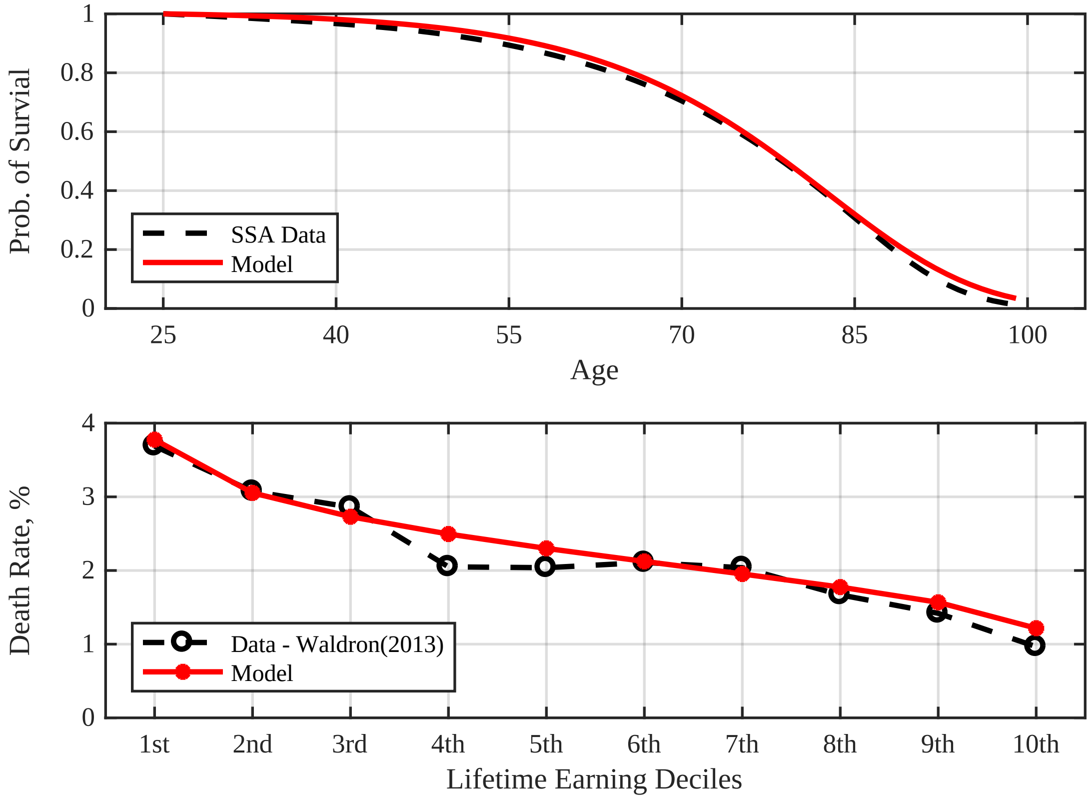

Figure 1: Fit of the mortality model. The top panel shows the average survival probability in the model vs. social security data. The bottom panel shows death rates at age 67 in the model vs. those reported in Waldron (2013).

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.[30] 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

[30]Our 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. See Table 23 in Appendix G for details.

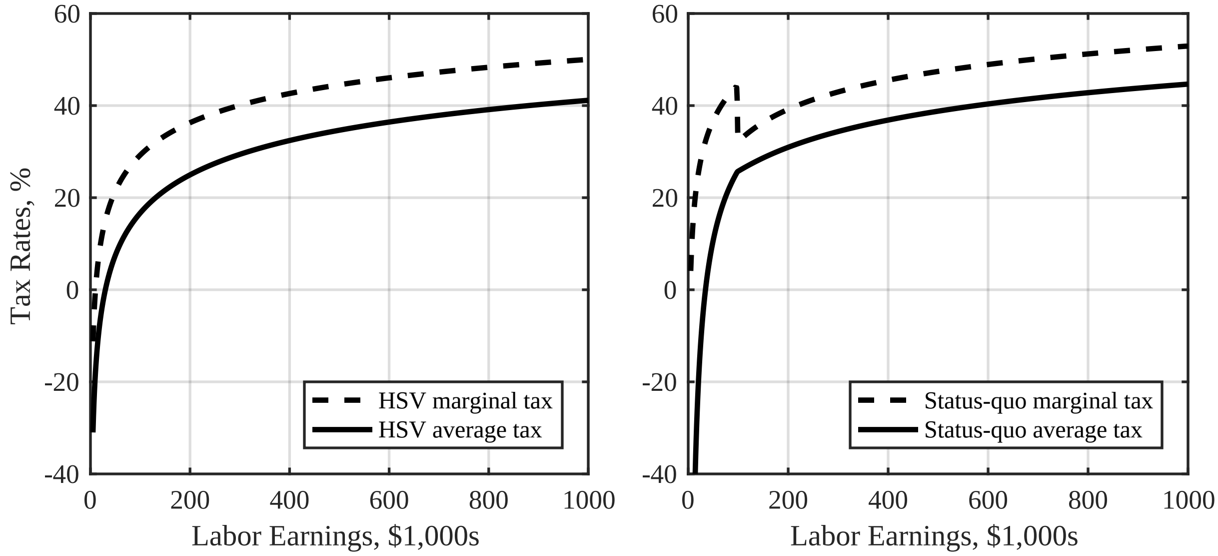

Figure 2: Tax functions. The left panel is the calibrated HSV tax function, $\mathcal{T}_{HSV}(\cdot)$. The right panel is the effective tax function (including HSV tax, payroll tax and transfers). The discontinuity is due to the social security cap on taxable earnings.

income and investment relative to GDP in the U.S. over the period 2000−2010.[31]

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:

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

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

[31] Our calibration procedure follows McGrattan and Prescott (2017). See Appendix G for details.

[32] The 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.

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

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 calculation of the capital share, government expenditure and government debt , see Appendix G.

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

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

[34]Our 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.

[35]This 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. See Appendix G for details.

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

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 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. [36]

Finally, we assume the government debt is 47 percent of GDP.[37] The transfers Tr are chosen 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

[36] In the Appendix, we report the results using an alternative tax function based on Congressional Budget Office's report on effective marginal tax rates (see Harris (2005)).

[37] This 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. See Appendix G for details.

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.[38] 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.[39] 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 accounts[40] (which are tax deductible and are treated as income during retirement), whose tax treatment is possibly progressive. On the other hand, richer individuals 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 Appendix D.6: an exercise with a flat and positive savings tax and one with a progressive savings tax.[41]

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

[38] This 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.

[39] A 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.

[40] See McGrattan and Prescott (2017)

[41] Various papers in the literature assume a flat savings tax schedule and use the estimate by McDaniel (2007). As we show in the appendix, our results on optimal policies are unchanged by this assumption, while the cost saving measures are magnified.

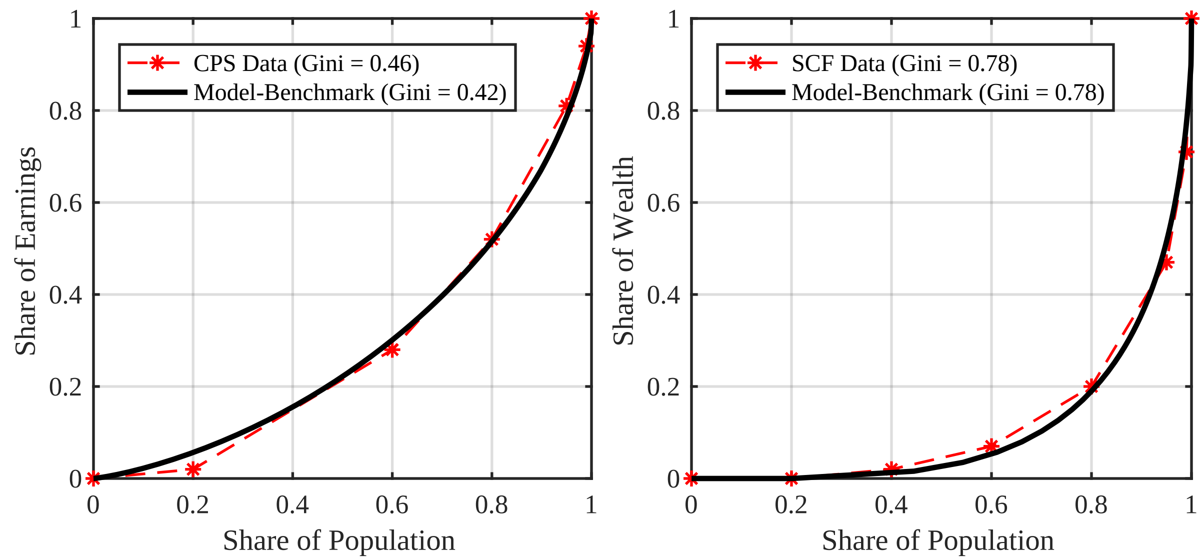

Figure 3: Fit of the distribution of earnings (left panel) and wealth (right panel).

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 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 Appendix E, 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 Appendix B 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

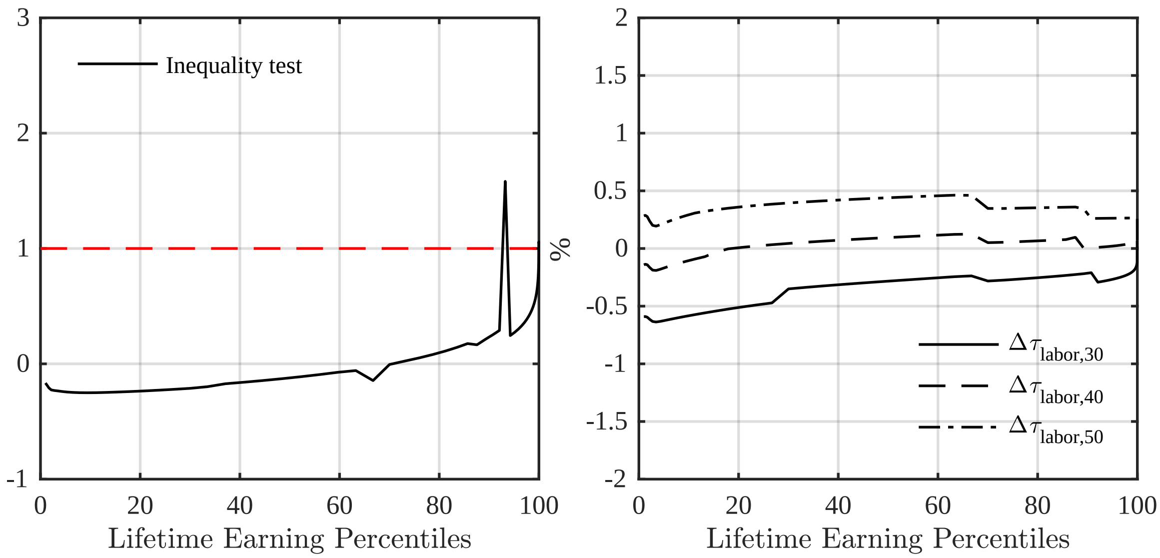

Figure 4: Test of Pareto optimality for status quo policies. The left panel plots the two sides of the inequality (18). The right panel depicts the change in earnings wedges required for (19) to hold.

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 economy. Figure 4 plots the performance of the tests for labor wedges.[42] 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

[42]Note 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).

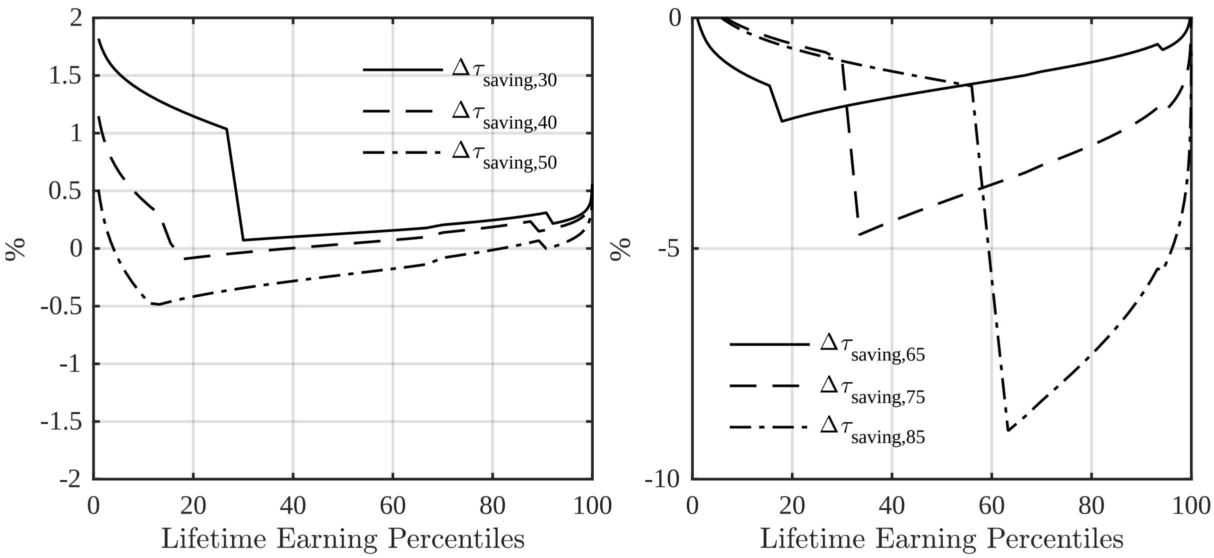

Figure 5: Test of Pareto optimality for status quo policies. The left panel depicts the required change in the savings wedge so that (16) holds at ages 30, 40 and 50. The right panel depicts the required change in the savings wedge so that (17) holds at ages 65, 75 and 85.

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.[43] For individuals with mid-values of lifetime earnings, the required change is 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

[43]Under 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.

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 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, $T_{a,j}\left(\left(1+r\right)a_j\right)$; (2) non-linear, age-dependent taxes on labor income, $T_{y,j}\left(y_j\right)$; (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.[44] 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.[45] 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 $j =$

[44] This feature resembles Ricardian equivalence.

[45] 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.

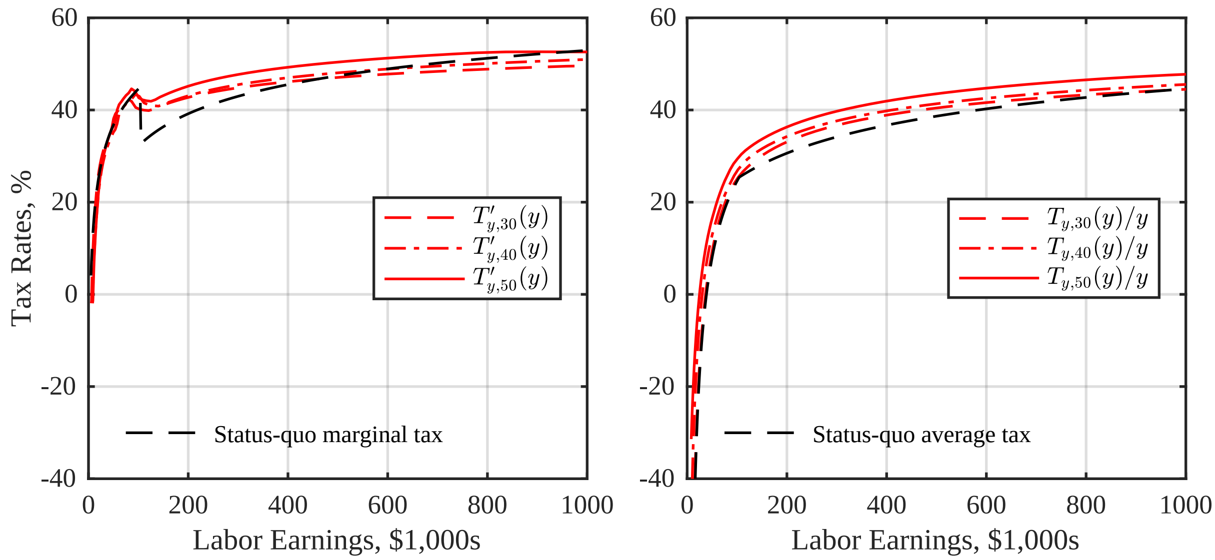

Figure 6: Optimal labor income tax functions. The left panel shows marginal taxes, and the right panel shows average taxes. The black dashed line is the effective status quo tax schedule.

$25, 35, 45$ (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 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.[46] 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).

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

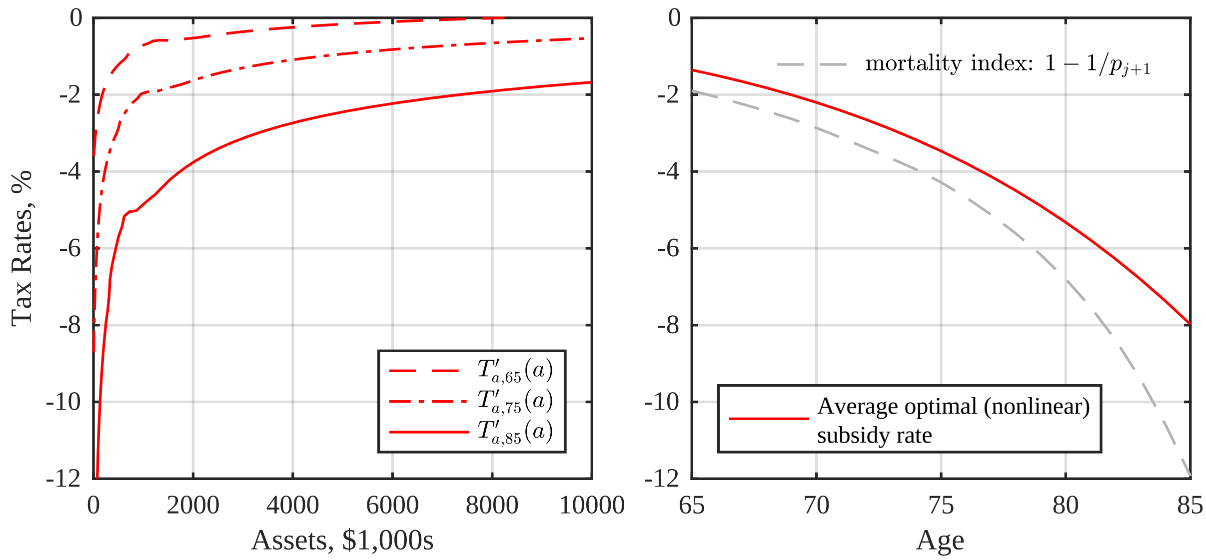

Figure 7: Optimal asset tax functions. The left panel shows the marginal taxes over all asset levels at ages 65, 75 and 85, while the right panel shows the average marginal rates at each age from 65 to 85. The dashed line is the population mortality index.

As before, the implied magnitudes of asset subsidies and their progressivity confirms the 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)).[47]

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

[47] To 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).

Table 4: Sources of retirement income

	Share of public transfers in retirement income (%)			
			Optimal reform	
Income quartiles	Data[a]	Status quo	(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

[a]Source: Table 6 in Poterba (2014).

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, 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

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

	Current U.S.	Optimal Reform	
	(1)	(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).

large difference in cost savings can be accounted for entirely by the fall in the interest rate.[48]

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.

[48]As 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.

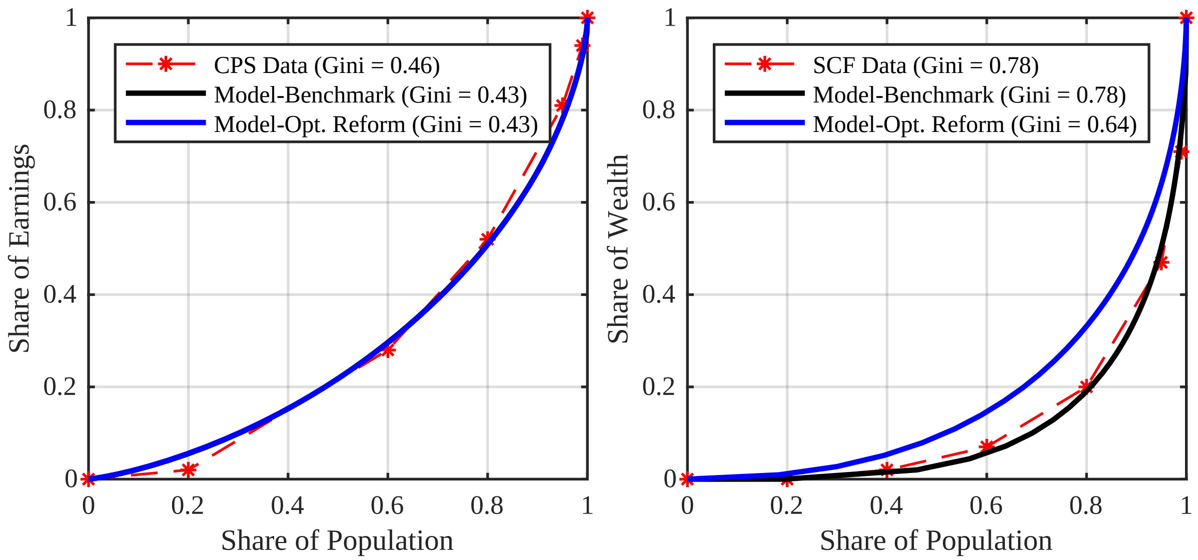

Figure 8: Distribution of earnings and wealth: status quo vs. optimal reform. The black line shows the results in the calibrated economy with current U.S. status quo policies. The blue line shows the results under Pareto optimal policies (for current U.S. demographic parameters and holding factor prices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

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. 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.

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.[49] 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 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.[50] 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

[49] We assume that the ratios of mortality among lifetime earnings deciles do not change.

[50] Our assumption regarding the continuation of status quo policies is similar to McGrattan and Prescott (2017) and many others who study policy reforms.

