Retirement Financing: An Optimal Reform Approach

Hosseini, Roozbeh and Shourideh, Ali

University of Georgia, Carnegie Mellon University

19 January 2016

Retirement Financing: An Optimal Reform Approach*

Roozbeh Hosseini
University of Georgia
roozbeh@uga.edu

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

February 7, 2017

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 earnings tax reforms cannot yield efficiency gains. On the other hand, progressive asset subsidies provide a powerful tool for Pareto optimal reforms. 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 5 percent in the steady state. These gains amount to a one-time lump-sum transfer to the initial generation equal to 9 percent of the GDP.

Keywords: retirement, optimal taxation, social security
JEL classification: H21, H55, E62

*We would like to thank Laurence Ales, Tony Braun, Berthold Herrendorf, Karen Kopeky, Dirk Krueger, Chris Sleet, Chris Telmer, Venky Venkateswaran, Shu Lin Wee, Sevin Yeltekin, Ariel Zetlin-Jones, and conference participants at Barcelona GSE workshop on Macro and Aging, Third Annual Macroeconomics and Business CYCLE Conference, SED 2016 and QSPS 2016, as well as seminar participants at CBO, Emory, FRB Atlanta, University of Calgary, University of Connecticut, University of Colorado-Boulder, University of Georgia, Virginia Commonwealth University and Washington State University. All errors are ours.
Please visit http://www.roozbeh-hosseini.com/ for updates.

1 Introduction

The government in the United States and 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. 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 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 vast 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 motivation for this assumption is the observed heterogeneity in savings rates across income

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

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 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 criteria for optimality for each generation. A tax system is optimal if it satisfies two sets of criteria, one for the earnings tax schedule and one for asset taxes. An earnings tax schedule meets the optimality criteria if and only if it satisfies a set of inequality conditions (one at each age). These inequalities relate the shape of the earnings tax schedule to the income distribution, elasticity of labor supply and distribution of consumption in the economy. They imply that earnings taxes are more likely to be inefficient (i.e., fail to satisfy the inequalities) when labor supply elasticity is high or earnings taxes are regressive (for some income range). The key idea is that if these inequalities are violated, it is possible to reduce the tax rate for some income groups without reducing tax revenue collections.[4] In particular, when the elasticity of labor supply is high or the tax schedule is regressive, a decline in the marginal tax rate for a small interval of the income distribution can lead to an increase in government revenue, which in turn can be used to improve everyone's welfare.

Optimality of the asset tax schedule is tied to the variability of mortality and discount factor

[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 various possible explanations.

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

[4]This is similar to a Laffer effect: when taxes are linear, if the marginal tax rate is above the peak of the Laffer curve, a decline in the tax rate could lead to an increase in government revenue.

across ability types, as well as the incompleteness in the annuity markets. In particular, 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 distribution of earnings and wealth. Our model can successfully match 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, our quantitative analysis shows that 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 levels of earning. On the other hand, our test shows 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 are a source of gains as opposed to earnings tax reforms.

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 and 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 15 percent. This is equivalent to a 2.21 percent reduction in the status quo consumption of all individuals, keeping their welfare unchanged.[5]

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 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 5 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 9 percent of the 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.

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.), as well as 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 the GDP). Moreover, these programs mostly benefit higher-income individuals.[6] 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 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

[5] 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.

[6] See Woo and Buchholz (2006).

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 (forthcoming) propose phasing out social security and Medicare benefits and removing payroll taxes. Blandin (2015) 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 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.[7]

Our paper also contributes to the literature on dynamic optimal taxation over the life cycle. Similar to Weinzierl (2011), Golosov et al. (forthcoming) 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

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

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 (2015), who focus on an optimal tax system that provides incentive for an efficient retirement age.

Finally, our paper is related to the literature on 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 precisely point 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.

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 techniques in Werning (2007) to an 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 mortality, θ is labor productivity, $u(\cdot)$ is strictly concave and $v(\cdot)$ is strictly convex.[8]

[8] We could also define utility over actual hours worked $l = y/\theta$. We sometimes adopt this notation in the main body of the paper.

Production is done using labor and capital, with the production function given by a $F(K, L)$, 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.

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

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, even though they could be heterogeneous in their survival probability. This covers two possible scenarios: first, annuities are non-existent or $q_t = 1$; second, annuities exist but the market suffers from adverse selection, in which case q_t is the average probability of survival weighted by total annuity purchases by the young individuals. These assumptions are consistent with the observation that annuity markets in the United States are very small. 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, as well as 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:

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

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

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* $t = 0, 1, \cdots$

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

subject to

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

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

Proof can be found in the appendix.

The above proposition is an extension of the results in Diamond (1965) to an environment with heterogeneity. It states that when the economy is dynamically efficient, $F_{K,t} > n$, then the possibility of a Pareto optimal reform depends on whether tax and transfer schemes exhibit inefficiencies within some generation. To the extent that dynamic efficiency seems to be the case in the data, the only possible Pareto optimal reforms can come from within-generation inefficiencies.[9] 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. We highlight the importance of this result by considering three examples of exercises commonly used in the literature.

Example 1. Suppose there is no heterogeneity in ability ($H(\theta)$ is degenerate) and all individuals survive to old age with probability 1 (therefore, there is no inefficiency associated with an

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

incomplete annuity market). Also, assume that taxes T_y and T_a are lump-sum taxes. This is the classic example of Diamond (1965). Any changes in lump-sum taxes only affect intergenerational transfers. In particular, removing these taxes causes the economy to converge to a new steady state with a higher stock of capital. All the individuals born in the steady state are better off. However, this comes at the cost of reducing the welfare of early generations. In fact, by Proposition 1, it is impossible to devise a Pareto improving transition, since there is no inefficiency that can be exploited.

Example 2. As before, consider an economy with degenerate θ but assume T_y is distortionary, $T_y = T_0 + \tau_y y$. In this case, a Pareto improving policy reform is feasible. For example, it is possible to reduce tax rate τ_y and adjust T_0 to ensure the status quo welfare is delivered. Removing inefficiencies due to distortionary taxes allows a policy to improve the welfare of all current and future generations. Many of the studies that show a Pareto improving transition from pay-as-you-go social security to a fully funded system are similar to this example (e.g., Conesa and Garriga (2008) and Birkeland and Prescott (2007)).

Example 3. Suppose there is heterogeneity in ability ($H(\theta)$ is not degenerate) and $P(\theta) = 1$ for all θ. Further, assume T_y is a nonlinear, increasing and smooth tax function, and T_a is lump-sum transfer. In other words, this is a heterogeneous-agent economy with nonlinear taxes and a pay-as-you-go social security. In this case, the only possible inefficiency can arise from inefficient nonlinear taxation. As is well-known from the public finance literature, the set of Pareto efficient tax functions is large.[10] This implies that distortionary taxes (payroll, earnings, etc.) cannot necessarily be removed, since they could satisfy the condition in Proposition 1.

We refer to the above proposition as the principle of no-free-lunch in policy reform. It implies that the possibility of a Pareto reform comes down to the efficiency of a tax and transfer schedule within each generation. To the extent that distortionary taxes (in the traditional sense of the word) are not necessarily inefficient, it is thus crucial to understand the determinants of an efficient tax schedule. The following proposition provides these conditions:

Proposition 2. *Suppose that $v(l) = \psi \frac{l^{1+1/\varepsilon}}{1+1/\varepsilon}$, where ε is the elasticity of labor supply. Then, a pair of policies $\tilde{T}_y$ and $\tilde{T}_a$ is efficient only if it satisfies the following relationships:*

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

$$\tilde{\tau}_{a,t}(\theta) = 1 - \frac{q_t}{P(\theta)} + \frac{q_t}{P(\theta)} \frac{\theta}{1+1/\varepsilon} \frac{\tilde{\tau}_{l,t}(\theta)}{1-\tilde{\tau}_{l,t}(\theta)} \left(\frac{\beta'(\theta)}{\beta(\theta)} + \frac{P'(\theta)}{P(\theta)} \right), \quad (2)$$

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

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

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

where the allocations are those induced by the policies.

In addition, if optimal allocations under the tax functions are fully characterized by an individual's first-order conditions, then the above are also sufficient for efficiency.

Proof can be found in the appendix.

The above proposition provides a test of Pareto optimality of a tax schedule. It extends the results in Werning (2007) for a static economy to our two-period OLG setup. First, consider the inequality (1). Intuitively, and as illustrated by Werning (2007), this inequality can be derived by a tax perturbation. In particular, consider a small decline in the marginal labor tax rate for an interval of the form $[y(\theta) - dy, y(\theta)]$ accompanied by an increase in the marginal tax rate of the same size for the interval $[y(\theta), y(\theta) + dy]$. Since such a tax perturbation improves everyone's utility, if a tax schedule is to be Pareto optimal, this perturbation must reduce government revenue.[11]

Consider the behavioral response of the individuals to such tax reform. The workers whose income is initially in $[y(\theta) - dy, y(\theta)]$ will increase their income, while the workers whose income is in the interval $[y(\theta), y(\theta) + dy]$ will decrease their income. The response of each set of workers depends on the elasticity of the labor supply and the curvature of the tax function. The resulting change in government revenue is negative when the increase in the bottom subinterval is smaller than the decrease in the top subinterval. Thus, the resulting change is more likely to be negative (1) the higher is the rate of change in the skill distribution, i.e., if there are more people in the interval $[y(\theta), y(\theta) + dy]$; (2) the higher is the slope of the marginal tax rate, which dampens the response of the bottom subinterval and increases the response of the individuals in the top subinterval; (3) the stronger is the income effect; and (4) the lower is the Frisch elasticity of labor supply. These forces can be identified in (1). An interesting observation is that when taxes become regressive, i.e., $\tau_l' < 0$, it is more likely that there is a Pareto improving reform.[12]

Equation (2) pertains to the optimality of asset taxes or intertemporal distortions induced by $\tilde{T}_a$. The formula has two components: a component that captures the inefficiencies arising from

[11]The perturbation we describe here is more suited for a static economy, since in our setup it results in changes in the agent's saving behavior and affects government revenue from taxes in the second period. While this means that the tax perturbation for the dynamic economy is more complicated, the intuition remains similar. We, thus, resort to the standard perturbation in Werning (2007), although our perturbation is slightly different in marginal tax rates as opposed to the tax level.

[12]As we will see in section 6.1, the main source of inefficiency in the earnings tax schedule results from the sudden drop of marginal tax rate around the social security maximum taxable earnings cap.

the incompleteness of annuity markets, $1 - q_t/P(\theta)$, and a component that captures the distortions to the annuity margin. The first component is standard, and it reflects the fact that in the absence of annuities, a subsidy to savings can provide annuity 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, taxation of second-period consumption can relax redistributive motives by the government, which in turn leads to lower taxes on earnings.[13]

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.

To summarize, our analysis in this section highlights the conditions required for the existence of a Pareto optimal reform. Proposition 1 states that in dynamically efficient economies, a policy is Pareto optimal if it is Pareto optimal within each generation. This implies that the problem of finding Pareto optimal policies can be solved separately across generations. Proposition 2 provides tests or conditions that any Pareto optimal policy must satisfy for each generation. In section 4.2, we provide an extension of these test to the case of a multi-period OLG economy. We then use this test in our quantitative model.

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 calculate the magnitudes of the amount 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 5 to match U.S. aggregate data and cross-sectional observations on earnings and asset distribution. In section 4, we show how this

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

model can be used to derive Pareto optimal policies.

3.1 Demographics, Preferences and Technology

Time is discrete and the economy is populated by $J+1$ overlapping generations. A cohort of individuals are born in each period $t = 0, 1, 2, \ldots$. The number of newborn 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 $h\left(\theta\right)$. This type 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 labor productivity of $\varphi_j(\theta)$ at age j. We assume that $\varphi_j'\left(\theta\right) > 0$ 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). 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

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

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

$$\sum_{j=0}^{J} \beta\left(\theta\right)^{j} P_{j,t}\left(\theta\right)\left[u\left(c_{j,t}\right) - v\left(l_{j,t}\right)\right]. \tag{3}$$

Here, $c_{j,t}\left(\theta\right)$ and $l_{j,t}\left(\theta\right)$ 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 $F\left(K_t, L_t\right)$. 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, $\varphi_j\left(\theta\right)$, multiplied by hours, $l_j\left(\theta\right)$. Its aggregate value is the sum of the units of effective labor across all individuals alive in each period. In other words,

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

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

3.2 Markets and Government

We assume that individuals supply labor in the labor market and earn wage w_t per unit of effective labor. In addition, individuals have access to a risk-free asset. The assets of the deceased in each period t convert to bequests and are distributed equally among the living population in period t.[14] 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.[15]

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:[16]

$$\begin{aligned}(1+\tau_c)\,c_j + a_{j+1} &= \left(w_{t+j}\varphi_j l_j - T_{y,j,t+j}\left(w_{t+j}\varphi_j l_j\right) + Tr_{j,t+j}\right)\mathbf{1}\left[j<R\right] \\ &\quad + \left(1+r_{t+j}\right)a_j - T_{a,j,t+j}\left(\left(1+r\right)a_j\right) + S_{j,t+j}\left(\mathcal{E}_t\right)\mathbf{1}\left[j\geq R\right] + B_{t+j}. \quad (4)\end{aligned}$$

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 retirement benefit from the government; and B_{t+j} is income earned from bequests. The dependence of retirement benefits on lifetime earnings is captured in $\mathcal{E}$, 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 $(1-\tau_K)\left(F_K\left(K_t,L_t\right)-\delta\right)$. [17] We assume that the government taxes households' holding of

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

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

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

[17]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

$$\int \sum_{j=0}^{J} \mu_t(\theta, j) Tr_{j,t} dH(\theta) + \int \sum_{j=R+1}^{J} \mu_t(\theta, j) S_{j,t}(\mathcal{E}_{t-j}(\theta)) dH(\theta) + G_t + (1+r_t) D_t =$$

$$\tau_C \int \sum_{j=0}^{J} \mu_t(\theta, j) c_{j,t-j}(\theta) dH(\theta) + \int \sum_{j=0}^{J} \mu_t(\theta, j) T_{y,j,t}(w_t \varphi_j(\theta) l_{j,t-j}(\theta)) dH(\theta) +$$

$$\int \sum_{j=0}^{J} \mu_t(\theta, j) T_{a,j,t}((1+r_t) a_{j,t-j}(\theta)) dH(\theta) + \tau_K (F_K(K_t, L_t) - \delta) + (1+\hat{n}_{t+1}) D_{t+1}, \quad (5)$$

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

$$\int \sum_{j=0}^{J} \mu_t(\theta, j) c_{j,t-j}(\theta) dH(\theta) + G_t + (1+n_{t+1}) K_{t+1} = F(K_t, L_t) + (1-\delta) K_t, \quad (6)$$

$$\int \sum_{j=0}^{J} \mu_t(\theta, j) p_{j+1,t-j}(\theta) a_{j+1,t-j}(\theta) dH(\theta) = (1+\hat{n}_{t+1})(K_{t+1} + D_{t+1}), \quad (7)$$

$$\int \sum_{j=0}^{J} \mu_t(\theta, j) (1 - p_{j+1,t-j}(\theta)) a_{j+1,t-j}(\theta) dH(\theta) = (1+\hat{n}_{t+1}) B_{t+1}. \quad (8)$$

Equilibrium. An equilibrium of this economy is defined as allocations where individuals maximize (3) subject to (4), while government budget constraint (5), market clearings (6), (7) and (8) 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 about the status quo policies yet. We will do that in section 5 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.

