Retirement Financing: An Optimal Reform Approach

Hosseini, Roozbeh and Shourideh, Ali

University of Georgia, University of Pennsylvania

19 January 2016

Retirement Financing: An Optimal Reform Approach*

Roozbeh Hosseini
Arizona State University
roozbeh.hosseini@asu.edu

Ali Shourideh
Wharton School
shouride@wharton.upenn.edu

January 19, 2016

Abstract

We study policy reforms aimed at overhauling retirement financing. We develop a novel approach by considering *optimal reforms*: policy reforms that minimize the cost for the government while respecting the distribution of welfare in the economy. Our model is an OLG model with life-cycle features and bequest motives where individuals are heterogeneous in their earning ability and mortality. Theoretically, we show that due to the negative correlation between earnings ability and mortality, post-retirement distortions to saving decisions are a robust feature of any optimal policy. We, then, use this framework to quantitatively analyze optimal reforms. Our quantitative exercise shows that an optimal reform relative to the status-quo must have three key features: First, post-retirement assets must be subsidized while bequests must be taxed. On average, optimal marginal subsidies on assets for individuals above age 65 is 3.2 percent, while optimal marginal tax on their bequest is 60 percent. Second, pre-retirement transfers must increase while social security benefits must become less generous in the aggregate and more progressive towards low income groups. Finally, earnings tax reform does not contribute to optimal reforms, i.e., optimal marginal taxes on earnings remain very close to the status-quo. The optimal policies reduce the present discounted value of net tax and transfers to each generation by 15 percent.

*We would like to thank Berthold Herrendorf valuable comments and discussions.
Please visit http://www.public.asu.edu/~rhossein for updates.

1 Introduction

Government in the United State has a big part in financing retirement consumption. Social security benefits are 40 percent of all income of the elderly. These benefits are the main source of income for half of the elderly population. [1] Currently, social security payouts are 30 percent of government expenditures and payroll taxes are 30 percent of total federal income taxes collected. The coming demographic shift will soon tilt the balance towards more outlays and less revenue. This will pose a significant fiscal challenge on the government budget. The severity of these costs together with an aging population has made reforms in the retirement system a necessity. Any reform, in order to be most effective, must contain three ingredients: 1. a reform of the tax and transfer system, 2. a reform of the saving incentives induced by the tax code, 3. respect distributional concerns in a politically viable fashion.

Various reforms by policy makers and academics have been proposed as policy alternatives that either reduce the cost of retirement financing, or increase revenues for the government. Some examples include: privatization of social security, lower capital and labor taxes and a phasing out of social security and medicare, higher cap on payroll, etc. to name a few.[2]

In this paper, we propose a new approach to policy reform regarding retirement financing by considering *optimal reforms*. To do so, we look for optimal government tax and transfer policies that respect the distributional concerns in the population. In other words, we find policies that minimize the present value of outlays net of receipts for the government while respecting the status-quo distribution of welfare. In doing this we do not impose any restriction on the set of policy instruments at the outset. This approach has two advantages. First, it allows us to identify which policies are the most important in a reform. Second, it allows us to separate the notion of policy reform from that of redistribution.

Our key findings regarding the main ingredients of an optimal reform are as follows. First, post retirement assets must be subsidized while bequests must be taxed. On average, optimal marginal subsidies on assets for individuals above age 65 is 3.2 percent, while optimal marginal tax on their bequest is 60 percent. Second, pre-retirement transfers must increase while social security benefits must become less generous in the aggre-

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

[2] Nishiyama and Smetters (2007) study the effect of social security privatization, McGrattan and Prescott (2013) study a proposal of phasing out the social security and medicare tax and transfers , Blandin (2015) studies the effect increasing maximum taxable earning cap.

gate and more progressive towards low income groups. Finally, earnings tax reform does not contribute to optimal reforms, i.e., optimal marginal taxes on earnings remain very close to the status-quo.

Our framework for optimal reform is an overlapping generations economy with life cycle features. In our model, individuals differ in two aspects: (a) productivity profiles that evolve with age and determine individuals earning ability at each age and (b) mortality profiles that determine probability of death at each age. The novel feature of our analysis is the negative correlation between income (or earning ability) and mortality. This assumption is motivated by a large number of studies documenting higher death rates at any age for individuals in lower income groups.[3] Other key features of our model are that individuals have preference over bequest and retirement age is exogenous.

As we show, this model is able to match basic characteristics of the distribution of consumption, earnings and bequests given the absence of annuity markets and the status-quo tax/transfer and social security system in the United States. [4] In our optimal reform exercise, we take as given the distribution of welfare implied by this model. We then solve a planning problem that minimizes the present discounted value of net transfers to each cohort in steady state subject to two constraints: i) individual's incentive compatibility (which is equivalent to saying that any policy we consider is consistent with individual optimization and their budget constraint); ii) the resulting allocation delivers life time welfare to each individual that is not lower than their lifetime welfare in the status-quo economy.

Our first main policy reform is a direct consequence of the negative correlation between mortality and productivity.[5] To understand the role of asset distortions late in life, consider first the hypothetical situation where markets are complete and individuals can purchase annuity and life insurance.[6] Due to a lower mortality, more productive individuals place more value on future consumption relative to less productive individuals. Therefore, if the annuity income of a low income individual is taxed, the high ability individuals choose to earn higher when young and benefit from lower (effective) taxes on their asset income in older ages when they survive with a higher probability.[7] Not

[3]See for example Waldron (2013) and Cristia (2009), Rogot et al. (1992), Duggan et al. (2007) and Brown (2001) among many others.

[4]The assumption about the annuity market is motivated by a large literature that documents the small private annuity market in the United States (see Benartzi et al. (2011) for an excellent review).

[5]In the absence of this assumption the usual Atkinson and Stiglitz (1976) uniform commodity taxation holds there is no inter-temporal distortion.

[6]In our model, life insurance demand comes from individuals' bequest motives.

[7]This result is similar to the insights from optimal commodity taxation as in Saez (2002) and Golosov et al. (2013).

surprisingly, distortions to the inter-temporal margin are larger late in life since the probability of death is higher and the heterogeneity in survival is more pronounced.

When markets are incomplete and individuals do not have access to annuities, the above mentioned distortions must be corrected to reflect the effect of missing annuity markets. In particular, individuals must receive subsidies on their savings upon survival. This subsidy reflects the fact that the non-contingent return on saving is too low. Hence, optimal asset taxes are a combination of the correctional subsidies and the taxes discussed in the previous paragraph. The former is higher the higher is mortality rate while the latter is higher the higher is mortality differentials. Since mortality differentials are lower in magnitude than mortality levels, assets must be subsidized at the optimum for individuals who survive. The logic for taxing bequests and life insurance is exactly the reverse. That is, in the absence of the annuity market, individuals hold too much asset due to precautionary motive. To reduce this incentive (without undoing the incentive and correction effect of asset subsidies), bequest must be taxed.

Guided by these qualitative analysis, we propose an optimal system of tax and transfers. The key ingredients of our proposed policy are nonlinear subsidies on assets as long as individuals are alive, and nonlinear taxes on their bequeathed assets when they die. As argued in the previous paragraph, mortality rate and mortality rate differentials are important determinants of these taxes and subsidies. Therefore, these optimal tax functions vary by age. We find that optimal asset subsidies are progressive. At any age, the subsidies that asset poor individuals receive (as fraction of their total asset) is higher than the subsidies that asset rich individuals receive. As an example, a 70 years old person with very low asset balance receives an average subsidy of 5 percent (on entire asset), while an asset rich individual at the same age, receives as low as 1 percent . The magnitude of subsidies increase with age. For example, average subsidies at the bottom and top of wealth distribution at age 80 is 15 percents and 2 percent respectively. Overall, the average marginal subsidy rate on assets range from 2 percent at age 65 to 5 percent at age 90.

On the flip side, optimal tax on bequests are regressive. In particular, the tax rates at the bottom of the wealth distribution (for all ages) equal to 100 percent and fall to lower levels for richer individuals (60 percent for 65 year olds at the top of wealth distribution and below 10 percent for same individuals in their 90s). Overall, the average marginal tax on bequest falls from about 90 percent at age 65 to 40 percent at age 90 (over all individual who die at each age).

Our main optimal policy reform has very little *direct* effect on government budget. Under the optimal policy, the government pays around 5 percent of GDP in asset subsidies, while it collects about 4 percent of GDP from taxes on bequests. However, the combination of the asset subsidies and bequest taxes greatly improves individuals' incentive to save. This results in a higher stock of capital in the economy which in turn generates more revenue from taxation of corporate income. This added revenue together with high taxes on bequests is enough to finance asset subsidies.

Our exercise also points toward the required reform in the government's tax and transfer policy. With regards to taxes, an interesting outcome is that earnings tax schedules almost remain unchanged between the status-quo economy and the solution of our optimal policy problem. That is given the observed preference for redistribution - as a consequence of the political process - labor taxes are not far from optimal. With regards to transfers, our model's main implication is that while the aggregate government transfers (to individuals pre- and post- retirement) should not change, its timing must change significantly. In particular, transfers pre-retirement must increase by 87 percent while social security benefits should decline by 14 percent.

Asset subsidies are central to our proposed optimal policy. These subsidies resemble some of the features of the US tax code, and indeed 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 few examples of such programs. The estimated cost of these programs is $367 billion in year 2005 (about 2.8 percent of GDP). Moreover, the benefit from these programs goes mostly to higher income individuals. [8] 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 current observed outcome). An important feature of the US tax code is that it penalizes accumulation of asset in tax deferred accounts beyond the age of $70\frac{1}{2}$. However, US tax code imposes no tax on transferring the balance on these accounts in the event of death. These features are completely at odd with optimal policy.

1.1 Related Literature

Our paper contributes to various strands in the literature on policy reform. We contribute the large growing literature that studies retirement financing. Most of this literature studies the implications of a specific set of policy proposals. For example, Nishiyama and Smetters (2007) studies the effect of privatization of social security, Kitao (2014) compares different combination of tax increase and benefit cuts within the current social security

[8] See Woo and Buchholz (2006).

system, McGrattan and Prescott (2013) propose phasing out social security and medicare tax and transfers, and Blandin (2015) studies the effect of increasing earnings cap. An important benefit of our optimal reform approach is that 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. For example, we find that changing the marginal tax rates on labor earnings is not a major contributor to an optimal policy reform. Almost, all other papers in the literature study policies that affect earnings taxes.

Our paper is also related to a large literature on optimal policy design. The common approach in this literature is to take stand on a specific social welfare criteria and find optimal polices that maximize the social. For example, Conesa and Krueger (2006) and Heathcote et al. (2014) study the optimal progressivity of tax formula for a paramedic 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 parametric restriction on policy instruments (similar to our paper). One drawback of this approach is that it relies on the choice of 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 holds distribution of welfare in the economy as 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. [9]

Our paper also makes a contribution to the literature on dynamic optimal taxation over the life cycle. Similar to Weinzierl (2011), Golosov et al. (forthcoming) and Farhi and Werning (2013) we provide analytical expressions for distortions and summarize insights from those expressions. However, unlike these cited works who focus on labor distortions over life cycle, our focus is on inter-temporal distortions. Furthermore, we emphasis the role for policy during the retirement period. This also relates our work to Golosov and Tsyvinski (2006), who study the optimal design of disability insurance system, and Shourideh and Troshkin (2015), who focus on an optimal tax system that provides incentive for efficient retirement age.

Golosov et al. (2013), Saez (2002) and Piketty and Saez (2013) argue that when high earners prefer certain types of goods, taxation of these goods can contribute positively to redistribution and should be used. While our result has a similar flavor, there is a key difference. That mortality not only affects preferences (through calculation of future expected utility) but also it affects resources. In particular, in our model, absent heterogeneity in labor productivity, there will be no inter-temporal distortion, even in presence of mortality differentials. This is not necessarily true when the only difference across in-

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

dividuals is preference heterogeneity. In this regard our paper is close to Bellofatto (2015) who uses the correlation between ability and life expectancy in a dynastic framework to study the optimal estate taxation and inter-generational transfers.