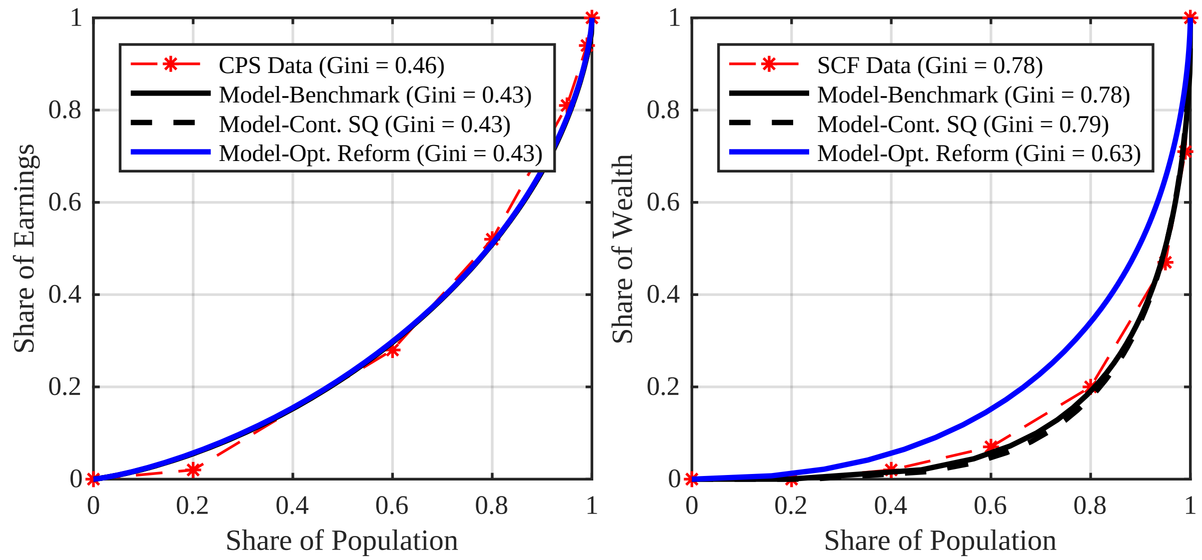

Figure 9: Distribution of earnings and wealth: status quo vs. optimal reform. The black solid line shows the results in the calibrated economy with current U.S. status quo policies. The black dashed line shows the steady state results with projected demographic parameters and continuation of status quo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

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.[51]

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 exisiting generations 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, 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. [52] 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 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

[51]Note 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.

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

Table 6: Aggregate effects of demographic transition and policy change

	Current U.S.	Continue	Optimal reform		
	(1)	(2)	(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).

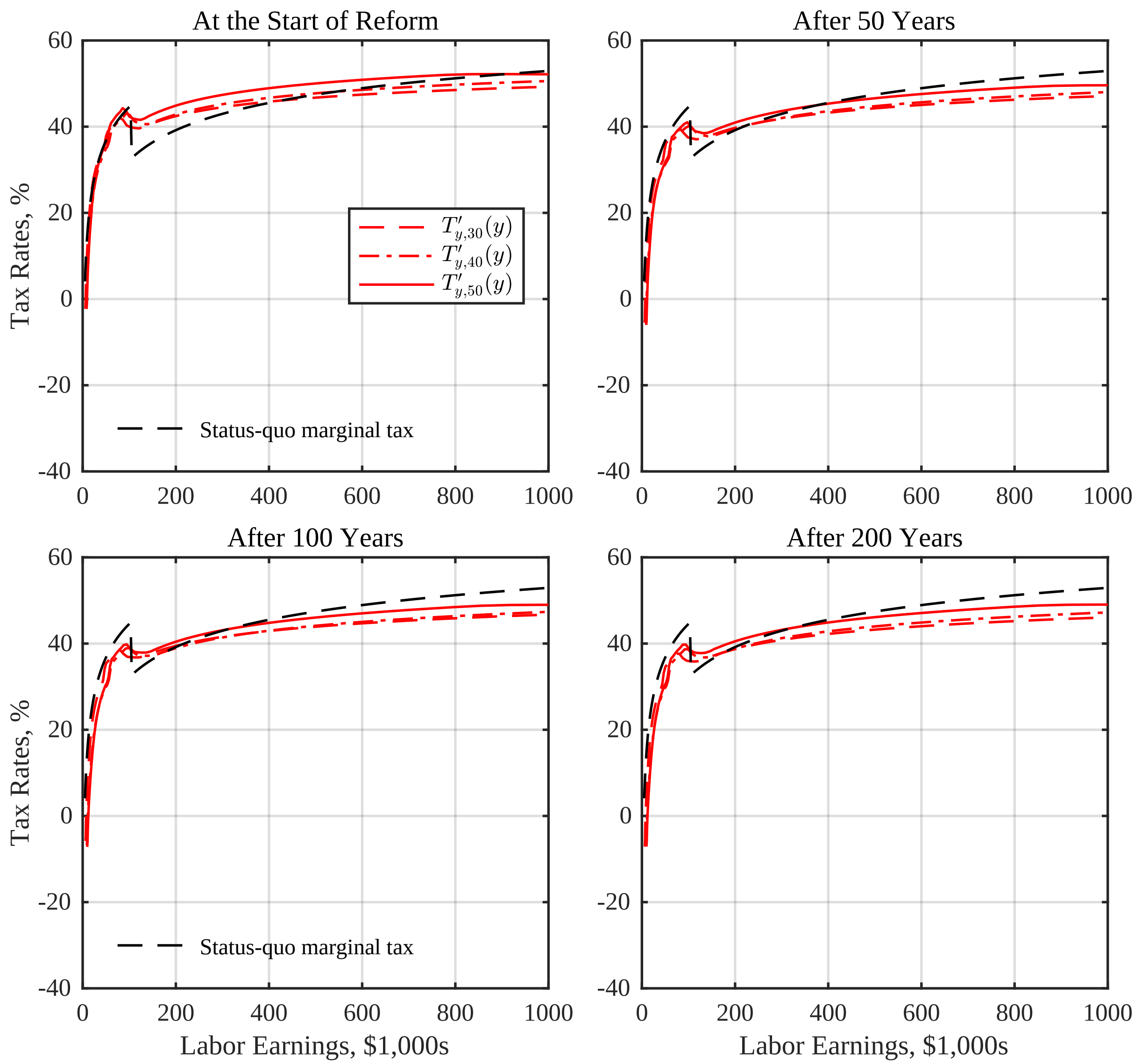

Figure 10: Evolution of optimal marginal labor income tax functions over the transition

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 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.

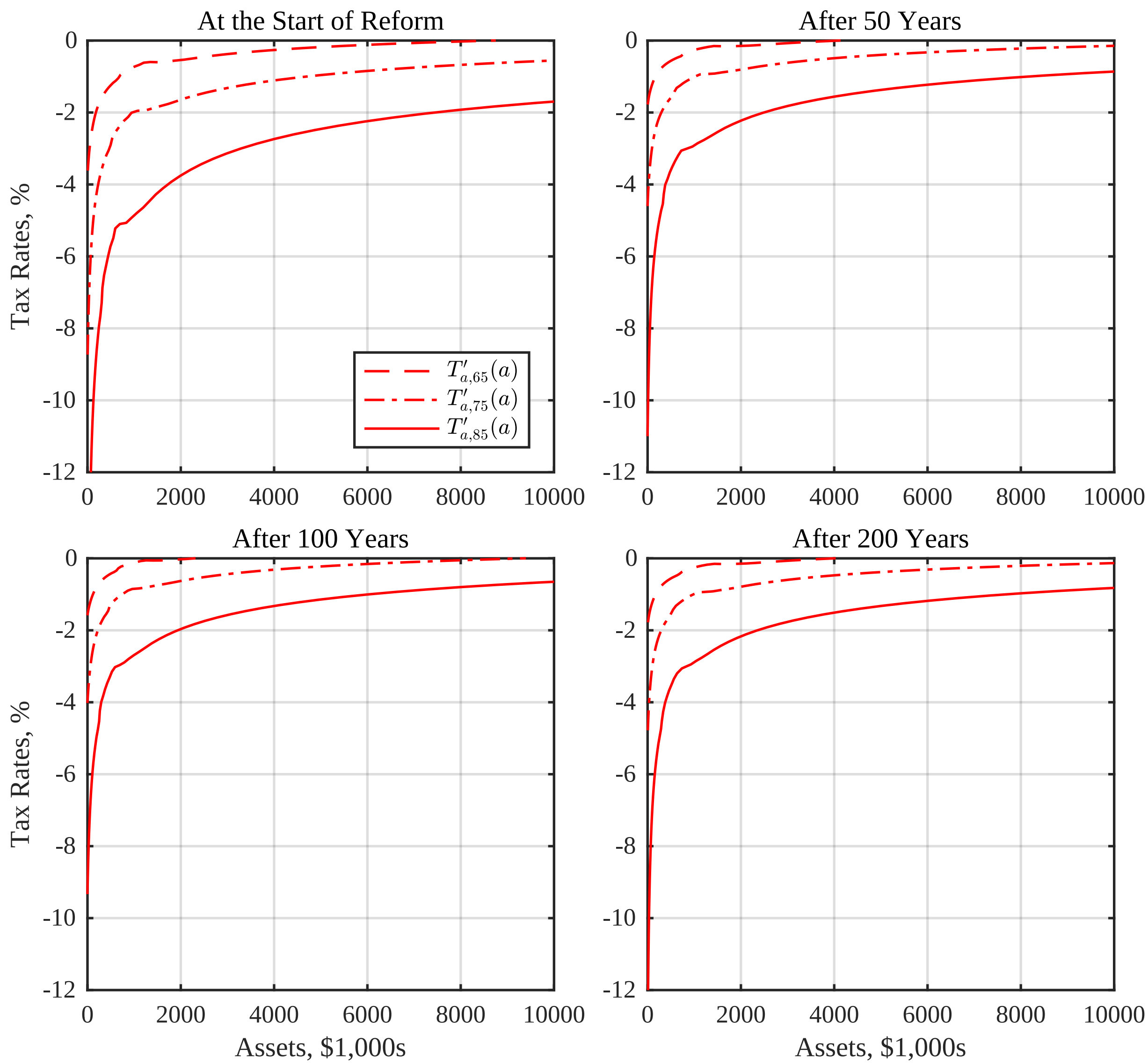

Figure 11: Evolution of optimal marginal asset tax functions over the transition

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

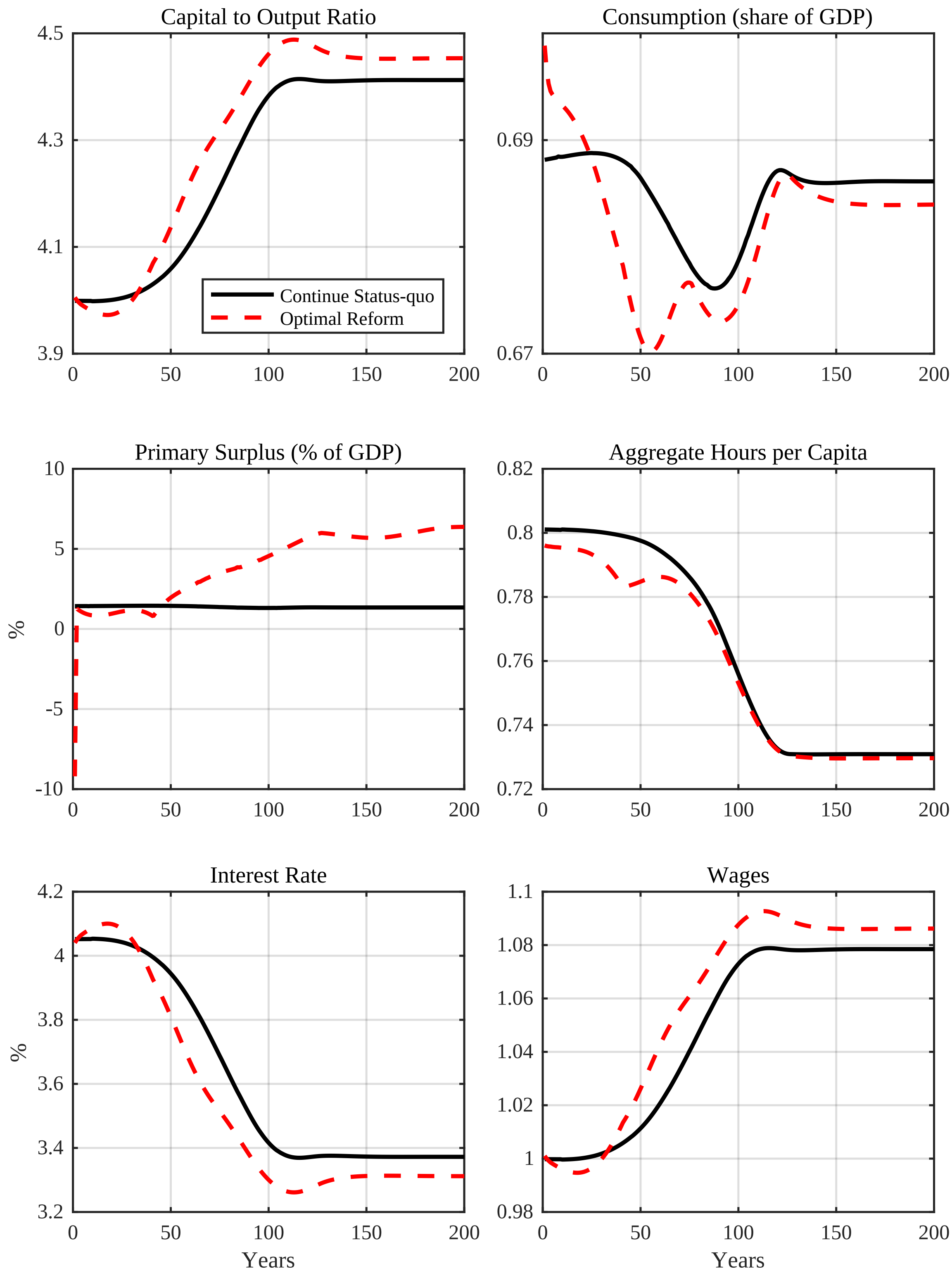

Figure 12: Evolution of aggregates along the transition

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.[53] 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 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.

Table 8 in Appendix D.1.1 shows the changes in the aggregate variables. Note that 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 allocation that it raises the costs of delivering utilities in the steady state. This highlights the importance of subsidies in any reform.[54]

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 Appendix D.1.2, 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.

[53] See Nishiyama and Smetters (2007) and McGrattan and Prescott (2017) for example.

[54] In Appendix D.2, 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.

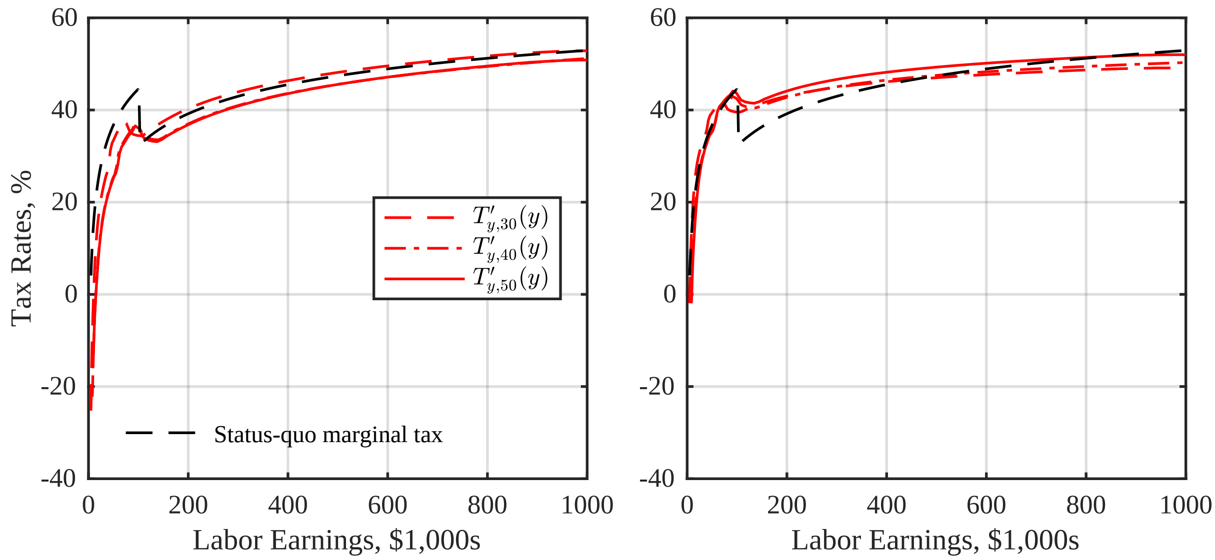

Figure 13: Optimal labor income tax functions with *privatization* (no old-age transfers and no asset subsidies). The left panel is optimal marginal taxes under *privatization*, while the right panel shows the same for the benchmark calibration. The black dashed line is the effective status quo tax schedule.

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.[55] This tax approximation includes many intricate features of the tax code including EITC phase in and phase out, AMT, CTC and itemized deductions (see Figure 26 in the appendix).

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

[55] Harris (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.

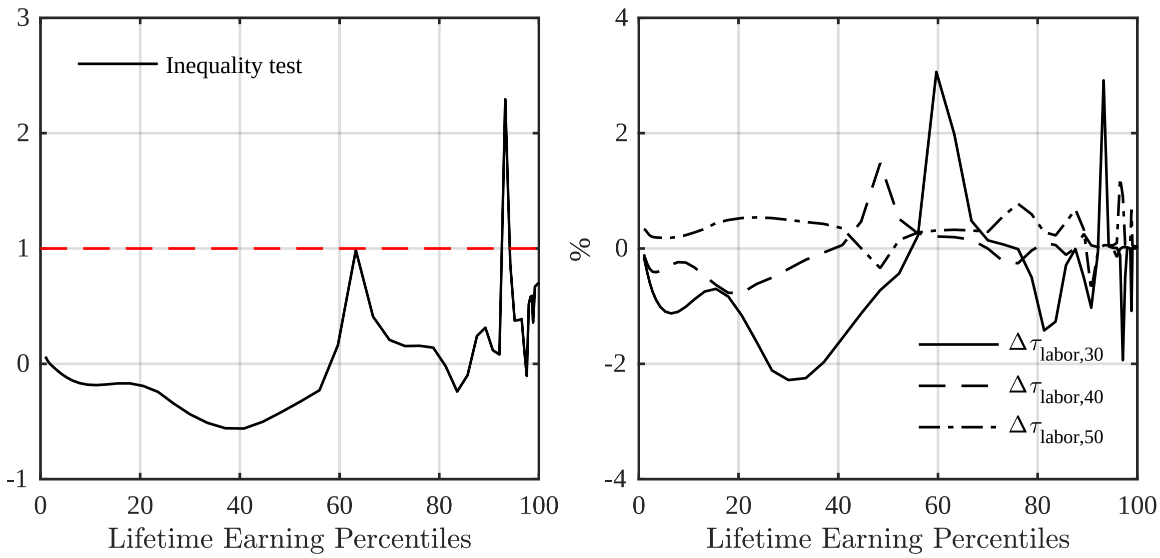

Figure 14: Test of Pareto optimality for status quo policies with a CBO effective tax rates. The left panel plots the two sides of the inequality (18). The right panel depicts the change in earnings wedges required for equation (19) to hold.

(which implies an effective marginal rate of 31 percent) to the 15 percent bracket. Furthermore, 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.[56]

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.

[56]Further quantitative results related to this exercise are provided in the Appendix D.3.

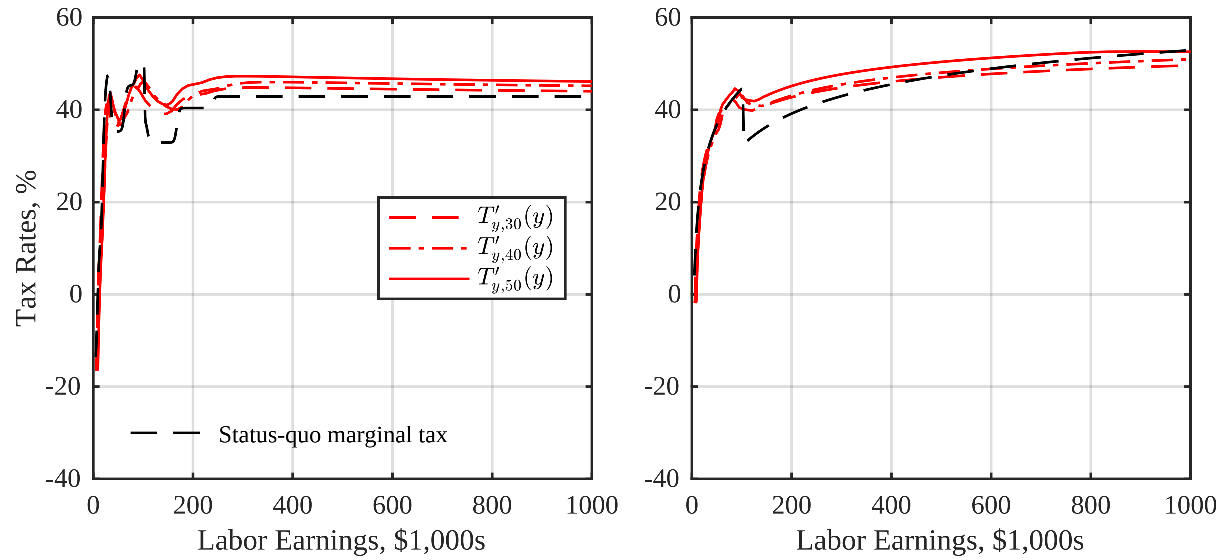

Figure 15: Optimal labor income tax functions for the alternative status quo tax policy. The left panel shows optimal marginal taxes and compares them with CBO effective tax rates. The right panel shows the same for the benchmark economy (with HSV tax function).

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.[57] 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

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

In other words, individuals have preferences over non-medical expenditure and hours worked. [58] The rest of the model is identical to what we described in Section 3. The formulation of

[57] 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.