4 Optimal Policy Reform: Theoretical Framework

Our optimal policy-reform exercise 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 utility profile $\{W^{sq}(\theta)\}_{\theta\in\Theta}$, given the assumed set of policies and to a particular representative cohort of individuals. The key benefit of this approach is that it allows us not to take a stand on the redistributional concerns inherent in the policy-making process. Thus the change in the cost of such reforms is purely in terms of efficiency gains and not redistributional gains. 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.

4.1 A Planning Problem

The set of policies that we allow for in our optimal reform are very similar to those described in section 3. In particular, we allow for non-linear and age-dependent taxation of assets. Moreover, we allow for non-linear and age-dependent taxation of earnings together with flat social security benefits (i.e., social security benefits are independent of lifetime earnings). Therefore, given any tax and benefit structure, each individual maximizes utility (3) subject to the budget constraints (4).

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.

Implementability. We use the primal approach a la Lucas and Stokey (1983) to solve the optimal reform problem. The primal approach transforms the problem of finding optimal policies to that of finding optimal allocations. However, due to the optimizing behavior and the fact that policies cannot depend on individual characteristics, the optimizing behavior by individuals imposes a constraint on the planning problem. This constraint can be written solely in terms of individual allocations and is described by the following lemma:

Lemma 1. *Consider an allocation $\{c_j(\theta), l_j(\theta)\}$ that maximizes individual preferences (3) subject to the constraints (4) for a given set of taxes. Let $U(\theta)$ be the utility associated with such an allocation.*

Then we must have

$$U'(\theta) = \sum_{j=0}^{J} \beta(\theta)^j P_j(\theta) \left[\frac{\varphi_j'(\theta) l_j(\theta)}{\varphi_j(\theta)} v'(l_j(\theta)) \right] + \sum_{j=0}^{J} \left(\frac{j\beta'(\theta)}{\beta(\theta)} + \frac{P_j'(\theta)}{P_j(\theta)} \right) \beta(\theta)^j P_j(\theta) \left[u(c_j(\theta)) - v(l_j(\theta)) \right]. \tag{9}$$

The proof can be found in the appendix.

Equation (9) is the envelope condition associated with the individual optimization problem of maximizing (3) subject to the budget constraint (4). We refer to (1) as the *implementability constraint*. This constraint is reminiscent of the local incentive compatibility constraint in the New Dynamic Public Finance and mechanism design literature. In fact, if instead of solving the policy problem described above, one solves the mechanism design problem when θ is private information, the exact same constraint is derived.

However, it remains to be shown that if an allocation satisfies (9), it must be the solution of individual optimization given some set of tax functions. Unfortunately, we cannot theoretically prove this result. In the appendix, we show that if an allocation satisfies (9) and certain monotonicity conditions, then a set of tax functions can be constructed so that the allocations satisfy the first-order conditions associated with the individual optimization and budget constraints.

Planning Problem. Our planning problem maximizes the revenue from delivering a steady-state allocation subject to the implementability constraint (9) and a minimum utility requirement given by

$$\max \int \sum_{j=0}^{J} \frac{P_j(\theta)}{(1+r)^j} \left[\varphi_j(\theta) l_j(\theta) - c_j(\theta) \right] dH(\theta) \tag{10}$$

subject to

$$U(\theta) = \sum_{j=0}^{J} \beta(\theta)^j P_j(\theta) \left[u(c_j(\theta)) - v(l_j(\theta)) \right] \tag{11}$$

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

$$U(\theta) \geq W^{sq}(\theta). \tag{13}$$

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.[18]

4.2 Efficient Distortions

Our analysis of optimal taxes can be informed by studying the wedges implied by the solution to the above planning problem. By this, we mean the magnitude of distortions to the individuals' trade-off between consumption and earnings, and consumption across periods. These are useful statistics about optimal allocation that inform us about the properties of optimal taxes.

The distortion to the consumption-earning margin or the *labor wedge* for each individual of type θ is defined by

$$\tau_{\text{labor},j}(\theta) = 1 - \frac{v'(l_j(\theta))}{\varphi_j(\theta)\, u'(c_j(\theta))}. \tag{14}$$

Intuitively, $\tau_{\text{labor},j}(\theta)$ is the fraction of earnings on the margin that is taken away from the individual in terms of period j consumption.

The wedge to the allocation of consumption across periods is less straightforward to define. In particular, the definition of distortions depends on the type of asset held by the individual. Two types of assets are of particular interest:

1. A non-contingent asset that pays a return $1 + r$ independent of the individual's survival, for which the wedge is defined by

$$\tau_{\text{savings},j}(\theta) = 1 - \frac{u'(c_j(\theta))}{\beta(\theta)(1+r)p_{j+1}(\theta)\, u'(c_{j+1}(\theta))}. \tag{15}$$

   We refer to this as *savings wedge*.

2. An annuity that pays a return $1 + r$ only in case of survival and is priced at an actuarially fair price, $p_{j+1}(\theta)$, given by

$$\tau_{\text{annuity},j}(\theta) = 1 - \frac{u'(c_j(\theta))}{\beta(\theta)(1+r)u'(c_{j+1}(\theta))}, \tag{16}$$

   which we refer to as the *annuity wedge*. The above can be interpreted as a tax imposed on income from an annuity purchased at the actuarially fair price of $\frac{p_{j+1}(\theta)}{1+r}$.

Note that the above wedges are hypothetical, since we do not allow for annuity holdings in our

[18] 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.

implementation. Nevertheless, they are informative in terms of separating the different roles that taxes play: incentive provision versus insurance.

The following lemma characterizes the optimal labor wedge:

Lemma 2. *The labor wedges implied by the efficient allocation are given by*

$$\frac{\tau_{labor,j}(\theta)}{1-\tau_{labor,j}(\theta)} = \frac{\varphi_j'(\theta)}{\varphi_j(\theta)} \frac{1-H(\theta)}{h(\theta)} \left(\frac{1}{\varepsilon_{F,j}(\theta)}+1\right) \frac{g(\theta)}{1-g(\theta)\frac{1-H(\theta)}{h(\theta)}\left(\frac{P_j'}{P_j}+j\frac{\beta'}{\beta}\right)} \tag{17}$$

where

$$g(\theta) = \int_\theta^{\bar{\theta}} \frac{u'(c_0(\theta))}{u'(c_0(\theta'))}\left[1-\gamma(\theta')\,u'(c_0(\theta'))\right]\frac{dH(\theta')}{1-H(\theta)}, \tag{18}$$

$\varepsilon_{F,j}(\theta) = \frac{v'(l_j(\theta))}{v''(l_j(\theta))l_j(\theta)}$ *is the Frisch elasticity of labor supply and* $\gamma(\theta)$ *is the multiplier on the constraint (13).*

Proof can be found in the appendix.

The above formula is the familiar one from the static optimal taxation literature as in Mirrlees (1971), Diamond (1998) and Saez (2001). The first term in (17) captures the tail property of the distribution at a given age, t. Intuitively, if the marginal tax for type θ increases at age j, it leads to a marginal output loss of $\varphi_j(\theta)\,h(\theta)$. However, it relaxes the incentive constraints on all the types above at age j (captured by $\varphi_j'(\theta)(1-H(\theta))$). The second term is capturing the behavioral response to taxes. The higher the Frisch elasticity of labor supply, the larger the response to higher taxes. Finally, the last term is the social marginal welfare weight (see Piketty et al. (2014)) and captures the redistributive motive of the government. The key difference between the above formula and the standard Mirrlees-Diamond-Saez formula is that the marginal social value of redistribution must be adjusted by mortality and discount factor heterogeneity. The intuition for this adjustment can be understood by inspecting the incentive constraint in (9). When survival and discount factor are positively correlated with labor productivity, utilities provided—either via leisure or consumption—in older ages tighten the incentive constraint more than those in earlier ages, due to heterogeneity in discount factor and survival probabilities. This, in turn, increases the dead-weight loss of redistributive taxation. As a result, distortions must grow with age in order to relax the incentive constraint. In other words, utilities should be front-loaded while distortions should be back-loaded. The denominator in the term

$$\frac{g(\theta)}{1-g(\theta)\frac{1-H(\theta)}{h(\theta)}\left(\frac{P_j'}{P_j}+j\frac{\beta'}{\beta}\right)}$$

captures this effect.

We now turn to the characterization of annuity wedges. Then, the characterization of the savings wedge will be straightforward. We have the following proposition:

Proposition 3. *The annuity wedge is given by*

$$\tau_{annuity,\, j}\left(\theta\right) = \left(\frac{p'_{j+1}\left(\theta\right)}{p_{j+1}\left(\theta\right)} + \frac{\beta'\left(\theta\right)}{\beta\left(\theta\right)}\right)\frac{1-H\left(\theta\right)}{h\left(\theta\right)}\frac{g\left(\theta\right)}{1-g\left(\theta\right)\frac{1-H(\theta)}{h(\theta)}\left(\frac{P'_j(\theta)}{P_j(\theta)}+j\frac{\beta'(\theta)}{\beta(\theta)}\right)}, \tag{19}$$

where $g\left(\theta\right)$ is given by (18). The annuity wedge is positive if survival and discount rates are positively correlated with labor productivity, i.e., $p'_{j+1}\left(\theta\right) > 0$, $\beta'\left(\theta\right) > 0$.

Proof can be found in the appendix.

Note we can rewrite this equation approximately as

$$\tau_{\text{annuity},\, j}\left(\theta\right) \approx \frac{1-H\left(\theta\right)}{h\left(\theta\right)} g\left(\theta\right)\left(\frac{p'_{j+1}\left(\theta\right)}{p_{j+1}\left(\theta\right)} + \frac{\beta'\left(\theta\right)}{\beta\left(\theta\right)}\right). \tag{20}$$

Intuitively, the idea behind the optimal tax on annuity income is similar to that of labor income taxes. That is, productive individuals have an incentive to under-save in response to transfers in the future. As a result, taxation of savings at lower levels prevents productive individuals from under-saving and leads to a lower dead-weight loss of redistribution.

The distortions to savings, risk-free and non-contingent on death, can be calculated from the annuity wedge provided above. In particular,

$$1-\tau_{\text{savings},j}\left(\theta\right) = \frac{\left(1-\tau_{\text{annuity},j}\left(\theta\right)\right)}{p_{j+1}\left(\theta\right)}.$$

We can write the above approximately as

$$\tau_{\text{savings},j} \approx \tau_{\text{annuity},j} - \left(1-p_{j+1}\right). \tag{21}$$

The above formula illustrates the forces for optimal tax or subsidies on savings. In other words, the presence of survival and discount rate heterogeneity creates forces toward taxation of savings, while market incompleteness leads to subsidizing savings. Our quantitative exercise in section 6 clarifies the magnitude of each of these forces.

4.3 Test of Pareto Optimality

We now extend the results of Proposition 2 in section 2 and derive the conditions for a Pareto optimality of any tax system. For any tax policy (not necessarily optimal), let the intratemporal distortion $\tau_{\text{labor},j}(\theta)$ be defined by equation (17). Also let the distortion to (non-contingent) savings $\tau_{\text{saving},j}(\theta)$ be defined by equation (21). Let ε be the elasticity of labor supply and σ be the coefficient of risk aversion. For ease of exposition, define

$$
\begin{aligned}
A_j(\theta) &= \frac{\frac{\tau_{\text{labor},j}(\theta)}{1-\tau_{\text{labor},j}(\theta)}\frac{\varphi_j(\theta)}{\varphi_{j,\theta}(\theta)}\frac{\varepsilon}{1+\varepsilon}}{1+\left(\frac{\tau_{\text{labor},j}(\theta)}{1-\tau_{\text{labor},j}(\theta)}\frac{\varphi_j(\theta)}{\varphi_{j,\theta}(\theta)}\frac{\varepsilon}{1+\varepsilon}\right)\left(\frac{P_j'(\theta)}{P_j(\theta)}+j\frac{\beta'(\theta)}{\beta(\theta)}\right)},\\
B_0(\theta) &= \frac{h'(\theta)}{h(\theta)}+\frac{\tau_{\text{labor},0}'(\theta)}{(1-\tau_{\text{labor},0}(\theta))\,\tau_{\text{labor},0}(\theta)}+\frac{\varphi_{0,\theta}(\theta)}{\varphi_0(\theta)}-\frac{\varphi_{0,\theta\theta}(\theta)}{\varphi_{0,\theta}(\theta)}+\sigma\frac{c_0'(\theta)}{c_0(\theta)},\\
C_j(\theta) &= \frac{P_{j+1}(\theta)}{P_j(\theta)},\\
D_j(\theta) &= \frac{\frac{1-\tau_{\text{labor},0}(\theta)}{\tau_{\text{labor},0}(\theta)}\frac{\varphi_{0,\theta}(\theta)}{\varphi_0(\theta)}\frac{1+\varepsilon}{\varepsilon}-\left(\frac{P_{j+1}'(\theta)}{P_{j+1}(\theta)}+(j+1)\frac{\beta'(\theta)}{\beta(\theta)}\right)}{\frac{1-\tau_{\text{labor},0}(\theta)}{\tau_{\text{labor},0}(\theta)}\frac{\varphi_{0,\theta}(\theta)}{\varphi_0(\theta)}\frac{1+\varepsilon}{\varepsilon}-\left(\frac{P'j(\theta)}{P_j(\theta)}+j\frac{\beta'(\theta)}{\beta(\theta)}\right)}.
\end{aligned}
$$

The following proposition provides the conditions under which a given set of policies are efficient.

Proposition 4. *A tax policy $\{T_{y,j}, T_{a,j}, S_j\}$ is efficient only if it satisfies the following conditions:*

$$-A_j(\theta)\cdot B_0(\theta) \leq 1, \tag{22}$$

$$C_j(\theta)\left(1-\tau_{saving,j}(\theta)\right) = D_j(\theta)\,. \tag{23}$$

In addition, if optimal allocations under the tax functions are fully characterized by individuals' first-order conditions, then the above conditions are also sufficient for efficiency.

The proof is in the appendix.

Note that, for $j = 0$, inequality (22) is identical to (1), derived in section 2. Therefore, the intuition for the test of the earnings taxes is similar to the one provided in section (2). As for the test of asset taxes, the ideas are similar to the two-period model. On the one hand, asset taxes must reflect distributional concerns that lead to positive taxes; this is captured in the term $D_j(\theta)$.[19] On the other hand, asset taxes must address the incompleteness of the annuity market and thereby subsidize savings; this is captured by the term $C_j(\theta)$.

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

[19] As mentioned in Section (2), this is similar to the effect identified by Saez (2002) and Golosov et al. (2013) among others.

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 these conditions are indeed satisfied. Needless to say, in all of our simulations the monotonicity constraints are satisfied.

5 Calibration

In order to conduct our policy experiments, we need parametric specifications and parameter values for the model described in section 3. We will estimate some of the parameters independently (e.g., wage/productivity profiles or mortality profiles). We choose the rest of the parameters (e.g., discount factor) so that the model matches some targets in the U.S. data. We describe these details below.