Finally, this paper is related to literature that studies the role of social security in providing longevity insurance. Hubbard and Judd (1987), İmrohoroğlu et al. (1995) and Hong and Ríos-Rull (2007) and Hosseini (2015) (among many other) have examined the welfare enhancing role of providing annuity income through social security when there is imperfections in the private annuity insurance market. Caliendo et al. (2014) point out that because social security does affect individual's inter-temporal trade-offs, its welfare enhancing role in providing annuitization is limited. In the current paper, we precisely point to the optimal distortions and policies that address this shortcoming in the system by pointing out that any optimal retirement system (whether public, private or mixed) must include features that affect individuals' inter-temporal decisions on the margin. In our proposed implementation that takes the form a nonlinear subsidy on assets.

2 Model

In this section, we develop a heterogenous-agent overlapping generations model suitable for our policy analysis. We calibration this model in Section 4 and show that it is consistent with US aggregate data and cross section observations on earnings and asset distribution.

2.1 Demographics and Endowment

Time is discrete and the economy is populated by T overlapping generations. The population grows at a constant rate n. A cohort of individuals are born in each period. Upon birth, each individual draws a type $\theta \in \Theta = \left[\underline{\theta}, \overline{\theta}\right]$ from a continuous distribution $F(\theta)$ that have density $f(\theta)$ This type parameter determines the individual's labor productivity profile over the course of the life-cycle together with the distribution of survival. In particular, an individual of type θ, has labor productivity of $\varphi_t(\theta)$ at age t. We often refer to θ as life-time productivity. Everyone retires at age R and $\varphi_t(\theta) = 0$ for $t > R$. [10]
An individual of type θ has survival rate $p_{t+1}(\theta)$ (this is the probability of being alive in age $t+1$, conditioned on being alive at age t). Nobody survives past age T (with

[10] Most of our analysis is done for an economy in steady state. Therefore, we drop the dependence of allocations to calendar date and t represents age of the individual.

$p_{T+1}(\theta) = 0$ for all θ). Unconditional probability of survival to each age t is

$$P_t(\theta) = \Pi_{s=0}^{t} p_s(\theta).$$

Let $\mu_t(\theta)$ be fraction of type of θ population that has age t and is alive. Therefore, Type θ at age t makes up fraction $f(\theta) \cdot \mu_t(\theta)$ of the population total population

$$\mu_{t+1}(\theta) = \frac{p_{t+1}(\theta)}{1+n} \cdot \mu_t(\theta),$$

$$\sum_{t=0}^{T} \int_{\theta} f(\theta) \cdot \mu_t(\theta) d\theta = 1,$$

$$\sum_{t=0}^{T} \mu_t(\theta) = 1 \quad \forall \theta.$$

2.2 Preferences and Technology

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

$$\sum_{t=0}^{T} \beta^t P_t(\theta) \left[u(c_t) - v(l_t) + \beta \left(1 - p_{t+1}(\theta)\right) w(b_{t+1}) \right], \tag{1}$$

where β is the subjective discount factor, $P_t(\theta)$ is probability of survival to age t and $1 - p_{t+1}(\theta)$ is mortality rate at end of age t. Individual who is alive at age t, enjoy living bequest on b_{t+1} if he dies at the end of period.

We assume that the economy-wide production function uses capital and labor and is linear. That is, output at each date is given by $(\tilde{r} + \delta)K + L$ where K is aggregate stock of capital and L is the aggregate effective units of labor.[11] Effective labor is defined as labor productivity, $\varphi_t(\theta)$ multiplied by l_t while its aggregate value is the sum of effective labors across all individuals alive in each period. In other words,

$$L = \int \sum_{t=0}^{T} \mu_t(\theta) \varphi_t(\theta) l_t(\theta) dF(\theta),$$

where $l_t(\theta)$ is the labor supply by an individual of type θ at age t.

[11] We normalize the wage per effective unit of labor to 1.

2.3 Markets and Government

We assume that individuals supply labor in the labor market and earn wage $w = 1$ per unit of effective labor. In addition, individuals have access to a risk-free asset. Upon their death, the risk-free assets convert to bequests. In other words, we assume that annuity markets do not exist. This assumption is in line with the observed low volume of trade in annuity markets in the United States and other countries.

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

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

$$\begin{aligned}
(1+\tau_c)\,c_t + a_{t+1} &= \left(\varphi_t l_t - T_y^s(\varphi_t l_t) + Tr_t^s\right)\mathbf{1}\,[t<R] \\
&\quad + (1+\tilde{r}\,(1-\tau_k))\,a_t + S_t^s(Y_t)\,\mathbf{1}\,[t\geq R]\,, \qquad (2) \\
b_t &= (1+\tilde{r}\,(1-\tau_k))\,(a_{t+1}+B) - T_b^s\left((1+\tilde{r}\,(1-\tau_k))\,a_{t+1}\right). \qquad (3)
\end{aligned}$$

where $T_y^s(\cdot)$ is the income tax function on earnings from labor, Tr_t^s are transfers to working individuals, S_t^s is retirement benefit from the government, and $T_b^s(\cdot)$ is taxes on bequests – superscript s stands for status-quo. We assume that bequests are collected and distributed as lump-sum transfer B to the entire population. The dependence of retirement benefits on life-time earnings is captured in Y_t which is given by

$$Y_t = \frac{1}{R+1}\sum_{t=0}^{R}\varphi_t l_t.$$

The rate τ_K is the effective marginal tax rate on capital gains. Since, $\tilde{r}$ is the return on capital net of depreciation before any taxation in the corporate sector, τ_K should be thought of as the combination of 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 given the after tax interest rate earned on all types of assets which we refer to as r. We assume that the government taxes household's holding of government debt at an equal rate and therefore, the interest paid on government debt is r.

Given the above assumptions, government budget constraint is given by

$$
\begin{aligned}
\tau_C \int \sum_{t=0}^{T} \mu_t(\theta) c_t(\theta) dF(\theta) + \tau_K \tilde{r} K + \int \sum_{t=0}^{T} \mu_t(\theta) T_y^s(\varphi_t(\theta) l_t(\theta)) dF(\theta) & \\
+ \int \sum_{t=0}^{T} \mu_t(\theta)(1 - p_{t+1}(\theta)) T_b^s((1+r) a_{t+1}(\theta)) dF(\theta) &= \\
\int \sum_{t=0}^{R} \mu_t(\theta) Tr_t^s dF(\theta) + \int \sum_{t=R+1}^{T} \mu_t(\theta) S_t^s(Y_t(\theta)) dF(\theta) + G + (r-n) D. & \quad (4)
\end{aligned}
$$

In addition, G is per-capita government purchases while D is per-capita government debt – thus its interest should be adjusted for the fact that population is growing at rate n. Finally goods and asset market clearing implies that

$$
\begin{aligned}
\int \sum_{t=0}^{T} \mu_t(\theta) c_t(\theta) dF(\theta) + G + nK &= \tilde{r} K + \int \sum_{t=0}^{T} \mu_t(\theta) \varphi_t(\theta) l_t(\theta) dF(\theta) & (5) \\
K &= \int \sum_{t=0}^{T} \mu_t(\theta)(a_t(\theta) + B) dF(\theta) - D & (6) \\
B &= \int \sum_{t=0}^{T} \mu_t(\theta)(1 - p_{t+1}(\theta))(a_{t+1}(\theta)) dF(\theta) & (7)
\end{aligned}
$$

where K is the steady-state level of per capita capital.[12]

Equilibrium. A steady-state equilibrium of this economy is thus defined as allocations where individuals maximize 1 subject to (2) and (3) while government budget constraint (4), market clearings (5) and (6) must hold.

This sums up our description of the economy. In the next section, we describe our approach to analyze optimal reform within the framework specified above. Note that at this point, we have not specified any details about the status-quo policies yet. We will do that in Section 4 where we impose detailed parametric specifications of the US tax and social security policies and calibrate this model to the US 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 by $W^s(\theta)$.

[12]Equation (5) is the national accounting identity with the term δK is omitted from both sides.

3 Optimal Policy Reform: Theoretical Framework

Our optimal policy reform exercise builds on the positive description of the economy in Section 2. In particular, we use the distribution of welfare implied by the model in Section 2 and consider a planning problem that minimizes cost of delivering this distribution of welfare, the utility profile $\{W^s(\theta)\}_{\theta\in\Theta}$, while allowing for a general set of policies. This approach allows us to separate distributional concerns from efficiency concerns.

3.1 A Planning Problem

The set of policies that we allow for in our optimal reform are very similar to those described in Section 2. In particular, we allow for non-linear and age-dependent taxation of assets as well as that of bequests. Moreover, we allow for non-linear and age-dependent taxation of labor income together with flat social security benefits, i.e., social security benefits are independent of life-time earnings. Therefore, given this tax and benefit structure, each individual maximizes utility (1) subject to the following budget constraints:

$$(1+\tau_c)\,c_t + a_{t+1} = \left(\varphi_t l_t - T_{y,t}\left(\varphi_t l_t\right)\right)\mathbf{1}\left[t<R\right] + (1+r)\,a_t - T_{a,t}\left((1+r)\,a_t\right) + S_t, \quad (8)$$

$$b_t = (1+r)\,a_{t+1} - T_{b,t}\left((1+r)\,a_{t+1}\right). \quad (9)$$

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, feasibility of allocations, and the requirement that each individuals' utility must be above $W^s(\theta)$. We also focus on the steady problem for the government and ignore issues related to transition.[13]

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 $\{c_t(\theta), l_t(\theta), b_t(\theta)\}$ that maximizes individual preferences (1) subject to the constraints (8) and (9) for a given set of taxes. Let $U(\theta)$ be the utility associated*

[13] We will consider the transition from one steady-state to another in the future revisions.

with such an allocation. Then we must have

$$U'(\theta) = \sum_{t=0}^{T} \beta^t P_t(\theta) \left[\frac{\varphi_t'(\theta) l_t(\theta)}{\varphi_t(\theta)} v'(l_t(\theta)) + \frac{P_t'(\theta) - P_{t+1}'(\theta)}{P_t(\theta)} \beta w(b_{t+1}(\theta)) \right] + \sum_{t=0}^{T} \beta^t P_t'(\theta) \left[u(c_t(\theta)) - v(l_t(\theta)) \right]. \tag{10}$$

The proof can be found in the appendix.

Equation (10) is the envelope condition associated with the individual optimization problem. Mechanically, one can replace $y_t = \varphi_t l_t$ in the budget constraints, (8) and (9). This implies that the budget constraints are independent of θ. Thus, the envelope condition associated with the optimization problem can be derived by simply differentiating the objective with respect to θ. We refer to (1) as the *implementability constraint*.

The above implementability constraint is reminiscent of the local incentive compatibility constraint in the New Dynamic Public Finance and the 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.

It, however, remains to be shown that if an allocation satisfies (10), it must be the solution of individual optimization given some tax function. Unfortunately, we cannot theoretically show that this result is true. In section 3.3, we show that if an allocation satisfies (10) and certain monotonicity constraints are satisfies, then a set of tax functions can be constructed so that the allocations satisfy first order conditions associated with the individual optimization and the budget constraints.

Planning Problem. Our planning problem maximizes the revenue from delivering a steady-state allocation subject to the implementability constraint (10) and a minimum utility requirement given by

$$\max \int \sum_{t=0}^{T} \frac{P_t(\theta)}{(1+r)^t} \left[\varphi_t(\theta) l_t(\theta) - c_t(\theta) - \frac{1 - p_{t+1}(\theta)}{1+r} b_{t+1}(\theta) \right] dF(\theta) \tag{11}$$

subject to

$$
\begin{aligned}
U(\theta) &= \sum_{t=0}^{T} \beta^t P_t(\theta) \left[u(c_t(\theta)) - v(l_t(\theta)) + \beta\left(1 - p_{t+1}(\theta)\right) w(b_{t+1}(\theta))\right] && (12)\\
U'(\theta) &= \sum_{t=0}^{T} \beta^t P_t(\theta) \left[\frac{\varphi_t'(\theta)\, l_t(\theta)}{\varphi_t(\theta)} v'(l_t(\theta)) + \frac{P_t'(\theta) - P_{t+1}'(\theta)}{P_t(\theta)} \beta w(b_{t+1}(\theta))\right] && (13)\\
&\quad + \sum_{t=0}^{T} \beta^t P_t'(\theta)\left[u\left(c_t(\theta)\right) - v\left(l_t(\theta)\right)\right]. \\
U(\theta) &\geq W_s(\theta) && (14)
\end{aligned}
$$

The objective in the above optimization problem is proportional to the present value of government receipts net of outlays, i.e., its primary surplus. [14]

3.2 Efficient Distortions

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

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

$$
\tau_{\text{labor},t}(\theta) = 1 - \frac{v'\left(l_t(\theta)\right)}{\varphi_t(\theta)\, u'\left(c_t(\theta)\right)}.
$$

Intuitively, $\tau_{\text{labor},t}(\theta)$ is the fraction of earning on the margin that is taken away from the individual in terms of period t consumption. The wedge to 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. Three 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

$$
\tau_{\text{un},t}(\theta) = 1 - \frac{u'\left(c_t(\theta)\right)}{\beta(1+r)\left[p_{t+1}(\theta)\, u'\left(c_{t+1}(\theta)\right) + \left(1 - p_{t+1}(\theta)\right) w'\left(b_{t+1}(\theta)\right)\right]}
$$

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

Note that the proceed from this asset in case of death is left as bequests. We refer to this as *saving wedge*.