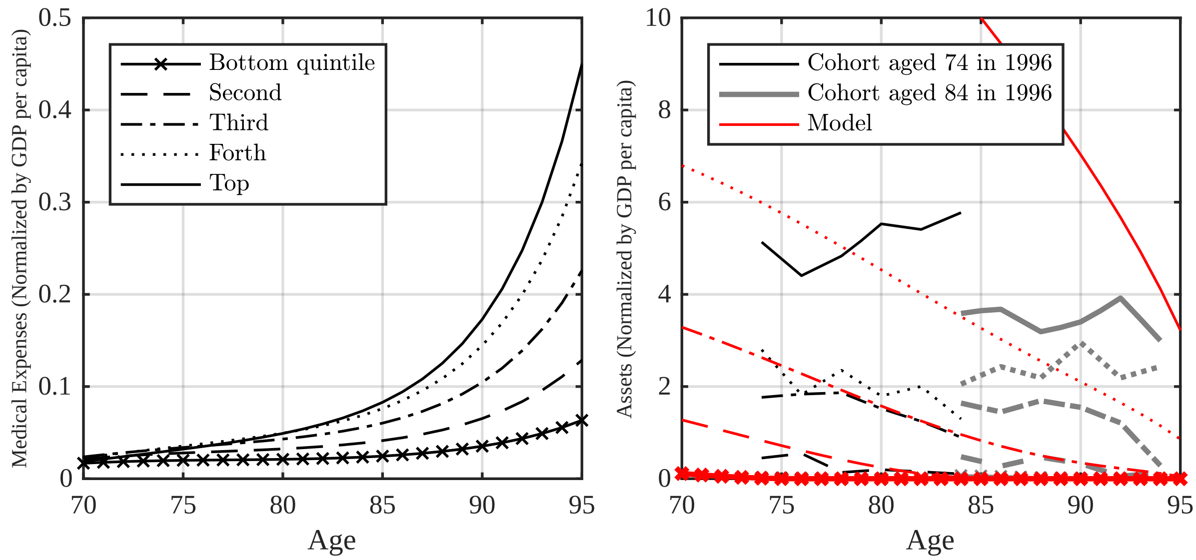

Figure 16: The left panel shows the average out-of-pocket medical expenditures by permanent income. The right panel shows the median assets by permanent income quintile in the model (solid line) and data (dashed line). Data source: De Nardi et al. (2010) calculations for AHEAD cohorts who were 74 and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

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

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

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 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

[58] Here, 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.

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 quantile, 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 deccumulation, 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 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

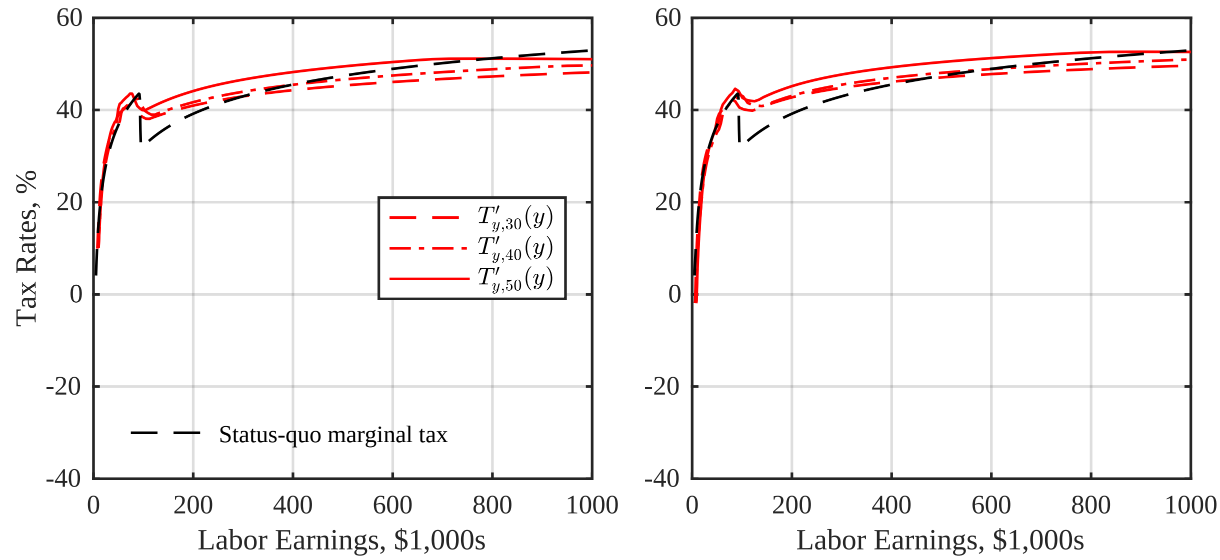

Figure 17: Optimal labor income tax functions with out-of-pocket medical expenditures. The left panel is optimal marginal taxes with out-of-pocket medical expenditures. The right panel shows the same for the benchmark model. The black dashed line is the effective status quo tax schedule.

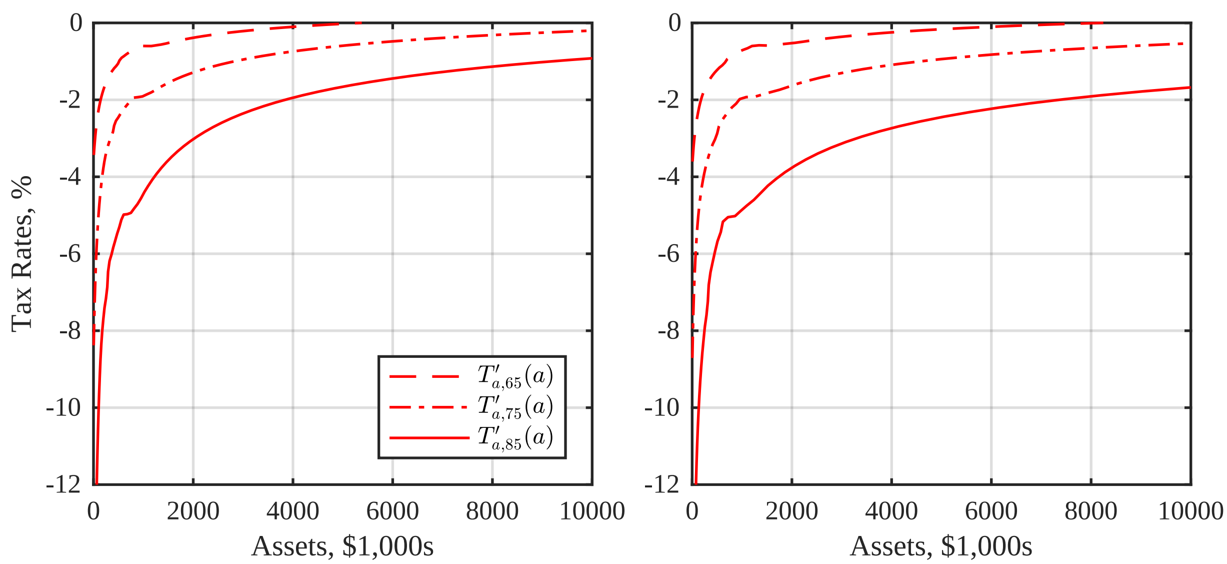

Figure 18: Optimal asset tax functions with out-of-pocket medical expenditures. The left panel shows optimal marginal taxes over all asset levels at ages 65, 75 and 85 for an economy with out-of-pocket medical expenditure. The right panel shows the same for the benchmark model.

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.

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

	Current U.S.	Continue	Optimal reform		
	(1)	(2)	(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).

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

$$\sum_{j=0}^{J} \beta(\theta)^{j} P_{j}(\theta)\left[u\left(c_{j}(\theta)-m_{j}(\theta)\right)-v\left(l_{j}(\theta)\right)+\beta\left(1-p_{j+1}(\theta)\right) w\left(b_{j+1}(\theta)\right)\right], \tag{21}$$

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.:

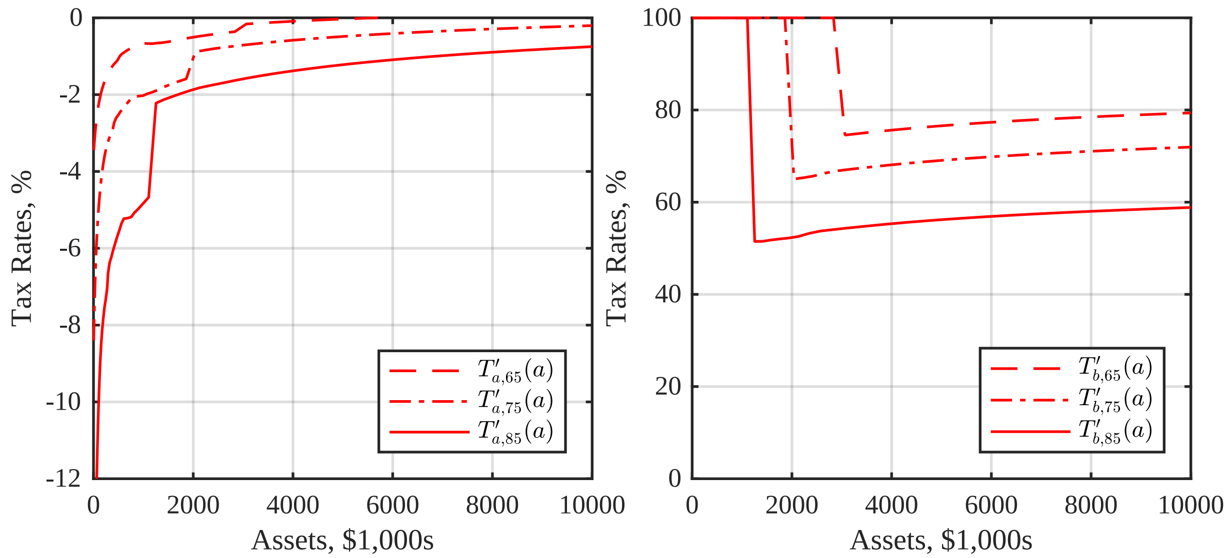

Figure 19: Optimal asset tax functions with out-of-pocket medical expenditures and bequest motives. The left panel shows the optimal marginal asset taxes over all asset levels for surviving individuals at ages 65, 75 and 85. The left panel shows the marginal bequest taxes for the same ages.

the fraction of the deceased individuals that leave no bequests and the ratio of bequests to wealth in the aggregate.

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 Appendix F.

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 Golosov et al. (2003) and Golosov 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.

Auten, G. and D. Joulfaian (1996): “Charitable contributions and intergenerational transfers,” *Journal of Public Economics*, 59, 55–68.

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.*

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.

Chetty, R. (2012): “Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply,” *Econometrica*, 80, 969–1018.

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): “Are micro and macro labor supply elasticities consistent? A review of evidence on the intensive and extensive margins,” *The American Economic Review*, 101, 471–475.

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. (2004): "Wealth inequality and intergenerational links," *The Review of Economic Studies*, 71, 743–768.

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," *Amrican 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.

Fernández-Villaverde, J. and D. Krueger (2007): "Consumption over the life cycle: Facts from consumer expenditure survey data," *The Review of Economics and Statistics*, 89, 552–565.

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*.

Gale, W. G. and J. K. Scholz (1994): "Intergenerational transfers and the accumulation of wealth," *The Journal of Economic Perspectives*, 145–160.

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.

Gourinchas, P.-O. and J. A. Parker (2002): "Consumption Over the Life Cycle," *Econometrica*, 70, 47–89.

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.

Hurd, M. and J. P. Smith (2002): "Expected bequests and their distribution," Tech. rep., National Bureau of Economic Research.

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.

Keane, M. P. (2011): "Labor supply and taxes: A survey," *Journal of Economic Literature*, 49, 961–1075.

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.

McDaniel, C. (2007): "Average tax rates on consumption, investment, labor and capital in the OECD 1950-2003." Arizona State Univeristy.

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.

POTERBA, J. M. AND A. A. SAMWICK (2003): "Taxation and Household Portfolio Composition: US Evidence from the 1980s and 1990s," *Journal of Public Economics*, 87, 5–38.

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.*

WORLD BANK, I. C. P. (2014): *Purchasing Power Parities and Real Expenditures of World Economies: Summary of Results and Findings of the 2011 International Comparison Program*, Washington, DC: World Bank Publications.

Appendices

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*

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

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

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

where $r_t = F_K\left(K_t, N_t \int y_t dH(\theta)\right)$. 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

$$\theta \in \arg\max_{\hat{\theta}\in\Theta} U\left(c_{1,t}\left(\hat{\theta}\right), c_{2,t}\left(\hat{\theta}\right), \frac{y_t\left(\hat{\theta}\right)}{\theta}\right)$$

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

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

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

subject to

$$\theta \in \arg\max_{\hat{\theta}\in\Theta} U\left(c_{1,t}\left(\hat{\theta}\right), c_{2,t}\left(\hat{\theta}\right), \frac{y_t\left(\hat{\theta}\right)}{\theta}\right)$$

$$U\left(c_{1,t}(\theta), c_{2,t}(\theta), \frac{y_t(\theta)}{\theta}\right) \geq W_t(\theta)$$

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

$$\theta \in \arg\max_{\hat{\theta} \in \Theta} U \left(c_{1,t}\left(\hat{\theta}\right), c_{2,t}\left(\hat{\theta}\right), \frac{y_t\left(\hat{\theta}\right)}{\theta} \right)$$

$$U \left(c_{1,t}(\theta), c_{2,t}(\theta), \frac{y_t(\theta)}{\theta} \right) \geq W_t(\theta)$$

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

$$\theta \in \arg\max_{\hat{\theta} \in \Theta} U \left(c_{1,t}\left(\hat{\theta}\right), c_{2,t}\left(\hat{\theta}\right), \frac{y_t\left(\hat{\theta}\right)}{\theta} \right)$$

$$U \left(c_{1,t}(\theta), c_{2,t}(\theta), \frac{y_t(\theta)}{\theta} \right) \geq W_t(\theta)$$

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, $N_t \lambda_t \to 0$ 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. □

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

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

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

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

Note that the FOC's also imply that

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

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. □

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

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

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 $u_2(\theta) = u(c_2(\theta))$ 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. □

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

$$U'(\theta) = \sum_{j=0}^{J} \left(\frac{j\beta'(\theta)}{\beta(\theta)} + \frac{P_j'(\theta)}{P_j(\theta)}\right) \beta(\theta)^j P_j(\theta) \left[u(c_j) - v\left(\frac{y_j}{\varphi_j(\theta)}\right)\right]$$
$$+ \sum_{j=0}^{J} \beta(\theta)^j P_j(\theta) \left[\frac{\varphi_j'(\theta) y_j}{\varphi_j(\theta)^2} v'\left(\frac{y_j}{\varphi_j(\theta)}\right)\right].$$

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

$$U'(\theta) = \sum_{j=0}^{J} \beta(\theta)^j P_j(\theta) \left[\frac{\varphi_j'(\theta) l_j(\theta)}{\varphi_j(\theta)} v'(l_j(\theta))\right]$$
$$+ \sum_{j=0}^{J} \left(\frac{j\beta'(\theta)}{\beta(\theta)} + \frac{P_j'(\theta)}{P_j(\theta)}\right) \beta(\theta)^j P_j(\theta) \left[u(c_j(\theta)) - v(l_j(\theta))\right].$$

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

$$
\begin{aligned}
1-\tau_{l,j} &= \frac{v_{l,j}}{\varphi_j u_{c,j}} \\
p_{j+1}\left(1-\tau_{a,j+1}\right) &= \frac{u_{c,j}}{\beta(1+r)u_{c,j+1}}
\end{aligned}
$$

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

$$
\begin{aligned}
\frac{1}{u_{c,0}} &= \eta \\
\frac{\varphi}{v'_{l,0}} &= \eta-\mu\frac{\varphi_{0,\theta}}{\varphi_0}\left(1+\frac{1}{\varepsilon}\right) \\
\frac{\tau_{l,0}}{1-\tau_{l,0}}\frac{1}{u_{c,0}} &= -\mu\frac{\varphi_{0,\theta}}{\varphi_0}\left(1+\frac{1}{\varepsilon}\right) \\
\mu &= -\frac{\tau_{l,0}}{1-\tau_{l,0}}\frac{1}{u_{c,0}}\frac{\varphi_0}{\varphi_{0,\theta}}\frac{\varepsilon}{1+\varepsilon}.
\end{aligned}
$$

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 $j=0$

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

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

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

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

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

Therefore

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

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. □

B Construction of Tax Schedules

In this section, we describe how to back out the optimal taxes from the optimal allocations and wedges discussed above.

The following lemma and its proof illustrate the construction of a tax and transfer schedule as in (7) such that individual optimizations' first order conditions are satisfied: (in what follows we adopt the following notation to avoid clutter; $u_{c,j}(\theta) \equiv u'(c_j(\theta))$ and $v_{l,j}(\theta) \equiv v'(l_j(\theta))$.)

Lemma 6. *Consider an allocation $\{c_j(\theta), l_j(\theta)\}$ that satisfies implementability constraint (13). Suppose that $(\varphi_j(\theta) l_j(\theta))' > 0$ and*

$$\sum_{s=j}^{J} \beta(\theta)^s P_s(\theta) \left[u_{c,s}(\theta) c_s'(\theta) - v_{l,s}(\theta) (\varphi_s(\theta) l_s(\theta))'\right] > 0.$$

Then tax and transfer functions $T_{a,j}(\cdot), T_{y,j}(\cdot), S_j$ together with asset holdings $a_j(\theta)$ exists so that the allocations $\{c_j(\theta), l_j(\theta), a_j(\theta)\}$ satisfy the budget constraints (7) and the first order conditions associated with the individual optimization.

Proof. We start by writing the first order conditions for the the maximization problem above for an individual of type θ

$$1 - T_{y,j}'(\varphi_j(\theta) l_j(\theta)) = \frac{v'(l_j(\theta))}{\varphi_j(\theta) u_{c,j}(\theta)}, \tag{28}$$

$$u_{cj} = \beta(1+r) p_{j+1}(\theta) \left(1 - T_{a,j+1}'\right) u_{cj+1} \tag{29}$$

Equation (28) is the individual intra-temporal optimality condition and equation (29) is the individual Euler equation.

We can use equation (28) to back out the optimal marginal taxes on labor earnings at each age. This is possible because the efficient allocations of consumption and hours come directly from solving the planning problem. Thus, the earnings taxes can simply be defined by integrating over the implied marginal rate in (28) - this is well-defined since output in each age is increasing in θ.

The calculation of optimal asset taxes, however, is not straight-forward. The level of assets a cannot be pinned down independent from the marginal taxes $T_{a,j+1}'$. Therefore, we are going to assume that asset holdings of the lowest type is zero for all ages. This implies that in the equilibrium that decentralizes efficient allocations, the poorest individual is hand-to-mouth in all ages. Given this restriction we can use the following procedure to find the optimal asset taxes.

We can combine the equations (28) and (29) together with (7) and use extensive algebra to show that the derivative of asset holdings with respect to θ, a_j', satisfies

$$a_j'(\theta) \quad = \quad \frac{1}{u_{c,j}(\theta)} \sum_{s=j}^{T} \beta^{s-j} \frac{P_s(\theta)}{P_j(\theta)} \left[u_{c,s}(\theta) c_s'(\theta) - v_{l,s}(\theta) (\varphi_s(\theta) l_s(\theta))' \right].$$

Since by assumption $a_j(\underline{\theta}) = 0$, the above determines the level of asset holdings at each age and for each type

$$a_j(\theta) = a_j(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} a_j'(\theta) \, d\theta.$$

Finally, using the Euler equation (29), we must have

$$1 - T_{a,j+1}' = \frac{u_{cj}}{\beta(1+r) p_{j+1} u_{cj+1}}.$$

The above formula determines the marginal tax rate on asset holdings and since $a_j' > 0$, a well-defined tax function on asset holdings must exist. This completes the construction. □

The construction of the taxes and asset holdings are somewhat standard. In particular, earnings and asset taxes can be constructed from integrating the labor and saving wedges as defined above. Furthermore, fixing the intercept of taxes at each age, determines the asset holdings of individuals. Finally, the assumptions imposed on allocations in the lemma ensure that assets and earnings at each age are increasing in θ and thus the tax functions constructed are well-defined.

We cannot derive a closed form formula for optimal taxes. However, our implementation procedure as discussed in the above paragraph provides a guideline on how to numerically compute the optimal tax functions. We present the results of these computations in Section 5.2.

Finally, note that the monotonicity constraints in Lemma 6 are necessary for existence of a tax function. While we have no way of theoretically checking that they are satisfied, our numerical simulations always involve a check that ensures that they are indeed satisfied. Needless to say, in all of our simulations the monotonicity constraints are satisfied.

C 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

C.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 period[59] 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

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

[59] Here, 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

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

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

$$\forall \hat{\theta} \in [\theta, \theta + d\theta_y], 1 - T_y'\left(y\left(\hat{\theta}\right) + \delta y\right) - d\tau_y = \frac{1}{\hat{\theta}} v'\left(\frac{y\left(\hat{\theta}\right) + \delta y}{\hat{\theta}}\right)$$

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

$$\begin{aligned}
-T_y''\left(y\left(\hat{\theta}\right)\right)\delta y - d\tau_y &= \frac{1}{\hat{\theta}^2} v''\left(y\left(\hat{\theta}\right)/\hat{\theta}\right)\delta y \\
-T_y''\left(y\left(\hat{\theta}\right)\right)\delta y - d\tau_y &= \frac{1}{\hat{\theta}^2} v'\left(y\left(\hat{\theta}\right)/\hat{\theta}\right)\frac{1}{\varepsilon\left(\hat{\theta}\right)}\frac{\hat{\theta}}{y\left(\hat{\theta}\right)}\delta y \\
-T_y''\left(y\left(\hat{\theta}\right)\right)\delta y - d\tau_y &= \left(1 - T_y'\left(y\left(\hat{\theta}\right)\right)\right)\frac{1}{\varepsilon\left(\hat{\theta}\right)}\frac{1}{y\left(\hat{\theta}\right)}\delta y \\
\frac{\delta y}{y\left(\hat{\theta}\right)}\frac{1 - T_y'}{d\tau_y} &= -\frac{1}{\frac{yT_y''}{1 - T_y'} + \frac{1}{\varepsilon(\theta)}}
\end{aligned}$$

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

$$T_y'\left(y\left(\theta\right)\right)\delta y d\theta_y h\left(\theta\right)=-\frac{T_y'}{1-T_y'}y\left(\theta\right)\frac{1}{\frac{yT_y''}{1-T_y'}+\frac{1}{\varepsilon(\theta)}}\frac{1}{y'\left(\theta\right)}h\left(\theta\right)d\tau_y dy$$

In the above, $d\theta_y h\left(\theta\right)$ 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

$$\forall\hat{\theta}\in\left[\theta,\theta+d\theta_a\right],1+T_a'\left(a\left(\hat{\theta}\right)+\delta a\right)+d\tau_a=\beta RPu'\left(Ra\left(\hat{\theta}\right)+\delta a\right)$$

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

$$\begin{aligned}
T_a''\left(a\left(\hat{\theta}\right)\right)\delta a+d\tau_a&=\beta RPu''\left(Ra\left(\hat{\theta}\right)\right)R\delta a\\
T_a''\left(a\left(\hat{\theta}\right)\right)\delta a+d\tau_a&=-\beta RPu'\left(Ra\left(\hat{\theta}\right)\right)\frac{-u''\left(Ra\left(\hat{\theta}\right)\right)Ra\left(\hat{\theta}\right)}{u'\left(Ra\left(\hat{\theta}\right)\right)}\frac{\delta a}{a\left(\hat{\theta}\right)}\\
T_a''\left(a\left(\hat{\theta}\right)\right)\delta a+d\tau_a&=-\left[1+T_a'\left(a\left(\hat{\theta}\right)\right)\right]\frac{-u''\left(Ra\left(\hat{\theta}\right)\right)Ra\left(\hat{\theta}\right)}{u'\left(Ra\left(\hat{\theta}\right)\right)}\frac{\delta a}{a\left(\hat{\theta}\right)}\\
\frac{\delta a}{a\left(\theta\right)}\frac{1+T_a'}{d\tau_a}&=-\frac{1}{\frac{aT_a''}{1+T_a'}+\frac{1}{\sigma(\theta)}}
\end{aligned}$$

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

$$\left[T_a'\left(a\left(\theta\right)\right)+\left(1-P\right)\right]\delta a d\theta_a h\left(\theta\right)=-\frac{T_a'+\left(1-P\right)}{1+T_a'}a\left(\theta\right)\frac{1}{\frac{aT_a''}{1+T_a'}+\frac{1}{\sigma(\theta)}}\frac{1}{a'\left(\theta\right)}h\left(\theta\right)d\tau_a da$$

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,

$$\left[1-\frac{P}{1+T_a'}\right]a\left(\theta\right)\frac{1}{\frac{aT_a''}{1+T_a'}+\frac{1}{\sigma(\theta)}}\frac{1}{a'\left(\theta\right)}h\left(\theta\right)=\frac{T_y'}{1-T_y'}y\left(\theta\right)\frac{1}{\frac{yT_y''}{1-T_y'}+\frac{1}{\varepsilon(\theta)}}\frac{1}{y'\left(\theta\right)}h\left(\theta\right)\tag{30}$$

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 (30), 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

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

which is the same as (3).

