Optimal Taxation of Inheritance and Retirement Savings

Yena Park
University of Rochester

October, 2019

[Preliminary and Incomplete]

Abstract

Can we find rationales for taxing or subsidizing inherited wealth without relying on the preferences of the society? To answer this question, we study optimal taxation of inheritance and retirement savings in a model where the bequest and saving motives are driven by the differences in medical expenses, mortality risk, and patience, as well as heterogeneous productivity. We show that the correlations between each of these factors and the earning productivities are the key for the marginal inheritance taxation. Positive inheritance taxes are optimal when rich people face higher medical expenses and are more patient. In the presence of heterogeneous mortality risk, longer life expectancy of more productive workers increases the tax on retirement savings and decreases the tax on inherited wealth.

Keywords: Optimal Taxation, Inheritance, Retirement Savings
JEL Classification: E21, D64, H21

1 Introduction

Optimal inheritance tax has been a highly controversial issue for both policy makers and economists. Studies on the desirability and the properties of optimal wealth transfer taxation have been developed, but there is very little consensus on the theoretical justification of taxing bequest. Celebrated uniform-taxation results (Atkinson and Stiglitz (1976)) imply zero bequest taxes if the social welfare function sorely takes the utility from the parent's perspective.[1]

The previous studies in the literature find reason for taxing/subsidizing bequest based on the preferences of the society. If the society attaches direct welfare weight to future generations, there are positive externalities of bequest which are not fully internalized by the parents' bequest decision in the most widely used model of bequest — altruistic bequest model or the warm-glow bequest model (Farhi and Werning (2010), Kopczuk (2010)). For example, with the altruistic bequest motive, the society directly values the children's utility from the bequest in addition to the indirect utility through the parent's altruism, thus the society wants to subsidize the inherited wealth. On the other hand, Piketty and Saez (2013) show that the optimal tax rate is positive if the society has a meritocratic preferences — society cares mostly about people receiving little inheritance. Farhi and Werning (2013a) provide another investigation with a social planner who has preferences for redistribution across heterogeneous altruism, and show that wide range of results can be optimal, from taxes to subsidies.

Although previous studies provide important implications, it is not satisfactory to find a rationales of the bequest taxation sorely based on the preferences of the society. Why does the society has additional redistribution objective across different inherited wealth? Choosing a specific social welfare function is unwelcome task in general. The main policy debate on bequest tax centers around the equity versus efficiency trade-off, thus we need to study this issue based on the key trade-off. The studies based on the meritocratic preferences and the Pareto weight for the future generation provide important rationales for taxing/subsidizing the bequest from the (normative) equity perspective, but we have very limited understanding on the efficiency cost of taxing/subsidizing bequest.

In this paper, we study the rationales for taxing/subsidizing inheritance based on the efficiency perspective. Without meritocratic preferences, does the society still want to redistribute beyond the income redistribution? That is, we analyze optimal inheritance tax in the world where there

[1] See Farhi and Werning (2010) for the detailed application of the uniform taxation to the bequest context with altruistic bequest motive.

is a theoretical justification of taxing/subsidizing bequest without assuming any additional redistributive preferences of the society — except for the traditional redistribution preferences across skills/income.

In the real world, the important factors determining the saving motives and the bequest motive in the later stage of life includes medical expenses, uncertain lifespan, and discounting factors. It is not so surprising that these factors are not completely independent of earnings ability. We show that positive (negative) correlations between these factors and earning abilities can generate rationales for positive (negative) tax on the inherited wealth. The correlation in these factors generate preferences for bequest that vary with ability. If the bequest is more (less) preferred by the high ability worker, taxing (subsidizing) bequest can improve the equity-efficiency trade-off. This is because leaving bequest becomes a source of indirect evidence about who has higher abilities and thus more efficient redistributive taxation can be achieved. Thus, even if we do abstract from additional redistributive preferences of the society — such as meritocratic preferences and the Pareto weight on the future generation, we can provide a theoretical justification for taxing (subsidizing) the inherited wealth.