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

$$\tau_{\text{annuity},t}(\theta) = 1 - \frac{u'(c_t(\theta))}{\beta(1+r)u'(c_{t+1}(\theta))},$$

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

3. A Life insurance contract that pays $1+r$ in the event of death. We can define a *life-insurance wedge* similar to the annuity wedge:

$$\tau_{\text{life insurance},t}(\theta) = 1 - \frac{u'(c_t(\theta))}{\beta(1+r)w'(b_{t+1}(\theta))}.$$

The above can be interpreted as a tax imposed on income from life insurance purchased at actuarially fair price of $\frac{1-p_{t+1}(\theta)}{1+r}$.

Note that the above wedges are hypothetical since we do not allow for annuity and life-insurance holdings in our implementation. Nevertheless, they are informative in terms of separating the different roles that taxes play: incentive provision (in case of above wedges) vs. completing the markets (as we describe later).

The following lemma characterizes optimal labor wedge:

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

$$\frac{\tau_{labor,t}(\theta)}{1-\tau_{labor,t}(\theta)} = \frac{\varphi_t'(\theta)}{\varphi_t(\theta)} \frac{1-F(\theta)}{f(\theta)} \left(\frac{1}{\varepsilon_{F,t}(\theta)}+1\right) \frac{g(\theta)}{1-g(\theta)\frac{1-F(\theta)}{f(\theta)}\frac{P_t'(\theta)}{P_t(\theta)}} \tag{15}$$

where

$$g(\theta) = \int_\theta^{\bar{\theta}} \frac{u'(c_0(\theta))}{u'(c_0(\theta'))}\left[1-\gamma(\theta')u'(c_0(\theta'))\right]\frac{dF(\theta')}{1-F(\theta)}, \tag{16}$$

$\varepsilon_{F,t}(\theta) = \frac{v'(l_t(\theta))}{v''(l_t(\theta))l_t(\theta)}$ *is the Frisch elasticity of labor supply and* $\gamma(\theta)$ *is the multiplier on the constraint (14).*

The above formula is the familiar formula from the static optimal taxation literature as in Mirrlees (1971), Diamond (1998) and Saez (2001). The first term in (15) captures the tail property of the distribution at a given age, t. Intuitively, if marginal tax for type θ

increases at age t, it leads to a marginal output loss of $\varphi_t(\theta) f(\theta)$. However, it relaxes the incentive constraints on all the type above at age t (captured by $\varphi'_t(\theta)(1-F(\theta))$). The second term is capturing the behavioral response to taxes. The higher the Frisch elasticity of labor supply, the larger is going to be the response to higher taxes. Finally, the last term is the social marginal welfare weight (see Piketty et al. (2014))and captures the re-distributive motive of the government. One way to think about the last term is that it is capturing the redistributive motives embedded in the status-quo tax and transfer schedule. Note that this term should be adjusted since individuals might not survive before retirement – the denominator in the last term $1-g(\theta)\frac{1-F(\theta)}{f(\theta)}\frac{P'_t(\theta)}{P_t(\theta)}$ is capturing this.

We now turn to characterization annuity and life-insurance wedges. Once we do that, the characterization of saving wedge will be easier. We have the following proposition:

Proposition 1. *The inter-temporal wedges are given by*

i. The annuity wedge:

$$\tau_{annuity,\,t}(\theta) = \frac{p'_{t+1}(\theta)}{p_{t+1}(\theta)}\frac{1-F(\theta)}{f(\theta)}\frac{g(\theta)}{1-g(\theta)\frac{1-F(\theta)}{f(\theta)}\frac{P'_t(\theta)}{P_t(\theta)}}, \tag{17}$$

where $g(\theta)$ is given by(16).

ii. The life-insurance wedge is given by

$$\tau_{life\ insurance,t}(\theta) = -\frac{p'_{t+1}(\theta)}{1-p_{t+1}(\theta)}\frac{1-F(\theta)}{f(\theta)}\frac{g(\theta)}{1-g(\theta)\frac{1-F(\theta)}{f(\theta)}\frac{P'_t(\theta)}{P_t(\theta)}}, \tag{18}$$

iii. The annuity (life-insurance) wedge is positive (negative), if and only if survival is positively correlated with labor productivity, i.e., $p'_{t+1}(\theta) > 0$.

The mechanics of the above results can be understood from inspecting the incentive constraint in (10). When survival is positively correlated with labor productivity, an increase in $c_{t+1}(\theta)$ leads to an increase in the right hand side of (10). This implies that such an increase is costly for redistribution since a government with redistributive motives desires utility profile $U(\theta)$ to be constant or decreasing. Thus second period consumption should be distorted downwards, i.e., annuitization margin should be taxed.

On the contrary, an increase in bequests, $b_{t+1}(\theta)$, decreases the right hand side of the incentive constraint in (10). As a result bequests are beneficial in that they make redistribution easier, since they relax the incentive constraint. Therefore, bequests should be distorted upwards, i.e., life-insurance margin should be subsidized.

Intuitively, the idea behind the optimal tax on annuity income is similar to that of labor income taxes. Since productive individuals have an incentive to under-report their productivity type, annuity purchases for any given type must be taxed so that under-reporting by higher productivity individuals are less attractive since they value future consumption at a higher rate. Similarly, bequests should be subsidized so that more productive individuals find under-reporting less attractive since they care less about bequests.

An alternative , and more relevant, measure of distortions to the saving decision is the saving wedge defined as the income lost from saving in the form of a risk-free asset, i.e., $\tau_{\text{un},t}(\theta)$ – a measure of saving distortions that is more commonly studied in the literature. Multiplying the formulas in (17) and (18) by $p_{t+1}(\theta)$ and $1 - p_{t+1}(\theta)$, respectively and summing the resulting equations implies that

$$\frac{\beta(1+r)}{u'(c_t(\theta))} = \frac{p_{t+1}(\theta)}{u'(c_{t+1}(\theta))} + \frac{1 - p_{t+1}(\theta)}{w'(b_{t+1}(\theta))}. \tag{19}$$

Equation (19) is known as *Inverse Euler Equation*. This equation describes a necessary condition that emerges in many dynamic incentive problems (see Kocherlakota (2010) for an extended discussion) in which the source of private information is a shock which does not directly affect marginal utility of consumption. To our knowledge, this is the first paper that derives this result as a necessary condition for efficiency in an environment with no shock in which source of private information directly affect marginal utility consumption (because of the term $P(\theta)u'(c_1(\theta))$). The intuition for this result is very close to its counterparts in the New Dynamic Public Finance literature and we refer the reader to associated papers for a detailed discussion of this relationship.

Proposition 2. *The optimal wedge on un-contingent saving is given by*

$$\frac{\tau_{un,t}(\theta)}{1-\tau_{un,t}(\theta)} = \frac{p_{t+1}(\theta)}{1 - \frac{p'_{t+1}(\theta)}{p_{t+1}(\theta)}\frac{1-F(\theta)}{f(\theta)}\frac{g(\theta)}{1-g(\theta)\frac{1-F(\theta)}{f(\theta)}\frac{P'_t(\theta)}{P_t(\theta)}}} + \frac{1-p_{t+1}(\theta)}{1 + \frac{p'_{t+1}(\theta)}{1-p_{t+1}(\theta)}\frac{1-F(\theta)}{f(\theta)}\frac{g(\theta)}{1-g(\theta)\frac{1-F(\theta)}{f(\theta)}\frac{P'_t(\theta)}{P_t(\theta)}}} - 1. \tag{20}$$

Moreover, the wedge is larger if $p'_{t+1}(\theta) > 0$ is larger.

The above results regarding taxation of annuities and un-contingent assets, points towards the key part of our policy reform. That is, saving distortions must be used specially when mortality differentials, as measured by $p'_{t+1}(\theta)$. As we discuss later - in Section 4, these mortality differentials are the largest for ages above 65, i.e., post retirement. Thus post-retirement asset taxes and subsidies could be a potential source of efficient provision of saving incentives and part of an optimal reform of retirement financing.

Note that since survival probability is bounded above by one, as productivity converges to ∞ survival differentials disappear. Hence, τ_a (and τ_b, τ_k) converges to zero as $\theta \to \infty$. This implies that in the limit, the top tax rate formula implied by the analysis in Saez (2001) holds in our model. This result suggests that mortality differentials, mostly affect optimal taxes for lower and the middle part of the earnings distribution and do not affect our understanding of factors affecting taxation of the rich.

The idea that future consumption (or savings) should be taxed in order to provide incentive for labor supply can be connected to the literature on optimal taxation and preference heterogeneity. Starting from Tuomala (1990) and extended by many others including Saez (2002), Golosov et al. (2013) and Piketty and Saez (2013), it has been argued that when high earners prefer certain types of goods, taxation of these goods can contribute positively to redistribution and should be used. While our result has a similar flavor, there is a key difference. It is that survival not only affects preferences but also it affects resources. In particular, in our model, absent heterogeneity in labor productivity, the annuity wedge must be zero even in presence of mortality differentials. This is not necessarily true when the only difference across individuals is preference heterogeneity.

3.3 Optimal Taxes

So far, we have mainly focused on optimal allocations. In this section, we describe how to back out the optimal taxes from the optimal allocations and wedges discussed above.

The following lemma establishes the possibility of constructing tax and transfer schedules as in (8) and (9) such that individual optimizations' first order conditions are satisfied: (in what follows we adopt the following notation to avoid clutter; $u_{c,t}(\theta) \equiv u'(c_t(\theta))$, $v_{l,t}(\theta) \equiv v'(l_t(\theta))$ and $w_{b,t}(\theta) \equiv w'(b_t(\theta))$.)

Lemma 3. *Consider an allocation $\{c_t(\theta), l_t(\theta), b_t(\theta)\}$ that satisfies implementability constraint (10), and such that $b_t'(\theta) > 0$, $(\varphi_t(\theta) l_t(\theta))' > 0$ and*

$$\sum_{s=t}^{T} \beta^s P_s(\theta) \left[u_{c,s}(\theta) c_s'(\theta) + \beta (1 - p_{s+1}(\theta)) w_{b,s+1}(\theta) b_{s+1}'(\theta) - v_{l,s}(\theta) (\varphi_s(\theta) l_s(\theta))' \right] > 0.$$

Then tax and transfer functions $T_{a,t}(\cdot), T_{b,t}(\cdot), T_{y,t}(\cdot), S_t$ together with asset holdings $a_t(\theta)$ exists so that the allocations $\{c_t(\theta), l_t(\theta), b_t(\theta), a_t(\theta)\}$ satisfy the budget constraints (8) and (9) 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

$$1 - T'_{y,t}\left(\varphi_t\left(\theta\right) l_t\left(\theta\right)\right) = \frac{v'\left(l_t\left(\theta\right)\right)}{\varphi_t\left(\theta\right) u_{c,t}\left(\theta\right)}, \tag{21}$$

$$u_{ct} = \beta\left(1+r\right)\left[p_{t+1}\left(1 - T'_{a,t+1}\right) u_{ct+1} + \left(1 - p_{t+1}\right)\left(1 - T'_{b,t+1}\right) w_{b,t+1}\right] \tag{22}$$

Equation (21) is the individual intra-temporal optimality condition and equation (22) is the individual euler equation. We know from discussion in Section 3.2 that this euler equation must be distorted at the efficient allocation. Therefore, optimal marginal taxes $T'_{a,t+1}$ and $T'_{b,t+1}$ are different from zero.

We can use equation (21) 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 (21) - this is well-defined since output in each age is increasing in θ.