C.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

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

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

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

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

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

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

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

or

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

and we can write the above in matrix form

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

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

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

and therefore

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

which becomes

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

which is the same as the desired equation.

D Extensions and Robustness

In this section, we provide supplementary material to the extensions discussed in section 7.

D.1 Alternative Policy Reforms

D.1.1 Optimal Privatization Reform

Table 8 describes the changes in aggregate variables under the privatization reform.

Table 8 shows the changes in the aggregate variables. Note that 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. These policies lead to a slightly lower stock of capital. Consumption and output are slightly higher due to higher labor supply. The labor supply will be higher because of the lower tax on labor income (see Figure 13).

We should note that a related paper McGrattan and Prescott (2017) show that a Pareto improving privatization policy does exist. Their model and calibration is different from this paper in many key dimensions. One of these differences is that they assume Frisch elasticity of labor supply of 2.5 which is much higher than 0.5 that we use. We repeat our optimal reform exercise using Frisch elasticity of labor supply of 2.5 and report results in appendix D.2. We find that in a model with higher labor supply elasticity optimal privatization policy is indeed Pareto improving. Therefore, we arrive at the same qualitative conclusion as McGrattan and Prescott (2017) when we adopt their value for labor supply elasticities. We should note while positive, the gains from privatization are significantly smaller than that of optimal reforms.

D.1.2 Optimal Linear Asset Subsidies

In this section, we illustrate the importance of the progressivity in subsidies. In particular, in our optimal policy problem, we impose that distortions to the intertemporal margin be equated across all individuals of the same age; variations with age are allowed.

The linearity restriction on asset taxes or subsidies imposes restrictions on the set of implementable allocations. To be more precise, in order for allocations to be implementable by linear asset taxes or subsidies, the following condition must hold

$$\frac{P_j(\theta)\, u'(c_j(\theta))}{P_{j+1}(\theta)\, \beta(\theta)\, (1+r)\, u'(c_{j+1}(\theta))} = \frac{P_j(\theta')\, u'(c_j(\theta'))}{P_{j+1}(\theta')\, \beta(\theta')\, (1+r)\, u'(c_{j+1}(\theta'))} \tag{31}$$

for all ages j and for all types θ and θ'. This equation simply implies that the intertemporal marginal rate of substitution must be equal across all types (and therefore, all asset levels).

In order to find the best policies that respect the linear asset tax or subsidy restrictions, we solve the planning problem with constraints (12), (13), (14) and the linear tax constraint (31). Imposing constraint (31) guarantees that the allocation can be implemented by a linear set of asset taxes or subsidies.

It is important to remember that in our model, even in the absence of differential mortality and differential discount factor, the optimal subsidies on assets are not 0. If there is no annuity market, the asset subsidies are needed to correct the inefficiency due to incomplete markets. In that case, these subsidies will be linear and the rate will be equal to the average population mortality at each age. In Figure 20 (right panel) we plot the optimal linear asset subsidies in our model along with the average marginal taxes in a fully optimal system (with nonlinear subsidies) and average population mortality index. The figure shows a large difference in these three measures. The optimal linear subsidies are much lower than the average mortality in the population. This implies that, even in deriving simple policies with linear subsidies, the differential mortality cannot be ignored. In other words, we still need to include these features in the model to correctly capture the effect of heterogeneity in mortality on optimal policies.

Figure 21 shows the marginal labor income tax functions. Note that the linearity restriction on asset taxes or subsidies results in a negative marginal tax on labor income for the poorest individuals (left panel). When subsidies are linear, the marginal rates are much lower for the poorest individuals (relative to the ones that result from fully optimal policies). Therefore, imposing restrictions on proper asset subsidies puts the burden of the redistribution on the labor income tax. In the absence of proper asset subsidies, the consumption for the poor is more front loaded. To accommodate high consumption, the labor income tax must be low (even negative). Also, note that the tax rates at the top are higher under policies with a linear subsidy restriction. Linear subsidies imply that high-income individuals receive too much asset subsidies (relative to the full optimal). Therefore, again, the burden of redistribution is on the labor income tax to correct this excess subsidization of the high-income workers.

Table 9 shows the effect on aggregate variables. Aggregate output, consumption and capital are affected similarly to the fully optimal reform. However, restricted policies only achieve a fraction of cost savings that is achieved by the fully optimal reform policies.

Table 8: Aggregate effects of privatization

	Current U.S.	Continue	Optimal privatization		
	(1)	(2)	(3)	(4)	(5)
Factor prices					
Interest rate (%)	4.05	3.37	4.05	4.06	3.44
Wage	1	1.08	1	0.99	1.07
Values relative to GDP					
Consumption	0.69	0.69	0.69	0.69	0.69
Capital	4.00	4.41	3.99	3.99	4.37
Tax revenue (total)	0.26	0.29	0.23	0.23	0.23
Earnings tax	0.14	0.14	0.11	0.11	0.10
Consumption tax	0.04	0.07	0.04	0.04	0.07
Capital (corporate) tax	0.08	0.07	0.08	0.08	0.08
Transfers	0.16	0.19	0.05	0.05	0.06
To retirees	0.08	0.12	0.00	0.00	0.01
To workers	0.08	0.07	0.05	0.05	0.05
Asset subsidies	0.00	0.00	0.00	0.00	0.00
Change (%) (relative to status quo)					
GDP	–	-2.13	0.38	0.47	-2.14
Consumption	–	-2.38	0.50	0.53	-2.06
Capital	–	7.96	0.07	0.33	6.88
Labor input	–	-9.26	0.62	0.57	-8.56
PDV of net resources	–	–	4.94	5.59	5.06
Consumption equivalence			-0.37	-0.41	-0.64

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 the government' budget constraint). Column (3) is the optimal privatization policies with prices and demographics fixed at the current U.S. values. Column (4) is the optimal privatization policies with equilibrium prices but fixed demographics (at current U.S. levels). Column (5) is the optimal privatization policies with equilibrium prices and future demographics. In column (3) and (4), the percentage change in PDV is calculated relative to column (1). In column (5), the percentage change in PDV is calculated relative to column (2).

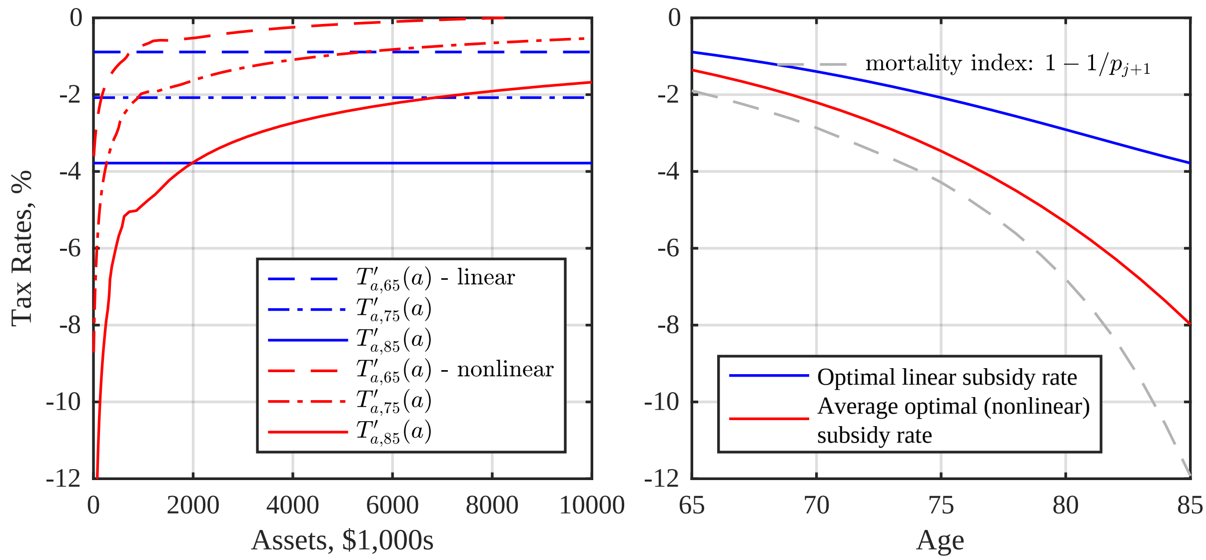

Figure 20: Optimal asset tax functions: linear subsidies vs. nonlinear subsidies. The left panel shows marginal taxes over all asset levels at ages 65, 75 and 85, while the right panel shows average marginal rates at each age from 65 to 85. Blue lines are optimal linear subsidies. Red lines are fully optimal nonlinear subsidies. The dashed line is the population average mortality index.

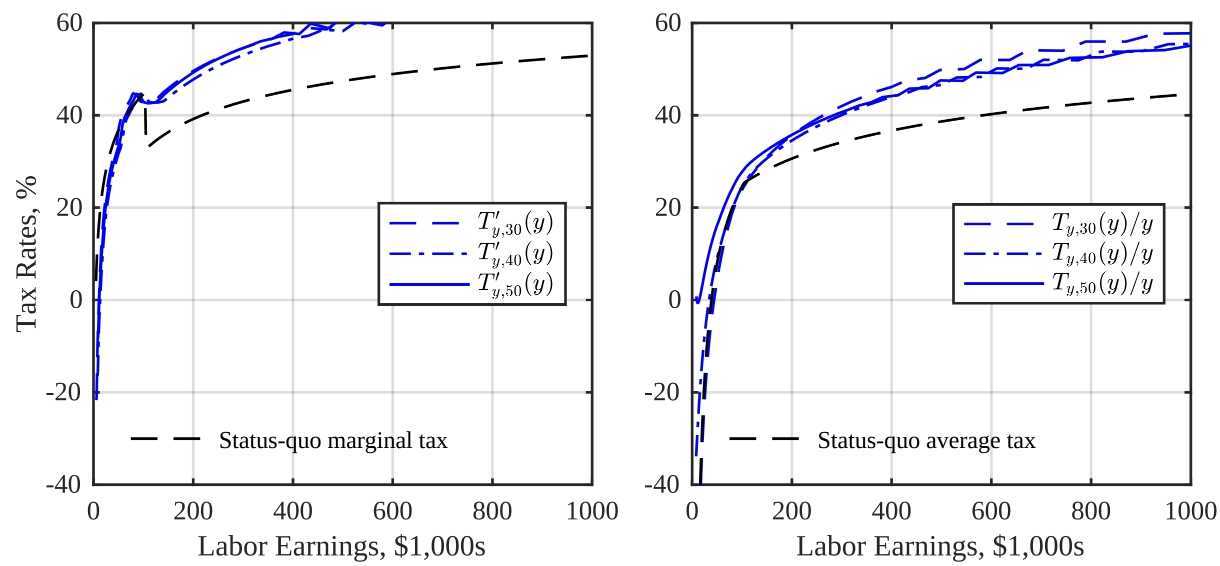

Figure 21: Optimal earnings tax functions with linear asset subsidies. The left panel is marginal taxes, while the right panel shows average taxes. The right panel is average taxes. The black dashed line is the effective status quo tax schedule.

Table 9: Aggregate effects of linear subsidies

	Current U.S.	Continue	Optimal linear subsidies		
	(1)	(2)	(3)	(4)	(5)
Factor prices					
Interest rate (%)	4.05	3.37	4.05	3.80	3.30
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	4.33	4.15	4.46
Tax revenue (total)	0.26	0.29	0.25	0.25	0.26
Earnings tax	0.14	0.14	0.12	0.13	0.11
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	0.09	0.09	0.08
To retirees	0.08	0.12	0.00	0.00	0.00
To workers	0.08	0.07	0.02	0.02	0.03
Asset subsidies	0.00	0.00	0.07	0.07	0.05
Change (%) (relative to status quo)					
GDP	–	-2.13	5.08	2.26	-1.19
Consumption	–	-2.38	2.19	1.02	-1.80
Capital	–	7.96	13.68	5.97	10.14
Labor input	–	-9.26	-1.55	-0.52	-9.12
PDV of net resources	–	–	-6.21	-26.05	-6.18
Consumption equivalence			0.46	1.93	0.79

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 the government' budget constraint). Column (3) is the optimal linear subsidy policies with prices and demographics fixed at the current U.S. values. Column (4) is the optimal linear subsidy policies with equilibrium prices but fixed demographics (at current U.S. levels). Column (5) is the optimal linear subsidy policies with equilibrium prices and future demographics. In column (3) and (4), the percentage change in PDV is calculated relative to column (1). In column (5), the percentage change in PDV is calculated relative to column (2).

D.2 High Labor Supply Elasticity

In our benchmark calibration we assume the Frisch elasticity of labor supply is $\varepsilon = 0.5$. This value is in the range estimated in micro studies and very common in quantitative life cycle macroeconomic models.[60] In this section we report our results for $\varepsilon = 2.5$, which is more in line with values calibrated using macro aggregates in the real business cycle studies. We re-calibrate our model using this value for ε. The calibrated parameters are presented in Table 10.

Table 10: Calibrated parameters—high labor supply elasticity

Parameters	Description		Values
β_0	discount factor: level		0.975
β_1	discount factor: elasticity w.r.t θ		0.009
ψ	weight on leisure		0.644
Targeted Moments		Data	Model
Wealth-income ratio		4.00	4.00
Wealth Gini		0.78	0.78
Average annual hours		2000	2000

We first check the optimality of the status quo policies using the conditions derived in Proposition 4. As we did in Section 5.1, we compute the intertemporal and intratemporal distortions for the status quo allocations and check how much the equations (16), (17), (19), and inequality (18) are violated.

The results are demonstrated in Figures 22 and 23 The left panel in Figure 22 shows the right hand side (black line) and left hand side (dashed red line) of inequality (18). Note that the inequality fails to hold not only at the social security earnings cap but also at the very top income levels. This is expected. When the elasticity of labor supply is high, it is more likely that for some workers there is a Laffer effect. Therefore, it is possible to reduce the tax rates for some individuals (and improve their welfare) while increasing tax revenue. However, this applies to a small fraction of the earnings distribution. As we see below, the optimal tax rates at the top are indeed much lower than the status quo.

The right panel of Figure 22 shows the changes in earnings wedges required for equation (19) to hold. As in the benchmark calibration, these deviations from efficiency are small. Figure 23 shows the changes in savings wedges required for equations (16) and (17) to hold. As in the baseline calibration, these figures show that the savings wedges are further away from efficiency, particularly at older ages.

Figures 24 and 25 shows the optimal asset taxes and optimal labor income taxes respectively. The only notation difference here is that optimal labor taxes on labor income deviations signifi-

[60] See Keane (2011), Chetty et al. (2011), and Chetty (2012).

cantly from status quo, particularly at the top. This is mainly due to a Laffer effect. Since labor supply is more elastic, it is possible to reduce tax rates, expand the tax base without sacrificing welfare.

We report the aggregate effects of full optimal reform in Table 11. To compare and contrast with the benchmark calibration, we also repeat privatization exercise. These results are reported in Table 12. As we see, when elasticity of labor supply is high enough even privatization yields efficiency gains. Although, the gains are smaller than those of the fully optimal reform.

To summarize, asset subsidies remain an integral part of Pareto optimal policies even in a specification with high labor supply elasticity. In this case, under optimal policies, major changes in earnings taxes applies only to top earning levels, while large asset subsidies apply to all individuals.

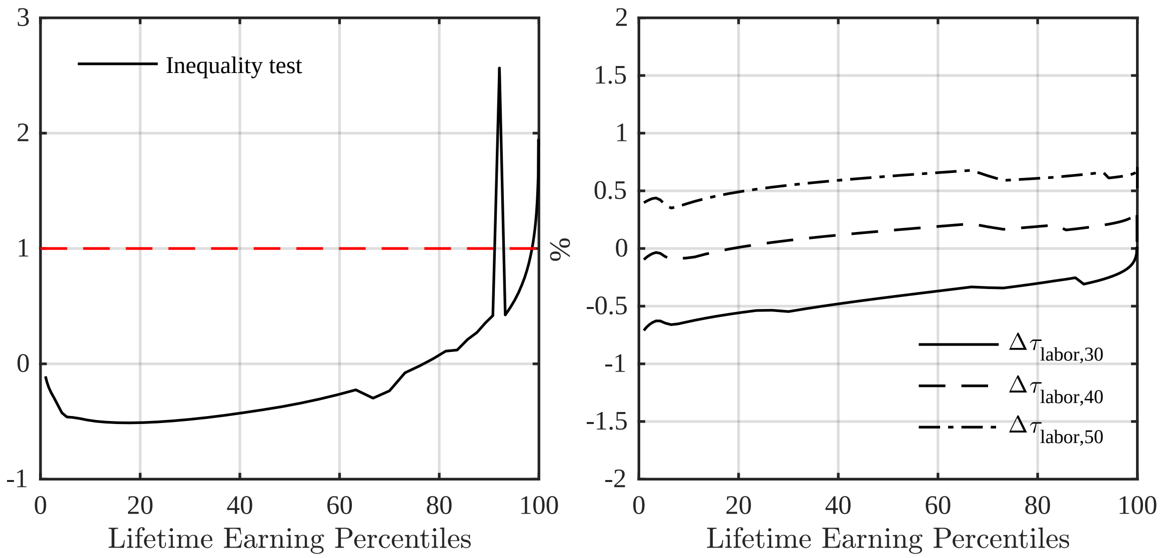

Figure 22: Test of Pareto optimality for status quo policies—high labor supply elasticity . The left panel plots the two sides of inequality (18). The right panel depicts the change in earnings wedges required for equation (19) to hold.

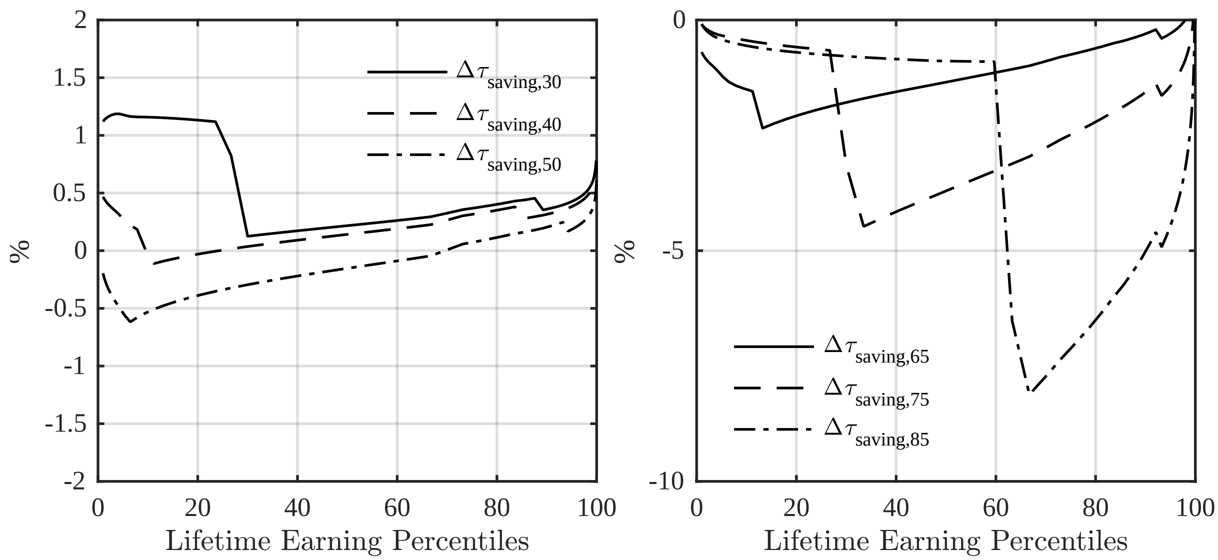

Figure 23: Test of Pareto optimality for status quo policies—high labor supply elasticity. The left panel depicts the required change in savings wedges so that (16) holds at ages 30, 40 and 50. The right panel depicts the required change in savings wedges so that (17) holds at ages 65, 75 and 85.