Earning ability profiles. Individual productivity $\varphi_j(\theta)$ has two components: a deterministic age-dependent component $\tilde{\varphi}_j$ and a type-dependent fixed effect θ. We assume

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

with the age-dependent component 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 (2015)) and use the effective reported labor earnings per hour 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. We use the data in Heathcote et al. (2010), who carefully address a number of well-known issues in the raw data. The 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 earning distribution. For this reason, it is commonly used in the literature (see Golosov et al. (forthcoming), Badel and Huggett

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 1000)	369	307	286	205	204	211	204	167	142	97

[a]source: Waldron (2013)

(2014) and Heathcote and Tsujiyama (2015)).[20] 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 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. 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

$$M_j(\theta) = \frac{\eta_0}{\theta^{\eta_1}} \left(\frac{\exp(\eta_2 j)}{\eta_2} - 1 \right). \tag{24}$$

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 (24) determines the changes in mortality by age and is common across all types. The first term is decreasing in θ and shifts mortality age profiles. 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 mortality across lifetime earning deciles reported in Waldron (2013). She uses Social Security Administration data to estimate mortality differentials at ages 67–71 by lifetime earnings decile. Table 1 shows the estimated annual mortality rates for 67- to 71-year-old males born in 1940. 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 $\eta_0 = 0.0006$,

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

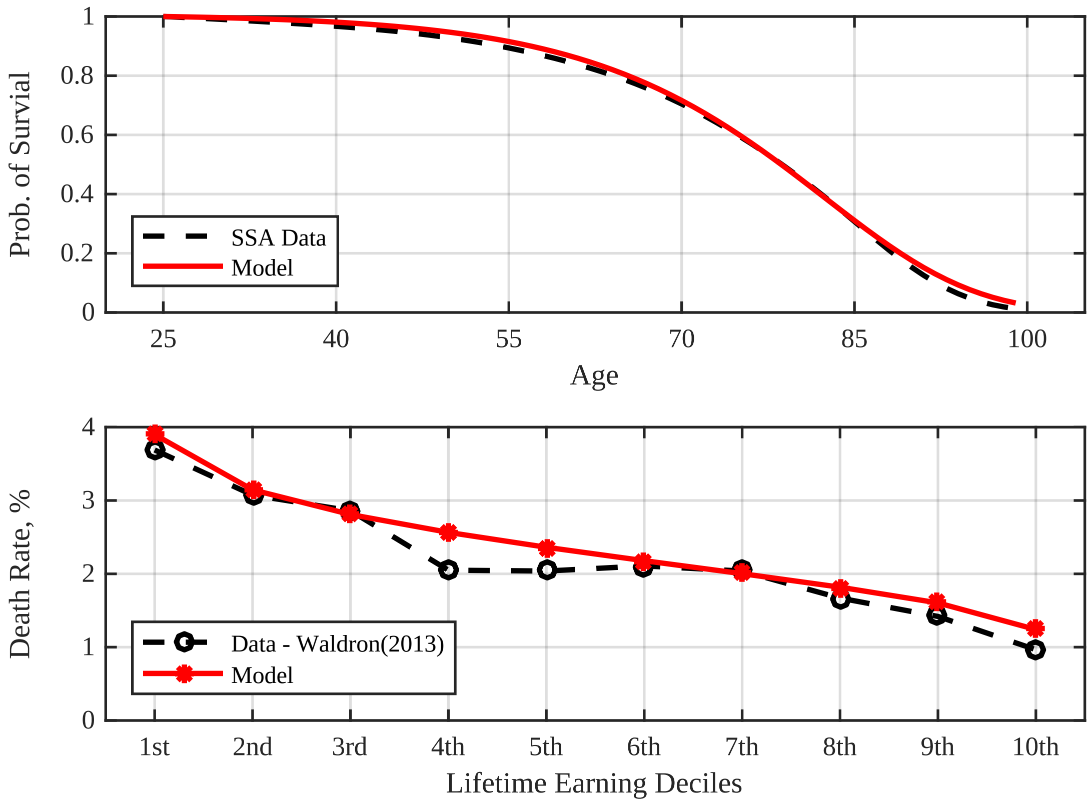

Figure 1: Fit of the mortality model. The top panel shows the average survival probability in the model vs. social security data. The bottom panel shows death rates at age 67 in the model vs. those reported in Waldron (2013).

$\eta_1 = 0.5545$ and $\eta_2 = 0.0855$. Figure 1 shows the fit of the model in terms of matching mortality across lifetime earnings decile 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, there are 4 workers per each retiree in the steady state. This is consistent with U.S. Census Bureau estimates.[21]

Preferences. 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 $\sigma = 1.5$ and the elasticity of labor supply is $\varepsilon = 0.5$. The weight of leisure in utility ψ is chosen so that, in the model, the average number of annual hours worked is 2000.

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

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

[21] http://www.census.gov/population/projections/data/national/2014.html

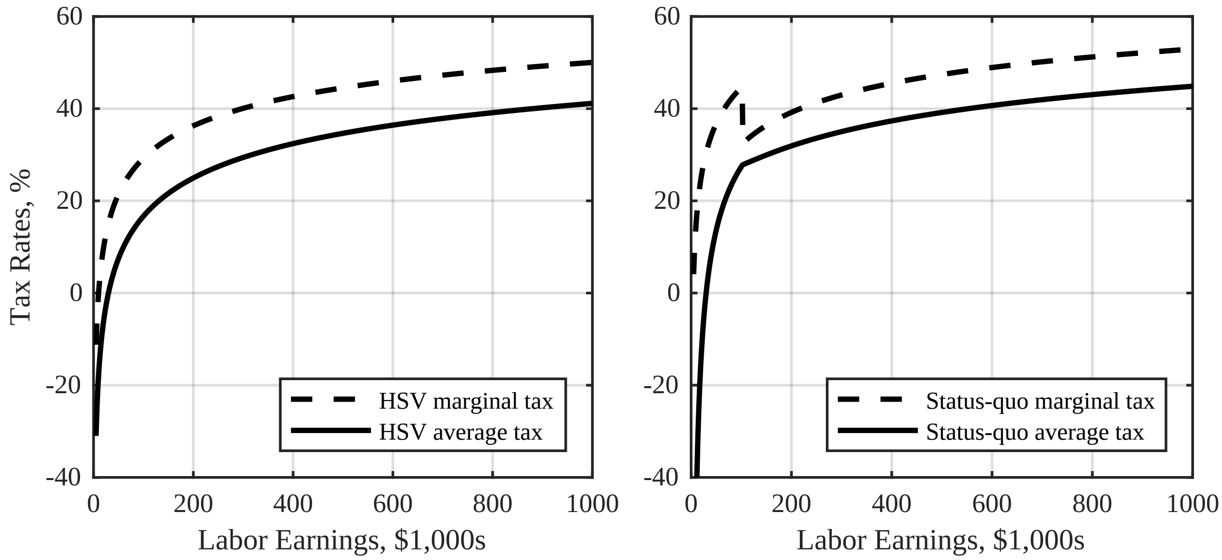

Figure 2: Tax functions. The left panel is the calibrated HSV tax function, $\mathcal{T}_{HSV}(\cdot)$. The right panel is the effective tax function (including HSV tax, payroll tax and transfers). The discontinuity is due to social security cap on taxable earnings.

We choose β_0 to match the wealth to income ratio of 3. The other parameter, β_1, determines the degree of heterogeneity in the discount factor. The larger β_1 is, the larger is the dispersion in the discount factor across ability types. We choose this parameter to match the wealth Gini index in SCF (1989) data.

The aggregate production function is Cobb-Douglas with a capital share parameter $\alpha = 0.36$, and the depreciation rate is $\delta = 0.06$.[22] There is a corporate income tax rate of $\tau_K = 0.33$ paid by the firms. Therefore, after tax, the return on assets is $r = 0.04$.[23] This is also the interest rate that government pays on its debt.

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.[24] 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 $\tau_{ss} = 0.124$.[25] 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

[22]Therefore, with a capital to output ratio of 3, the steady state ratio of investment to GDP is 0.21, which is aligned with the U.S. average over the period 2000–2010.

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

[24]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.

[25]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
$H_j(\theta)$	mortality hazard	see text
Preferences		
σ	risk aversion parameter	1.5
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	0.36
δ	depreciation rate	0.06
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[a]
G	government purchases	9% of GDP
D	government debt	50% of GDP

[a]Source: Heathcote et al. (2014)

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 that the aggregate Medicare benefits are 3 percent of the GDP.[26]

[26]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.

Table 3: Parameters Calibrated Using the Model

Parameters	Description		Values
β_0	discount factor: level		0.975
β_1	discount factor: elasticity w.r.t θ		0.01
ψ	weight on leisure		0.74
Targeted Moments		Data	Model
Wealth-income ratio		3	3
Wealth Gini		0.78	0.78
Average annual hours		2000	2000

Tax functions and government purchases. In addition to social security, the government has an exogenous spending G, which we assume to be 9 percent of the GDP.[27] There is a consumption tax τ_C and a nonlinear tax on labor income. We use 5.5 percent for consumption tax as calculated in McDaniel (2007). For the income tax function, we use

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

where y is the 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 ϕ is, the higher are the total tax revenues for a given τ). Heathcote et al. (2014) estimate a value of 0.151 for τ, based on PSID income data and income tax calculations using NBER's TAXSIM program. We use their estimated value for τ and choose λ. We refer to this tax function as HSV tax function. The left panel in Figure 2 illustrates the resulting marginal and average taxes as functions of annual earnings in constant 2000 dollars.

Finally, we assume government debt to be 50 percent of the GDP.[28] The transfers Tr are such that the government budget constraint (equation (5)) is satisfied in stationary equilibrium. In our calibrated model, we follow McGrattan and Prescott (forthcoming) and assume all individuals face the same after-tax interest rate regardless of their income (and therefore asset taxes are equal to

[27]This is the sum of all government consumption expenditure on national defense, general public service, public order and safety, and economic affairs in NIPA Table 3.16. We use the average over the period 2000 to 2010.

[28]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.

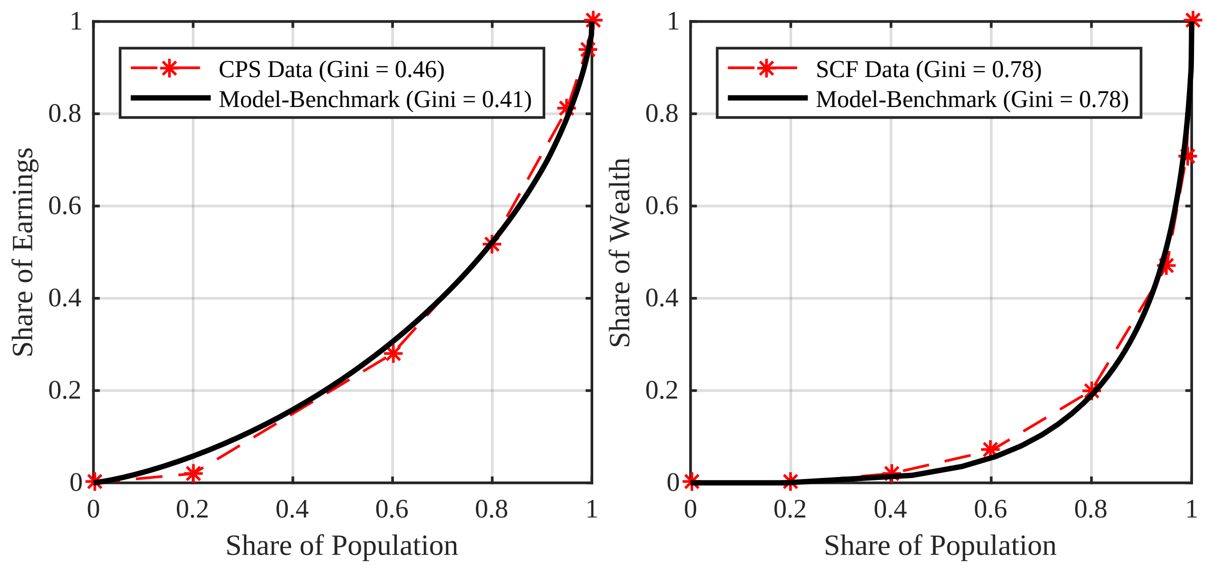

Figure 3: Fit of the distribution of earnings (left panel) and wealth (right panel).

0).[29]

To summarize, individuals face three different types of taxes on their income: 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.

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 parameter values. The bottom panel shows the targeted moments in data and 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 shows the commutative share of earnings at each commutative population share for individuals age 25 to 60 in CPS (1994). The solid line shows the same measure in the model. Overall, the model does a good job at capturing the extent of earnings inequality in the data. The Gini index of earnings is 0.41 in the model and 0.46 in the data. Moreover, the model is able to capture the

[29] As McGrattan and Prescott (forthcoming) point out, most of the capital income is earned on assets in untaxed or tax-deferred accounts.

concentration on 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 SCF (1989). 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.

6 Quantitative Results: Steady State

In this section we apply the tools developed in section 4 to our calibrated economy described in section 5. 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 section 4.3 to solve for optimal policies that implement efficient distortions in the economy. Finally, we report the effect that an optimal reform has on individual choices, macro aggregates and government budget. Note that our optimal policies minimize the present value of consumption net of labor income for each generation. We report the reduction in this cost as a measure of efficiency gains from optimal reform policies.

Two points are worth emphasizing about our exercise: First, the efficiency gains from our Pareto optimal policy reforms can be redistributed across individuals in various ways. In this section, we do not specify how the gains are distributed. In the next section, we provide one way to distribute these gains to a subset of the population. Second, since it is important to disentangle the partial and general equilibrium effects of the reform, in section (6.1)-(6.4), we assume that prices—interest rates and wages—are fixed at the status quo level. In sections 6.5 and 7, we report the results with endogenous factor prices.

6.1 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 the inequality (22) and equation (23).

The left hand side of inequality (22) is depicted in Figure 4 (left panel) for ages $j = 25, 35, 45$. As the figure illustrates, this term is below 1 for most ability types (each lifetime unit of earnings corresponds to an ability type). The inequality only fails to hold over a small range for all ages. This is where effective earnings taxes are regressive, due to the social security maximum taxable

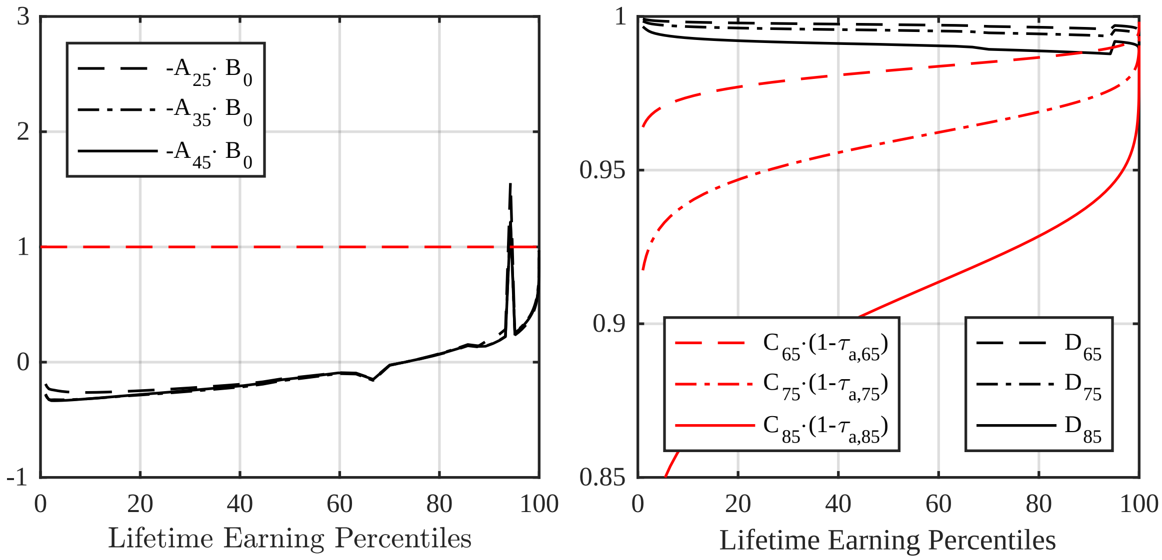

Figure 4: Test of Pareto optimality for status quo policies. The left panel is the test of earnings tax schedule. The black lines are the left hand side of inequality (22). The test fails only at the social security maximum taxable earnings' cap. The right panel is the test of asset tax. The black lines are the right hand side of Equation (23). The red lines are the left hand side of that equation (equal to survival rate). This test fails everywhere.