To show this, we extend the canonical Mirrleesian tax model — the model with heterogeneous earning productivity and the nonlinear labor income tax — by adding bequest decision. We investigate the constrained efficient allocation and its implications for the marginal inheritance tax.

First of all, we examine the role of the heterogeneous need for medical expenses on optimal inheritance taxation. It is now well known that the medical expenses are important for the elderly savings and the bequest. Especially near death, serious health shock and illness can easily deplete the wealth of most retirees and reduce the amount of bequest significantly. Another important feature of the medical expenses is that the average out-of-pocket medical expenditures rise very rapidly with age and income (De Nardi, French, and Jones (2010), De Nardi, French, and Jones (2016)). High ability individuals tend to live longer, and the medical expense risk gets higher with age, which generates a positive correlation between the ability and the medical expenses. We show that this positive correlation generates a higher valuation for the bequest for more skilled parents and thus taxing the inheritance can relax the cost of redistribution — the cost of redistributing across income generating abilities.

Heterogeneous mortality risk generates more subtle implications for the optimal tax system between the inheritance tax and the taxes on the retirement savings. In reality, parent's retirement savings can simultaneously serve both a precautionary savings for longer lifespan and a bequest to children, and thus the two motives of savings cannot be clearly distinguished (Dynan, Skinner, and Zeldes (2002), Lockwood (2014)). When the savings can serve for both the precautionary retirement saving and the bequest, we need to design the optimal taxation/subsidy on the inheritance and the retirement savings jointly. Empirical evidences strongly show that people with higher earning ability tend to live longer (e.g., Cristia (2009) and Waldron (2013)). This positive correlation between the earning abilities and longevity implies that more productive parents have relatively higher valuations for the consumption after retirement than those for the bequest. Then taxing the returns of the retirement savings and subsidizing the inherited wealth can contribute to more efficient redistributive taxation — taxation of redistributing across heterogeneous skills.

Finally, the heterogeneous discount factor can be another rationale of the positive bequest taxation. The intuition for this result is consistent with that in Saez (2002). Saez (2002) shows that when those with higher earning abilities save more, taxation of saving can help with the equity-efficiency trade-off because savings can be an indirect evidence for higher earning abilities. The same argument can be applied to the bequest. Parents with altruistic bequest motive values the bequest more when they have higher discount factor. If there is a positive correlation between the patience and productivities, then positive inheritance taxation can be optimal.

We extend the analysis to the infinite horizon overlapping generation model with realistic life-cycle. This analysis not only confirms the main mechanism of the inheritance tax in the two-period model, but also shows that age-dependent labor income tax and asset income tax can be also used to reduce the cost of redistribution.

The theoretical analysis shows that to determine the sign and shape of the inheritance taxation, we need to carefully set the relationship among the earning ability, medical expenses, mortality, discount factor. We plan to carry out careful quantitative investigation by carefully calibrating the model to the data.

The remainder of the paper is organized as follows. In Section 2, we study build a simple two period model to study the optimal taxation of inheritance and retirement savings. In section 3, we discuss the role of the key factors of the bequest and savings for determining the sign of the taxes. In section 4, we extend the analysis to the general social welfare function and infinite horizon economy. In section 5, we plan to present the quantitative investigation.

Literature Review

[List of papers that need to be discussed:]

- Piketty and Saez (2013)
- Farhi and Werning (2010) and Farhi and Werning (2013a)
- Pavoni and Yazici (2017)
- Laitner and Juster (1996)
- Lockwood (2014)
- Hosseini and Shourideh (2016)

2 The Model: Two Period Economy

In this section, we use a simple two period economy to provide a theoretical analysis of optimal taxation on bequest and retirement savings.

There are two generations, parents born at $t = 1$ and children born at $t = 2$. A continuum of parents live for at most two periods. Upon birth, each parent draws a type $\theta \in \Theta$ from a continuous distribution $F(\theta)$ that has density $f(\theta)$. The skill type θ determines labor productivity when parents are young ($t = 1$). As in the Mirrlees economy, the planner does have standard preferences of redistribution over the heterogeneous skill θ — the planner wants to transfer from the high skilled agents to the low skilled agents. However, the skill type is not observable, — or equivalently, the tax system cannot depend on the skill but only on income — which creates the standard efficiency cost of redistribution. Thus, the planner in this economy faces the traditional equity-efficiency trade-off of optimal labor income taxation.