The calculation of optimal asset taxes is not straight forward. More importantly, the level of assets a cannot be pinned down independent from the marginal taxes $T'_{a,t+1}$ and $T'_{b,t+1}$. 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 decentralized 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 (21) and (22) together with (8) and (9) and use extensive algebra to show that the derivative of asset holdings with respect to θ, a'_t, satisfies

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

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

$$b_t\left(\theta\right) = \left(1+r\right) a_t\left(\theta\right) - T_{b,t}\left(\left(1+r\right) a_t\left(\theta\right)\right) \tag{23}$$

Since $a_t\left(\theta\right)$ and $b_t\left(\theta\right)$ is determined in the optimal allocation, the above formula determines bequests taxes.

Finally, using (23) and the Euler equation (22), we must have

$$1 - T'_{a,t+1} = \frac{u_{ct}}{\beta\,(1+r)\,p_{t+1}u_{ct+1}} - \frac{1-p_{t+1}}{p_{t+1}}\frac{w_{b,t+1}}{u_{ct+1}}\frac{b'_{t+1}}{(1+r)\,a'_{t+1}}.$$

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

□

Unfortunately, we cannot derive a closed form formula for optimal taxes. However, our implementation procedure provides a guideline on how to numerically compute the optimal tax functions. We present these numerical derivations in Section 5.

Note that monotonicity constraints in Lemma 3 are necessary for existence of 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 constraint are satisfied.

4 Calibration

In order to be able to conduct our policy experiments we need parametric specifications and parameter values for the model described in Section 2. We can estimate parameters for some of the model ingredients independently (e.g., wage/productivity profiles or mortality profiles). However, in order to choose some of the model parameters (e.g, weight of bequest in the flow utility) we need to use our model to match some targets in the U.S. data. We describe these details below.

Earning ability profiles. Individual productivity depends on two component: a deterministic age-dependent component $\tilde{\varphi}(t)$ and type-dependent fix effect θ. Therefore, the natural logarithm of ability is

$$\log \varphi(\theta) = \log \theta + \log \tilde{\varphi}_t.$$

For the deterministic part we assume a polynomial

$$\log \tilde{\varphi}_t = \beta_0 + \beta_1 \cdot t + \beta_2 \cdot t^2 + \beta_3 \cdot t^3,$$

which we estimate using labor earnings data in the PSID. We follow a large part of the literature (e.g., Altig et al. (2001), Nishiyama and Smetters (2007) and Shourideh and

Table 1: Death Rates by Lifetime Earning Deciles for Male Age 67-71

	Lifetime Earning 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)

Troshkin (2015)) and use the logarithm of effective reported labor earnings per hour as a proxy for $\log \varphi_t(\theta)$. We calculate this as the ratio of all reported labor earnings to total reported hours. For labor earning 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 $\beta_0 = 0.879$, $\beta_1 = 0.1198$, $\beta_2 = -0.00171$ and $\beta_0 = 7.26 \times 10^{-6}$.

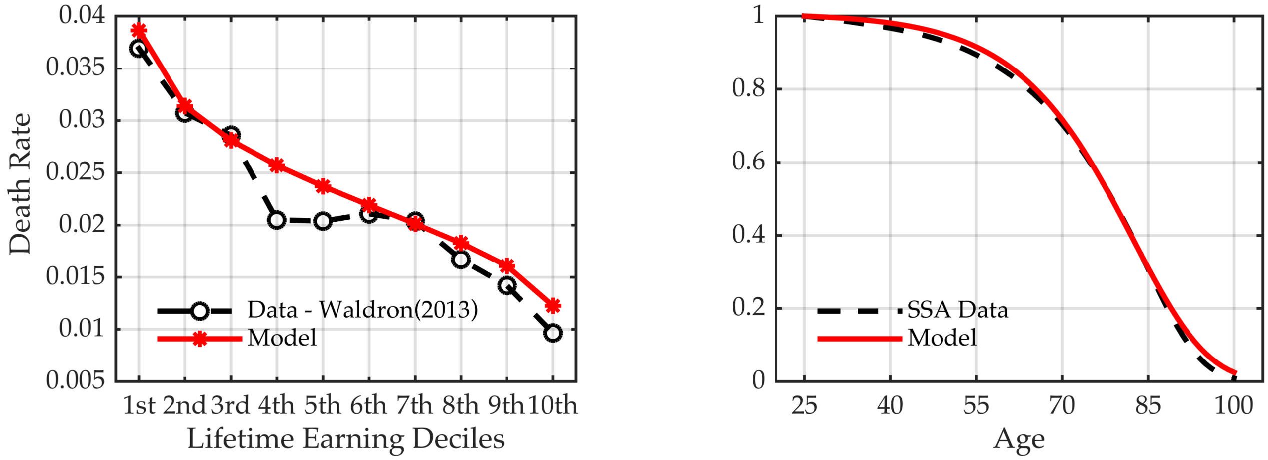
We assume the type-dependent fix effect θ, has a Pareto-Lognormal distribution with parameters $(\mu_\theta, \sigma_\theta, a_\theta)$. This distributional approximate a lognormal with parameters μ_θ and σ_θ at low incomes and a Pareto with parameter a_θ at high values. It therefore, allows for a heavy right tail at the top of ability and earning distribution. For this reason it is commonly used in the literature (see Golosov et al. (forthcoming), Badel and Huggett (2014) and Heathcote and Tsujiyama (2015)). [15] We choose the tail parameter and variance parameter to be $a_\theta = 3$ and $\sigma_\theta = 0.5$. The location parameter is set to $\mu_\theta = -1/a_\theta$ so that $\log \theta$ has mean zero. With these parameters the cross section variance of log hourly wage in the model is 0.36. Also, the ratio of median hourly wages to bottom decile of hourly wage is 2.3. These statistics are consistent with reported facts on cross section distribution of hourly wage reported in Heathcote et al. (2010). Figure 2a displays a sample of our productivity profiles.

Demographics and Mortality Profiles. We assume individuals start at age 25 and nobody survives beyond 100 years of old. Everyone retire at age 65. Individuals have Gompertz force of mortality

$$\lambda_t(\theta) = \frac{m_0}{\theta^{m_1}}(\exp(m_2 t)/m_2 - 1). \tag{24}$$

Gompertz distribution is widely used in the actuarial literature that model mortality (see Horiuchi and Coale (1982)). It is also used in Einav et al. (2010) to model differential

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



(a) (b)

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

mortality. 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 lower mortality at all ages. The key parameter is m_1 which determines how mortality varies with ability. To choose this parameter with use the data on mortality across lifetime earning deciles reported in Waldron (2013). She uses Social Security Administrative data to estimates 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 evidence points to large differences in death rates across different income groups with the poorest deciles almost 4 times more likely to die than he richest decile. We use this data to calibrate parameter m_1.

The parameter m_2 is chosen to match 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, m_0 is chosen so that mortality at age 25 is zero. The parameters that give the best fit to the mortality data in Table 1 and average mortality data are $\alpha_0 = 0.0006$, $\alpha_1 = 0.5545$ and $\alpha_2 = 0.0855$. Figure 1a shows the fit of the model in terms of matching mortality across lifetime earnings decile in Waldron (2013). Once w have mortality hazard $\lambda_t(\theta)$ we can find survival probability $P_t(\theta) = \exp(-\lambda_t(\theta))$. A sample of survival probabilities implied by our calibration is shown in Figure 2b.

Preferences. We assume constant relative risk aversion over consumption $u(c) = \frac{c^{1-\sigma}-1}{1-\sigma}$. However, for utility over bequest we follow De Nardi (2004), Ameriks et al. (2011), Pashchenko

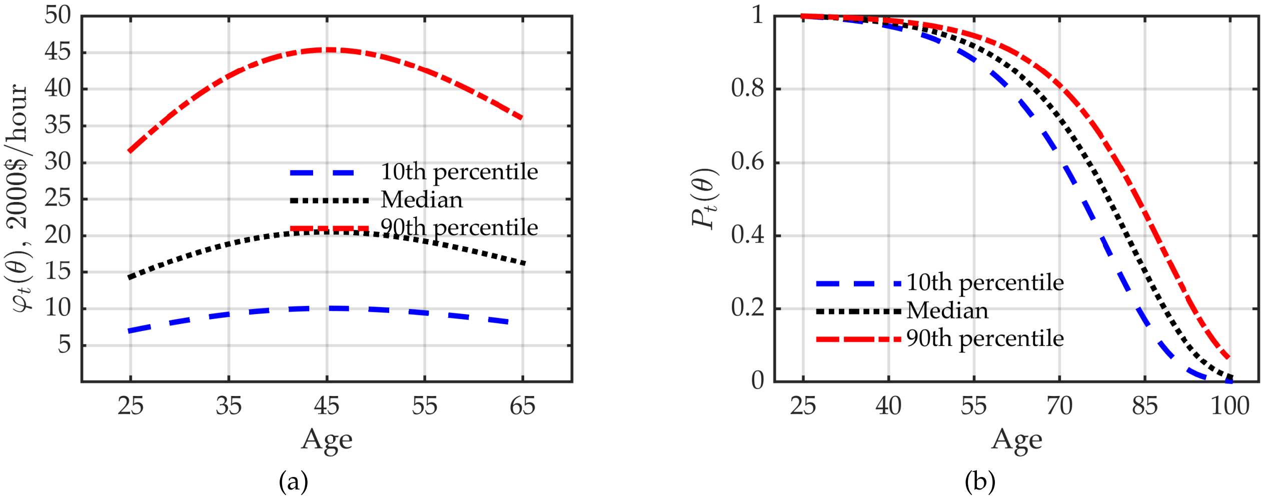


Figure 2: Sample profiles of labor productivities (2a) and survival probabilities (2b) for different θ types.

(2013) and De Nardi and Yang (forthcoming) and assume

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

Parameter χ determines the strength of bequest motive, while $\bar{B}$ reflects the extent to which bequests are luxury goods. If $\bar{B} > 0$, the marginal utility of bequest is bounded. At the same time, marginal utility of large bequest declines more slowly than the marginal utility of consumption. As a result a richer individual has stronger motive to leave bequest. [16] As noted in De Nardi (2004) a positive value for parameter $\bar{B}$ is needed to match the fraction of deceased who leave no bequest. [17]

We assume risk aversion parameter $\sigma = 1.5$. Strength of bequest χ is chosen to match the bequest to wealth ratio of 0.0118 as reported in Gale and Scholz (1994). To calibrate $\bar{B}$ we use data on distribution of bequest reported in Hurd and Smith (2002). We choose this parameter so that in the model 25 percent leave bequest of less than one third of median income.[18]

We assume constant Frisch elasticity for disutility over hours worked

[16] The wealth elasticity of both realized and anticipated bequests have both been estimated to be about 1.3 (see Auten and Joulfaian (1996) and Hurd and Smith (2002)). Among single Americans who were at least 70 years old in 1993 and died before 1995, the 30th percentile of the bequest distribution was just \$2 thousand, the median was \$42 thousand, and the mean was \$82 thousand (Hurd and Smith (2002)).

[17] To make computation easier we approximate the function above by a smooth function. Therefore, in our model everyone leaves bequest. However, for a large fraction the amount is very small.

[18] We normalize the data in Table 11.1 of Hurd and Smith (2002) by median household income in 1994 CPS which was \$32264.

$$v(l) = \psi \frac{l^{1+\frac{1}{\epsilon}}}{1+\frac{1}{\epsilon}}.$$

with elasticity of labor supply $\epsilon = 0.5$. The weight of leisure in utility ψ chosen so that the average annual hours worked in the model 2000. We choose discount factor β to match the wealth to income ratio of 3. Depreciation rate of capital is $\delta = 0.06$ and return on capital net of depreciation $\tilde{r}$ is 0.06.[19] There is a corporate income tax rate of $\tau_K = 0.33$ paid by the firms. Therefore, after tax return on asset is $r = 0.04$. [20] This is also the interest rate that government pays on its debt. We take the rerun on assets (and government debt) fixed throughout our analyses. [21]

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 average taxable earnings over lifetime.[22] Let e be labor earning and e_{max} be maximum taxable earning. We set e_{max} equal to 2.47 times the average earning in the economy. Social security tax rate is $\tau_{ss} = 0.124$. [23] There is also an medicare tax rate$\tau_m = 0.029$ which applies to entire earning.