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 (22) up. This figure illustrates one of our main findings: that earnings taxes are not a large source of inefficiency of the tax code.

We now examine asset taxes. We plot the left hand side (drawn in red) and right hand side (drawn in black) of Equation (23) separately in Figure 4 (right panel) for ages $j = 65, 75, 85$. With $\tau_{\text{savings},j}(\theta) = 0$, these are simply survival rates at each age. The right hand side of the equation is driven by the gradient of survival with respect to ability, $P'_j(\theta)$, and the gradient of discount factor with respect to ability, $\beta'(\theta)$. Note that in an efficient tax schedule, $C_j\left(1 - \tau_{\text{savings},j}\right)$ and D_j must coincide. The graph, thus, identifies large deviations from efficiency for asset taxes. Moreover, it suggests that a potentially large subsidy is required in order to equate $C_j\left(1 - \tau_{\text{savings},j}\right)$ with D_j. This implies that the market incompleteness effect (captured by C_j) is more important than the redistributional effect (captured by D_j) in shaping optimal asset taxes.

In summary, our tests provide two main results: first, earnings taxes pass the Pareto optimality tests to a great extent, except around the social security earnings cap; second, asset 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.

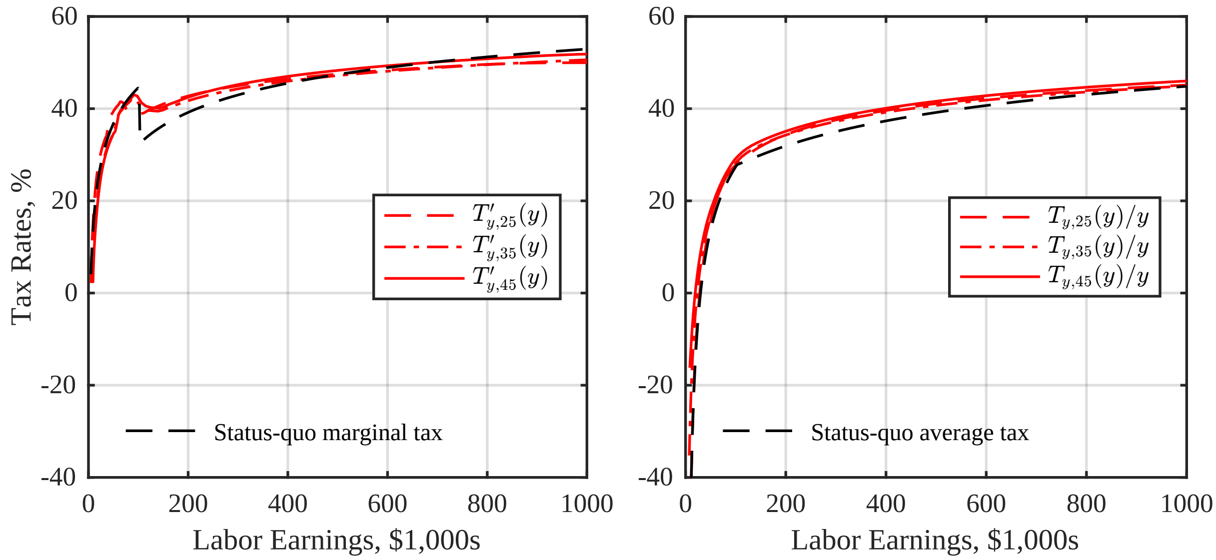

Figure 5: Optimal labor income tax functions. The left panel shows marginal taxes, while the right panel shows average taxes. The black dashed line is the effective status quo tax schedule.

6.2 Optimal Policies

We solve for optimal policies using the planning problem 10 outlined in section 4. 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.[30] As a result, we choose transfers such that the lowest-ability type chooses 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 makes comparing labor income taxes across economies straightforward.[31] Finally, we fix the corporate 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 optimal reform.

Figure 5 shows the optimal marginal and average labor income tax functions for ages $j = 25, 35, 45$ (solid lines). We also plot the status quo tax functions for comparison (dashed lines). Notice that except for the region where there is a sharp drop in the status quo tax rates (due to 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. These results

[30] This feature resembles Ricardian equivalence.

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

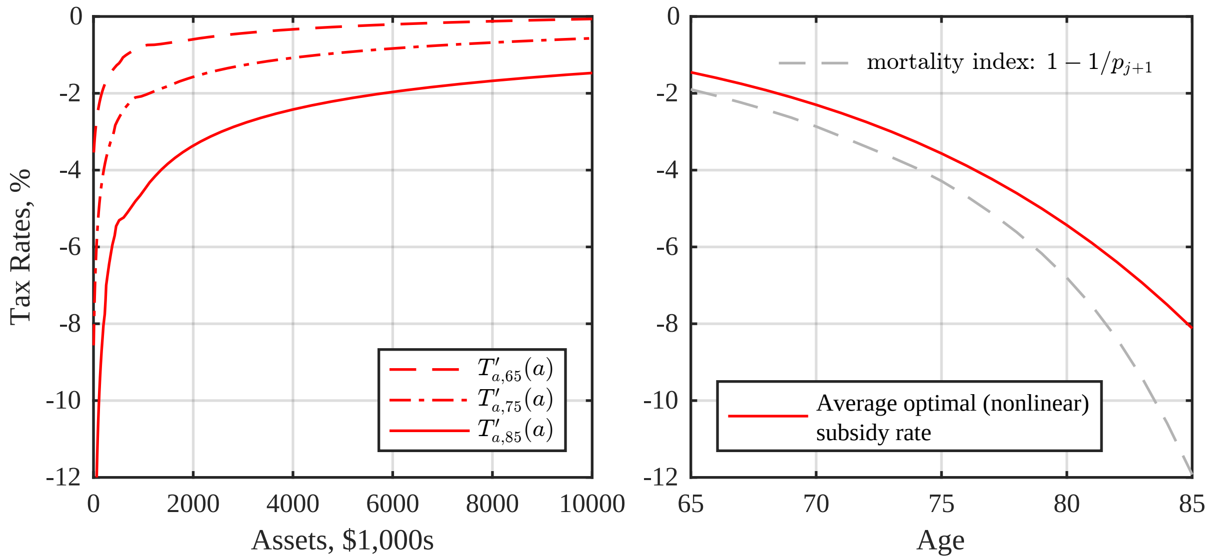

Figure 6: Optimal asset tax functions. The left panel shows marginal taxes over all asset levels at ages 65, 75 and 85, while the right panel shows average marginal rates at each age from 65 to 85. The dashed line is the population mortality index.

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 6.1 regarding earnings taxes.

The left panel of Figure 6 shows 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. On the other hand, 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 6 also illustrates that subsidies are large, around 5 percent, and thus can play an important role in 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 that 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 were mortality was observed by the government.

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.

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	81	52
2nd	90	97	71	31
3rd	67	88	64	20
4th	34	50	50	8

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

6.3 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 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 CPS data (reported in Poterba (2014)).[32]

The numbers in our status quo economy are close to the CPS data, particularly for the lower half of the income distribution. For the upper half, the discrepancy mainly stems from our assumption that retirement age is the same across all individuals, but in reality some individuals age 65 or older still work and do not collect social security benefits.

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

6.4 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.

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

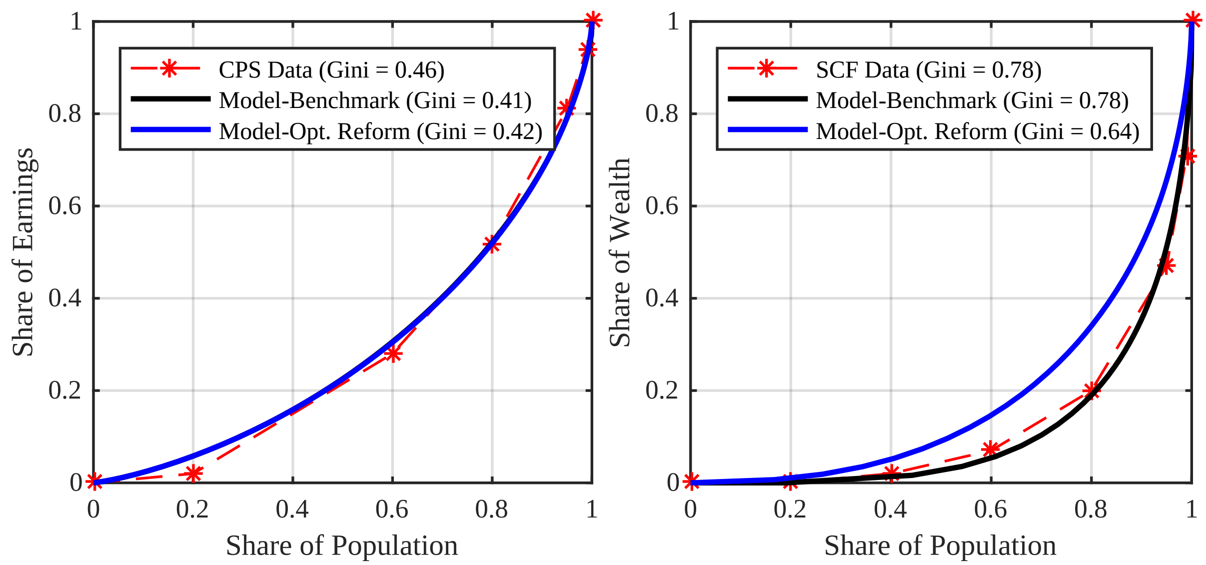

Figure 7: Distribution of earnings and wealth: status quo vs. optimal reform. The black line shows the results in the calibrated economy with current U.S. status quo policies. The blue line shows the results under Pareto optimal policies (for current U.S. demographic parameters and holding factor prices fixed).

Figure 7 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 government's tax revenue and transfers. There is little difference in total tax revenue and total transfers as a fraction of the GDP. However, the nature of transfers changes significantly in an optimal reform. Pure transfers before and after retirement fall as a percentage of the GDP and instead asset subsidies which amount to 7 percent of the 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.

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

	Current U.S.	Optimal Reform	
		Fixed Prices	Endogenous Prices
Factor prices			
Interest rate (%)	4	4	3.7
Wage	1	1	1.02
Values relative to GDP			
Consumption	0.70	0.67	0.69
Capital	3.00	3.43	3.13
Tax revenue (total)	0.25	0.26	0.26
Earnings tax	0.15	0.15	0.16
Consumption tax	0.04	0.04	0.04
Capital (corporate) tax	0.06	0.07	0.06
Transfers	0.14	0.13	0.10
To retirees	0.09	0.03	0.02
To workers	0.05	0.03	0.03
Asset subsidies	0.00	0.07	0.05
Changes relative to the status quo (%)			
GDP	–	7.15	2.17
Consumption	–	2.59	0.81
Capital	–	22.34	6.69
Labor input	–	-1.39	-0.29
PDV of net resources	–	-5.15	-15.36

6.5 Aggregate Effects of Reforms

Table 5 shows the summary statistics of 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 22 percent higher relative to the status quo. This is due to higher incentives to save provided by optimal asset subsidies. As a result, the GDP is higher by 7 percent and consumption by 2.6 percent relative to the status quo. However, consumption as share of the GDP falls slightly from 0.7 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 generation falls by 5.15 percent in the optimal reform relative to the status quo.

The third column in Table 5 shows aggregate quantities under Pareto optimal policies with endogenous factor prices. In this case, the capital stock is higher by only 6.7 percent. This is due to the general equilibrium effect of the lower real return (3.7 percent relative to 4 percent). The

GDP is higher by 2.2 percent and consumption by 0.8 percent relative to the status quo. The cost savings in this case are three times larger relative to the case with fixed factor prices. In other words, the present discounted value of consumption, net of labor income, for each a generation falls by 15.35 percent. This can be accounted for entirely by the fall in the interest rate.[33]

6.6 The Importance of Asset Subsidies

As shown above, asset subsidies play an important role in our Pareto optimal reforms. In particular, based on the optimality tests and the optimal reform exercise, 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 highlights the role of asset taxes and the insignificance of earnings taxes. We leave the details of the analysis to the appendix.

Our restricted reform is similar to privatization. More precisely, we present the best reform policies that remove old-age transfers but do not include any asset taxes or subsidies. In this regard, the efficiency gains from these policies can be viewed as an upper bound on privatization policies. Our exercise shows that the best privatization policy involves a decline in the earnings tax rate by around 3 percent. Furthermore, our calculations show that the cost of this reform, measured as the present value of consumption net of labor income, increases by 2.25 percent under fixed prices, compared to a decline of 5.15 percent in our optimal reform.[34] This result is mainly due to the absence of annuities. Without annuities and asset subsidies, the consumption of individuals declines as they age. This observation, combined with the fact that in the status quo economy, social security transfers partially substitute for annuities, implies that privatization cannot decrease the cost of delivering the status quo level of welfare.

Finally, in the appendix, we perform various robustness exercises on parameter values in order to test the robustness of our results. In all of the calculations, asset subsidies prove to be an integral part of the reforms, while earnings taxes are of less importance.[35]

[33] As we show in section 7, 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.

[34] This is in stark contrast to McGrattan and Prescott (forthcoming), who find a Pareto improving policy reform. Their finding relies on the choice of the elasticity of the labor supply. As we show in the appendix, when the elasticity of the labor supply is high, it is possible to reduce the resource cost through privatization and hence a Pareto improving privatization policy exists.

[35] In the appendix, we also describe an exercise that imposes linearity on asset subsidies. This constraint significantly lowers the gains of the reform.

7 Quantitative Results: Transition

The above analysis points toward the key policies 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 5. 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 (24) to match the differences in mortality rates among lifetime earning deciles reported in Waldron (2013), as well as the new population mortality rates.[36] 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. For the status quo, we assume the income tax schedules and social security benefit formula do not change. Moreover, the debt to GDP ratio is held constant at its initial calibrated value of 50 percent. Therefore, in order to satisfy government budget constraint, we adjust the transfers to workers. It is important to note that, due to political uncertainties, it is impossible to know how status quo policies evolve in response to demographic changes. Here, we use the simplest benchmark to conduct our analysis. However, our methodology can be applied to any alternative assumption for the future path of status quo policies.