The new feature of this economy which are the key for generating rationales for taxing/subsidizing human capital is that this skill type is correlated with important factors for the saving and bequest. We capture this correlation in the simplest possible way by assuming that the skill type also determines the following characteristics of the individuals : 1. θ also determines the average medical expenses near death $m(\theta)$, 2. θ also determines the mortality risk which is captured by the survival probability at $t = 2$, $P(\theta)$, 3. It also determines the average discount factor of future $\beta(\theta)$.

For the theoretical analysis, we only make assumptions on the sign of $\dot{m}(\theta)$, $\dot{P}(\theta)$, and $\dot{\beta}(\theta)$, which are the keys for the sign of the inheritance taxation and asset income taxation. First of all, rich people are healthier and tend to live longer, and thus we assume that the average probability of living long $P(\theta)$ is higher for parents with higher productivity ($\dot{P}(\theta) > 0$).

Next, another important factor for the retirement savings and the bequest is the near-death medical expenses. We assume that the average medical expenses $m(\theta)$ are increasing in skill θ because of the following reasons. It is now well known that the out-of-pocket medical expenditures for the elderly household is rapidly increasing with age (e.g., De Nardi, French, and Jones (2010), De Nardi, French and Jones (2016)). Older individuals are much more likely to be in a nursing home, and they use more expensive medical services at older ages. Combined with the longer lifespan of the high skilled people, high-skilled agents tend to spend more medical expenditures at the end of their life. In sum, we capture the effects of skill type on the medical expenditure by assuming $\dot{m}(\theta) > 0$.

There can be counter-effects of high skills on the medical expenditures. Empirical evidences show that rich people tend to be healthier at given age, which might lead to lower medical cost for the high skilled people. However, we assume that age effects dominate and thus $\dot{m}(\theta) > 0$. More precisely, the average medical expenditure of the elderly (e.g. above age 65) is

$$m(\theta) = \sum_{j \geq 65} m_j(\theta) P_j(\theta),$$

where $m_j(\theta)$ is the medical expenditure of type θ at age j and $P_j(\theta)$ is the survival probability of type θ at age j. Thus, the medical expenditure depends on the skill type through the health effects and age effects:

$$m'(\theta) = \underbrace{\sum_{j \geq 65} m_j'(\theta) P_j(\theta)}_{\text{health effects}} + \underbrace{\sum_{j \geq 65} m_j(\theta) P_j'(\theta)}_{\text{age effects}}.$$

The productivity θ captures comprehensive income-generating ability including the health status of the individual. If the good health of high skilled people implies less medical expenditures at given age ($m_j'(\theta) < 0$), there can be counter effects of the skill. However, as long as the age effects dominate, medical expenditure increases in the skill. The empirical evidence of Jones, De Nardi, French, McGee, and Kirschner (2018) supports the dominance of the age effects — People initially in good health have higher lifetime expenditures than those initially in bad health due to longer life expectation and higher medical costs at higher age.[2]

[2] Jones, De Nardi, French, McGee, and Kirschner (2018) also document that although high income people are less likely to be in a nursing home at any given age, they live longer, and older individuals are much more likely to be in a nursing home.

We also want to remark that the key correlation for taxing/subsidizing inherited wealth is the correlation between the skill and the medical expenditure, not the correlation between the income and the medical expenditure. It is well know that medical expenditure is increasing with income due to the progressive Medicare and Medicaid program of the US government. The out-of-pocket medical expenditure is much higher for the higher income household. This correlation between income and medical expenditures created by the government medical policy cannot generate the rationales for taxing/subsidizing the bequest, because the individuals optimally adjust their labor supply considering this progressive government policy unless the households have some myopic perspective in their labor supply decision. This will be more clearly explained when we show the optimal tax result.