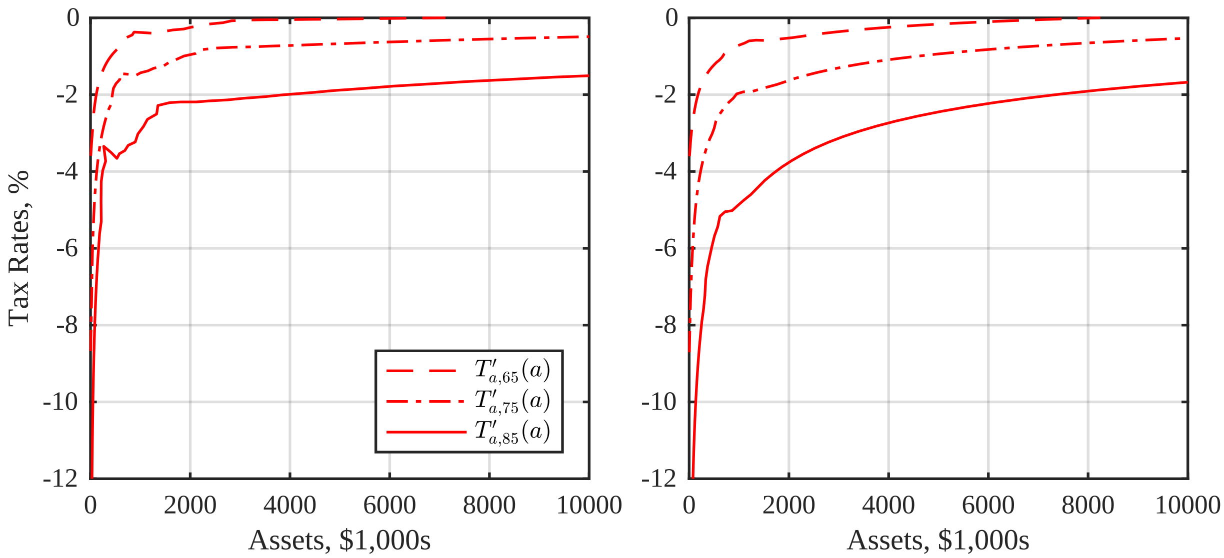

Figure 24: Optimal asset tax functions – high labor supply elasticity. The left panel shows optimal marginal taxes with labor supply elasticity of 2.5 over all asset levels at ages 65, 75 and 85. The right panel shows the same for the benchmark model.

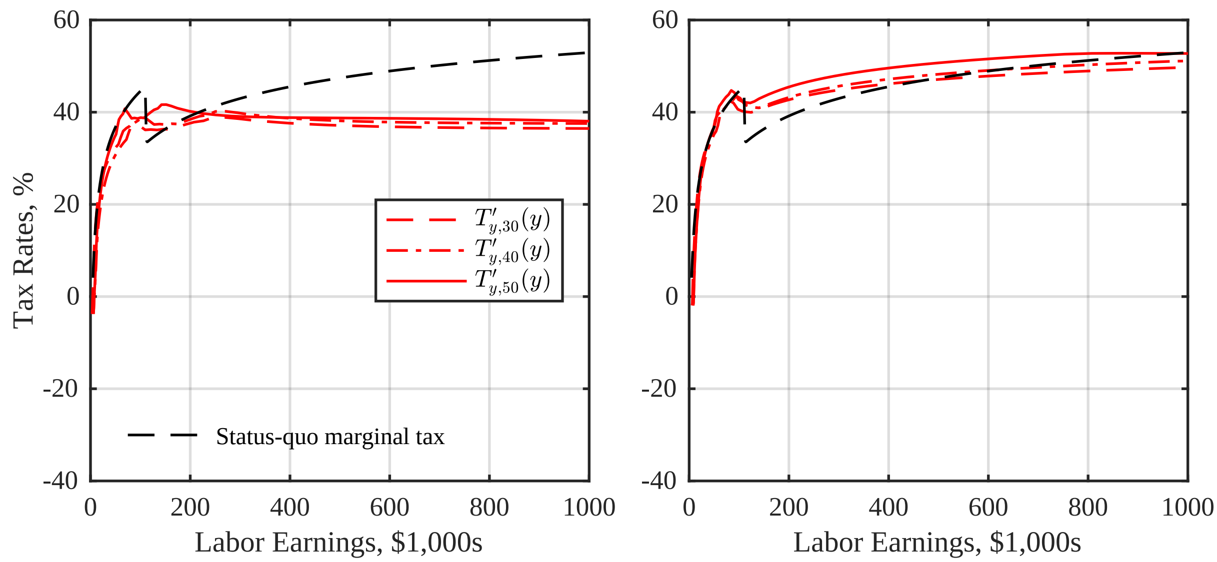

Figure 25: Optimal earnings tax functions – high labor supply elasticity. The left panel shows optimal marginal taxes with labor supply elasticity of 2.5. The right panel shows the same for the benchmark model. The black dashed line is the effective status quo tax schedule.

Table 11: Aggregate effects of optimal policies – high labor supply elasticity

	Current U.S.	Continue	Optimal reform		
	(1)	(2)	(3)	(4)	(5)
Factor prices					
Interest rate (%)	4.05	3.48	4.05	3.89	3.48
Wage	1.00	1.06	1.00	1.02	1.06
Values relative to GDP					
Consumption	0.69	0.69	0.67	0.68	0.69
Capital	4.00	4.34	4.27	4.09	4.34
Tax revenue (total)	0.26	0.29	0.26	0.27	0.27
Earnings tax	0.14	0.15	0.14	0.15	0.13
Consumption tax	0.04	0.07	0.04	0.04	0.07
Capital (corporate) tax	0.08	0.08	0.09	0.08	0.08
Transfers	0.17	0.20	0.14	0.14	0.13
To retirees	0.08	0.12	0.02	0.03	0.04
To workers	0.09	0.08	0.07	0.07	0.06
Asset subsidies	0.00	0.00	0.05	0.05	0.03
Change (%) (relative to status quo)					
GDP	–	-0.56	4.44	2.16	0.87
Consumption	–	-0.29	2.03	1.38	1.16
Capital	–	7.94	11.57	4.46	9.41
Labor input	–	-6.66	-1.05	0.42	-5.26
PDV of net resources	–	–	-8.46	-17.76	-5.01
Consumption equivalence			0.83	1.73	0.73

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 the 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 PDV is calculated relative to column (1). In column (5), the percentage change in PDV is calculated relative to column (2).

Table 12: Aggregate effects of optimal privatization policies – high labor supply elasticity

	Current U.S.	Continue	Optimal privatization		
	(1)	(2)	(3)	(4)	(5)
Factor prices					
Interest rate (%)	4.05	3.35	4.05	3.81	3.449
Wage	1.00	1.06	1.00	1.03	1.06
Values relative to GDP					
Consumption	0.69	0.69	0.68	0.69	0.69
Capital	4.00	4.34	4.12	4.04	4.34
Tax revenue (total)	0.26	0.29	0.25	0.25	0.25
Earnings tax	0.14	0.15	0.13	0.13	0.13
Consumption tax	0.04	0.07	0.04	0.04	0.07
Capital (corporate) tax	0.08	0.08	0.08	0.08	0.08
Transfers	0.17	0.20	0.12	0.12	0.13
To retirees	0.08	0.12	0.01	0.01	0.03
To workers	0.09	0.08	0.11	0.11	0.10
Asset subsidies	0.00	0.00	0.00	0.00	0.00
Change (%) (relative to status quo)					
GDP	–	-0.56	3.38	2.49	1.63
Consumption	–	-0.29	2.36	2.15	1.96
Capital	–	7.94	6.41	3.49	10.16
Labor input	–	-6.66	1.05	1.73	-4.48
PDV of net resources	–	–	-2.66	-6.42	-1.67
Consumption equivalence			0.26	0.63	0.25

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 the government' budget constraint). Column (3) is the optimal privatization policies with prices and demographics fixed at the current U.S. values. Column (4) is the optimal privatization policies with equilibrium prices but fixed demographics (at current U.S. levels). Column (5) is the optimal privatization policies with equilibrium prices and future demographics. In column (3) and (4), the percentage change in PDV is calculated relative to column (1). In column (5), the percentage change in PDV is calculated relative to column (2).

D.3 Alternative status quo tax function

In this section we report our results with a non-smooth approximation of the status quo tax function. We use effective marginal federal (and state and local) tax rates for head of household with one child as reported in Harris (2005). We plot these effective marginal rates and implied average tax rates in Figure 26. The left panel shows the tax rates excluding payroll taxes. The right panel includes payroll taxes. In contrast to the benchmark tax function that we use, we see many jumps in tax rates, specially at lower incomes.

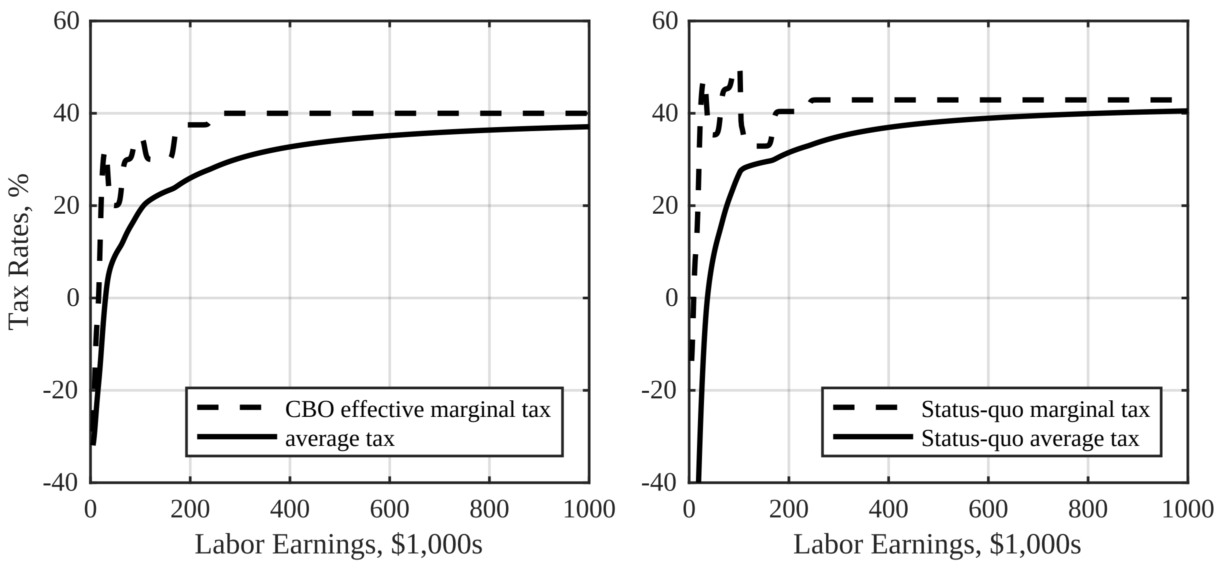

Figure 26: Alternative tax functions. The left panel is the effective marginal (and average) tax rates for head of household with one child as reported in Harris (2005), excluding payroll taxes. The right panel is the total tax effective tax rates including payroll taxes.

Figures 27 and 28 shows the result of efficiency test. Left panel of Figure 27 shows the inequality (28). Despite a non-smooth status quo tax function, the earnings tax function fails the inequality (28) only at the social security maximum taxable income. However it comes close to being violated at a lower income level as well. This is the large drop in the effective marginal tax rate due to transition from EITC phase out (which implies the effective marginal rate of 31%) to the 15% bracket (see Harris (2005) for more details). Also, the deviations from equations the tax smoothing equation for saving and earnings taxes, equations (16), (17), and (19) vary a lot more relative to the benchmark. Figure 29 depicts the optimal savings taxes. There is no significant change in policy with regard to asset subsidies.

Finally, Table 13 shows the aggregate implications of the reform under various scenarios regarding partial or general equilibrium and current vs. future demographics. The aggregate implications are very similar to the ones reported in Section 6 for the main model.

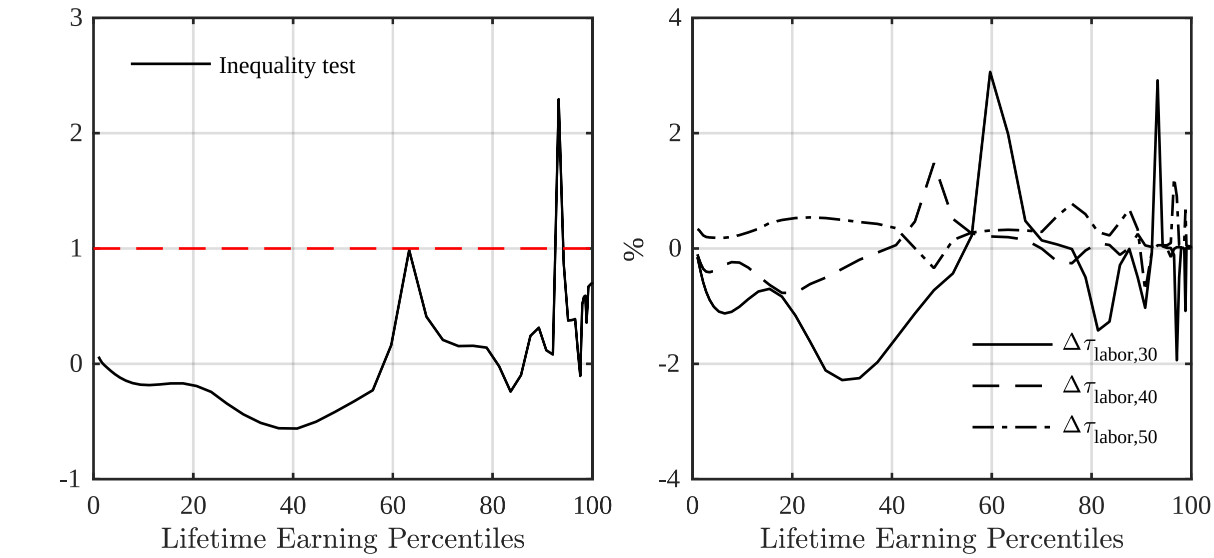

Figure 27: Test of Pareto optimality for status quo policies with CBO effective tax rates. The left panel plots the two sides of the inequality (18). The right panel depicts the change in earnings wedges required for (19) to hold.

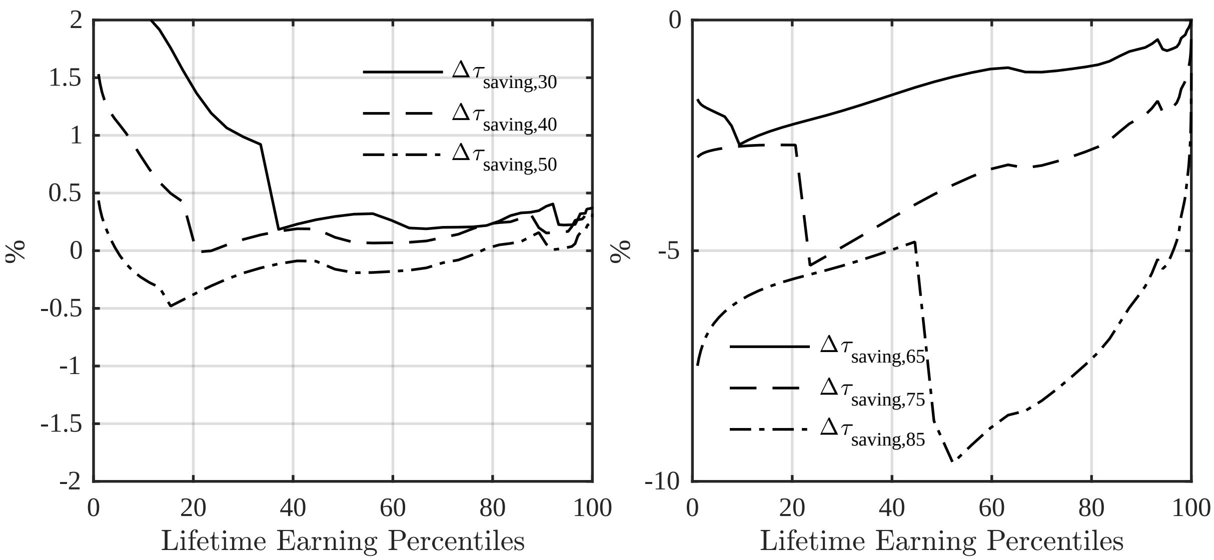

Figure 28: Test of Pareto optimality for status quo policies with CBO effective tax rates. Test of Pareto optimality for status quo policies. The left panel depicts the required change in savings wedges so that (16) holds at ages 30, 40 and 50. The right panel depicts the required change in savings wedges so that (17) holds at ages 65, 75 and 85.

Table 13: Aggregate effects of optimal policies—alternative (status quo) earnings tax function

	Current U.S.	Continue	Optimal reform		
	(1)	(2)	(3)	(4)	(5)
Factor prices					
Interest rate (%)	4.05	3.38	4.05	3.83	3.33
Wage	1	1.08	1	1.02	1.08
Values relative to GDP					
Consumption	0.69	0.69	0.67	0.68	0.68
Capital	4.00	4.40	4.29	4.13	4.44
Tax revenue (total)	0.26	0.29	0.27	0.27	0.28
Earnings tax	0.14	0.14	0.14	0.15	0.14
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	0.14
To retirees	0.08	0.12	0.02	0.02	0.03
To workers	0.08	0.07	0.05	0.05	0.05
Asset subsidies	0.00	0.00	0.08	0.08	0.06
Change (%) (relative to status quo)					
GDP	–	-2.13	4.07	1.64	-1.41
Consumption	–	-2.30	1.57	0.59	-1.87
Capital	–	7.67	11.45	4.76	9.39
Labor input	–	-9.08	-1.63	-0.70	-8.98
PDV of net resources	–	–	-10.81	-28.48	-7.17
Consumption equivalence			0.79	2.09	0.92

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 the 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 PDV is calculated relative to column (1). In column (5), the percentage change in PDV is calculated relative to column (2).

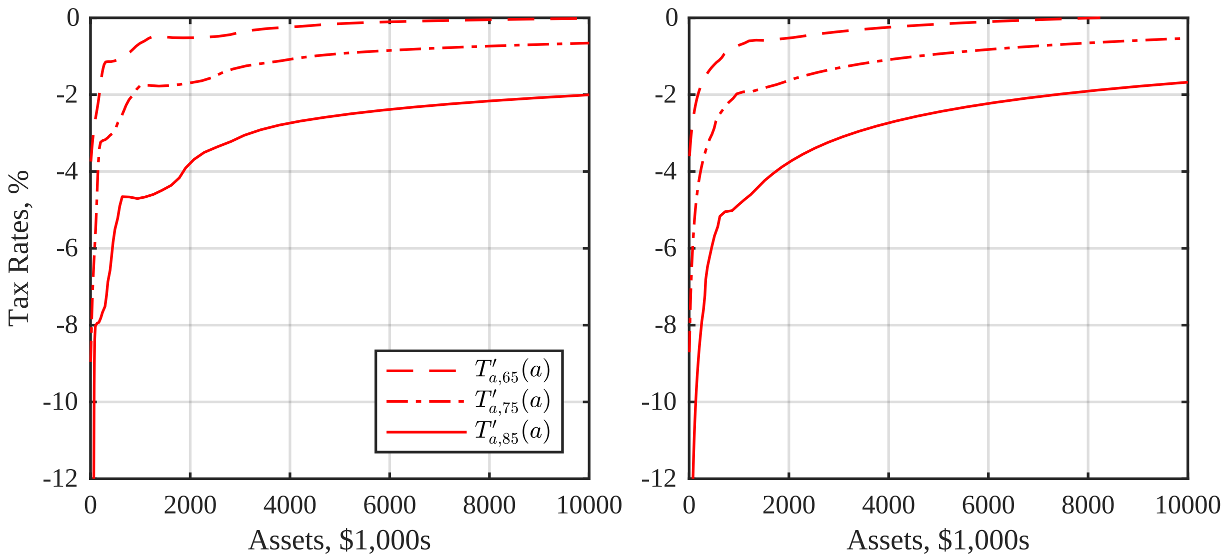

Figure 29: Optimal asset tax functions for the alternative status quo tax policy (CBO effective tax rates). The left panel shows the optimal marginal taxes for calibration with CBO effective tax rates at ages 65, 75 and 85. The right panel shows the same for the benchmark model.

D.4 Small Open Economy

In Section 6, we described the transition from the current steady state in the U.S. economy to a new steady state with new demographic parameters. In performing that exercise, we maintained the close economy assumption and determined factor prices in general equilibrium. In this section we report the steady state result under a small open economy (fixed factor price assumptions). Figure 30 and 31 show the optimal earnings and asset tax functions at the new steady state, respectively. Table 14 reports the aggregate implications. There is no significant difference between these numbers and the ones reported for the main exercise.