The second column in Table 6 shows how the demographic change and continuation of status quo policies affect the aggregates. Since mortality is lower, individuals live longer and therefore have a higher demand for saving. This in turn increases the stock of capital by 3.4 percent. However, due to the lower number of workers as share of population, the labor input falls by 7.7 percent, resulting in a 3.9 percent decline in the GDP.

While continuation of the status quo policies does not change the tax revenue as percentage

[36] We assume that the ratio of mortality among lifetime earning deciles does not change.

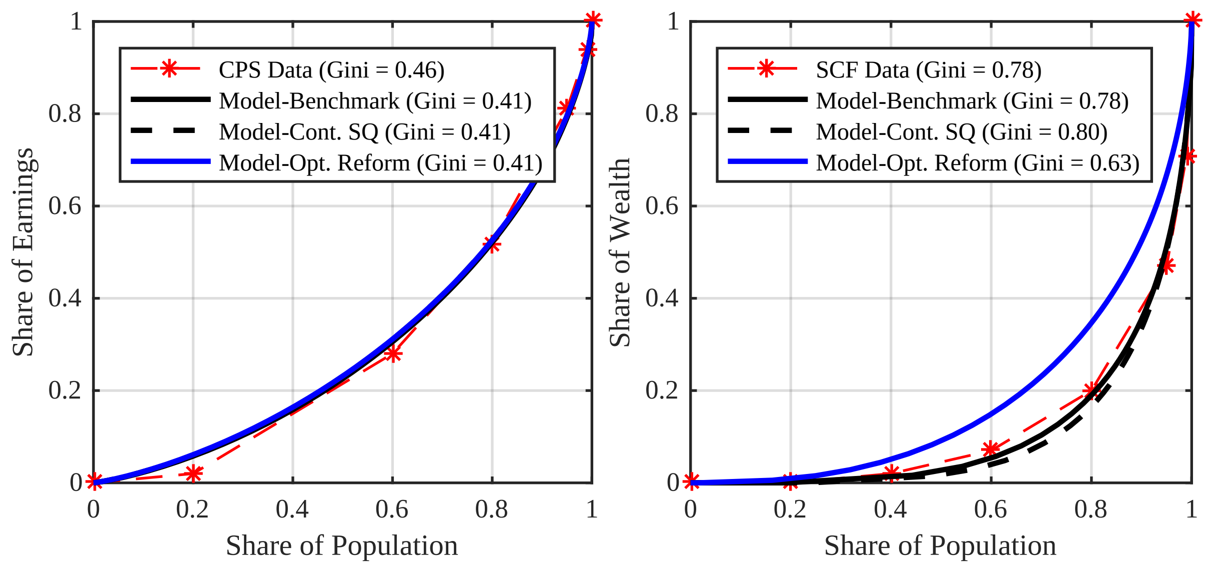

Figure 8: Distribution of earnings and wealth: status quo vs. optimal reform. The black solid line shows the results in the calibrated economy with current U.S. status quo policies. The black dashed line shows the steady state results with projected demographic parameters and continuation of status quo policies. The blue line shows the steady state results under Pareto optimal policies for projected demographic parameters. Factor prices are endogenous.

of the GDP, there is a significant increase in old-age transfers as percentage of the GDP. This is due to the fact that 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 transfer to workers drops to 1 percent of the GDP (from the initial 5 percent). This decline in transfers contributes to an increase in inequality in the economy. To be able to quantify the changes in inequality, we report the "90 to 10 consumption equivalence". This is the percentage of the lifetime consumption that a person in the 90th percentile of the ability distribution is willing to give up, in order to be indifferent between his own consumption allocation and that of a person in the 10th percentile of the ability distribution. As we see in the first and second column, the inequality increases according to this measure. We also report a cross-sectional distribution of earnings and wealth for the new steady state in Figure 8. There is no change in the distribution of earnings. However, the distribution of wealth becomes more unequal (the wealth Gini index rises from 0.78 to 0.8).

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. There are two main subtleties that we need to take a stand on in this exercise: First, we need to take a stand on what can be done with the initial generations (i.e., the generations that are alive at the time of reform). Second, we need to take a stand on how the gains from the reform are distributed.

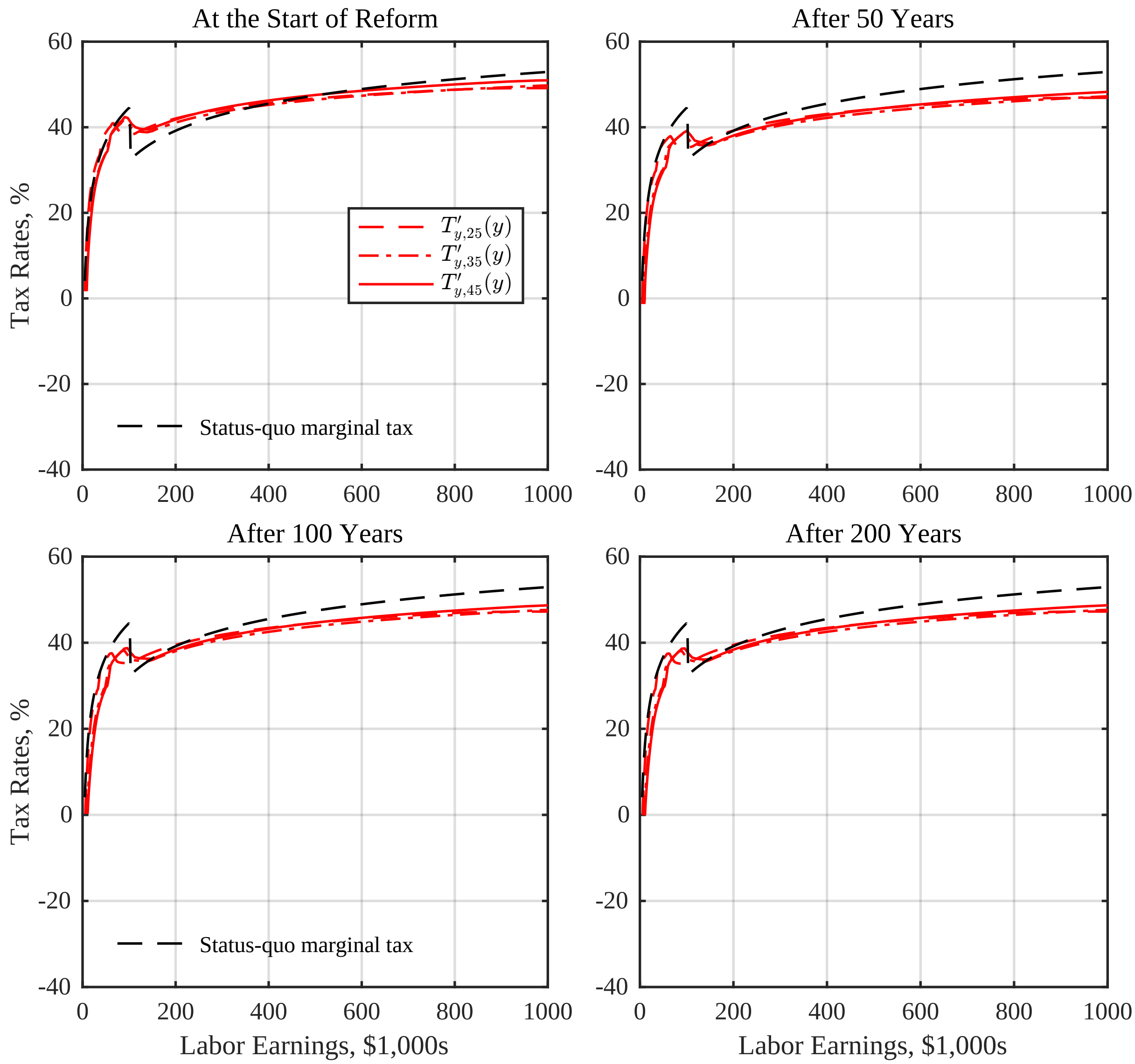

Figure 9: Evolution of optimal marginal labor income tax functions over the transition.

On the first issue, 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. We sidestep this issue by assuming that the initial generations face the same policies as the status quo economy. In other words, the policy reform only applies to those who are born after the start of the reform.

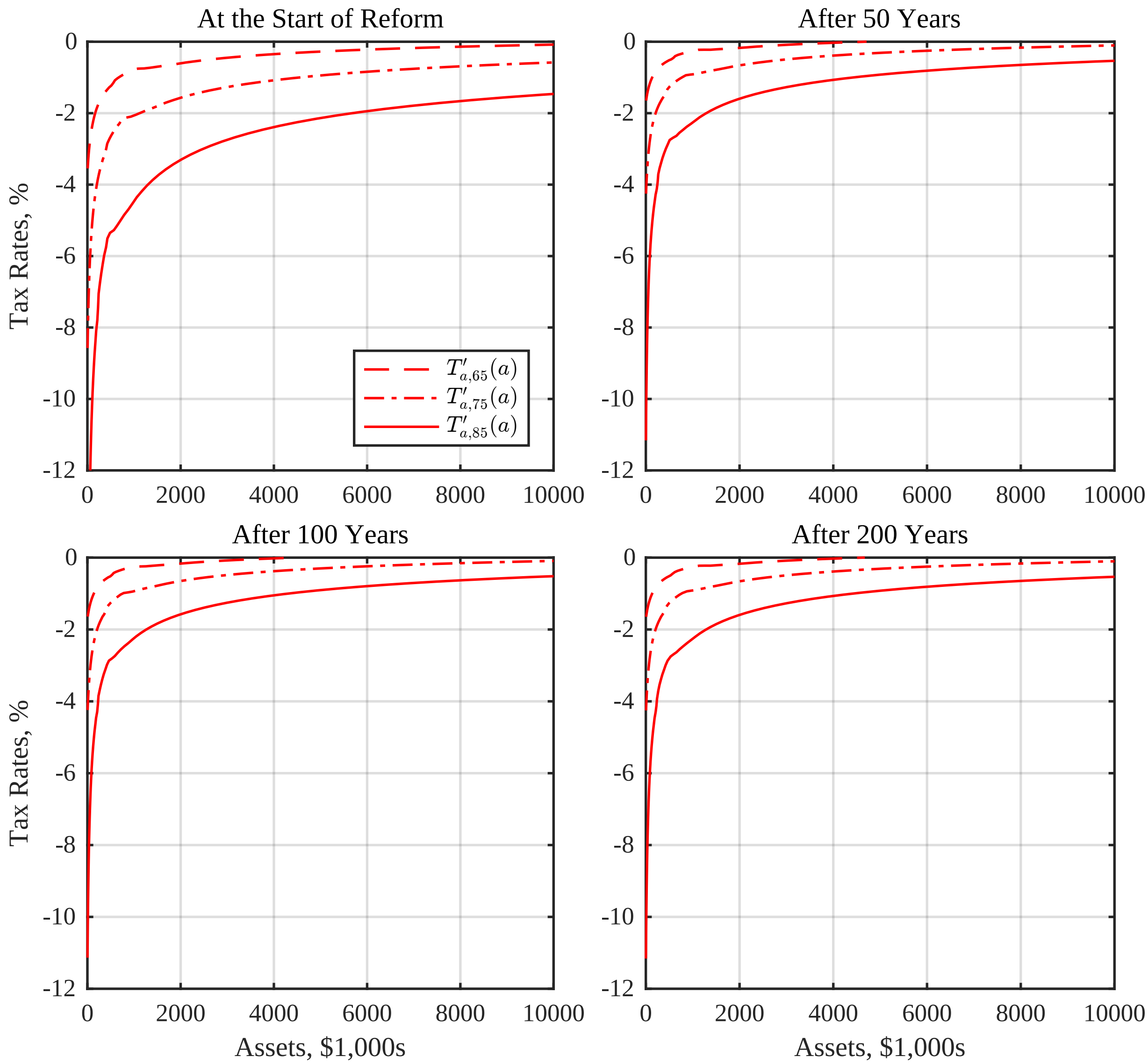

Figure 10: Evolution of optimal marginal asset tax functions over the transition.

On the second issue, there are many ways to distribute the welfare gains from switching to optimal reforms. Given the above discussion, we choose one where everything is given to the initial generations (i.e., those alive at the time of implementing the reform). While this is one of many ways to do the reform, we think it is perhaps more politically viable than others, since all of the gains are given to the initial generations, who will be voting for such a reform.

To summarize, any person who is alive at the beginning of the reform ($t = 0$) will face the status quo policies and will receive an additional one-time lump-sum transfer. All other individuals will face optimal reform policies.

Optimal Reforms. Our quantitative exercise for the transition mainly confirms our previous

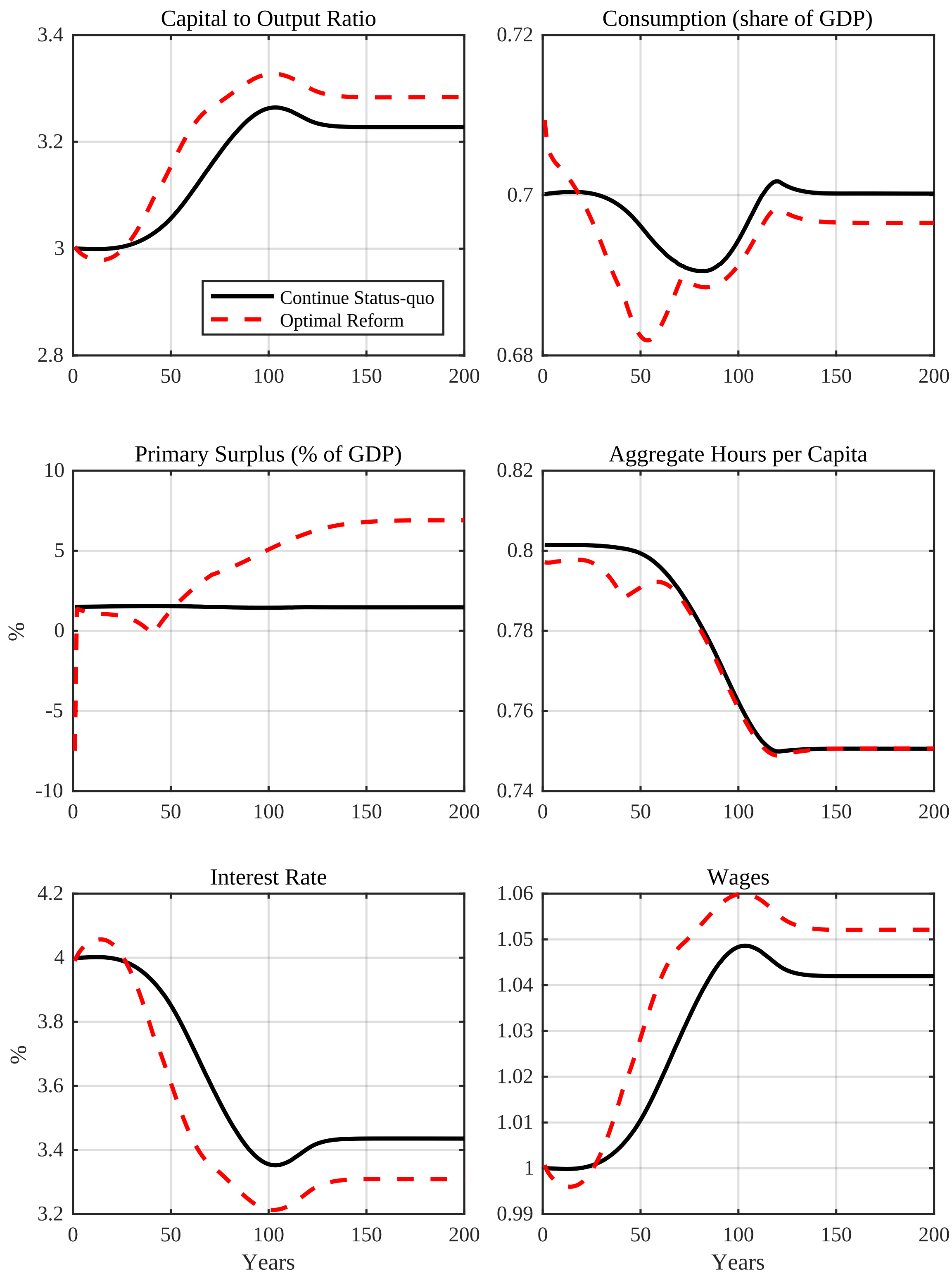

Figure 11: Evolution of aggregates along the transition.