Finally, the empirical evidences on the relationship between the discount factor and the productivity are more subtle, because there are many other factors than can make rich people save more and bequest more. For most of the theoretical analysis, we assume that $\dot{\beta}(\theta) > 0$, but we also discuss other possibilities below.

For the theoretical analysis, we do not have to make strong assumptions on the shape of these functions $m(\theta)$, $P(\theta)$, and $\beta(\theta)$. However, to get the sharp policy implications, we need to discipline the slope of these functions based on the empirical evidences. Later, in the quantitative analysis, the shape of $m(\theta)$, $P(\theta)$, and $\beta(\theta)$ will be determined to match the empirical correlation between the earnings productivity and medical expenses, the correlation between the income and the mortality, and the correlation between the income and the savings/bequest.

In the two-period life-time model, each period represent young and old, respectively. At $t = 1$, parents first learn their type θ and then provide $l(\theta)$ units of work effort to produce efficiency units of labor $y(\theta) = \theta \cdot l(\theta)$. With probability $P(\theta)$, the parents with productivity θ live long — survive in period 2. Near death, parents face negative health shock which causes the average medical expenses $m(\theta)$. After the parents die, remaining wealth will be inherited to the child, but the medical expenses near death will reduce the inherited wealth to the child. The utility of the parent with productivity θ is given by

$$U(c_1, c_2, c_c, y; \theta) \quad = \quad u(c_1) - \psi\left(\frac{y}{\theta}\right) + P(\theta)\beta(\theta)\left[u(c_2) + \beta(\theta)V_c(c_{c3}^*)\right] + (1 - P(\theta))\beta(\theta)V_c(c_{c2}^*),$$

where c_1 is consumption when parents are young and c_2 are the consumption after the retirement if they live long, and c_{c2}^* and c_{c3}^* are the consumption of children. We assume that the flow utility from parent's consumption $u(\cdot)$ and the child's utility $V_c(\cdot)$ are strictly concave and the disutility of parent's working $\psi(\cdot)$ is strictly convex.

In the presence of uncertainty in mortality, parents' savings in period $t = 1$ serve both purposes of savings — precautionary savings for the retirement and savings for leaving the bequests. Parents save against the future risk of living a long time after the retirement, which happens with probability $P(\theta)$. In the event they do not survive, however, bequests are given to their children, and the parents value the bequest. Thus, in this model, the parents have an altruistic bequest motive, but the motive is operative only in the state of the death. This implies that the precautionary savings motive for the life-cycle after retirement and the bequest motives are overlapping and cannot be clearly distinguished, and thus optimal tax system on the bequest and retirement savings should be designed together. This operative bequest motive was first pointed out by Dynan, Skinner, and Zeldes (2002) and has been adopted in the papers studying the elderly savings and bequest (De Nardi, French, and Jones (2010),Lockwood (2014)). Lockwood (2014) has shown that this overlapping bequest motives can explain retiree's puzzling decisions on savings and insurance. In this paper, we study the optimal inheritance taxation in the economy where the two motives of the bequest are overlapping due to the uncertainty in life-cycle.

In addition to the production, there is an endowment of goods I_1 and I_2 to each parent in period $t = 1$ and in period $t = 2$ (if they survive). Goods can be transferred between periods $t = 1$ and $t = 2$ with a savings technology with rate of return $R > 0$. The total cost for the child's consumption and the medical expenses are denoted by $c_{c2}(\theta) = c^*_{c2}(\theta) + m(\theta)$ and $c_{c3}(\theta) = c^*_{c3}(\theta) + m(\theta)$ respectively. Then, an allocation is resource feasible if

$$\int \left[c_1(\theta) + P(\theta) \left\{ \frac{1}{R} c_2(\theta) + \frac{1}{R^2} c_{c3}(\theta) \right\} + (1 - P(\theta)) \frac{1}{R} c_{c2}(\theta) \right] dF(\theta) \leq \int y(\theta) dF(\theta) + I, \quad (1)$$

where $I = I_1 + \frac{I_2}{R}$.