Each individual's benefit is a function of his average life time earning (up to e_{max}). We denote this by $\bar{e}$. We use the same benefit formula that the U.S. Social Security Administrations uses to determine the primary insurance amount (PIA) for retires:

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

To account for medicare benefits, we assume each individual in retirement will receive an additional transfer independent of their earning history. We choose this value so that

[19] Therefore, with capital to output ratio of 3, the steady state ratio of investment to GDP is 0.21 which is aligned with the US average over 2000 to 2010.

[20] 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.

[21] In this regard, our analysis can be viewed as a partial equilibrium analysis since we do not allow return on capital adjust as stock of capital changes. In future drafts we will endogenies return on assets and report results in a full blown general equilibrium model.

[22] Social security administration uses only the highest 35 years of earning to calculate the average life earning. We use entire earning history for easier computation.

[23] We account for disability insurance tax and benefits by bunching them together with social security.

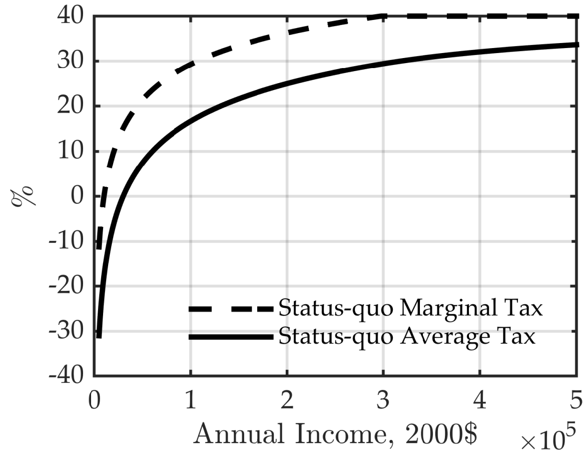


Figure 3: Calibrated effective income tax function, $\mathcal{T}(\cdot)$.

the aggregate medicare benefits are 3 percent of GDP. [24]

Tax function and government purchases. In addition to social security, government has an exogenous spending G, which we assume to be 9% of GDP. [25] There is a consumption tax τ_C and a nonlinear tax on labor income income. We use 5.5% for consumption tax as calculated in McDaniel (2007) . For income tax function we use

$$\mathcal{T}(y) = y - \phi y^{1-\tau}.$$

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

Tax function of this form are extensively used to approximate the effective income taxes in the United States. The parameter τ determines the progressivity of the tax function, while ϕ determines the level (the lower ϕ is, the higher is the total tax revenues for a given τ). Heathcote et al. (2014) estimate value of 0.151 for τ, using PSID income data and income tax calculations using NBER's TAXSIM program. We use their estimated value for

[24]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.

[25]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.

Table 2: Exogenous Parameters Chosen Outside the Model

Parameter	Description	Values/source
Demographics		
T	maximum age	75 (100 years old)
R	retirement age	40 (65 years old)
$\lambda_t(\theta)$	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	0.5
a_θ	pareto-lognormal tail parameter	3
μ_θ	pareto-lognormal location parameter	-0.33
Technology		
r	return on assets	0.04
Government policies		
τ_{SS}, τ_m	social security and Medicare tax rates	0.124,0.029
S^s	social security benefit formula	see text
τ_c	consumption tax	0.055
τ, ϕ	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)

τ and choose ϕ. Figure 3 illustrates the resulting marginal and average taxes as functions of annual earnings in constant 2000 dollars. [26]

Finally, we assume government debt to be 50 percent of GDP. [27] To make sure the flow government budget constraint holds in steady state, we assume that there are lump sum transfers Tr made to all workers before retirement.

Tables 2 and 3 show the calibration summary. Tables 2 lists parameters that are either taken from other studies or estimated/calculated independent of the model structure. Their sources and estimation/calculation procedures are outlined in 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 shows the moments targets in data and resulting values in the model. The bottom pane lists the parameter values. Note that in some cases

[26]We cap the marginal tax rate at 40 percent to be consistent with top federal marginal tax rate in the US.

[27]This is sum of the state and local municipal securities and federal treasury securities. We use the average over 2000 to 2010.

Table 3: Parameters Calibrated Using the Model

Parameters	Description		Values
β	discount factor		0.98
χ	strength of bequest		24.18
$\bar{B}$	bequest utility shifter		430410
ψ	weight on leisure		5.3×10^{-13}
Moments		Data	Model
Wealth-income ratio		3	3
Bequest-wealth ratio		0.0118	0.013
Fraction with almost no bequest[a]		0.2	0.17
Average annual hours		2000	2000

[a]This the fraction of bequests that are less than a third of median income

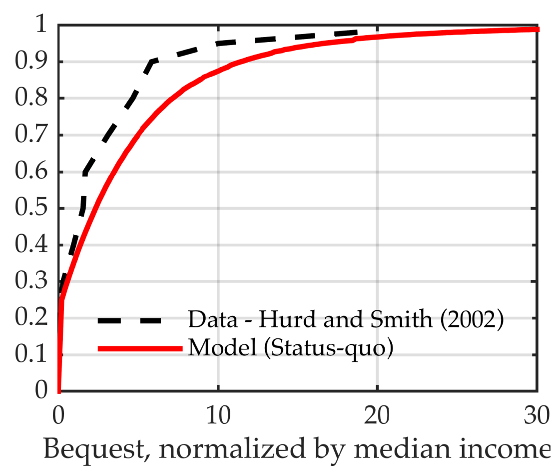


Figure 4: Cumulative distribution of bequest, model vs. Hurd and Smith (2002) data.

the moments are no matched exactly, although the errors are small. It is important to remind our readers that the model is very stylized. The only motive for saving in this model is life cycle motive as well as bequest motive. Therefore, a subjective discount rate low enough to generate a capital-output ratio of 3, will encourage all individuals to save higher. This in turn reduces the fraction of people who die with no asset. [28]

Figure 4 displays the cumulative distribution distribution of bequests in the model and the distribution that Hurd and Smith (2002) report from AHEAD data for decedents. The size of bequest is normalized by median income. Despite the model inability to hit the calibration target with regard to the fraction zero bequests, the overall distribution of bequest compares very well to the AHEAD data.

[28]In contract, a model with mensurable income uncertainty the link between discount rate and saving is weaker. Therefore, models that include this feature (e.g. De Nardi (2004)) are able to hit all the targets.

Table 4: Distribution of Wealth

		percentage of wealth in the top					
	Wealth Gini	1%	5%	20%	40%	60%	80%
U.S. data (SCF (1989))	0.78	29	53	80	93	98	100
Model (status-quo)	0.60	10	29	61	82	94	99

Table 5: Distribution of Earnings

		percentage of earnings in the top				
	Earnings Gini	1%	5%	20%	40%	80%
U.S. data (CPS (1994))	0.46	6	19	48	72	98
Model (status-quo)	0.48	5	16	44	68	96

The model also does a reasonable job at matching distribution of wealth, except maybe at the very top. The first line in Table 4 displays Gini coefficient and the distribution of wealth in the U.S. from SCF (1989). The second line in the Table shows the implied distribution from the model. The difficulties of economic models in generating high concentration of the wealth at the top is well known. In particular, in a model with no entrepreneurial risk (e.g., Cagetti and De Nardi (2009)), no extreme income risk (e.g., Castaneda et al. (2003)) and no direct intergenerational link (e.g., De Nardi (2004)) the share of wealth at the top is far below the data. However, despite that, the distribution of wealth in our model compares well with data, in particular for the poorer individuals. [29]

The first line in Table 5 displays the distribution of earnings for individuals age 25 to 60 in CPS. The second line is the distribution of earnings implied by model. As it is evident, the model matches distribution of earning quite well.

5 Optimal Policy Reform: Numerical Results

We now use the the calibrated model to solve numerically for the allocation of consumption, hours and bequests for each productivity type θ in steady state under the status-quo policies described in Section 4. We use these allocations to compute the lifetime welfare under the status-quo policies. This the present discounted value of expected utility from consumption, bequest and hours that each productivity type θ experiences under the status-quo policies. These status-quo life time welfares are the inputs to the planning problem.

[29] Adding more features to the model in the interest of getting a better match of wealth distribution is straight-forward. However, this will make our normative analysis much more complicated and it is outside the scope of this paper.

Given the status-quo lifetime welfares, using the parametrization described in Section 4, we can solve for the efficient allocations. These are the allocations that solve the problem (11) described in Section 3.1.

Our goal in this section is to first make a case for policy reforms by comparing intra-temporal and inter-temporal distortions in the status-quo economy and the efficient allocation. Motivation for policy reform arises from the differences between distortions under status-quo policies under efficient allocations. We can then, use the procedure outlined in Section 3.3 to solve for optimal policies that implement efficient distortions in the economy. Finally, we report the effect that optimal reform has on individual choices, macro aggregates and government budget.

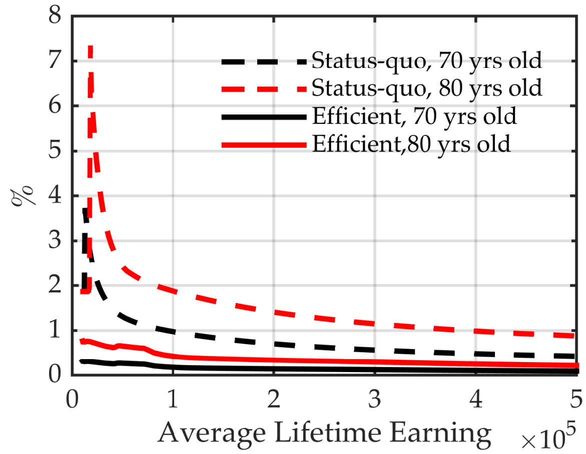
5.1 Distortions: status-quo vs efficient allocations

As we argued in Section 3.2 one property of efficient allocation is that the inter-temporal margins (as well al intra-temporal margins) are distorted. In other words, marginal rate substitution (between consumption in consecutive periods, or between consumption and bequest) is not equal to actuarially fair prices. We call these distortions *wedges*. We use the following formula, to compute these wedges for each productivity type θ at each age.

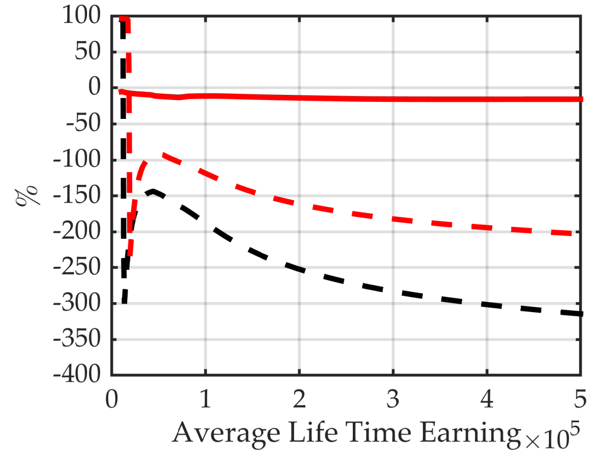
$$\begin{aligned}
\tau_{\text{annuity},t}(\theta) &= 1 - \frac{u'(c_t(\theta))}{\beta(1+r)\,u'(c_{t+1}(\theta))}, \\
\tau_{\text{life insurance},t}(\theta) &= 1 - \frac{u'(c_t(\theta))}{\beta(1+r)\,w'(b_{t+1}(\theta))}, \\
\tau_{\text{labor},t}(\theta) &= 1 - \frac{v'(l_t(\theta))}{\varphi_t(\theta)\,u'(c_t(\theta))}.
\end{aligned}$$

We do the calculations both for efficient allocations and the status-quo allocations. The resulting wedges are plotted in Figure 5.