Table 6: Aggregate effects of demographic transition and policy change

	Current Demographics	Future Demographics	
	Status quo	Continue Status quo	Optimal Reform
Factor prices			
Interest rate (%)	4	3.4	3.3
Wage	1	1.04	1.05
90-10 consumption equivalence (%)	3.27	3.70	3.70
Values relative to GDP			
Consumption	0.70	0.70	0.70
Capital	3.00	3.23	3.28
Tax revenue (total)	0.25	0.25	0.24
Earnings tax	0.15	0.16	0.15
Consumption tax	0.04	0.04	0.04
Capital (corporate) tax	0.06	0.05	0.05
Transfers	0.14	0.15	0.08
To retirees	0.09	0.14	0.03
To workers	0.05	0.01	-0.01
Asset subsidies	0.00	0.00	0.06
Change relative to current U.S. (%)			
GDP	–	-3.9	-2.7
Consumption	–	-3.8	-3.1
Capital	–	3.4	6.5
Labor input	–	-7.7	-7.5

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 9 shows the changes in the earnings taxes. Marginal earnings taxes decline slightly, since inequality is growing over time. Note that in our approach, an increase in inequality in the status quo economy increases the lifetime utility that must be delivered to the most productive individuals. This in turn makes the promise-keeping constraint (inequality (13)) on these individuals tighter, thereby increasing their implicit welfare weights. Higher welfare weight on these individuals implies lower distortion. As a result, they face lower marginal taxes. This decline, however, is not large (around 3–5 percent). Furthermore, asset subsidies are still significant although slightly lower, due to the decline in the mortality rate (Figure 10).

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 the GDP and aggregate consumption. [37] Figure 11 shows the path of the

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

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 4.9 percent lower relative to the status quo in the steady state. 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 9 percent of the GDP in the initial steady state.

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

8 Conclusion

In this paper, we have provided a theoretical and quantitative analysis of Pareto optimal policy reforms aimed at financing retirement, i.e., 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. (forthcoming)) 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 are 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.

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.

Birkeland, K. and E. C. Prescott (2007): "On the needed quantity of government debt," *Federal Reserve Bank of Minneapolis Quarterly Review*, 31, 2–15.

Blandin, A. (2015): "Reforming Social Security: The Earnings Cap and Endogenous Human Capital," *Unpublished.*

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.

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

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.

Fukushima, K. (2011): "Quantifying the welfare gains from flexible dynamic income tax systems," *manuscript.*

Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70, 569–587.

Golosov, M., M. Troshkin, and A. Tsyvinski (forthcoming): "Redistribution and social insurance," *American Economic Review.*

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.

HEATHCOTE, J., F. PERRI, AND G. L. VIOLANTE (2010): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic dynamics*, 13, 15–51.

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2014): "Optimal tax progressivity: An analytical framework," Tech. rep., National Bureau of Economic Research.

HEATHCOTE, J. AND H. TSUJIYAMA (2015): "Optimal Income Taxation: Mirrlees Meets Ramsey," *Research Department Staff Report 507.*

HONG, J. AND J.-V. RÍOS-RULL (2007): "Social Security, Life Insurance and Annuities for Families," *Journal of Monetary Economics*, 54, 118–40.

HORIUCHI, S. AND A. J. COALE (1982): "A simple equation for estimating the expectation of life at old ages," *Population Studies*, 36, 317–326.

HOSSEINI, R. (2015): "Adverse Selection in the Annuity Market and the Role for Social Security," *Journal of Political Economy*, 123, 941–984.

HUBBARD, R. G. AND K. L. JUDD (1987): "Social Security and Individual Welfare: Precautionary Saving, Borrowing Constraints, and the Payroll Tax," *American Economic Review*, 77, 630–46.

HUGGETT, M. AND J. C. PARRA (2010): "How Well Does the U.S. Social Insurance System Provide Social Insurance?" *Journal of Political Economy*, 118, 76–112.

HUGGETT, M. AND G. VENTURA (1999): "On the distributional effects of social security reform," *Review of Economic Dynamics*, 2, 498–531.

İMROHOROĞLU, A., S. İMROHOROĞLU, AND D. H. JOINES (1995): "A Life Cycle Analysis of Social Security," *Economic Theory*, 6, 83–114.

JAMES, E. AND D. VITTAS (2000): *Annuity markets in comparative perspective: Do consumers get their money's worth?*, vol. 2493, World Bank Publications.

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.

Lucas, R. E. and N. L. Stokey (1983): "Optimal fiscal and monetary policy in an economy without capital," *Journal of monetary Economics*, 12, 55–93.

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 (forthcoming): "On financing retirement with an aging population," *Quantitative Economics*.

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

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.

Piketty, T., E. Saez, and S. Stantcheva (2014): "Optimal Taxation of Top Incomes: A Tale of Three Elasticities," *American Economic Journal: Economic Policy [Internet]*, 6, 230–271.

Poterba, J. (2001): "The History of Annuities in the United States," in *The Role of Annuity Markets in Financing Retirement*, ed. by J. Brown, O. Mitchell, J. Poterba, and M. Warshawsky, Cambridge: MIT Press, 23–56.

Poterba, J. M. (2014): "Retirement security in an aging population," *The American Economic Review*, 104, 1–30.

Reed, W. J. and M. Jorgensen (2004): "The double Pareto-lognormal distribution–a new parametric model for size distributions," *Communications in Statistics-Theory and Methods*, 33, 1733–1753.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–229.

——— (2002): "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes," *Journal of Public Economics*, 83, 217–230.

Shourideh, A. and M. Troshkin (2015): "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.

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*, 44, 45.

WOO, L. AND D. BUCHHOLZ (2006): "Subsidies for Assets: A New Look at the Federal Budget," in *Assets Learning Conference – A Lifetime of Assets, sponsored by the Federal Reserve and CFED.*

Appendix

A Proofs

A.1 Proof of Proposition 1

We first show the following lemma:

Lemma 3. *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*

$$\theta \in \arg\max_{\hat{\theta}} U\left(c_{1,t}\left(\hat{\theta}\right), c_{2,t}\left(\hat{\theta}\right), \frac{y\left(\hat{\theta}\right)}{\theta}\right) \tag{25}$$

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 an 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. □

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 (25)), 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 convex, in the utility space. This property allows 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

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

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

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

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

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

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 $\lambda_t = \prod_{s=0}^{t}\left(1 + F_{K,s}\right)^{-1}$. 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). If we replace incentive compatibility constraint with its associated first order condition

$$U'\left(\theta\right) = \left(\frac{\beta'\left(\theta\right)}{\beta\left(\theta\right)} + \frac{P'\left(\theta\right)}{P\left(\theta\right)}\right)\beta\left(\theta\right) P\left(\theta\right) u\left(c_1\left(\theta\right)\right) + \frac{y\left(\theta\right)}{\theta^2} v'\left(\frac{y\left(\theta\right)}{\theta}\right)$$

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

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

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

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

The above implies that

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

When $\gamma \geq 0$, we must have that

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

which is the inequality stated in the proposition.

Note that the FOC's also imply that

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

which completes the proof. □

A.3 Proof of Lemma 1

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

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

subject to (4).Note that θ does not appear in the budget constraint.

Now take envelope condition with respect to θ

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

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

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

□

A.4 Proof of Lemma 2

Proof. Let $dH(\theta) = h(\theta)\,d\theta$ 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 (11), (12) and (13) respectively. The first order conditions for this problem are

$$\left(\eta(\theta) + \mu(\theta)\left(\frac{P_j'(\theta)}{P_j(\theta)} + j\frac{\beta'(\theta)}{\beta(\theta)}\right)\right) u'(c_j(\theta)) = \frac{1}{\beta^j(1+r)^j} \tag{26}$$

$$\left(\eta(\theta) - \mu(\theta)\frac{\varphi_j'(\theta)}{\varphi_j(\theta)}\left(1 + l_j(\theta)\frac{v''(l_j(\theta))}{v'(l_j(\theta))}\right) + \mu(\theta)\left(\frac{P_j'(\theta)}{P_j(\theta)} + j\frac{\beta'(\theta)}{\beta(\theta)}\right)\right) v'(l_j(\theta)) = \frac{\varphi_j(\theta)}{\beta^j(1+r)^j} \tag{27}$$

$$\eta(\theta) - \mu(\theta)\frac{h'(\theta)}{h(\theta)} - \gamma(\theta) = \mu'(\theta), \tag{28}$$

and the boundary conditions

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

Combine Equations (26) and (27) and let $\varepsilon_{F,j}(\theta) = \frac{v'(l_j(\theta))}{v''(l_j(\theta))l_j(\theta)}$.

$$\begin{aligned}\frac{v'(l_j(\theta))}{\varphi_j(\theta)\,u'(c_j(\theta))} &= \frac{\eta(\theta) + \mu(\theta)\left(\frac{P_j'(\theta)}{P_j(\theta)} + j\frac{\beta'(\theta)}{\beta(\theta)}\right)}{\eta(\theta) + \mu(\theta)\left(\frac{P_j'(\theta)}{P_j(\theta)} + j\frac{\beta'(\theta)}{\beta(\theta)}\right) - \mu(\theta)\frac{\varphi_j'(\theta)}{\varphi_j(\theta)}\left(1 + l_j(\theta)\frac{v''(l_j(\theta))}{v'(l_j(\theta))}\right)} \\ &= \frac{\eta(\theta) + \mu(\theta)\left(\frac{P_j'(\theta)}{P_j(\theta)} + j\frac{\beta'(\theta)}{\beta(\theta)}\right)}{\eta(\theta) + \mu(\theta)\left(\frac{P_j'(\theta)}{P_j(\theta)} + j\frac{\beta'(\theta)}{\beta(\theta)}\right) - \mu(\theta)\frac{\varphi_j'(\theta)}{\varphi_j(\theta)}\left(1 + 1/\varepsilon_{F,j}(\theta)\right)}.\end{aligned}$$

Therefore,

$$\begin{aligned}\tau_{\text{labor},j}(\theta) &= 1 - \frac{v'(l_j(\theta))}{\varphi_j(\theta)\,u'(c_j(\theta))} \\ &= \frac{-\mu(\theta)\frac{\varphi_j'(\theta)}{\varphi_j(\theta)}\left(1 + 1/\varepsilon_{F,j}(\theta)\right)}{\eta(\theta) + \mu(\theta)\left(\frac{P_j'(\theta)}{P_j(\theta)} + j\frac{\beta'(\theta)}{\beta(\theta)}\right) - \mu(\theta)\frac{\varphi_j'(\theta)}{\varphi_j(\theta)}\left(1 + 1/\varepsilon_{F,j}(\theta)\right)}\end{aligned}$$

and

$$\frac{\tau_{\text{labor},j}(\theta)}{1 - \tau_{\text{labor},j}(\theta)} = \frac{-\frac{\mu(\theta)}{\eta(\theta)}\frac{\varphi_j'(\theta)}{\varphi_j(\theta)}\left(1 + 1/\varepsilon_{F,j}(\theta)\right)}{1 + \frac{\mu(\theta)}{\eta(\theta)}\left(\frac{P_j'(\theta)}{P_j(\theta)} + j\frac{\beta'(\theta)}{\beta(\theta)}\right)} \tag{29}$$

Now, note that from Equation (26), $\eta\left(\theta\right)=\frac{1}{u'\left(c_0\left(\theta\right)\right)}$. Form (28) we can solve for $\mu\left(\theta\right)$

$$\begin{aligned}
\mu\left(\theta\right) &= -\frac{1}{h\left(\theta\right)}\int_{\theta}^{\bar{\theta}}\left(\eta\left(\tilde{\theta}\right)-\gamma\left(\tilde{\theta}\right)\right)dH\left(\tilde{\theta}\right)\\
&= -\frac{1}{h\left(\theta\right)}\int_{\theta}^{\bar{\theta}}\frac{1}{u'\left(c_0\left(\tilde{\theta}\right)\right)}\left(1-\gamma\left(\tilde{\theta}\right)u'\left(c_0\left(\tilde{\theta}\right)\right)\right)dH\left(\tilde{\theta}\right)\\
&= -\frac{1-H\left(\theta\right)}{h\left(\theta\right)}\int_{\theta}^{\bar{\theta}}\frac{1}{u'\left(c_0\left(\tilde{\theta}\right)\right)}\left(1-\gamma\left(\tilde{\theta}\right)u'\left(c_0\left(\tilde{\theta}\right)\right)\right)\frac{dH\left(\tilde{\theta}\right)}{1-H\left(\theta\right)}\\
&= -\eta\left(\theta\right)\frac{1-H\left(\theta\right)}{h\left(\theta\right)}\int_{\theta}^{\bar{\theta}}\frac{u'\left(c_0\left(\theta\right)\right)}{u'\left(c_0\left(\tilde{\theta}\right)\right)}\left(1-\gamma\left(\tilde{\theta}\right)u'\left(c_0\left(\tilde{\theta}\right)\right)\right)\frac{dH\left(\tilde{\theta}\right)}{1-H\left(\theta\right)}.
\end{aligned}$$

Therefore,

$$-\frac{\mu\left(\theta\right)}{\eta\left(\theta\right)}=\left(\frac{1-H\left(\theta\right)}{h\left(\theta\right)}\right)g\left(\theta\right),$$

where

$$g\left(\theta\right)=\int_{\theta}^{\bar{\theta}}\frac{u'\left(c_0\left(\theta\right)\right)}{u'\left(c_0\left(\tilde{\theta}\right)\right)}\left(1-\gamma\left(\tilde{\theta}\right)u'\left(c_0\left(\tilde{\theta}\right)\right)\right)\frac{dH\left(\tilde{\theta}\right)}{1-H\left(\theta\right)}.$$

By replacing these back into (29) we get Equation (17). □

A.5 Proof of Proposition 3

Proof. From Equation (26)

$$\frac{u'\left(c_j\left(\theta\right)\right)}{\beta(1+r)u'\left(c_{j+1}\left(\theta\right)\right)}=\frac{\eta\left(\theta\right)+\mu\left(\theta\right)\left(\frac{P_j'(\theta)}{P_j(\theta)}+j\frac{\beta'(\theta)}{\beta(\theta)}\right)}{\eta\left(\theta\right)+\mu\left(\theta\right)\left(\frac{P_j'(\theta)}{P_j(\theta)}+j\frac{\beta'(\theta)}{\beta(\theta)}\right)}$$