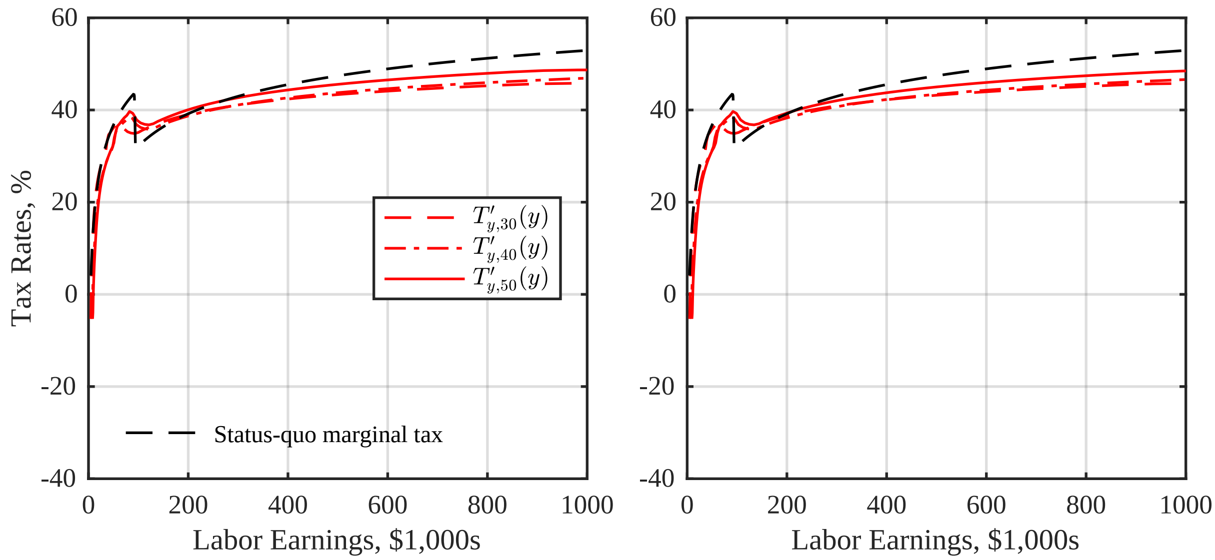

Figure 30: Optimal labor income tax functions–small open economy assumption. The left panel shows the marginal taxes in a small open economy, while the right panel shows the same for the benchmark model. The black dashed line is the effective status quo tax schedule. Both panels show the tax functions in the new steady state after demographics transition

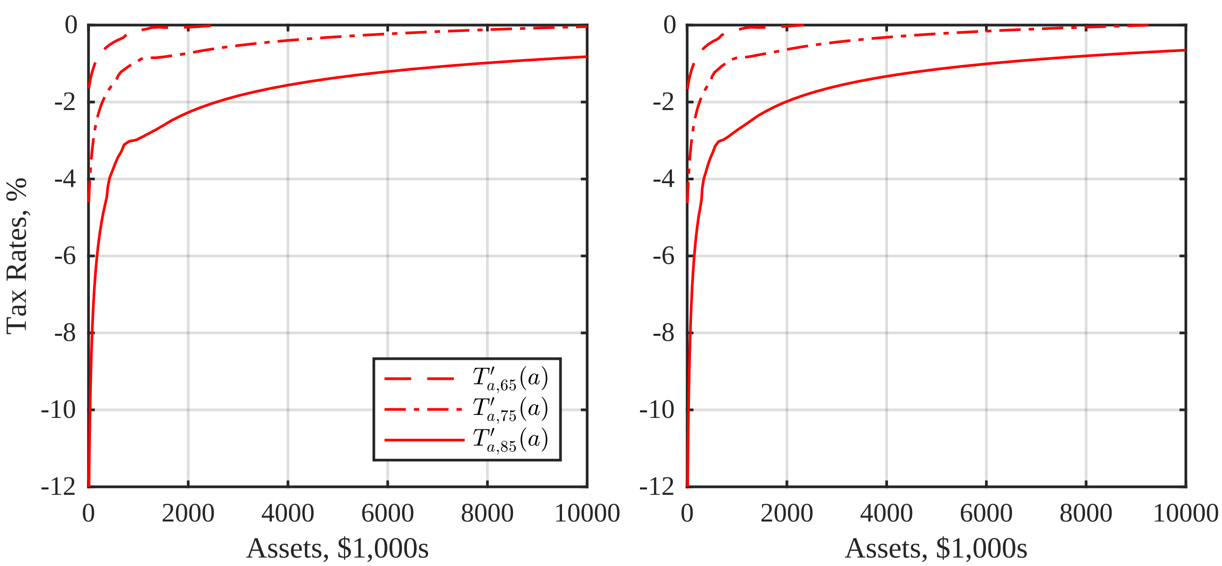

Figure 31: Optimal asset tax functions–small open economy assumption. The left panel shows the marginal taxes in a small open economy over all asset levels at ages 65, 75 and 85, while the right panel shows the same for the benchmark calibration. Both panels show the tax functions in the new steady state after demographics transition.

Table 14: Aggregate effects of optimal policies–small open economy

	Current U.S.	Continue	Optimal reform	
	(1)	(2)	(3)	(4)
Factor prices				
Interest rate (%)	4.05	4.05	4.05	4.05
Wage	1.00	1.00	1.00	1.00
Values relative to GDP				
Consumption	0.69	0.80	0.67	0.82
Capital	4.00	4.00	4.31	4.00
Tax revenue (total)	0.26	0.30	0.27	0.31
Earnings tax	0.14	0.14	0.14	0.14
Consumption tax	0.04	0.09	0.04	0.09
Capital (corporate) tax	0.08	0.08	0.09	0.08
Transfers	0.16	0.21	0.15	0.18
To retirees	0.08	0.13	0.02	0.04
To workers	0.08	0.08	0.05	0.05
Asset subsidies	0.00	0.00	0.08	0.09
Change (%) (relative to status quo)				
GDP	–	-10.78	4.33	-11.39
Consumption	–	3.40	1.66	5.52
Capital	–	-10.78	12.29	-11.39
Labor input	–	-10.78	-1.80	-11.39
PDV of net resources	–	–	-11.08	-5.59
Consumption equivalence			0.82	0.56

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 the government's budget constraint). Column (3) is the optimal reform policies with demographics fixed at the current U.S. values. Column (4) is the optimal reform with future demographics. In column (3), the percentage change in PDV is calculated relative to column (1). In column (4), the percentage change in PDV is calculated relative to column (2).

D.5 Out-of-pocket Medical Expenditures

In this section, we describe the details of our estimation of out-of-pocket medical expenditure, as well as the calibration and optimal policy exercise in this extension of our model.

Out-of-Pocket Medical Expenditure. The estimation of out-of-pocket medical expenditure profiles closely follows De Nardi et al. (2010), who use a sample of 3259 single retired individuals in the AHEAD survey between 1996 and 2006. [61] All individuals in the sample are aged 70 or older in 1994 (the year they enter the AHEAD survey). The measure of medical expenditure includes out-of-pocket expenditures on insurance premia, drug costs, hospital stays, nursing home care, doctor visits, dental visits, and outpatient care. The measure of income includes the value of social security benefits, defined pension benefits and annuities, veteran's benefit, welfare and food stamps. The permanent income is defined as the average of all income that an individual receives over the period he or she is in the sample. The measure of permanent income is then used to compute each person's rank in the income distribution (which we refer to as permanent income ranking). In the estimation, this variable is used to capture the dependence of medical expenditure on income.

The estimation procedure is as follows. The mean of logged medical expenditure is modeled as a polynomial in age, a quadratic in the individual's permanent income ranking, and permanent income ranking interacted with age. We estimate the profiles using a fixed-effect estimator.[62] The resulting estimation gives us the following model for the log of medical expenditure:

$$\log\left(\text{medical expenditure at age } j\right) = \text{polynomial in } j + f(\text{permanent income ranking}, j),$$

where $f\left(.,.\right)$ is the sum of a quadratic function of permanent income ranking and permanent income ranking interacted with age.

We can now use our estimation to generate the profiles of age-dependent medical expenditures $m_j\left(\theta\right)$. In our framework, there is a one-to-one correspondence between permanent (or lifetime) income and ability type θ. Therefore, for each θ we use $H\left(\theta\right)$ as the permanent income ranking. The left panel in Figure (16) shows the simulated mean medical expenditure profiles for each permanent income quintile normalized by GDP per capital in 1998. This corresponds to Figure 3 in De Nardi et al. (2010).

Preferences. In order to better capture the pattern of asset decumulation, we assume a curvature of $\sigma = 2$ for utility over consumption. Finally, we fix the value β_1, the gradient of the

[61] The replication data, codes and estimation results for De Nardi et al. (2010) are available at http://users.nber.org/∼denardim/research/De_Nardi_French_Jones.zip. We would like to thank John Jones for providing details and patiently walking us through the estimation procedure.

[62] Here we depart from De Nardi et al. (2010) and do not include gender and health status in the estimation since we do not have these variables in our model.

Table 15: Parameters calibrated using the model—with out-of-pocket medical expenditures

Parameters	Description		Values
β_0	discount factor: level		0.987
ψ	weight on leisure		0.037
Targeted Moments		Data	Model
Capital-output ratio		4.00	4.00
Average annual hours		2000	2000

discount factor with respect to ability θ, at the benchmark calibrated value and, therefore, we do not target the wealth Gini in cross section.[63] As before, we choose β_0 (the location parameter for the discount factor) to match a capital to output ratio of 4. We also choose ψ (disutility of work) to match the average annual hours worked of 2000. Table 15 shows the new calibration results.

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 median assets by permanent income quintile in the AHEAD data. The data are based on De Nardi et al. (2010) calculations for the 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. However, under this calibration, the wealth Gini is 0.67. Therefore, the model fails to capture the cross section of wealth inequality.

Using the calibrated model we compute the optimal earnings tax and asset subsidies. These are presented in Figure 32 and 33. As these figures demonstrate, there are no significant differences between the optimal policies derived in the model with out-of-pocket medical expenditures and those in our main exercise.

Table 16 shows the effect of the optimal policies on aggregate quantities. The last two rows present the efficiency gains. The second to last row is a decline in the present discounted value of lifetime consumption net of labor income for each cohort. The last row is the percentage decline that is required in non-medical consumption under the status quo policies to make the cost of allocations (in net PDV) equal across the two economies. As we see, the magnitude of cost savings is not very different than the ones in the main exercise. For the partial equilibrium calculation (column 3), they are lower, indicating that the value consumption smoothing over life cycle is lower in this model. However, in the general equilibrium setup with new demographics (column 5), the magnitudes are slightly greater. This is because with an aging population, medical expenditures provide even a stronger motive for saving. This leads to high stock of capital and low interest rate. This in effect magnifies the present discounted values.

In summary, the inclusion of out-of-pocket medical expenditure results in a richer model that

[63]Our attempt to match the wealth Gini leads to a very slow decumulation of wealth at the top income quintile.

is able to capture more details in patterns of asset decumulation at older ages. However, the model's implication for an optimal policy does not change. Moreover, the efficiency gains from implementing optimal policies are still significant but lower compared to the benchmark model without out-of-pocket medical expenditure.

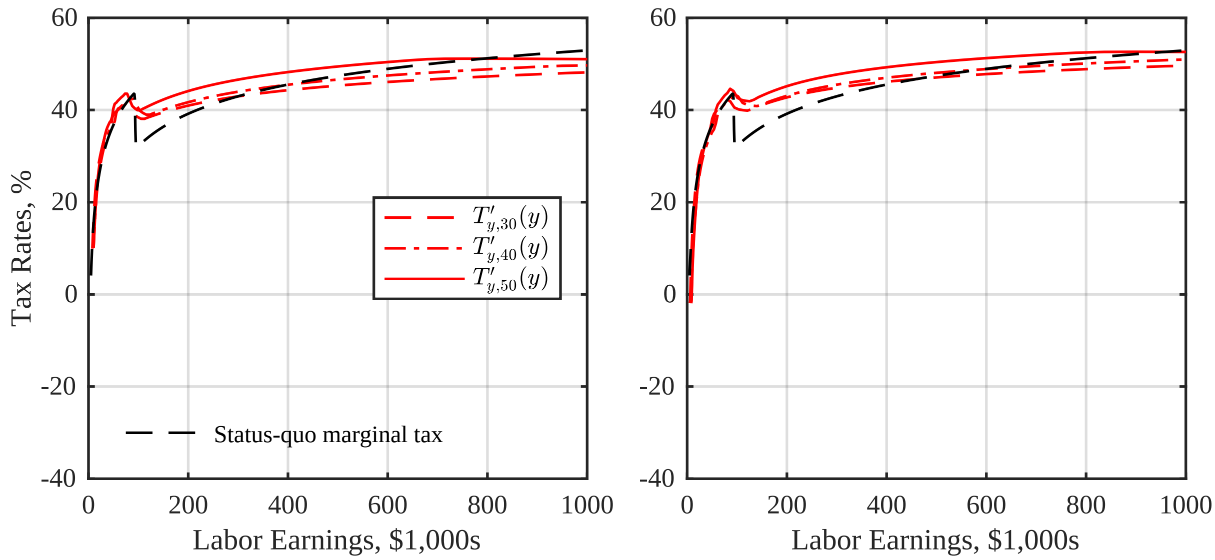

Figure 32: Optimal labor income tax functions with out-of-pocket medical expenditures. The left panel is optimal marginal taxes with out-of-pocket medical expenditure. The right panel shows the same in the benchmark model. The black dashed line is the effective status quo tax schedule.

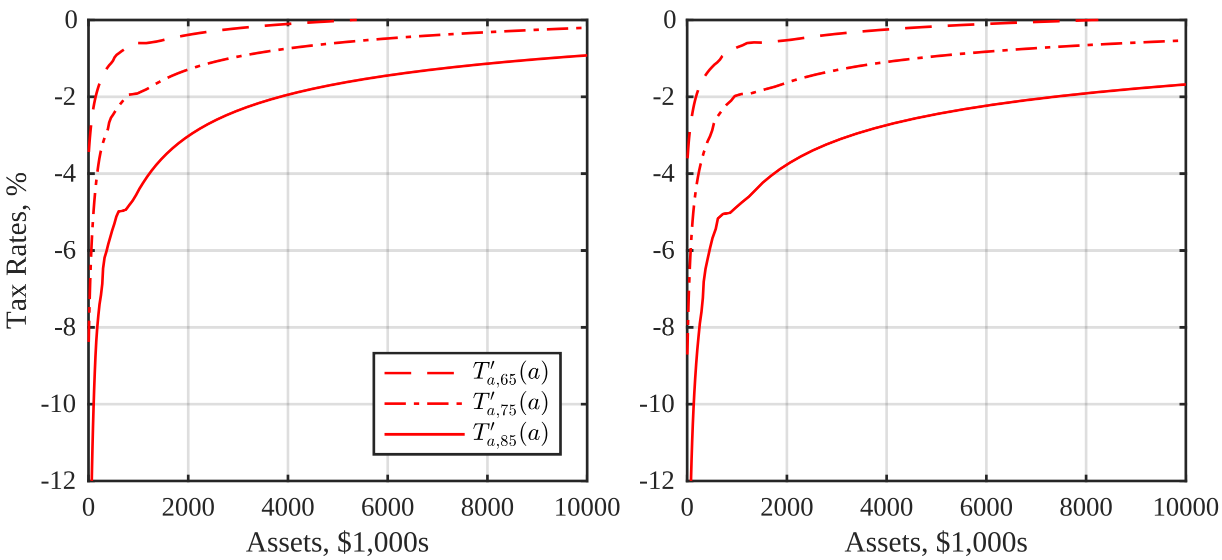

Figure 33: Optimal asset tax functions with out-of-pocket medical expenditures. The left panel shows the optimal marginal taxes with out-of-pocket medical expenditures at ages 65, 75 and 85. The right panel shows the same in the benchmark model.

Table 16: Aggregate effects of optimal policies—with out-of-pocket medical expenditures

	Current U.S.	Continue	Optimal reform		
	(1)	(2)	(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
Values relative to GDP					
Consumption	0.69	0.68	0.68	0.68	0.67
Capital	4.00	4.57	4.21	4.12	4.63
Tax revenue (total)	0.24	0.27	0.26	0.26	0.27
Earnings tax	0.12	0.13	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	0.15	0.18	0.18	0.17	0.16
To retirees	0.08	0.12	0.05	0.04	0.06
To workers	0.07	0.06	0.06	0.06	0.04
Asset subsidies	0.00	0.00	0.07	0.07	0.06
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 the 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 PDV is calculated relative to column (1). In column (5), the percentage change in PDV is calculated relative to column (2).

D.6 Asset Tax Paid Directly by Households

In this section, we consider an alternative calibration in which households directly pay taxes on their assets. We consider two alternatives: a flat tax on asset income and a progressive tax on asset income.

D.6.1 Flat Tax on Asset Income

In this section, we assume that taxes on capital are not entirely paid by firms, and a portion is paid by individuals. In particular, there is a flat asset tax rate of 27 percent on individual asset income, while the corporate income tax rate is 8.2 percent. These numbers are chosen so that the after-tax return on savings is the same as in the benchmark model, and that corporate income tax revenue is 2 percent of GDP. As in the benchmark model, we reform the savings taxes, but not the corporate income tax rate. The rest of the model parameters are not changed.

Within this setup we solve for optimal policies. In Figure 34, we compare the optimal asset subsidies under this setup (left panel) to those in the main exercise (right panel). Notice that at the bottom of the asset distribution they are very similar. Inclusion of the asset tax in the calibration does not affect very low income individuals under the status quo, since they do not hold any asset (regardless of whether taxes are paid by them or by the firm). In the middle range, however, the removal of the asset tax raises the after-tax return on assets. This achieves some of the saving motives that the asset subsidy is supposed to provide. Therefore, the required optimal subsidy is smaller. Overall, however, two key features of asset subsidies are robust to this variation. They are large and they are progressive.

The result regarding optimal earnings taxes is less robust. The left panel in Figure 35 shows the optimal marginal taxes on earnings, and it compares them to those in the main exercise. The optimal marginal tax rates on earnings are much higher in this exercise, except at the bottom and the very top of the income distribution. Note that except for a fraction of people at the bottom, all individuals pay asset taxes under the status quo. An optimal reform removes this tax and instead subsidizes asset accumulation at older ages, which raises consumption at older ages. In order to reduce the cost of delivering the status quo welfare, optimal policies need to reduce consumption at younger ages. The only instrument the government has to achieve this is earnings tax. Therefore, earnings taxes must increase. Note, however, that in this model removing the asset tax generates huge efficiency gains. Table 17 shows the effect on aggregates is quite large. Specifically, the bottom two rows demonstrate a massive decline in cost measures, both in terms of present discounted value of net resources allocated to a cohort and flow consumption.

In summary, the qualitative implication of the policy regarding the optimality of asset subsidies is robust. What changes, however, is that implementing these subsidies (while maintaining the status quo welfare) implies a massive reform of earnings taxes. This reform, however, goes in the opposite direction of the most proposed tax reforms in that it implies higher tax rates for those with middle to upper middle income, while leaving the same tax rates for those at the bottom and the very top of the income distribution.

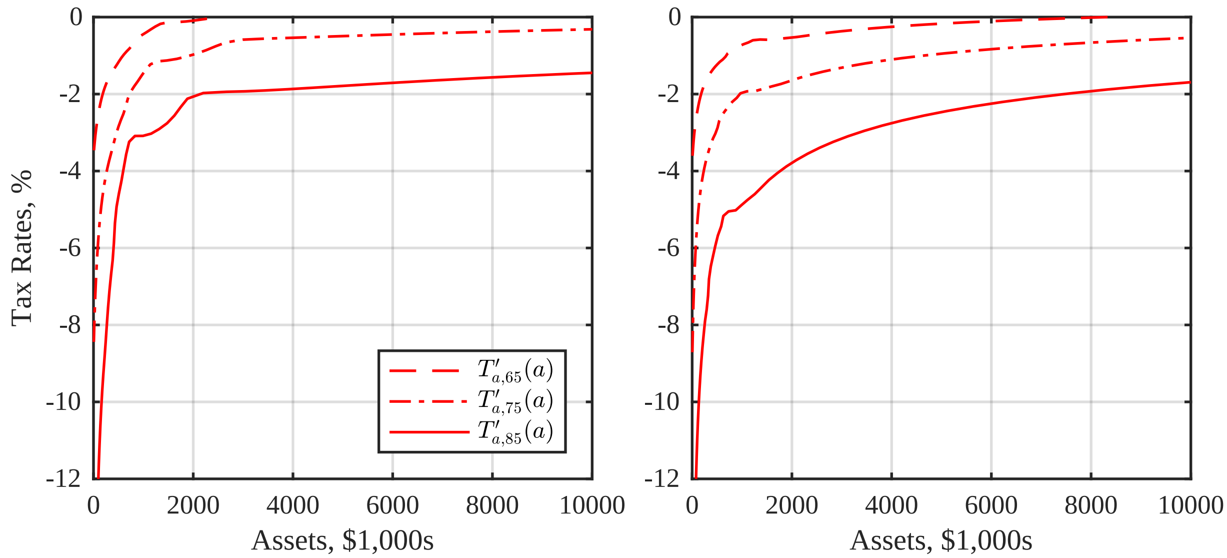

Figure 34: Optimal asset tax functions in the model with a flat (status quo) asset tax. The left panel is optimal marginal taxes in the model with a flat (status quo) asset tax. The right panel shows the same for the benchmark model.

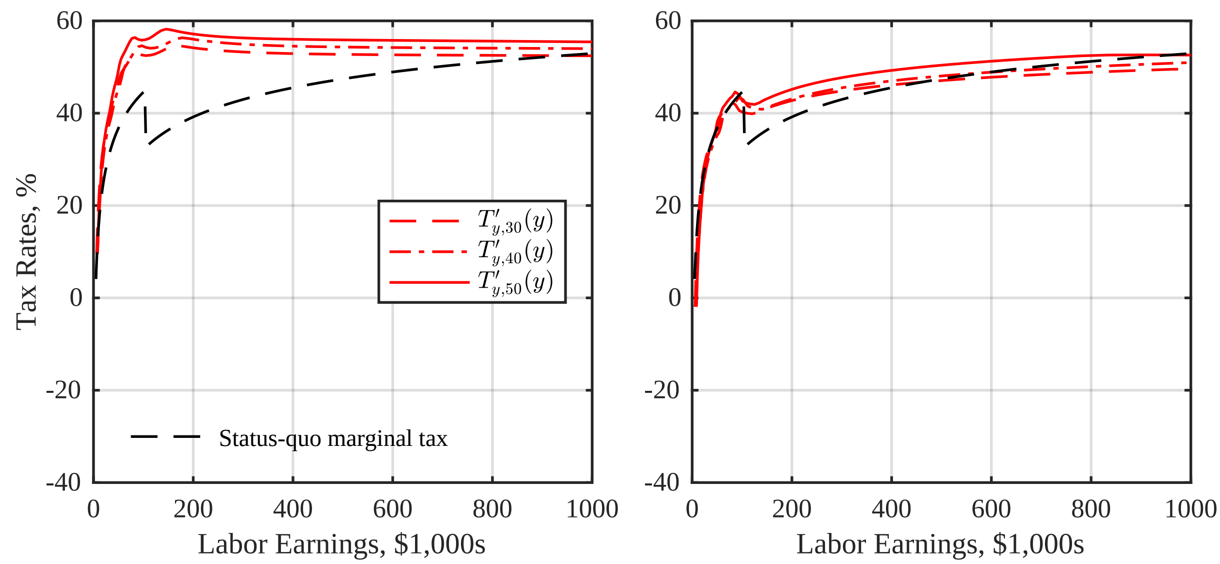

Figure 35: Optimal labor income tax functions in the model with a flat (status quo) asset tax. The left panel is optimal marginal taxes in the model with a flat (status quo) asset tax. The right panel shows the same for the benchmark model.

Table 17: Aggregate effects of optimal policies—with flat (status quo) asset tax

	Current U.S.	Continue	Optimal Reform		
	(1)	(2)	(3)	(4)	(5)
Factor prices					
Interest rate (%)	5.58	4.64	5.58	4.38	3.79
Wage	1	1.08	1	1.10	1.16
Values relative to GDP					
Consumption	0.69	0.69	0.61	0.66	0.66
Capital	4.00	4.41	5.36	4.54	4.87
Tax revenue (total)	0.32	0.36	0.21	0.25	0.25
Earnings tax	0.20	0.20	0.15	0.19	0.17
Consumption tax	0.04	0.07	0.03	0.04	0.07
Capital (corporate) tax	0.08	0.09	0.03	0.02	0.02
Transfers	0.16	0.19	0.13	0.13	0.11
To retirees	0.08	0.12	0.03	0.02	0.03
To workers	0.08	0.07	0.03	0.04	0.03
Asset subsidies	0.00	0.00	0.07	0.07	0.05
Change (%) (relative to status quo)					
GDP	–	-2.13	25.72	7.49	3.95
Consumption	–	-2.39	11.35	2.75	0.01
Capital	–	7.96	68.35	22.07	26.52
Labor input	–	-9.24	-7.10	-2.54	-10.40
PDV of net resources	–	–	-24.01	-61.26	-35.24
Consumption equivalence			4.04	10.30	8.13

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 the 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 PDV is calculated relative to column (1). In column (5), the percentage change in PDV is calculated relative to column (2).