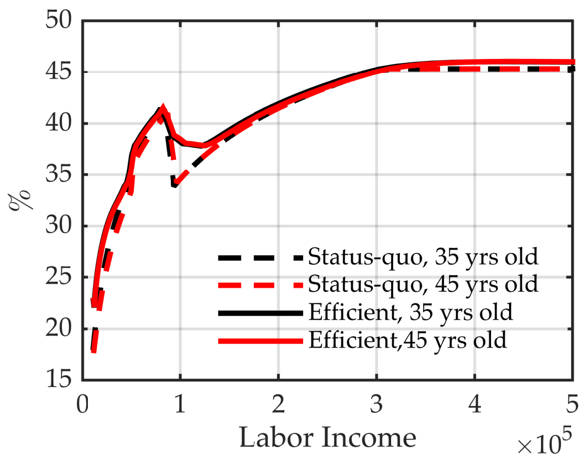
Figure 5a shows the inter-temporal distortions on the annuitization margin (marginal rate of substitution between consumption at age t and $t+1$). There are three important observations. First, this margin is positively distorted (effectively taxed) both in status-quo economy and efficient allocations. Second, these distortions are large. Third, there is a rather large difference between efficient distortions and distortions in the status-quo economy. This difference is the key to understanding the sign and magnitude of optimal taxes. Note that if the distortions in status-quo economy and efficient allocation were equal, then there would be no justification for government intervention in the asset market (i.e. taxing or subsidizing the asset accumulation). Annuitization margin is distorted



(a) Inter-temporal (annuity) distortions



(b) Inter-temporal (life insurance) distortions



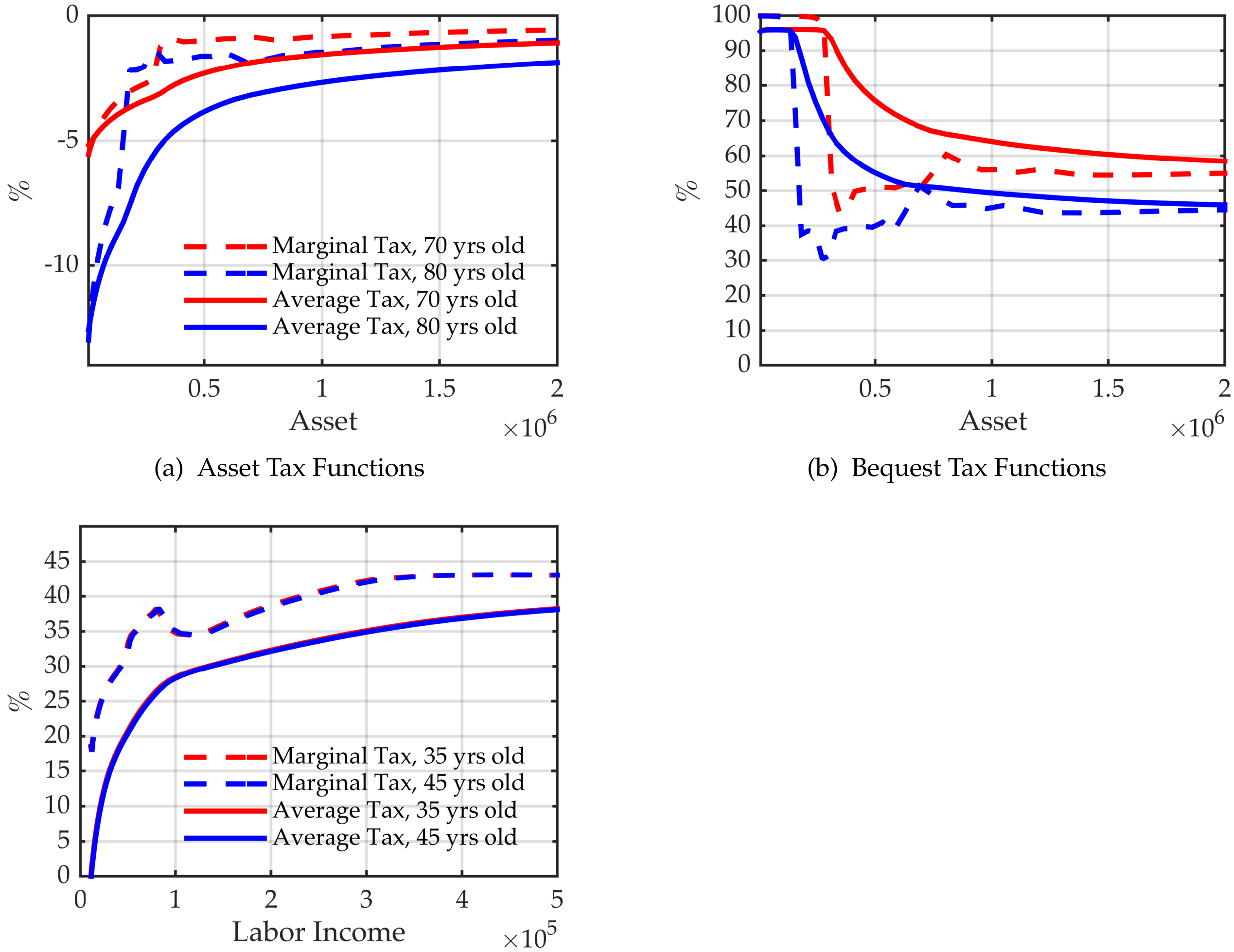
(c) Intra-temporal (labor) distortions

Figure 5: Status quo distortions vs efficient distortions.

in the status-quo economy due to missing annuity markets. [30] For each type θ, there is a gap between the rate of return on assets in market, r, and the actuarially fair return on annuity for that type, $r + 1 - \frac{p_{t+1}(\theta)}{p_t(\theta)}$. Since mortality is higher for individuals with lower ability, this gap is larger for these individuals. The dashed lines in 5a show these distortions for two difference ages.

The solid lines in Figure 5a show distortions that result from the efficient allocations. The reason that annuitization margins are distorted under the efficient allocation is the following. In this environment, individuals with higher labor productivity have lower mortality. As the result, they place more weight on future consumption (than a lower

[30] Our findings do not depend on strong assumption of missing annuity market. These findings are valid as long as there is imperfection in the annuity market and individuals are not paid the actuarially fair rates on their annuities.



(a) Asset Tax Functions

(b) Bequest Tax Functions

(c) Labor Tax Functions

Figure 6: Optimal Tax Functions

productivity individual). Taxing the annuity income of the lower ability individuals reduce the incentive for higher ability to shirk. Instead, they will have incentive to work harder, earn more income when young and benefit from lower distortion on the annuitization when they are old. This, in turn, lowers the distortionary cost of taxation for the government. The size of these distortions depend on mortality differentials (the differences in mortality across different productivity types).

The discussion above highlights the deep reasons why there is inter-temporal distortions under efficient allocation and under the status-quo polices and why they are different. In our model the main rationale for asset taxes is to close this gap. Since status-quo distortions are alway higher than the efficient distortions, the optimal policy has a negative sign, i.e. they are subsidies.

Figure 5b shows the distortions along the life insurance margin (marginal rate of substitution between consumption at age t and bequest at age $t+1$). The dashed lines are distortions under status-quo economy. Solid lines are distortions under efficient allocation. The observations here are similar to the ones discussed in the previous paragraph, with opposite signs. Under status-quo, due to missing annuity market, individuals die with more than desired assets (due to precautionary saving motive). This reflects in large negative wedge (effective subsidy) on life insurance margin. Under efficient allocation, the distortions along life insurance margin are smaller, i.e. (effective subsidies to leave bequest are smaller). [31] Therefore, there is need for government to intervene with a positive bequest tax.

Finally, Figure 5c shows distortions along intra-temporal margin (marginal rate of substation between consumption and hours worked). In the status-quo economy these distortions arise from the distortionary tax on labor income. [32] Under the efficient allocations they depend on elasticity of labor supply, tail properties of ability distribution and the planner's redistribution motive. Since, by construction, we do not alter the distribution of welfare in the economy, the overall shape of the distortion under efficient allocation is very similar to the status-quo. This implies that under optimal reform policies, marginal tax on labor income is very similar to status-quo marginal taxes.

5.2 Optimal Policies

We now follow the procedure outlined in Section 3.3 to solve for system of optimal tax and transfers. These are: 1) non-linear, age dependent tax on assets upon survival $T_{a,t}\left(\left(1+r\right)a_t\right)$, 2) non-linear, age dependent tax on bequests upon death $T_{b,t}\left(\left(1+r\right)a_t\right)$, 3) non-linear, age dependent tax on labor income $T_{y,t}\left(y_t\right)$, 3) Transfer to workers before retirement Tr_t, and 4) Transfer to workers after retirement, S_t. Note that transfer are the independent of individual choices but they dependent on age.

Figure 6 show resulting marginal and average tax functions for select ages. As we argued in Section 5.1, in the absence of annuity market individuals do not receive actuarially fair return on their assets. In that regard assets are too expensive, and more so for poorer individuals who have higher mortality. Therefore, marginal subsidies on asset must be higher for poorer individuals. Therefore, asset taxes are progressive (Figure 6a). On the other hand, bequest taxes are regressive (Figure 6b) and assets of the poor is subject to up to 100 percent taxes when they die. This arise from the luxury good prop-

[31] The reasoning on why they are needed under efficient allocations are analogous to the discussion for annuitization distortions.

[32] The nonlinear income tax together with social security and medicate tax on earnings, up to a maximum taxable earning.

Table 6: Sources of Retirement Income

	Share of public transfers in retirement income (%)		
Income Quartiles	Data[a]	Status-quo	Efficient
1st	95	94	89
2nd	90	84	77
3rd	67	75	64
4th	34	53	34

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

erty of bequests. Poorer individuals have no demand for bequest (or they have very little demand). However, they must accumulate assets to finance retirement. The high tax on their bequests ensures that they do not leave too much bequest.

Figure 6c shows marginal and average tax on labor income. As argued before, these are very similar to tax functions in the status-quo economy.

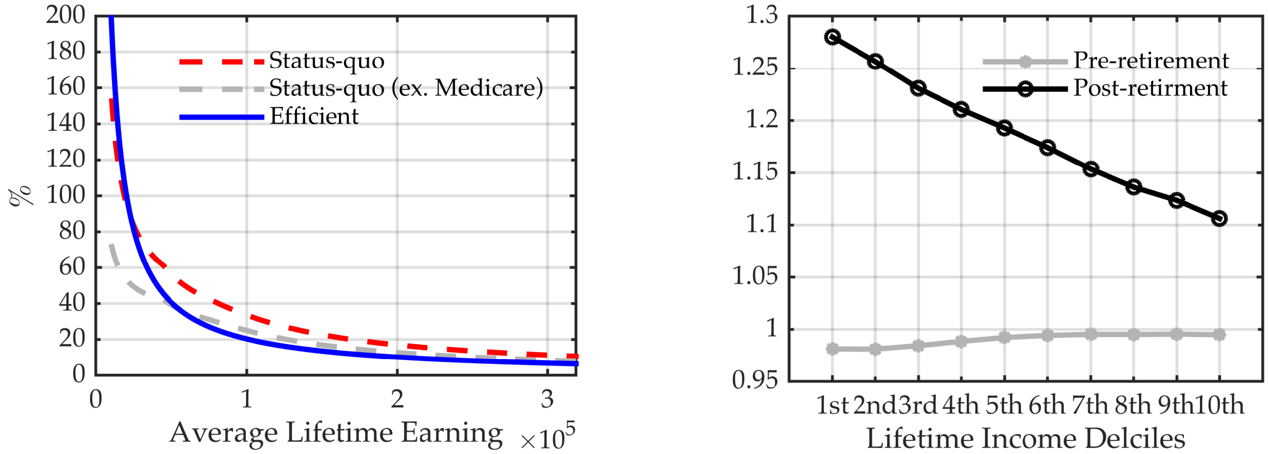
5.3 Sources of retirement income

The composition of retirement income under optimal policy are not very different from status-quo economy. Table 6 summarizes the share of social security income form total retirement income under status-quo and optimal reform. For comparisons we also include this ratio in CPS data (reported in Poterba (2014)).[33] The ratio in the model is close to the CPS data, particularly for the lower half of the income distribution. Under the optimal reform, the ratio stays very close to the status-quo.

One feature of the optimal reform policies is the uniform social security benefits that are independent of history of earning. To show how these benefit compare the status-quo economy we plot the replacement ratios under two systems in Figure 7a. These are the ratio of retirement benefits to the average lifetime earnings (equivalent to social security's average indexed monthly earnings, AIME). The replacement ratios under optimal reform are lower than status-quo for most of the population, except for bottom 20 percent of lifetime earning.

Figure 7b shows the ratio of pre and post retirement consumption between efficient allocation and status-quo. The pre retirement consumption falls a little under efficient allocation while post retirement consumption increases by up to 30 percent for the lowest income decile.

[33]To make these number comparable to our model we exclude labor earning. We also aggregate pension income together with asset income.



(a) Replacement ratio

(b) Ratio of consumption

Figure 7: Figure 7a shows the replacement ratios under status-quo and efficient allocation. Figure 7b is the ratio of consumption to status-quo consumption in pre and post retirement, by lifetime income deciles.