Therefore,

$$\begin{aligned}
\tau_{\text{annuity},j}(\theta) &= 1-\frac{u'(c_j(\theta))}{\beta(1+r)u'(c_{j+1}(\theta))}\\
&= \frac{\frac{\mu(\theta)}{\eta(\theta)}\left(\frac{P_j'(\theta)}{P_j(\theta)}-\frac{P_{j+1}'(\theta)}{P_{j+1}(\theta)}+\frac{\beta'(\theta)}{\beta(\theta)}\right)}{1+\frac{\mu(\theta)}{\eta(\theta)}\left(\frac{P_j'(\theta)}{P_j(\theta)}+j\frac{\beta'(\theta)}{\beta(\theta)}\right)}\\
&= \frac{-\frac{\mu(\theta)}{\eta(\theta)}\frac{p_{j+1}'(\theta)}{p_{j+1}(\theta)}}{1+\frac{\mu(\theta)}{\eta(\theta)}\left(\frac{P_j'(\theta)}{P_j(\theta)}+j\frac{\beta'(\theta)}{\beta(\theta)}\right)}\\
&= \frac{p_{j+1}'(\theta)}{p_{j+1}(\theta)}\left(\frac{1-H(\theta)}{h(\theta)}\right)\frac{g(\theta)}{1-\left(\frac{1-H(\theta)}{h(\theta)}\right)g(\theta)\frac{P_j'(\theta)}{P_j(\theta)}}.
\end{aligned}$$

The second equality is true because $\frac{P_j'(\theta)}{P_j(\theta)}-\frac{P_{j+1}'(\theta)}{P_{j+1}(\theta)}=\frac{p_{j+1}'(\theta)}{p_{j+1}(\theta)}$. □

A.6 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. Consider first order conditions (26), (27) and (28). Also, recall that

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

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 $j = 0$

$$1 \geq -\frac{\tau_{l,0}}{1-\tau_{l,0}} \frac{\varphi_0}{\varphi_{0,\theta}} \frac{\varepsilon}{1+\varepsilon} \left(\frac{h'}{h} + \frac{\tau'_{l,0}}{(1-\tau_{l,0})\,\tau_{l,0}} + \frac{\varphi_{0,\theta}}{\varphi_0} - \frac{\varphi_{0,\theta\theta}}{\varphi_{0,\theta}} + \sigma \frac{c'_0}{c_0} \right), \tag{30}$$

where $\sigma = -\frac{u_{cc,0}c_0}{u_{c,0}}$.

Note, also that combining (26) and (27) we get

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

Therefore

$$\frac{\tau_{l,j+1}}{1-\tau_{l,j+1}} = \frac{\tau_{l,j}}{1-\tau_{l,j}} \frac{\beta(1+r)u_{c,j+1}}{u_{c,j}} \frac{\varphi_{j+1,\theta}/\varphi_{j+1}}{\varphi_{j,\theta}/\varphi_j}. \tag{31}$$

Replace for

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

we get

$$\frac{\tau_{l,j+1}}{1-\tau_{l,j+1}} = \frac{\tau_{l,j}}{1-\tau_{l,j}} \frac{1}{p_{j+1}\left(1-\tau_{a,j+1}\right)} \frac{\varphi_{j+1,\theta}/\varphi_{j+1}}{\varphi_{j,\theta}/\varphi_j}. \tag{32}$$

Note also that

$$\begin{aligned} p_{j+1}\left(1-\tau_{a,j+1}\right) &= \frac{u_{c,j}}{\beta(1+r)u_{c,j+1}} \\ &= (\beta R) \frac{\eta + \mu\left(\frac{\dot{P}_{j+1}}{P_{j+1}} + (j+1)\frac{\dot{\beta}}{\beta}\right)}{\eta + \mu\left(\frac{\dot{P}_j}{P_j} + j\frac{\dot{\beta}}{\beta}\right)} \\ &= \frac{\frac{1-\tau_{l,0}}{\tau_{l,0}} \frac{\varphi_{0,\theta}}{\varphi_0} \frac{1+\varepsilon}{\varepsilon} - \left(\frac{\dot{P}_{j+1}}{P_{j+1}} + (j+1)\frac{\dot{\beta}}{\beta}\right)}{\frac{1-\tau_{l,0}}{\tau_{l,0}} \frac{\varphi_{0,\theta}}{\varphi_0} \frac{1+\varepsilon}{\varepsilon} - \left(\frac{\dot{P}_j}{P_j} + j\frac{\dot{\beta}}{\beta}\right)} \end{aligned} \tag{33}$$

Use this to replace $(1-\tau_{a,j+1})$ in equation (32) and combine the result with (30) we get

$$1 \geq -\frac{\frac{\tau_{l,j}}{1-\tau_{l,j}}\frac{\varphi_j}{\varphi_{j,\theta}}\frac{\varepsilon}{1+\varepsilon}}{1+\left(\frac{\tau_{l,j}}{1-\tau_{l,j}}\frac{\varphi_j}{\varphi_{j,\theta}}\frac{\varepsilon}{1+\varepsilon}\right)\left(\frac{\dot{P}_j}{P_j}+j\frac{\dot{\beta}}{\beta}\right)}\left(\frac{h'}{h}+\frac{\tau_{l,0}'}{(1-\tau_{l,0})\,\tau_{l,0}}+\frac{\varphi_{0,\theta}}{\varphi_0}-\frac{\varphi_{0,\theta\theta}}{\varphi_{0,\theta}}+\sigma\frac{c_0'}{c_0}\right),$$

which is the inequality (22). Also, the equation (23) is given by (33). □

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 (4) 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 4. *Consider an allocation $\{c_j(\theta), l_j(\theta)\}$ that satisfies implementability constraint (9). 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 (4) 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

$$1 - T_{y,j}'(\varphi_j(\theta) l_j(\theta)) = \frac{v'(l_j(\theta))}{\varphi_j(\theta) u_{c,j}(\theta)}, \tag{34}$$

$$u_{cj} = \beta(1+r) p_{j+1}(\theta)\left(1 - T_{a,j+1}'\right) u_{cj+1} \tag{35}$$

Equation (34) is the individual intra-temporal optimality condition and equation (35) is the individual Euler equation. We know from discussion in Section 4.2 that this Euler equation must be distorted at the efficient allocation. Therefore, optimal marginal taxes $T_{a,j+1}'$ are different from zero.

We can use equation (34) to back out the optimal marginal taxes on labor earning 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 (34) - 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 (34) and (35) together with (4) 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 (35), 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 6.2.

Finally, note that the monotonicity constraints in Lemma 4 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 Alternative Policy Reforms

In this section, we repeat our Pareto optimal reform exercise for a restricted set of policies. We first solve for optimal policies that do not include any old age transfers or any asset subsidies. This can be thought of as the best possible outcome that can be achieved through phasing out of social security and medicare. The goal of this exercise is to examine the importance of asset subsidies.

In the second exercise we solve for optimal polices that are restricted to only linear asset subsidies. The gaol of this exercise is to examine the importance of progressivity in asset subsidies.

Similar to our main exercise, both of these alternative policy reforms are implemented in an economy with current U.S. demographics and fixed factor prices (at the calibrated status quo values).

C.1 Privatization: no old age transfers and no asset subsidies

One particular proposal that has received considerable attention in the literature is *privatization* of social security. More precisely, this is the proposal to eliminate social security retirement benefits and reduce payroll taxes and move towards a save-for-retirement system.[38] These *privatization* polices differ from our optimal reform policy in two very important ways. First, our optimal reform policies do not involve a major adjustment in labor income taxes. Second, our optimal reform policies rely crucially on asset subsidies.

In this section we examine the importance of the asset subsidies. More precisely, we find the best reform policies that feature no old age transfers and no asset taxes/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.

To carry out this exercise we need to put restrictions on the type of policies available to the government. First, note that in the absence of asset taxes and subsidies, the individuals' consumption allocations must satisfy the following equation

$$\frac{P_j(\theta)\, u'(c_j(\theta))}{P_{j+1}(\theta)\, \beta(\theta)\, (1+r)\, u'(c_{j+1}(\theta))} = 1 \tag{36}$$

all ages j and for all ability types θ. This equation is simply the restriction that individuals do not face taxes/subsidies on their risk free asset returns.

In order to find the best policies that respect the no tax/subsidy restrictions, we solve planning problem (10) subject to constraints (11), (12), (13) and the no tax/subsidy constraint (36). Imposing constraint (36) guarantees that the allocation can be implemented without the need for asset

[38] See Nishiyama and Smetters (2007) and McGrattan and Prescott (forthcoming) for example.

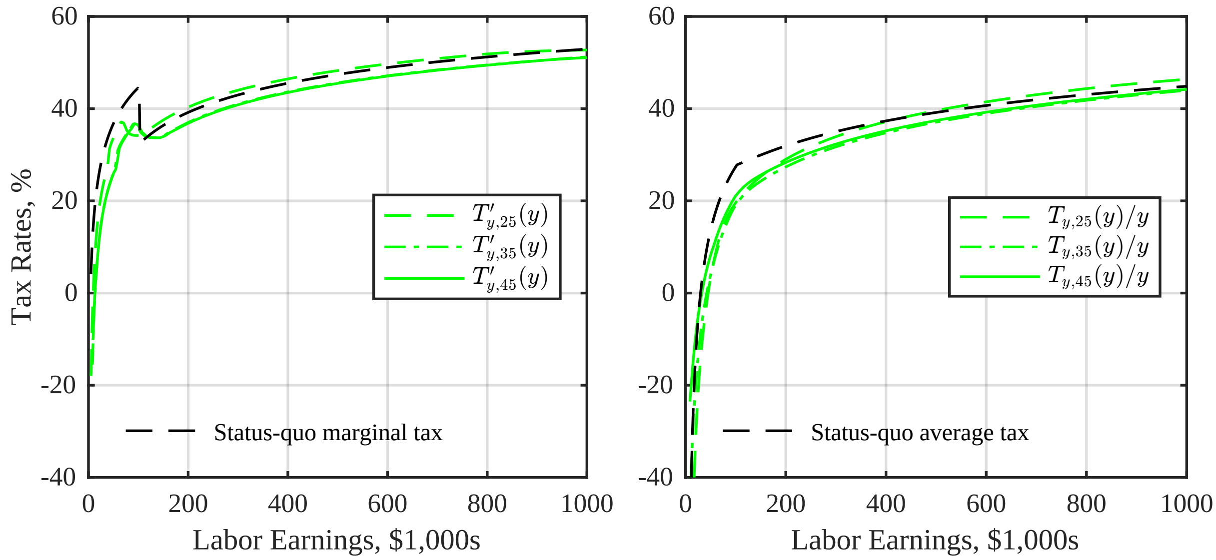

Figure 12: Optimal labor income tax functions with *privatization* (no old age transfers and no asset subsidies). The left panel are marginal taxes, while the right panel shows average taxes. The black dashed line is the effective status quo tax schedule.

taxes/subsidies. However, these allocations cost more than the allocations that result under fully optimal reform policies discussed in the main text. Note that we only impose restrictions on asset taxes and do not impose any restriction on earnings taxes.

We start by comparing labor income tax functions. Figure 12 (left panel) shows the optimal marginal taxes under privatization policies. Note that marginal rates are lower than status quo for almost all income levels. Moreover, for most income levels the drop in marginal taxes match 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. Right panel in Figure 12 shows the optimal average taxes.

Table 7 shows changes in the aggregate variables. Note that under privatization policies present discounted value of consumption, net of labor income, rises relative to status quo. [39]

Finally, these policies lead to slightly lower stock of capital. Consumption and output are

[39]This is in contrast to McGrattan and Prescott (forthcoming) who find a Pareto improving privatization policy. Their finding relies on the choice of elasticity of labor supply. As we show in appendix D.1, when elasticity of labor supply is high, it is possible to reduce resource cost through privatization and hence a Pareto improving privatization policy exists.

Table 7: Aggregate effects of privatization

	Current U.S.	Optimal Reform	Privatization
Values relative to GDP			
Consumption	0.70	0.67	0.70
Capital	3.00	3.43	2.99
Tax revenue (total)	0.25	0.26	0.23
Earnings tax	0.15	0.15	0.13
Consumption tax	0.04	0.04	0.04
Capital (corporate) tax	0.06	0.07	0.06
Transfers	0.14	0.13	0.04
To retirees	0.09	0.03	0.00
To workers	0.05	0.03	0.04
Asset subsidies	0.00	0.07	0.00
Change relative to status quo (%)			
GDP	–	7.15	0.15
Consumption	–	2.59	0.22
Capital	–	22.34	-0.08
Labor input	–	-1.39	-0.28
PDV of net resources	–	-5.15	2.25

slightly higher due to higher labor supply. The labor supply will be higher because of the lower tax on labor income (see Figure 12).

C.2 Optimal linear asset subsidies

The central feature of our optimal reform policies is the nonlinear subsidies on asset. There are two reasons for nonlinearity of these subsidies. As we showed in equation (21), the marginal rates can be decomposed to two terms. One is the type specific mortality rate, and the other is the efficient inter-temporal (annuity) wedge. Both these terms crucially depend on mortality heterogeneity. To gauge the importance of mortality heterogeneity – and therefore the nonlinearity in assets subsidies – we experiment with the set of policies that only use linear asset subsidies (or taxes).

Linearity restriction on asset tax/subsidies impose restrictions on set of implementable allocations. To be more precise, in order for allocations to be implementable by linear asset tax/subsidies, the following condition must hold

$$\frac{P_j(\theta)\, u'(c_j(\theta))}{P_{j+1}(\theta)\, \beta(\theta)\,(1+r)\, u'(c_{j+1}(\theta))} = \frac{P_j(\theta')\, u'(c_j(\theta'))}{P_{j+1}(\theta')\, \beta(\theta')\,(1+r)\, u'(c_{j+1}(\theta'))} \tag{37}$$

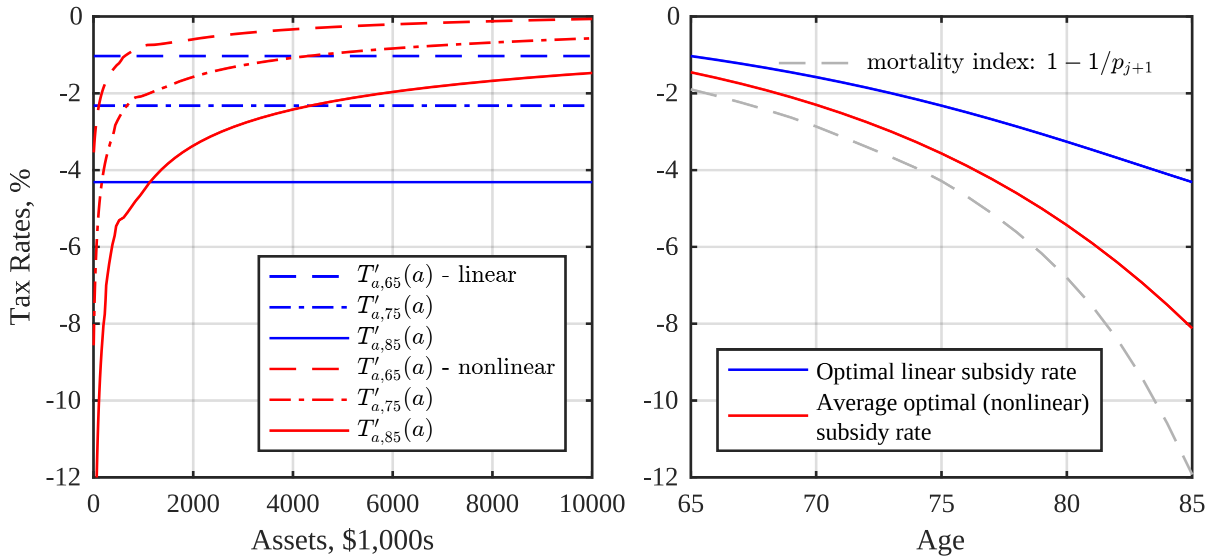

Figure 13: Optimal asset tax functions: linear subsidies vs nonlinear subsidies. The left panel shows marginal taxes over all asset levels at ages 65, 75 and 85, while the right panel shows average marginal rates at each age from 65 to 85. Blue lines are optimal linear subsidies. Red lines on fully optimal nonlinear subsidies. The dashed line is the population mortality index.