D.6.2 Progressive Tax on Asset Income

Here, we repeat the previous exercise in a setting where the status quo asset tax is not flat but progressive. For this exercise we keep the corporate income tax rate at 8.2 percent. However, we assume households face a progressive tax schedule on their asset income. We calibrate this tax schedule using Poterba and Samwick (2003) calculations of the effective tax rates on household portfolios. More precisely, we assume households face an incremental tax schedule with tax rates of 0, 15, 28 and 33 percent on their asset income. We choose the increment thresholds so that the distribution of tax rates in the model matches those computed by Poterba and Samwick (2003) for 1989 (Figure 1 in their paper).[64] Figure 36 shows the asset tax function (left panel) and the cross sectional distribution of tax rates (right panel).

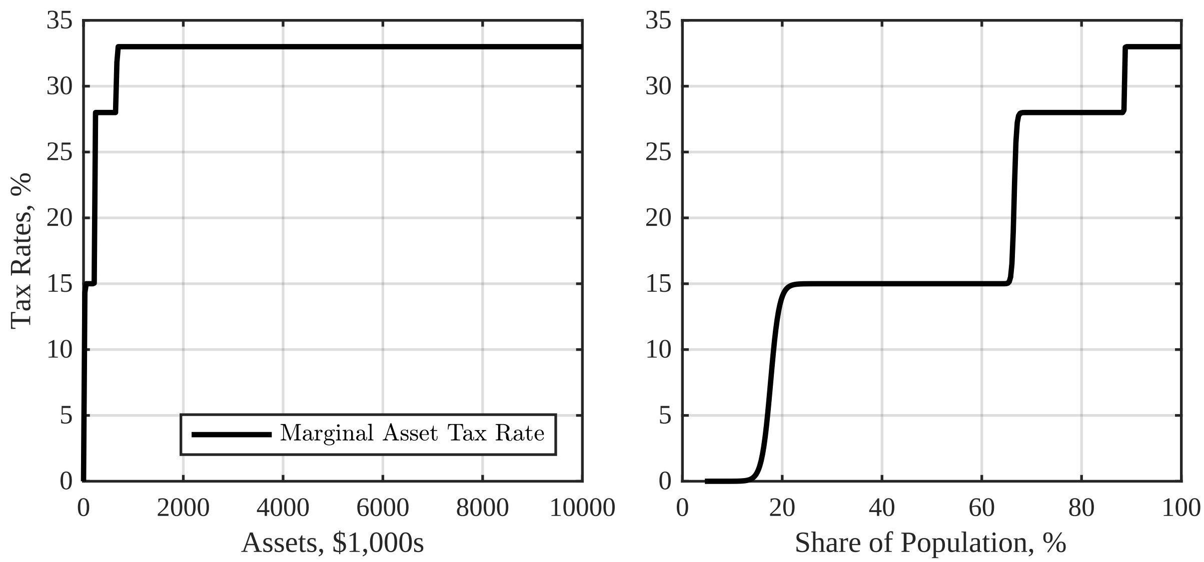

Figure 36: Nonlinear (status quo) asset tax. The left panel shows the marginal asset tax rates on asset income as a function of assets. The right panel plots these tax rates against the fraction of the population who pays these tax rates.

Even if in this exercise we use a progressive asset tax schedule, the results regarding optimal asset subsidies and earnings taxes remain the same as those of the model with a flat savings tax. Figure 37 shows the optimal asset subsidy in this setup (left panel) and that of the main exercise. Figure 38 shows the optimal earnings taxes (left panel) and those in the main exercise. Table 18 shows the aggregate implications.

Overall, as in the case with a flat asset tax, the optimal reform yields massive efficiency gains.

[64]The calculations by Poterba and Samwick (2003) ignore the implicit subsidies present in individual retirement accounts as well as the mortgage subsidies and the ability of rich individuals to work around savings taxes. We think this is one of the main reasons that reduce asset taxes for the whole population.

To achieve these efficiency gains, optimal policies require implementing large progressive subsidies. However, unlike in our main exercise, to be able to implement those subsidies and achieve the cost savings, the earnings tax rates must rise for those with a middle to upper income.

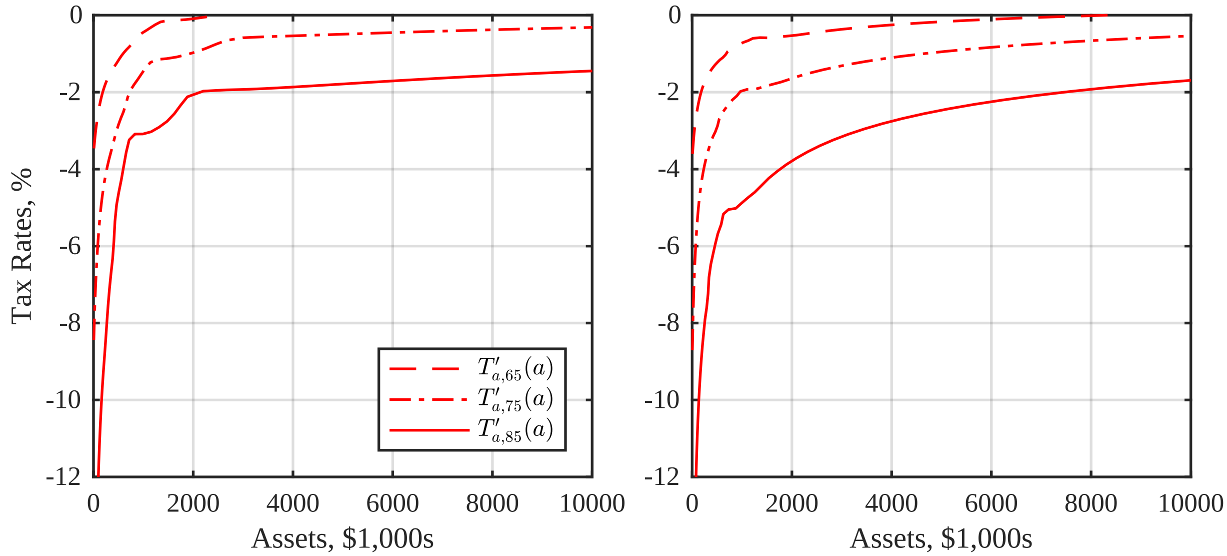

Figure 37: Optimal asset tax functions in the model with a progressive (status quo) asset tax. The left panel is optimal marginal taxes in the model with a progressive (status quo) asset tax. The right panel shows the same for the benchmark model.

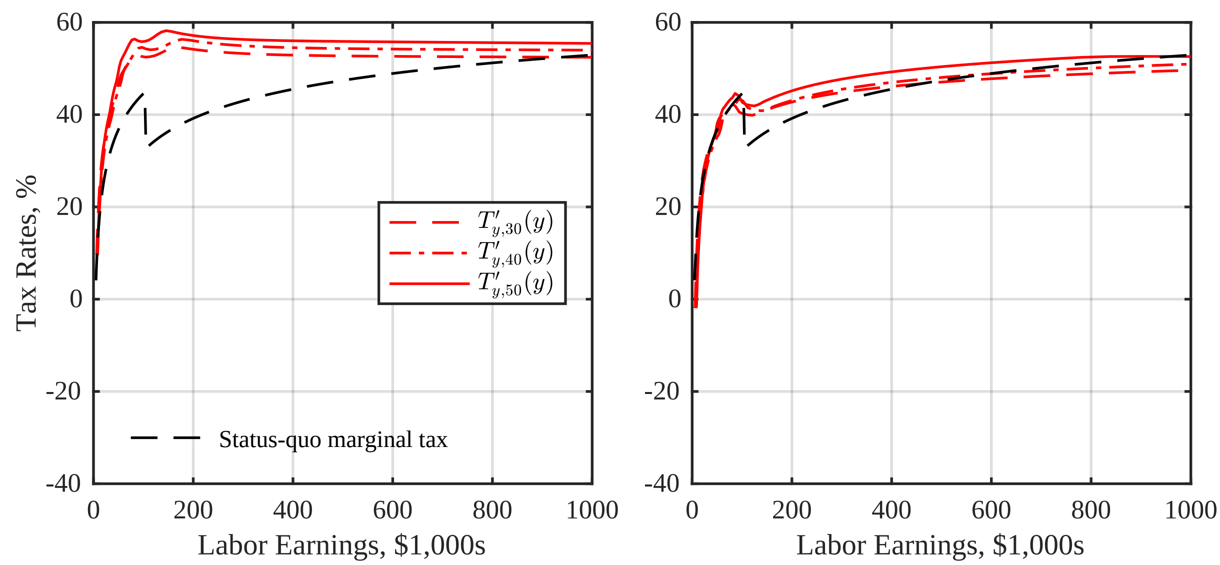

Figure 38: Optimal labor income tax functions in a model with a progressive (status quo) asset tax. The left panel is optimal marginal taxes in the model with a progressive (status quo) asset tax. The right panel shows the same for the benchmark model.

Table 18: Aggregate effects of optimal policies—with progressive (status quo) asset tax

	Current U.S.	Continue	Optimal Reform		
	(1)	(2)	(3)	(4)	(5)
Factor prices					
Interest rate (%)	5.58	4.63	5.58	4.44	3.39
Wage	1	1.16	1	1.10	1.15
Values relative to GDP					
Consumption	0.69	0.69	0.61	0.66	0.66
Capital	4.00	4.42	5.32	4.51	4.84
Tax revenue (total)	0.31	0.34	0.21	0.24	0.25
Earnings tax	0.20	0.20	0.15	0.18	0.16
Consumption tax	0.04	0.07	0.03	0.04	0.07
Capital (corporate) tax	0.08	0.07	0.03	0.02	0.02
Transfers	0.15	0.19	0.12	0.12	0.11
To retirees	0.08	0.12	0.03	0.02	0.03
To workers	0.07	0.06	0.05	0.03	0.03
Asset subsidies	0.00	0.00	0.07	0.07	0.06
Change (%) (relative to status quo)					
GDP	–	-1.99	25.44	7.27	3.77
Consumption	–	-2.29	11.49	2.65	0.11
Capital	–	8.20	66.80	20.97	25.44
Labor input	–	-10.32	-6.43	-2.21	-10.32
PDV of net resources	–	–	-20.48	-54.77	-31.20
Consumption equivalence			3.61	9.65	7.45

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 the 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 PDV is calculated relative to column (1). In column (5), the percentage change in PDV is calculated relative to column (2).

E Consumption Profiles

To demonstrate where the efficiency gains (cost savings) come from, we plot the sample consumption and earnings profiles in Figure 39. The left panel shows the 25th, 50th and 75th percentile of consumption at each age under both status quo policies and optimal reform policies. The right panel displays the same moments for earnings (before taxes and transfers) under both policies.

Note first that the status quo consumption profiles are hump shaped. This is consistent with evidence on consumption over lifecycle as documented by Gourinchas and Parker (2002) and Fernández-Villaverde and Krueger (2007) (among others). However, these profiles are steeper and peak later in life relative to the estimated profiles in CEX data. This is mainly because matching the capital-output ratio in our model requires a higher value for the discount factor. This leads to steeper profiles and delaying the peak, especially for higher-ability groups (who also have lower mortality).

Consumption under the optimal reform policies closely follows the status quo consumption at younger ages. It is only at older ages that the two values of consumption differ. This is when mortality becomes large (particularly for poorer types) and providing annuitization is valuable. Note that this increase in consumption in old ages is compensated by a reduction in consumption when individuals are young. When evaluating the present discounted value of consumption for a cohort, this young-age consumption is less discounted. Therefore, this reduction in consumption in young age is the main reason the consumption cost of delivering the status quo welfare goes down. Moreover, as we see in the right panel of Figure 39, there is little change in earnings. Therefore, the present value of consumption next to labor income falls.

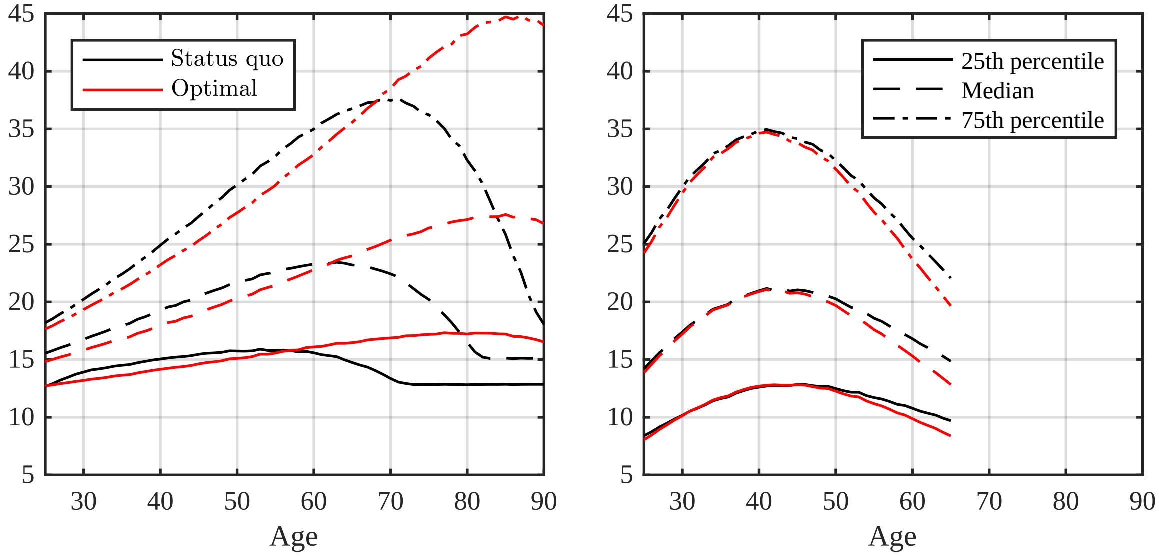

Figure 39: The left panel is the 25th, 50th and 75th percentile of consumption at each age under the status quo (black) and optimal reform policies (red). The right panel shows the same moments for earnings (before tax and transfers).

F Model with Bequest Motives

In this section, we describe the details of the model with bequest motives. [65]

$$P_j(\theta) = \Pi_{s=0}^{j} p_s(\theta).$$

Individuals have preferences over consumption and leisure, a joy-of-giving bequests motive and a time (and state) separable utility function

$$\sum_{j=0}^{J} \beta(\theta)^j P_j(\theta) \left[u(c_j(\theta)) - v(l_j(\theta)) + \beta(1 - p_{j+1}(\theta)) w(b_{j+1}(\theta))\right], \tag{32}$$

where $\beta(\theta)$ is the subjective discount factor, $P_j(\theta)$ is probability of survival to age t and $1 - p_{j+1}(\theta)$ is mortality rate at the end of age j. An individual who is alive at age j enjoys leaving bequest b_{j+1} if he or she dies at the end of period j. As in the main model we assume there is no market for survival contingent assets (i.e., annuities and life insurance policies). Therefore, there are both voluntary and accidental bequests in the model.

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 government uses the revenue from taxation 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 an individual's lifetime income. The difference with the benchmark model is that we allow for the government to tax the assets of the deceased in a distortionary fashion. We call this a bequest tax and denote it by $T_{b,j}^s(\cdot)$ to distinguish it from the asset tax or subsidy that an individual pays or receives upon survival $T_{a,j}^s(\cdot)$. The superscript s stands for status quo.

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

$$\begin{aligned} (1+\tau_c) c_j + a_{j+1} &= \left(w\varphi_j l_j - T_{y,j}^s(w\varphi_j l_j) + Tr_j^s\right) \mathbf{1}[t < R] \\ &\quad + (1+r) a_j - T_{a,j}^s((1+r) a_j) + S_j^s(Y_j) \mathbf{1}[j \geq R] + B_j, \qquad (33) \\ b_{j+1} &= (1+r) a_{j+1} - T_{b,j}^s((1+r) a_{j+1}), \qquad (34) \end{aligned}$$

where $T_{y,j}^s(\cdot)$ is the income tax function on earnings from labor, Tr_j^s are transfers to working individuals, S_j^s is the retirement benefit from the government. We assume that bequests are collected and distributed as a lump-sum transfer B to the entire population. The dependence of

[65] Extending this setup to one that includes both bequest motives and exogenous out-of-pocket medical expenditure is straightforward.

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

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

The rest of the model (government budget constraint, equilibrium conditions, etc.) is identical to that in section 3. Note that even though we allow for bequest motives, we abstract from direct intergenerational transfers. Allowing for it introduces analytical and computational complications to the optimal policy exercise, and is outside the scope of this paper.

F.1 Optimal Policies

The set of policies that we allow for in our optimal reform are very similar to those described in the main text. The only addition is that we allow for a nonlinear tax on bequest.

Planning Problem. Our planning problem maximizes the revenue from delivering a steady-state allocation subject to the implementability constraint (37) and a minimum utility requirement given by

$$\max \int \sum_{j=0}^{J} \frac{P_t(\theta)}{(1+r)^j} \left[w\varphi_j(\theta)\, l_j(\theta) - c_j(\theta) - \frac{1-p_{j+1}(\theta)}{1+r} b_{j+1}(\theta) \right] dH(\theta) \tag{35}$$

subject to

$$\begin{aligned}
U'(\theta) \;&=\; \sum_{j=0}^{J} \beta(\theta)^j P_j(\theta) \frac{\varphi_j'(\theta)}{\varphi_j(\theta)} \psi l_j(\theta)^\gamma & (36)\\
&\quad + \sum_{j=0}^{J} \beta(\theta)^j P_j(\theta) \left(\frac{P_j'(\theta)}{P_j(\theta)} + j \frac{\beta'(\theta)}{\beta(\theta)} \right) \left[u(c_j(\theta)) - \psi \frac{l_j(\theta)^\gamma}{\gamma} \right] & (37)\\
&\quad + \sum_{j=0}^{J} \beta(\theta)^{j+1} \left[P_j(\theta) - P_{j+1}(\theta) \right] \left(\frac{P_j'(\theta)}{P_j(\theta)} - \frac{p_{j+1}'(\theta)}{1-p_{j+1}(\theta)} + (j+1) \frac{\beta'(\theta)}{\beta(\theta)} \right) w\left(b_{j+1}(\theta)\right) & (38)\\
U(\theta) \;&\geq\; W_s(\theta) . & (39)
\end{aligned}$$

The above optimal allocations can be used to construct optimal policies. However, the mapping from allocations to policies is less straightforward. This is because in our implementation markets are incomplete. Nevertheless, we are able to construct the tax schedules, as described in (33) and (34), for this incomplete market economy in the following lemma.

: (in what follows we adopt the following notation to avoid clutter: $u_{c,j}(\theta) \equiv u'(c_j(\theta))$, $v_{l,j}(\theta) \equiv v'(l_j(\theta))$ and $w_{b,j}(\theta) \equiv w'(b_j(\theta))$.)

Lemma 7. *Consider an allocation* $\{c_j(\theta), l_j(\theta), b_j(\theta)\}$ *that satisfies the implementability constraint (37) such that* $b'_j(\theta) > 0$, $(\varphi_j(\theta) l_j(\theta))' > 0$ *and*

$$\sum_{s=j}^{j} \beta^s P_s(\theta) \left[u_{c,s}(\theta) c'_s(\theta) + \beta(1 - p_{s+1}(\theta)) w_{b,s+1}(\theta) b'_{s+1}(\theta) - v_{l,s}(\theta)(\varphi_s(\theta) l_s(\theta))'\right] > 0.$$

Then tax and transfer functions $T_{a,j}(\cdot), T_{b,j}(\cdot), T_{y,j}(\cdot), S_j$ *together with asset holdings* $a_j(\theta)$ *exist so that the allocations* $\{c_j(\theta), l_j(\theta), b_j(\theta), a_j(\theta)\}$ *satisfy the budget constraints (33) and (34) and the first order conditions associated with the individual optimization.*

Proof. We start by writing the first order conditions for the maximization problem above for an individual of type θ

$$1 - T'_{y,j}(\varphi_j(\theta) l_j(\theta)) = \frac{v'(l_j(\theta))}{\varphi_j(\theta) u_{c,j}(\theta)}, \tag{40}$$

$$u_{cj} = \beta(1+r)\left[p_{j+1}\left(1 - T'_{a,j+1}\right) u_{cj+1} + (1 - p_{j+1})\left(1 - T'_{b,j+1}\right) w_{b,j+1}\right]. \tag{41}$$

Equation (40) is the individual intratemporal optimality condition, and equation (41) is the individual Euler equation.

We can use equation (40) to back out the optimal marginal taxes on labor earnings at each age. This is possible because the efficient allocations of consumption and hours come directly from solving the planning problem. Thus, the earnings taxes can simply be defined by integrating over the implied marginal rate in (40). This is well-defined since output in each age is increasing in θ.

The calculation of optimal asset taxes is not straightforward. More importantly, the level of assets a cannot be pinned down independent from the marginal taxes $T'_{a,j+1}$ and $T'_{b,j+1}$. Therefore, we assume that asset holdings of the lowest type are 0 for all ages. This implies that in the equilibrium which decentralizes an incentive compatible allocation, the poorest individual is hand-to-mouth in all ages. Given this restriction we can use the following procedure to find the optimal asset taxes.

We can combine equations (40) and (41) together with (33) and (34) and use extensive algebra to show that the derivative of asset holdings with respect to θ, a'_j, satisfies

$$\begin{aligned} a'_j(\theta) \quad = \quad & \frac{1}{u_{c,j}(\theta)} \sum_{s=j}^{T} \beta^{s-j} \frac{P_s(\theta)}{P_j(\theta)} \left[u_{c,s}(\theta) c'_s(\theta) + \beta(1 - p_{s+1}(\theta)) w_{b,s+1}(\theta) b'_{s+1}(\theta)\right. \\ & \left. - v_{l,s}(\theta)(\varphi_s(\theta) l_s(\theta))'\right]. \end{aligned}$$

Since by assumption $a_j(\underline{\theta}) = 0$, the above determines the level of asset holdings at each age and for each type. Additionally, taxes on bequests must satisfy

$$b_j(\theta) = (1+r)\,a_j(\theta) - T_{b,j}\left((1+r)\,a_j(\theta)\right). \tag{42}$$

Since $a_j(\theta)$ and $b_j(\theta)$ are determined in the optimal allocation, the above formula determines the bequests taxes.