Table 7: Earnings Distribution

		percentage of earnings in the top				
	Earnings Gini	1%	5%	20%	40%	80%
U.S. data (CPS (1994))	0.46	6	19	48	72	98
Model (status-quo)	0.48	5	16	44	68	96
Model (efficient)	0.49	5	17	46	69	96

5.4 Distributional, aggregate and budgetary effects

Tables 7 and 8 report the distribution of earning and assets. As we see the optimal reform policies do not have large effect on the distribution of earnings and assets in the economy. The distribution of bequest, however, changes considerably (Figure 8). In particular, the fraction of individual who leave zero bequest increases to about 60 percent. The reason is that bequests are luxury good and they are not an efficient way to deliver utility to poorer individual. Therefore, heavy tax on them are optimal.

Table 8: Wealth Distribution

		percentage of wealth in the top					
	Wealth Gini	1%	5%	20%	40%	60%	80%
U.S. data (SCF (1989))	0.78	29	53	80	93	98	100
Model (status-quo)	0.60	10	29	61	82	94	99
Model (efficient)	0.60	10	29	62	83	94	99

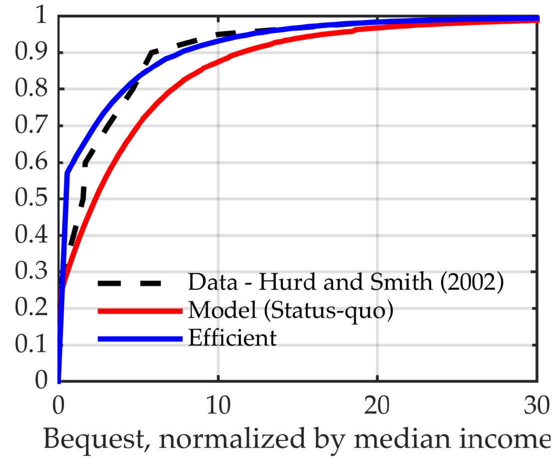


Figure 8: Distribution of bequest: status-quo vs efficient vs Hurd and Smith (2002) data

Table 9 shows how the optimal reform affect aggregate quantities. The first column shows the steady state values of aggregate quantities relative to GDP in the status-quo economy. The second column shows the same quantities under optimal reform.

The optimal reform reduces present discounted value of all transfers net of taxes to a generation by 15 percent. This extra resource is used in the economy to accumulate capital and running down government debt. This is almost by assumption. We do not take a stand on how the society decides to spend its free resources. They can, in principle, be transferred back to individuals. In that case the government debt does not fall. However, the increased incentive for saving will still induce more capital accumulation.[34] As a result of more capital accumulation, GDP goes up by about 8 percent. However, labor earnings fall about 1.5 percent due to fall in hours worked. [35]

6 Conclusion

In this paper, we analyzed optimal policy reforms of retirement financing, i.e., reform that intend to separate the efficiency of such schemes from its distributional consequences. Our optimal reform approach points towards the importance of saving distortions - taxation of bequests and subsidization of asset holdings, late in life. The important ingredient of our analysis is the fact that individuals have differential mortalities and this is correlated with their earning ability.

[34]The stock of capital is uniquely determined. Its ownership depends on the details of transfers.

[35]Some of these effects are due the fact that we do not let factor prices adjust. In future versions of the paper we will report numbers with endogenous wages and return on capital.

Table 9: Aggregate effect of policy reforms

		value relative to GDP	
		Status-quo	Reform (efficient)
Consumption		0.70	0.67
Capital		3.00	3.46
Government Debt		0.50	-0.06
Net Worth		3.53	3.43
Total Tax Revenue		0.23	0.23
	Labor income	0.14	0.13
	Consumption	0.04	0.04
	Capital	0.06	0.07
	Bequest	0.00	0.04
	Asset Subsidy	0.00	-0.05
Total Transfers		0.13	0.14
	Post retirement	0.09	0.07
	Pre retirement	0.04	0.07

To keep our analysis tractable we have focused on permanent ability types and abstracted form idiosyncratic shocks that are the focus of most of 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 saving 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 large body of evidence that points to a strong correlation between socio-economic 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

ALES, L., R. HOSSEINI, AND L. E. JONES (2014): "Is There" Too Much" Inequality in Health Spending Across Income Groups?" *Unpublished*. 35

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. 19

AMERIKS, J., A. CAPLIN, S. LAUFER, AND S. VAN NIEUWERBURGH (2011): "The joy of giving or assisted living? Using strategic surveys to separate public care aversion from bequest motives," *The journal of finance*, 66, 519–561. 21

ATKINSON, A. B. AND J. E. STIGLITZ (1976): "The design of tax structure: direct versus indirect taxation," *Journal of public Economics*, 6, 55–75. 3

AUTEN, G. AND D. JOULFAIAN (1996): "Charitable contributions and intergenerational transfers," *Journal of Public Economics*, 59, 55–68. 22

BADEL, A. AND M. HUGGETT (2014): "Taxing top earners: a human capital perspective," *Federal Reserve Bank of St. Louis Working Paper Series*. 20

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. 21

BELLOFATTO, A. A. (2015): "Wealth Taxation and Life Expectancy," *Unpublished*. 7

BENARTZI, S., A. PREVITERO, AND R. H. THALER (2011): "Annuitization puzzles," *The Journal of Economic Perspectives*, 143–164. 3

BLANDIN, A. (2015): "Reforming Social Security: The Earnings Cap and Endogenous Human Capital," *Unpublished*. 2, 6

BROWN, J. R. (2001): "Private Pensions, Mortality Risk, and The Decision to Annuitize," *Journal of Public Economics*, 82, 29–62. 3

CAGETTI, M. AND M. DE NARDI (2009): "Estate Taxation, Entrepreneurship, and Wealth," *American Economic Review*, 99, 85–111. 27

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. 7

CASTANEDA, A., J. DÍAZ-GIMÉNEZ, AND J.-V. RÍOS-RULL (2003): "Accounting for the US earnings and wealth inequality," *Journal of Political Economy*, 111, 818–857. 27

CONESA, J. C. AND D. KRUEGER (2006): "On the optimal progressivity of the income tax code," *Journal of Monetary Economics*, 53, 1425–1450. 6

CRISTIA, J. P. (2009): "Rising mortality and life expectancy differentials by lifetime earnings in the United States," *Journal of Health Economics*, 28, 984–995. 3

DE NARDI, M. (2004): "Wealth inequality and intergenerational links," *The Review of Economic Studies*, 71, 743–768. 21, 22, 26, 27

DE NARDI, M. AND F. YANG (forthcoming): "Wealth inequality, family background, and estate taxation," *Journal of Monetary Economics*. 22

DIAMOND, P. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 83–95. 14

DUGGAN, J. E., R. GILLINGHAM, AND J. S. GREENLEES (2007): "Mortality and lifetime income: evidence from US social security records," . 3

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. 20

FARHI, E. AND I. WERNING (2012): "Capital taxation: Quantitative explorations of the inverse Euler equation," *Journal of Political Economy*, 120, 398–445. 35

——— (2013): "Insurance and Taxation over the Life Cycle," *The Review of Economic Studies*, 80, 596–635. 6

FUKUSHIMA, K. (2011): "Quantifying the welfare gains from flexible dynamic income tax systems," *manuscript*. 6

GALE, W. G. AND J. K. SCHOLZ (1994): "Intergenerational transfers and the accumulation of wealth," *The Journal of Economic Perspectives*, 145–160. 22

GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal Indirect and Capital Taxation," *Review of Economic Studies*, 70, 569–587. 35

GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (forthcoming): "Redistribution and social insurance," *American Economic Review*. 6, 20, 35

GOLOSOV, M., M. TROSHKIN, A. TSYVINSKI, AND M. WEINZIERL (2013): "Preference heterogeneity and optimal capital income taxation," *Journal of Public Economics*, 97, 160–175. 3, 6, 17

GOLOSOV, M. AND A. TSYVINSKI (2006): "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy*, 114, 257–269. 6

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. 20

HEATHCOTE, J., K. STORESLETTEN, AND G. L. VIOLANTE (2014): "Optimal tax progressivity: An analytical framework," Tech. rep., National Bureau of Economic Research. 6, 24, 25

HEATHCOTE, J. AND H. TSUJIYAMA (2015): "Optimal Income Taxation: Mirrlees Meets Ramsey," *Research Department Staff Report 507*. 6, 20

HONG, J. AND J.-V. RÍOS-RULL (2007): "Social Security, Life Insurance and Annuities for Families," *Journal of Monetary Economics*, 54, 118–40. 7

HORIUCHI, S. AND A. J. COALE (1982): "A simple equation for estimating the expectation of life at old ages," *Population Studies*, 36, 317–326. 20

HOSSEINI, R. (2015): "Adverse Selection in the Annuity Market and the Role for Social Security," *Journal of Political Economy*, 123, 941–984. 7

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. 7

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. 6, 13

HUGGETT, M. AND G. VENTURA (1999): "On the distributional effects of social security reform," *Review of Economic Dynamics*, 2, 498–531. 24

HURD, M. AND J. P. SMITH (2002): "Expected bequests and their distribution," Tech. rep., National Bureau of Economic Research. 22, 26, 34

İMROHOROĞLU, A., S. İMROHOROĞLU, AND D. H. JOINES (1995): "A Life Cycle Analysis of Social Security," *Economic Theory*, 6, 83–114. 7

KITAO, S. (2014): "Sustainable Social Security: Four Options," *Review of Economic Dynamics*, 17, 756–779. 5

KOCHERLAKOTA, N. R. (2010): *The New Dynamic Public Finance*, Princeton University Press. 16

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. 11

MCDANIEL, C. (2007): "Average tax rates on consumption, investment, labor and capital in the OECD 1950-2003." Arizona State Univeristy. 24

MCGRATTAN, E. R. AND E. C. PRESCOTT (2013): "On financing retirement with an aging population," Tech. rep., National Bureau of Economic Research. 2, 6

MIRRLEES, J. (1971): "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 38, 175–208. 14

NISHIYAMA, S. AND K. SMETTERS (2007): "Does Social Security Privatization Produce Efficiency Gains?" *Quarterly Journal of Economics*, 122, 1677–1719. 2, 5, 19

PASHCHENKO, S. (2013): "Accounting for Non-Annuitization," *Journal of Public Economics*, 98, 53–67. 21

PIKETTY, T. AND E. SAEZ (2013): "A theory of optimal inheritance taxation," *Econometrica*, 81, 1851–1886. 6, 17

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. 15

POTERBA, J. M. (2014): "Retirement security in an aging population," *The American Economic Review*, 104, 1–30. 2, 32

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. 20

ROGOT, E., P. D. SORLIE, AND N. J. JOHNSON (1992): "Life Expectancy by Employment Status, Income, and Education in the National Longitudinal Mortality Study," *Public Health Reports*, 104, 457–461. 3

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–229. 14, 17

——— (2002): "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes," *Journal of Public Economics*, 83, 217–230. 3, 6, 17

SHOURIDEH, A. AND M. TROSHKIN (2015): "Providing Efficient Incentives to Work: Retirement Ages and the Pension System," *Available at SSRN 2062355*. 6, 19

SIEGEL, J. J. AND D. G. COXE (2002): *Stocks for the long run*, vol. 3, McGraw-Hill New York. 23

TUOMALA, M. (1990): *Optimal Income Tax and Redistribution*, Oxford University Press. 17

WALDRON, H. (2013): "Mortality Differentials by Lifetime Earnings Decile: Implications for Evaluations of Proposed Social Security Law Changes," *Social Security Bulletin*, 73. 3, 20, 21

WEINZIERL, M. (2011): "The Surprising Power of Age-Dependent Taxes," *Review of Economic Studies*, 78, 1490–1518. 6

WERNING, I. (2007): "Pareto Efficient Income Taxation," *Massachusetts Institute of Technology, Manuscript*, 44, 45. 6

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*. 5

Appendix

A Proofs

A.1 Proof of Lemma 1

Proof. Consider the individual maximization problem for type θ where hours $l_t(\theta)$ are replaced by $y_t(\theta) = \varphi_t(\theta)\, l_t(\theta)$.