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

In order to find the best policies that respect the linear asset tax/subsidy restrictions, we solve planning problem (10) subject to constraints (11), (12), (13) and the linear tax constraint (37). Imposing constraint (37) guarantees that the allocation can be implemented by a linear set of asset tax/subsidies.

It is important to remind the reader that in our model, even in the absence of differential mortality and differential discount factor, the optimal subsidies on assets are not zero. If there is no annuity market, the asset subsidies are needed to correct the inefficiency due to incomplete market. In that case these subsidies will be linear and the rate will be equal to average population mortality at each age. In Figure 13 (right panel) we plot the optimal linear asset subsidies in our model along with 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 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 14 shows the marginal labor income tax functions. Note that linearity restriction on asset taxes/subsidies results in negative marginal tax on labor income for poorest individuals (left

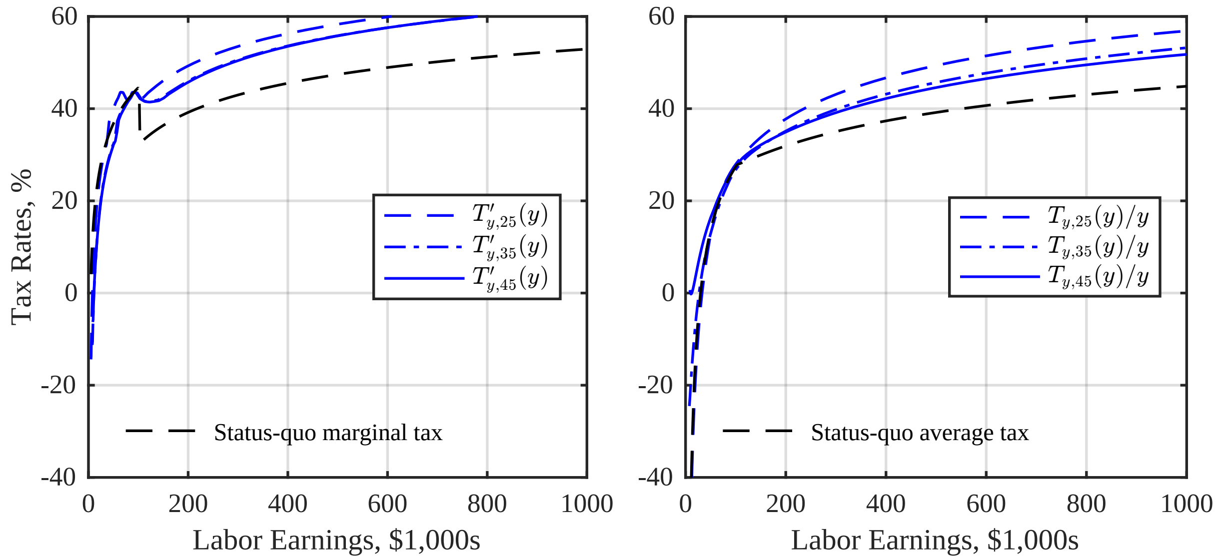

Figure 14: Optimal labor income tax functions with linear asset subsidies. The left panel are marginal taxes, while the right panel shows average taxes. Right panel are average taxes. The black dashed line is the effective status quo tax schedule.

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, 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 labor income tax must be low (even negative). Also, note that tax rates at the top is higher under policies with linear subsidy restriction. Linear subsidies imply that high income individuals receive too much asset subsidies (relative to full optimal). Therefore, again, the burden of redistribution in on the labor income tax to correct this excess subsidization of the high income workers.

Table 8 shows the effect on aggregate variables. Aggregate output, consumption and capital is affected similarly to the fully optimal reform. However, restricted policies only achieve a fraction of cost saving that is achieved by the fully optimal reform policies.

D Robustness

In this section we examine the robustness of our findings with regards to the choice of labor supply elasticity. All calculations are performed in an economy with current U.S. demographics and fixed factor prices (at the calibrated status quo values).

Table 8: Aggregate effects of linear subsidies

	Current U.S.	Optimal Reform	Linear subsidies
Values relative to GDP			
Consumption	0.70	0.67	0.67
Capital	3.00	3.43	4.45
Tax revenue (total)	0.25	0.26	0.25
Earnings tax	0.15	0.15	0.14
Consumption tax	0.04	0.04	0.04
Capital (corporate) tax	0.06	0.07	0.07
Transfers	0.14	0.13	0.08
To retirees	0.09	0.03	0.00
To workers	0.05	0.03	0.01
Asset subsidies	0.00	0.07	0.08
Changes relative to status quo (%)			
GDP	–	7.15	7.88
Consumption	–	2.59	3.02
Capital	–	22.34	24.09
Labor input	–	-1.39	-1.23
PDV of net resources	–	-5.15	-3.19

D.1 Labor supply elasticity

In our benchmark calibration we assume the Frisch elasticity of labor supply to be $\varepsilon = 0.5$ which is in the range estimated in micro studies and very common in quantitative life cycle macroeconomic models.[40] 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 (see McGrattan and Prescott (forthcoming) for example). We re-calibrate our model using this value for ε. The calibrated parameter are presented in Table 9.

We first check the optimality of status quo policies using the conditions derived in Proposition 4. The results are demonstrated in Figure 15. Note that the earning tax schedule fails the test (left panel) not only at the social security earning cap, but also at the very top income levels. This is expected. When 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 tax rates for some individuals (and improve their welfare) while increasing tax revenue. However, this applies to a small fraction of earnings distribution.

As in the benchmark case, the status quo asset taxes (or absence of them) also fails the efficiency test. However, notice that the term $D_j(\theta)$ (drawn in black) are larger here (relative to

[40] See Keane (2011) and Chetty et al. (2011); Chetty (2012).

Table 9: Calibrated Parameters – high labor supply elasticity

Parameters	Description		Values
β_0	discount factor: level		0.974
β_1	discount factor: elasticity w.r.t θ		0.006
ψ	weight on leisure		0.716
Targeted Moments		Data	Model
Wealth-income ratio		3	3
Wealth gini		0.78	0.78
Average annual hours		2000	2000

benchmark – compare to Figure 4). This means that we should expect the optimal asset subsidies be smaller in the case where elasticity of labor supply is high. We plot these optimal asset taxes in Figure 16.

Finally, Figure 17 shows the optimal labor income taxes. The left panel are optimal taxes in the full optimal reform. The right panel are optimal taxes in the restricted reform that does not feature asset subsidies and old age transfer (i.e., privatization). Two observations here. First, the optimal labor income taxes are much lower that status quo, especially for high income level. This is a contrast to the benchmark calculations which feature lower elasticity of labor supply. Second, the optimal tax rates in the full optimal reform and privatization reform are very similar (although, not identical).

We report the aggregate effects of full optimal reform and privatization reform in Table 10. With high labor supply elasticities gains in efficiency are larger. Moreover, there are gains even in the privatization reform where asset subsidies are restricted to be equal zero (and there are no old age transfers). However, without utilizing the optimal asset subsidies, only a third of the cost saving can be achieved.

These findings are in line with McGrattan and Prescott (forthcoming) who report the existence of a Pareto improving privatization reform for their benchmark model that has elasticity of labor supply equal to 2.6. They also report results with labor supply elasticity of 0.5 and find that their privatization reform is not Pareto improving. This is consistent with our findings in appendix C.1.

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.

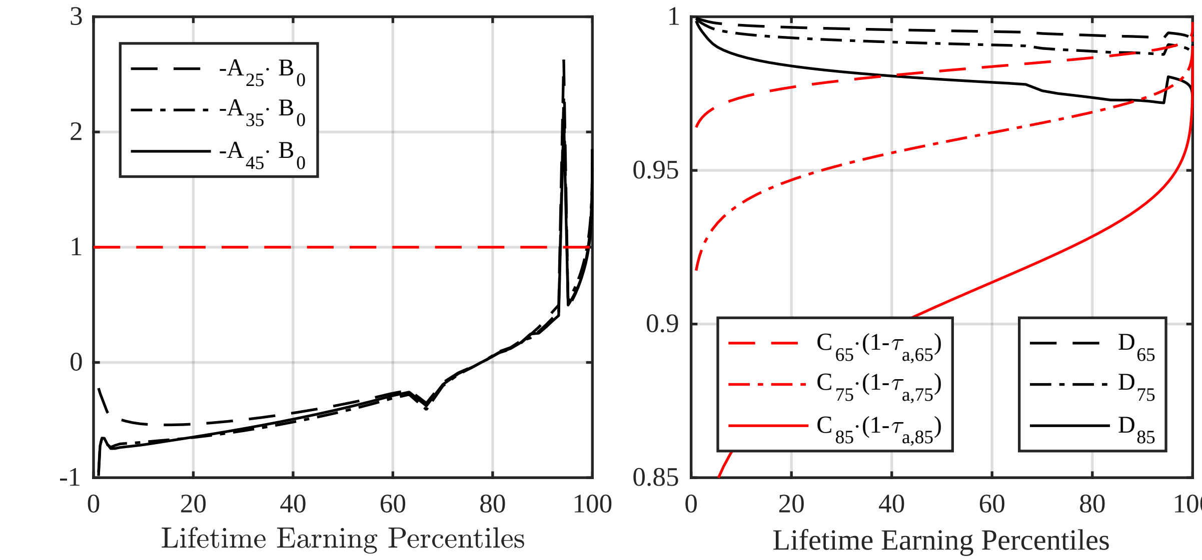

Figure 15: Test of Pareto optimality for status quo policies – high elasticity of labor supply. The left panel is the test of earnings tax schedule. The black lines are the left hand side of inequality (22). The test fails only at the social security maximum taxable earnings' cap. The right panel is the test of asset tax. The black lines are the right hand side of Equation (23). The red lines are the left hand side of that equation (equal to survival rate). This test fails everywhere.

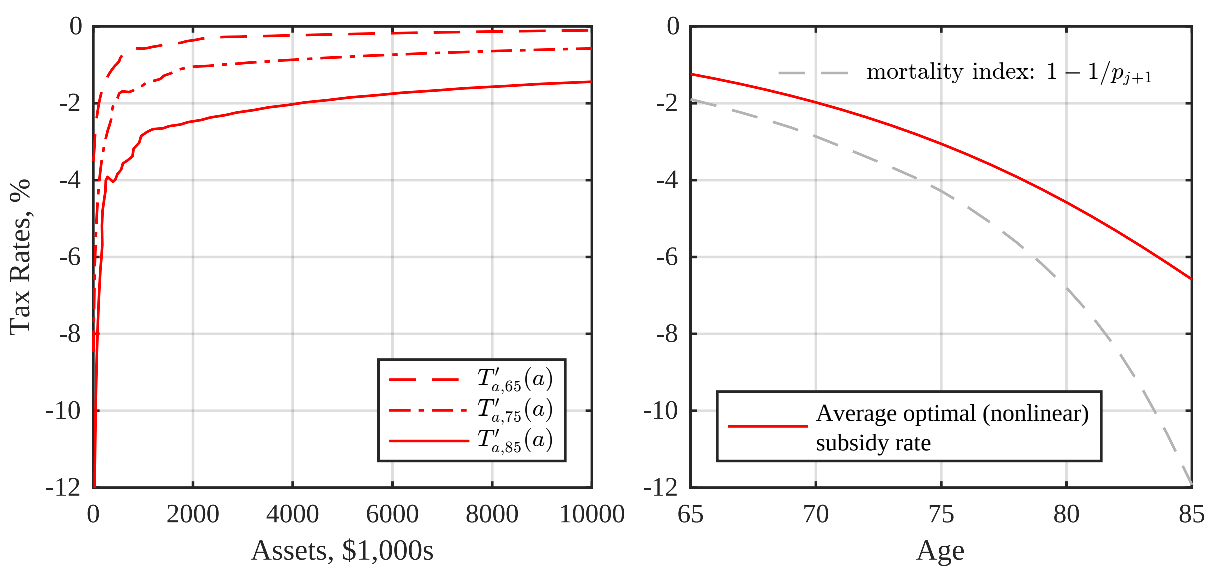

Figure 16: Optimal asset tax functions – high elasticity of labor supply. The left panel shows marginal taxes over all asset levels at ages 65, 75 and 85, while the right panel shows average marginal rates at each age from 65 to 85. The dashed line is the population mortality index.

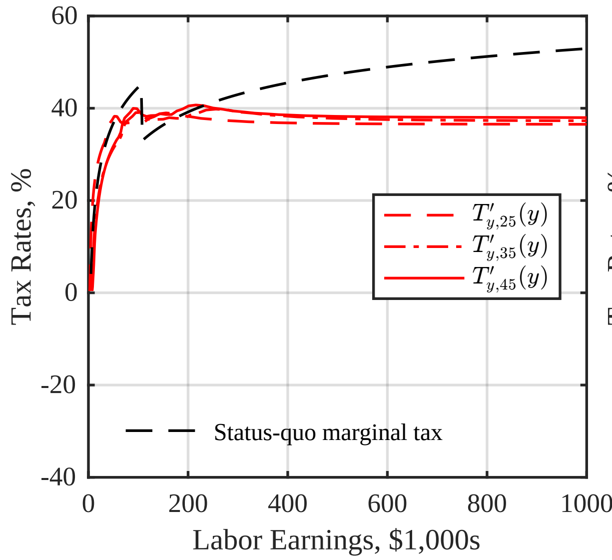

Figure 17: Optimal asset tax functions – high elasticity of labor supply. The left panel shows optimal marginal taxes under fully optimal reform. The right panel shows optimal marginal taxes under privatization. The dashed line is the status quo effective tax rate.

Table 10: Aggregate effects of reforms – high labor supply elasticity

	Current U.S.	Optimal Reform	Privatization
Values relative to GDP			
Consumption	0.70	0.67	0.69
Capital	3.00	3.43	3.16
Tax revenue (total)	0.26	0.26	0.24
Earnings tax	0.16	0.15	0.14
Consumption tax	0.04	0.04	0.04
Capital (corporate) tax	0.06	0.07	0.06
Transfers	0.15	0.13	0.08
To retirees	0.09	0.03	0.01
To workers	0.06	0.05	0.07
Asset subsidies	0.00	0.05	0.00
Changes relative to status quo (%)			
GDP	–	7.13	3.32
Consumption	–	2.54	1.72
Capital	–	22.42	8.67
Labor input	–	-1.48	0.31
PDV of net resources	–	-5.15	-1.93