Finally, using (42) and the Euler equation (41), we must have

$$1 - T'_{a,j+1} = \frac{u_{cj}}{\beta(1+r)\,p_{j+1}u_{cj+1}} - \frac{1-p_{j+1}}{p_{j+1}}\frac{w_{b,j+1}}{u_{cj+1}}\frac{b'_{j+1}}{(1+r)\,a'_{j+1}}. \tag{43}$$

The above formula determines the marginal tax rate on asset holdings, and since $a'_j > 0$, a well-defined tax function on asset holdings must exist. This completes the construction. □

Unfortunately, we cannot derive a closed-form formula for optimal taxes. However, our implementation procedure provides a guideline on how to numerically compute the optimal tax functions.

Note that monotonicity constraints in Lemma 7 are necessary for the existence of the tax function. While we have no way of theoretically checking that they are satisfied, our numerical simulations always involve a check that ensures that they are. Needless to say, in all our simulations, the monotonicity constraint is satisfied.

Finally, it is worth noting that in the model with bequests, the degree of market incompleteness in the presence of risk-free assets depends on the strength of the bequest motive. In particular, when individuals put a high valuation on bequests relative to assets upon survival, a risk-free asset comes very close to implementing efficient allocations. As a result, the strength of the subsidy depends on the strength of the bequest motive. In general, we can use a derivation similar to that of (4) and write (43) as

$$\begin{aligned}T'_{a,j+1} =& 1 - \frac{1}{p_{j+1}} + \frac{1}{p_{j+1}}\left(\frac{p'_{j+1}(\theta)}{p_{j+1}(\theta)} + \frac{\beta'(\theta)}{\beta(\theta)}\right)\frac{\tau_{l,j}(\theta)}{1-\tau_{l,j}(\theta)}\frac{\varphi_j(\theta)}{\varphi'_j(\theta)}\left(\frac{1+\varepsilon_{F,j}(\theta)}{\varepsilon_{F,j}(\theta)}\right)\\ &+ \frac{1-p_{j+1}}{p_{j+1}}\frac{w_{b,j+1}}{u_{cj+1}}\frac{b'_{j+1}}{(1+r)\,a'_{j+1}}\end{aligned}$$

The above formula highlights the role of bequests in affecting the optimal savings taxes. The first term is the standard terms associated with market incompleteness. The second term is the redistributive motives identified in section 2. The last term is related to the strength of the bequest motive. For example, when there is no mortality heterogeneity and bequest motives are sufficiently strong, this term becomes close to $\frac{1-p_{j+1}}{p_{j+1}}$, which then cancels out the market incompleteness effect.

Table 19: Parameters calibrated using the model with bequest motive

Parameters	Description	Values
β_0	discount factor: level	0.985
ψ	weight on leisure	0.037
χ	strength of bequest motive	50
$\bar{B}$	bequest utility shifter	600

Targeted moments	Data	Model
Capital-output ratio	4.00	4.00
Average annual hours	2000	2000
Bequest-wealth ratio	0.0118	0.013
Fraction who leave no bequest	0.25	0.23

When bequests are luxury goods—see our calibration of the function $w(b)$ below—it is efficient for low-income individuals not to leave any bequests. In this case, $b'_{j+1}(\theta) = 0$, and the above becomes identical to (4).

F.2 Calibration

Calibration of the model with bequests follows the baseline calibration whenever possible. For status quo policies, we assume that there are no bequest taxes in the status quo model. [66] All other policies are the same as in the main model.

Bequest motives are captured by the following utility function

$$w(b) = \chi \frac{\left(b + \bar{B}\right)^{1-\sigma}}{\sigma}.$$

Parameter χ determines the strength of the bequest motive, while $\bar{B}$ reflects the extent to which bequests are luxury goods. If $\bar{B} > 0$, the marginal utility of bequests is bounded. At the same time, the marginal utility of large bequests declines more slowly than the marginal utility of consumption. As a result, richer individuals have stronger motives to leave bequests. [67] We follow De Nardi (2004) and choose the value for parameter $\bar{B}$ to match the fraction of the deceased individuals who leave no bequest. We assume the risk aversion parameter $\sigma = 2$ (the same as we did for the curvature of the utility function in section 7.3.1). The strength of bequest χ is chosen to match the bequest to wealth ratio of 0.0118, as reported in Gale and Scholz (1994). To calibrate $\bar{B}$, we use data on the distribution of bequests reported in Hurd and Smith (2002). We choose

[66] Bequest and estate tax affect only a small portion of the richest U.S. tax payers.

[67] The wealth elasticity of realized and anticipated bequests has been estimated to be about 1.3 (see Auten and Joulfaian (1996) and Hurd and Smith (2002)). Among single Americans who were at least 70 years old in 1993 and died before 1995, the 30th percentile of the bequest distribution was just \$2000, the median was \$42000, and the mean was \$82000 (Hurd and Smith (2002)).

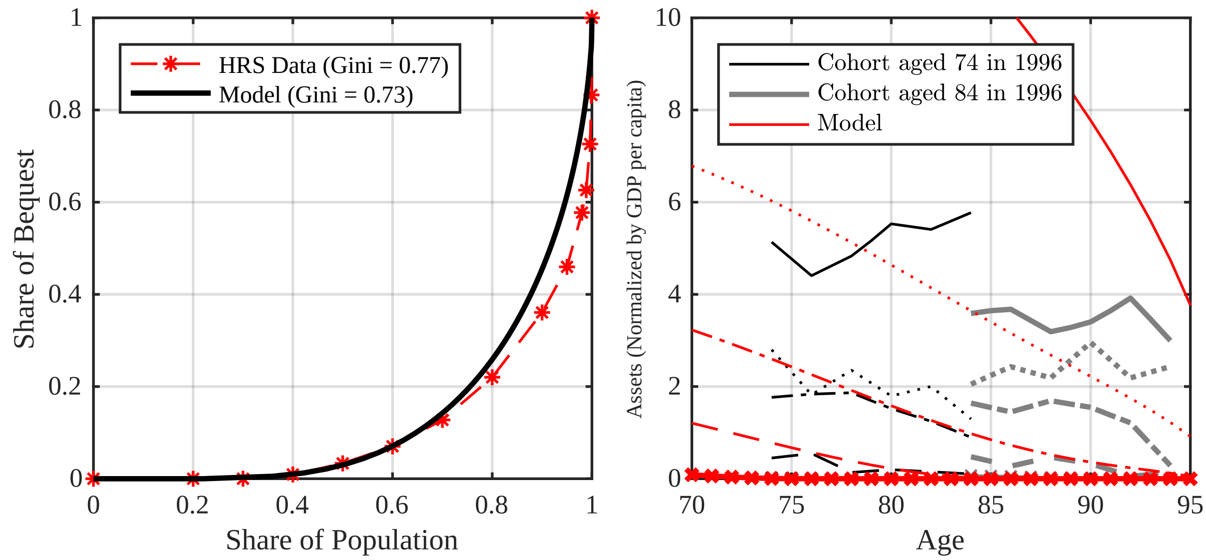

Figure 40: The left panel shows the distribution of bequests left by the deceased in the model (solid line) vs. in the data (dashed line). The right panel shows median assets by permanent income quintile in the model (solid line) and in the data (dashed line). Data source: De Nardi et al. (2010) calculations for AHEAD cohorts who were 74 and 84 years old in 1996.

this parameter so that in the model 25 percent leave no bequest. As in the case with medical expenditure, we hold parameter β_1 (gradient of discount factor with respect to ability type θ) at the benchmark level. The calibrated parameters are reported in Table 19.

F.3 Results

In Figure 40, we show how well the model captures the pattern of dissaving as well as the distribution of bequests. The left panel in Figure 40 shows the distribution of bequests in the model and in the HRS data as reported in Hurd and Smith (2002). The right panel shows the average assets by permanent income quintile in the model and in the data. The model does a reasonable job at capturing dispersion in bequests. It also captures the pattern of asset deccumulation, except perhaps at the highest income quintile.

Figure 19 reports the optimal asset taxes in the optimal reform. As mentioned before, since leaving a bequest is an active decision by individuals, an optimal policy requires the introduction of a new instrument, i.e., a tax on the assets of the deceased. In other words, an optimal policy has two components of asset taxation. As before, there is a subsidy on the assets of individual who survive. The idea behind this policy is the same as the one described throughout the paper. We plot these asset subsidies in the left panel of Figure 19. The new policy, tax on bequest, accomplishes two tasks. On one hand, it deters poor individuals from leaving bequests. Since

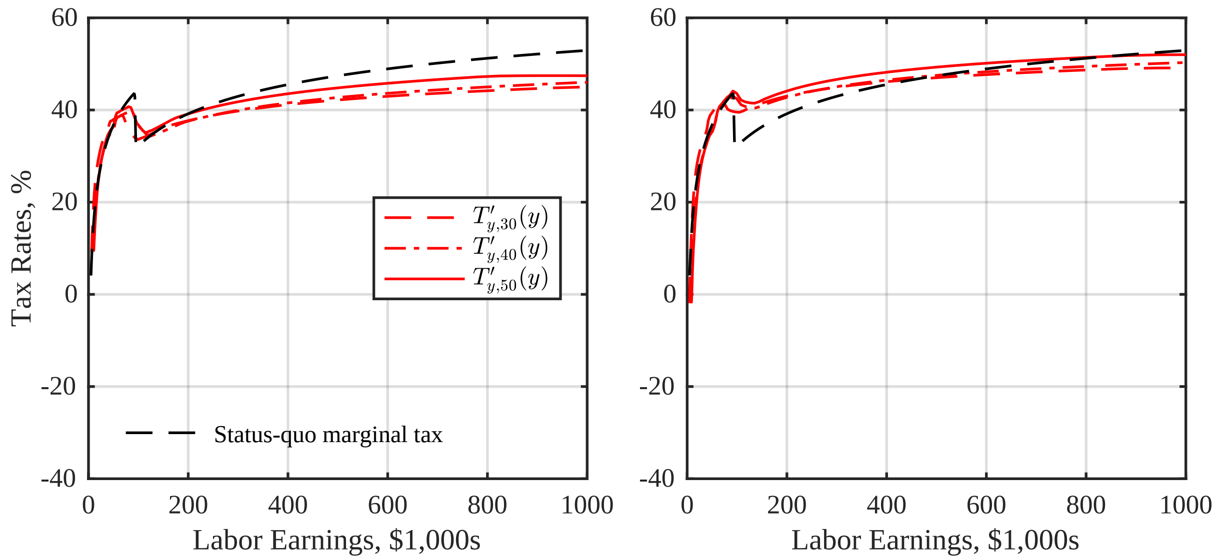

Figure 41: Optimal labor income tax functions with out-of-pocket medical expenditures and bequest motives. The left panel is optimal marginal taxes with out-of-pocket medical expenditure and bequest motives. The right panel shows the same in the benchmark model. The black dashed line is the effective status quo tax schedule.

a bequest is a luxury good, delivering utils to poor individuals via bequests is not efficient. As the right panel of Figure 19 demonstrates, the bequest is taxed at 100 percent for asset-poor individuals (they instead receive utils via high asset subsidies and therefore they have higher consumption while they are alive). Also, for those who are rich enough, leaving bequests is valuable and therefore there is a sharp drop in bequest taxes. For these individuals, bequest taxes spread the dead-weight-loss of taxation between the state of survival and death.

Figure 41 shows the optimal earnings taxes. There is no significant difference relative to the previous analysis here. The earnings tax reform continues to be a rather inessential part of the reform.

Finally, Table 20 reports the aggregate effect of an optimal policy when implemented under the current U.S. demographics. It is important to highlight the large efficiency gains resulting from these policies. Under optimal policies there is a massive efficiency gain from reducing the bequest left by low-income individuals. This reduces the cost of delivering the status quo welfare to the current cohort. The mechanism is the following. Due to market incompleteness (lack of survival contingent assets) poor individuals leave too much bequest. However, this bequest contributes to their lifetime welfare, although at a high cost. Under Pareto optimal reform policies, low-income individuals leave no bequest due to 100 percent bequest tax. Instead they receive status quo welfare entirely through consumption while they are alive (i.e., asset subsidies). This leads to an efficient delivery of status quo welfare and contributes to the large efficiency gains reported in Table 20.

Table 20: Aggregate effects of optimal policies—with out-of-pocket medical expenditures and bequest motives

	Current U.S.	Optimal reform	
	(1)	(2)	(3)
Factor prices			
Interest rate (%)	4.05	4.05	3.76
Wage	1.00	1.00	1.03
Values relative to GDP			
Consumption	0.69	0.67	0.68
Capital	4.00	4.29	4.17
Tax revenue (total)	0.24	0.33	0.33
Earnings tax	0.12	0.13	0.14
Consumption tax	0.04	0.04	0.04
Bequest tax	0.00	0.07	0.07
Capital (corporate) tax	0.08	0.09	0.08
Transfers	0.15	0.24	0.24
To retirees	0.08	0.06	0.06
To workers	0.07	0.11	0.11
Asset subsidies	0.00	0.07	0.07
Change (%) (relative to status quo)			
GDP	–	3.93	2.25
Consumption	–	1.44	0.83
Capital	–	11.28	6.47
Labor input	–	-2.30	-1.348
PDV of net resources	–	-276.88	-389.29
Consumption equivalence		4.32	6.09

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 the 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 PDV is calculated relative to column (1). In column (5), the percentage change in PDV is calculated relative to column (2).

G Calibration: Calculating U.S. Aggregates

In this section we describe the calculation of capital income share, investment expenditure as share of GDP, government expenditure as share of GDP, debt as share of GDP, and capital to output ratio. We use our calculated values as calibration targets. The main data sources for our calculations are the U.S. National Income and Production Account, the Fixed Asset Tables (compiled by the Bureau of Economic Analysis) and several balance sheet items from the Flow of Funds of the United States (compiled by the Federal Reserve Board of Governors). All the collocations are done using data from year 2000 to 2010. [68]

Income Categories

We closely follow McGrattan and Prescott (2017) and adjust NIPA's measure of income to conform to our theoretical model. These adjustments are mainly subtraction of sales taxes and addition of imputed capital service of consumer durables and government capital. Our discussion here is brief. For details, refer to McGrattan and Prescott (2017).

We start with income data from Table 1.10 in the NIPA data. We categorize the compensation of employees and 70 percent of proprietors' income as labor income. The rest of the income is categorized as capital income after the following adjustments. First, we subtract taxes other than property tax from NIPA's measure of "taxes on production and imports". Second, we impute capital services for consumer durables, which we treat as investment, and government capital. The imputed services are estimated to be 4 percent times the current-cost net stock of consumer durable goods and government fixed assets (both of these stocks are reported in BEA's Fixed Asset Tables). In addition, we include depreciation of consumer durables from the Flow of Funds accounts. After these adjustments, the capital income is the sum of corporate profits, 30 percent of proprietors' income, surplus of government enterprises, rents, net income, property taxes, depreciation of capital, and imputed capital services. This sum amounts to 43.5 percent of the adjusted GDP on average between 2000 and 2010. We use this figure as our target for the capital share of income. Table 21 shows the breakdown of income and its components relative to GDP. All numbers are averages between 2000 and 2010.

Expenditure Categories

We divide expenditures into three categories: government spending, investment and consumption. Table 22 shows each expenditure category relative to GDP. We define "government spending" as the sum of defense expenditure (both consumption and gross investment), general public

[68] All calculations are based on the January 26, 2018 release of the data.

service, and public order and safety. These add up to about 8 percent of GDP. The rest of government expenditures is either treated as investment or consumption expenditures. The "investment" category consists of NIPA's gross private domestic investment, net export, income from the rest of the world, consumer durable goods (net of imputed sales tax), and government non-defense investment expenditures. Therefore, investment relative to GDP is 23.2 percent. "Consumption" expenditure consists of NIPA's consumption of non-durables and services, imputed capital services of consumer durables and government capital, consumer durable depreciations, and other government consumption expenditures which are included in NIPA but are not included in our measure of "government spending". These are transportation and other economic affairs, housing and community services, health, recreation and culture, education and welfare, which are mostly services and/or transfers that are close substitutes to private consumption. We assume that these are effectively lump-sum transfers to households.[69]

Physical Capital

We present two different approaches to measure the stock of physical capital, one based on NIPA and Fixed Asset Tables and one based on the Flow of Funds accounts. Our first approach closely follows McGrattan and Prescott (2017). We define physical capital as the sum of fixed and private assets, stock of consumer durables, stock of inventories and land. This approach yields a measure of physical capital that is about 4.07 times GDP (average between 2000 and 2010). The top panel of Table 23 lays out the detailed calculations with sources for each subcategory.

Alternatively, we can measure the stock of physical capital as the total sum of all non-financial assets in the Flow of Funds accounts. These include household and nonprofits, non-financial corporates, non-financial non-corporates and government. This approach results in a stock of physical capital that is 3.97 times GDP. For our calibration we use a capital to output ratio of 4 as our target, which is a round number that is close to both these measures.

Stock of National Debt

Our measure of government debt includes state and local municipal securities, federal treasury securities, and federal budget securities. To account for the fact that a portion of this debt is held by government agencies, we subtract government debt held by the Social Security Administration. This results in a debt to adjusted GDP ratio of 0.47. See Table 24 for details.

[69] As McGrattan and Prescott (2017) point out, this is consistent with the accounting of the World Bank (2014) that assumes "actual individual consumption comprises all the goods and services that households consume to meet their individual needs . . . whether they are purchased by households or are provided by general government and nonprofit institutions service households" (p. 9).

Table 21: Income categories relative to GDP, 2000–2010

Total income	1.000
Labor income	0.565
Compensation of employees (NIPA 1.10)	0.516
Wages and salary accruals (NIPA 1.10)	0.418
Supplements to wages and salaries (NIPA 1.10)	0.098
70% of proprietors' income with IVA, CCadj (NIPA 1.10)	0.048
Capital income	0.435
Corporate profits with IVA and CCadj (NIPA 1.10)	0.072
30% of proprietors' income with IVA, CCadj (NIPA 1.10)	0.021
Rental income of persons with CCadj (NIPA 1.10)	0.018
Surplus on government enterprises (NIPA 1.10)	-0.000
Net interest and misc. payments, domestic industries (NIPA 1.10)	0.051
Indirect business taxes	0.069
Taxes on production and imports (NIPA 1.10)	0.066
Less: Subsidies (NIPA 1.10)	0.004
Business current transfer payments (NIPA 1.10)	0.007
Less: Sales tax	0.041
Federal excise taxes (NIPA 3.5)	0.005
Federal customs duties (NIPA 3.5)	0.002
State and local sales taxes (NIPA 3.5)	0.028
Motor vehicle licenses (NIPA 3.5)	0.001
Severance taxes (NIPA 3.5)	0.001
Special assessments (NIPA 3.5)	0.000
Other taxes on production and imports (NIPA 3.5)	0.003
Net income, rest of world (NIPA 1.13)	0.007
Consumption of fixed capital (NIPA 1.10)	0.144
Consumer durable depreciation (FOF F.10)	0.058
Imputed capital services	0.038
Consumer durable services	0.012
Government capital services	0.026
Statistical discrepancy (NIPA 1.10)	-0.003

Note: Data sources are in parenthesis. IVA, inventory valuation adjustment; CCadj, capital consumption adjustment; NIPA, national income and product accounts; FoF, the flow of funds of the United States.

Table 22: Expenditure categories relative to GDP, 2000–2010 (Cont.)

Total expenditure	1.000
Consumption	0.688
Personal consumption expenditures (NIPA 1.1.5)	0.632
Less: Consumer durable goods (NIPA 1.1.5)	0.078
Less: Imputed sales tax, nondurables and services	0.036
Plus: Imputed capital services, durables	0.012
Government consumption expenditures, other (NIPA 3.9.5)	0.064
Plus: Imputed capital services, government capital	0.026
Consumer durable depreciation (FOF F.10)	0.058
Government spending	0.080
Government consumption expenditures, national defense (NIPA 3.9.5)	0.034
Government gross investment, national defense (NIPA 3.9.5)	0.009
General public service (NIPA 3.9.5)	0.016
Public order and safety (NIPA 3.9.5)	0.021
Investment	0.232
Gross private domestic investment (NIPA 1.1.5)	0.164
Consumer durable goods (NIPA 1.1.5)	0.078
Less: Imputed sales tax, durables	0.005
Government gross investment, non-defense (NIPA 3.9.5)	0.029
Net exports of goods and services (NIPA 1.1.5)	-0.040
Net income rest of world (NIPA 1.13)	0.007

Note: Data sources are in parenthesis. NIPA, national income and product accounts; FoF, the flow of funds of the United States.

Table 23: Stock of physical capital (top panel) and non-financial assets (bottom panel), averages relative to GDP, 2000–2010

Physical capital	4.072
Fixed private assets (FA 1.1)	2.123
Fixed government assets (FA 1.1)	0.637
Consumer durables (FA 1.1)	0.294
Inventories (NIPA 5.8.5B)	0.130
Land	0.887
Households and non-profits (FOF B.101)	0.519
Non-financial corporate (FOF B.103)	0.103
Non-financial non-corporate (FOF B.104)	0.265
Stock of non-financial assets	3.974
Households and non-profits (FOF B.101)	1.727
Non-financial corporate (FOF B.103)	0.989
Non-financial non-corporate (FOF B.104)	0.620
Government (FOF B.1)	0.637

Note: Data sources are in parenthesis. FA, fixed asset tables; NIPA, national income and product accounts; FoF, the flow of funds of the United States.

Table 24: Stock of government debt relative to GDP, 2000–2010

Government consolidated debt	0.467
State and local municipal securities (FOF L.105)	0.165
Federal Treasury securities (FOF L.105)	0.434
Federal budget agency securities (FOF L.105)	0.002
Less: Government debt held by SSA	0.134

Note: Data sources are in parenthesis. NIPA, national income and product accounts; FoF, the flow of funds of the United States; SSA, Social Security Administration.