$$U(\theta) = \max_{c_t, y_t, a_{t+1}, b_{t+1}} \sum_{t=0}^{T} \beta^t P_t(\theta) \left[u(c_t) - v\left(\frac{y_t}{\varphi_t(\theta)}\right) + \beta\left(1 - p_{t+1}(\theta)\right) w(b_{t+1}) \right]$$

subject to

$$\begin{aligned} \left(1+\tau_{c}\right) c_{t}+a_{t+1} &= \left(y_{t}-T_{y,t}\left(y_{t}\right)\right) \mathbf{1}\left[t<R\right]+\left(1+r\right) a_{t}-T_{a,t}\left(\left(1+r\right) a_{t}\right)+S_{t}, \\ b_{t} &= \left(1+r\right) a_{t+1}-T_{b,t}\left(\left(1+r\right) a_{t+1}\right). \end{aligned}$$

Note that θ does not appear in the budget constraint. Now take envelope condition with respect to θ

$$\begin{aligned} U'\left(\theta\right) &= \sum_{t=0}^{T} \beta^{t} P_{t}'\left(\theta\right) \left[u(c_{t})-v\left(\frac{y_{t}}{\varphi_{t}\left(\theta\right)}\right)+\beta\left(1-p_{t+1}\left(\theta\right)\right) w(b_{t+1})\right] \\ &\quad + \sum_{t=0}^{T} \beta^{t} P_{t}\left(\theta\right) \frac{\varphi_{t}'\left(\theta\right) y_{t}}{\varphi_{t}\left(\theta\right)^{2}} v'\left(\frac{y_{t}}{\varphi_{t}\left(\theta\right)}\right) \\ &\quad - \sum_{t=0}^{T} \beta^{t+1} P_{t}\left(\theta\right) p_{t+1}'\left(\theta\right) w(b_{t+1}) \\ &= \sum_{t=0}^{T} \beta^{t} P_{t}'\left(\theta\right) \left[u(c_{t})-v\left(\frac{y_{t}}{\varphi_{t}\left(\theta\right)}\right)\right] \\ &\quad + \sum_{t=0}^{T} \beta^{t} P_{t}\left(\theta\right) \left[\frac{\varphi_{t}'\left(\theta\right) y_{t}}{\varphi_{t}\left(\theta\right)^{2}} v'\left(\frac{y_{t}}{\varphi_{t}\left(\theta\right)}\right)+\left(\frac{P_{t}'\left(\theta\right)-P_{t+1}'\left(\theta\right)}{P_{t}\left(\theta\right)}\right) \beta w(b_{t+1})\right] \end{aligned}$$

where the last equality is because $P_{t+1}'\left(\theta\right)=\left(P_{t}\left(\theta\right)\cdot p_{t+1}\left(\theta\right)\right)'=P_{t}'\left(\theta\right)\cdot p_{t+1}\left(\theta\right)+P_{t}\left(\theta\right)\cdot p_{t+1}'\left(\theta\right)$. Now, replace $l_{t}=\frac{y_{t}}{\varphi_{t}(\theta)}$ and evaluate at the solution $\left\{c_{t}\left(\theta\right), l_{t}\left(\theta\right), b_{t}\left(\theta\right)\right\}$

$$\begin{aligned} U'\left(\theta\right) &= \sum_{t=0}^{T} \beta^{t} P_{t}'\left(\theta\right) \left[u(c_{t}\left(\theta\right))-v\left(l_{t}\left(\theta\right)\right)\right] \\ &\quad + \sum_{t=0}^{T} \beta^{t} P_{t}\left(\theta\right) \left[\frac{\varphi_{t}'\left(\theta\right) l_{t}\left(\theta\right)}{\varphi_{t}\left(\theta\right)} v'\left(l_{t}\left(\theta\right)\right)+\left(\frac{P_{t}'\left(\theta\right)-P_{t+1}'\left(\theta\right)}{P_{t}\left(\theta\right)}\right) \beta w(b_{t+1}\left(\theta\right))\right] \end{aligned}$$

□

A.2 Proof of Lemma 2

Proof. Let $dF\left(\theta\right)=f\left(\theta\right) d\theta$ where $f\left(\theta\right)$ is the density function. Let $\eta\left(\theta\right) f\left(\theta\right)$, $\mu\left(\theta\right) f\left(\theta\right)$ and $\gamma\left(\theta\right) f\left(\theta\right)$ be multipliers on equations (12), (13) and (14) respectively. The first order conditions for this problem are

$$\left(\eta(\theta)+\mu(\theta)\frac{P_t'(\theta)}{P_t(\theta)}\right)u'(c_t(\theta)) = \frac{1}{\beta^t(1+r)^t} \tag{25}$$

$$\left(\eta(\theta)+\mu(\theta)\frac{P_t'(\theta)-P_{t+1}'(\theta)}{P_t(\theta)-P_{t+1}(\theta)}\right)w'(b_{t+1}(\theta)) = \frac{1}{\beta^{t+1}(1+r)^{t+1}} \tag{26}$$

$$\left(\eta(\theta)-\mu(\theta)\frac{\varphi_t'(\theta)}{\varphi_t(\theta)}\left(1+l_t(\theta)\frac{v''(l_t(\theta))}{v'(l_t(\theta))}\right)+\mu(\theta)\frac{P_t'(\theta)}{P_t(\theta)}\right)v'(l_t(\theta)) = \frac{\varphi_t(\theta)}{\beta^t(1+r)^t} \tag{27}$$

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

and the boundary conditions

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

Combine Equations (25) and (27) and let $\varepsilon_{F,t}(\theta) = \frac{v'(l_t(\theta))}{v''(l_t(\theta))l_t(\theta)}$.

$$\begin{aligned}\frac{v'(l_t(\theta))}{\varphi_t(\theta)u'(c_t(\theta))} &= \frac{\eta(\theta)+\mu(\theta)\frac{P_t'(\theta)}{P_t(\theta)}}{\eta(\theta)+\mu(\theta)\frac{P_t'(\theta)}{P_t(\theta)}-\mu(\theta)\frac{\varphi_t'(\theta)}{\varphi_t(\theta)}\left(1+l_t(\theta)\frac{v''(l_t(\theta))}{v'(l_t(\theta))}\right)} \\ &= \frac{\eta(\theta)+\mu(\theta)\frac{P_t'(\theta)}{P_t(\theta)}}{\eta(\theta)+\mu(\theta)\frac{P_t'(\theta)}{P_t(\theta)}-\mu(\theta)\frac{\varphi_t'(\theta)}{\varphi_t(\theta)}\left(1+1/\varepsilon_{F,t}(\theta)\right)}.\end{aligned}$$

Therefore,

$$\begin{aligned}\tau_{\text{labor},t}(\theta) &= 1-\frac{v'(l_t(\theta))}{\varphi_t(\theta)u'(c_t(\theta))} \\ &= \frac{-\mu(\theta)\frac{\varphi_t'(\theta)}{\varphi_t(\theta)}\left(1+1/\varepsilon_{F,t}(\theta)\right)}{\eta(\theta)+\mu(\theta)\frac{P_t'(\theta)}{P_t(\theta)}-\mu(\theta)\frac{\varphi_t'(\theta)}{\varphi_t(\theta)}\left(1+1/\varepsilon_{F,t}(\theta)\right)}\end{aligned}$$

and

$$\frac{\tau_{\text{labor},t}(\theta)}{1-\tau_{\text{labor},t}(\theta)} = \frac{-\frac{\mu(\theta)}{\eta(\theta)}\frac{\varphi_t'(\theta)}{\varphi_t(\theta)}\left(1+1/\varepsilon_{F,t}(\theta)\right)}{1+\frac{\mu(\theta)}{\eta(\theta)}\frac{P_t'(\theta)}{P_t(\theta)}} \tag{29}$$

Now, note that from Equation (25), $\eta(\theta) = \frac{1}{u'(c_0(\theta))}$. Form (28) we can solve for $\mu(\theta)$

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

Therefore,

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

where

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

By replacing these back into (29) we get Equation (15). □

A.3 Proof of Proposition 1

Proof. From Equation (25)

$$\frac{u'(c_t(\theta))}{\beta(1+r)u'(c_{t+1}(\theta))} = \frac{\eta(\theta) + \mu(\theta) \frac{P'_{t+1}(\theta)}{P_{t+1}(\theta)}}{\eta(\theta) + \mu(\theta) \frac{P'_t(\theta)}{P_t(\theta)}}$$

Therefore,

$$\begin{aligned}
\tau_{\text{annuity},t}(\theta) &= 1 - \frac{u'(c_t(\theta))}{\beta(1+r)u'(c_{t+1}(\theta))} \\
&= \frac{\frac{\mu(\theta)}{\eta(\theta)} \left(\frac{P'_t(\theta)}{P_t(\theta)} - \frac{P'_{t+1}(\theta)}{P_{t+1}(\theta)}\right)}{1 + \frac{\mu(\theta)}{\eta(\theta)} \frac{P'_t(\theta)}{P_t(\theta)}} \\
&= \frac{-\frac{\mu(\theta)}{\eta(\theta)} \frac{p'_{t+1}(\theta)}{p_{t+1}(\theta)}}{1 + \frac{\mu(\theta)}{\eta(\theta)} \frac{P'_t(\theta)}{P_t(\theta)}} \\
&= \frac{p'_{t+1}(\theta)}{p_{t+1}(\theta)} \left(\frac{1-F(\theta)}{f(\theta)}\right) \frac{g(\theta)}{1 - \left(\frac{1-F(\theta)}{f(\theta)}\right) g(\theta) \frac{P'_t(\theta)}{P_t(\theta)}}.
\end{aligned}$$

The second equality is true because $\frac{P_t'(\theta)}{P_t(\theta)} - \frac{P_{t+1}'(\theta)}{P_{t+1}(\theta)} = \frac{p_{t+1}'(\theta)}{p_{t+1}(\theta)}$. The derivation of Equation (18) is similar. □

A.4 Proof of Proposition 2

Proof. Recall that

$$\begin{aligned}
\tau_{\text{un},t}(\theta) &= 1 - \frac{u'(c_t(\theta))}{\beta(1+r)\left[p_{t+1}(\theta)\, u'(c_{t+1}(\theta)) + (1-p_{t+1}(\theta))\, w'(b_{t+1}(\theta))\right]} \\
&= 1 - \frac{1}{\beta(1+r)\left[p_{t+1}(\theta)\frac{u'(c_{t+1}(\theta))}{u'(c_t(\theta))} + (1-p_{t+1}(\theta))\frac{w'(b_{t+1}(\theta))}{u'(c_t(\theta))}\right]} \\
&= 1 - \frac{1}{\frac{p_{t+1}(\theta)}{1-\tau_{\text{annuity},t}(\theta)} + \frac{1-p_{t+1}(\theta)}{1-\tau_{\text{life insurance},t}(\theta)}}.
\end{aligned}$$

Therefore,

$$\frac{\tau_{\text{un},t}(\theta)}{1-\tau_{\text{un},t}(\theta)} = \frac{p_{t+1}(\theta)}{1-\tau_{\text{annuity},t}(\theta)} + \frac{1-p_{t+1}(\theta)}{1-\tau_{\text{life insurance},t}(\theta)} - 1.$$

Now replace formulas for the annuity and life insurance wedges

$$\frac{\tau_{\text{un},t}(\theta)}{1-\tau_{\text{un},t}(\theta)} = \frac{p_{t+1}(\theta)}{1-\frac{p_{t+1}'(\theta)}{p_{t+1}(\theta)}\frac{1-F(\theta)}{f(\theta)}\frac{g(\theta)}{1-g(\theta)\frac{1-F(\theta)}{f(\theta)}\frac{P_t'(\theta)}{P_t(\theta)}}} + \frac{1-p_{t+1}(\theta)}{1+\frac{p_{t+1}'(\theta)}{1-p_{t+1}(\theta)}\frac{1-F(\theta)}{f(\theta)}\frac{g(\theta)}{1-g(\theta)\frac{1-F(\theta)}{f(\theta)}\frac{P_t'(\theta)}{P_t(\theta)}}} - 1.$$

Note that $p_{t+1}(\theta)\left(1-\tau_{\text{annuity},t}(\theta)\right) + (1-p_{t+1}(\theta))\left(1-\tau_{\text{life insurance},t}(\theta)\right) = 1$. Therefore, for larger $p_{t+1}'(\theta)$ the denumerator in the right hand side becomes a mean preserving spread. As a result, the expected value of its inverse increases as spread (i.e., $p_{t+1}'(\theta)$) increases - under the assumption that everything else remains constant. □