We assume that each parent's θ type is private information. We can restrict attention to direct mechanisms applying the revelation principle — parents report their type and receive allocation as a function of this report. An allocation is incentive compatible if

$$u(c_1(\theta)) - \psi\left(\frac{y(\theta)}{\theta}\right) + P(\theta)\beta(\theta)\left[u(c_2(\theta)) + \beta(\theta)V_c(c^*_{c3}(\theta))\right] + (1 - P(\theta))\beta(\theta)V_c(c^*_{c2}(\theta)) \geq$$
$$u(c_1(\theta')) - \psi\left(\frac{y(\theta')}{\theta}\right) + P(\theta)\beta(\theta)\left[u(c_2(\theta')) + \beta(\theta)V_c(c^*_{c3}(\theta'))\right] + (1 - P(\theta))\beta(\theta)V_c(c^*_{c2}(\theta')), \quad \forall \theta, \theta'. \quad (2)$$

An allocation is incentive feasible if it satisfies the resource constraint (1) and the incentive compatibility constraint (2).

In the benchmark analysis, we assume that the social planner only cares for the utility of the parents, and the utility of the children are valued by the planner only indirectly through the altruism of the parents. Later we will extend our analysis to the case where the social planner values the children's utility directly. We consider a weighted Utilitarian criterion:

$$\int \lambda(\theta)U(c_1, c_2, c_c, y; \theta)f(\theta)d\theta, \tag{3}$$

where $\lambda(\theta)$ is the weight on the parents of type θ. We can interpret $\lambda(\theta)$ as the Pareto weight, and by varying the weight $\lambda(\theta)$, we can trace out the Pareto frontier. Later, in the quantitative analysis, we will consider a class of Pareto weight function that can invert the U.S. policymakers' tastes for redistribution from the current US tax system. Note that the social welfare function excludes any meritocratic preferences of the society and does not put additional Pareto weights for the future generation. The planner in the benchmark analysis has the redistributive preferences across the skill only, not across the bequest.

We analyze optimal taxation following a Mirrleesian approach. That is, we do not impose any restriction on the policy instruments of the government except that the type θ is private information. Our first goal is to study the constrained efficient allocation to wedges implied by the optimal allocation. The social planner solves the following mechanism design problem of maximizing the social welfare function (3) subject to the resource constraint (1) and the incentive constraints (2).

We now simplify the incentive compatibility constraint (2) to rewrite the planner's problem. A necessary condition for an allocation (c_1, c_2, c_c, y) to be incentive compatible is given by the envelope condition:

$$\begin{aligned} \dot{v}(\theta) \quad = \quad & \psi'\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta^2} + \dot{P}(\theta)\left[\beta(\theta)u(c_2(\theta)) + \beta(\theta)^2 V_c(c_{c3}^*(\theta))\right] - \dot{P}(\theta)\beta(\theta)V_c(c_{c2}^*) \\ & - P(\theta)\beta(\theta)V_c'(c_{c3}^*(\theta))\dot{m}(\theta) - (1-P(\theta))\beta(\theta)V_c'(c_{c2}^*(\theta))\dot{m}(\theta) \\ & + P(\theta)\dot{\beta}(\theta)\left[(\theta)u(c_2(\theta)) + 2\beta(\theta)V_c(c_{c3}^*(\theta))\right] + (1-P(\theta))\dot{\beta}(\theta)V_c(c_{c2}^*(\theta)), \quad \forall\theta, \end{aligned} \tag{4}$$

where $v(\theta)$ is the associated indirect utility function. As in the standard Mirrleesian taxation, we can show that the envelope condition (4) is also sufficient for an allocation to be incentive compatible, if the allocation satisfies additional monotonicity conditions.[3]

[3] In the standard Mirrleesian taxation literature, this is the case under the preferences that satisfy the single-crossing property. Note that in our simple model, the single crossing is always satisfied because we only consider the additively separable utility functions. Our analysis can be extended to general utility functions with the additional single crossing conditions.

Condition 1. ***(MON)***

$c_2(\theta)$, $c^*_{c3}(\theta)$, and $y(\theta)$ are nondecreasing in θ, and $c^*_{c2}(\theta)$ satisfies the following:

$$\begin{cases} \dot{c}^*_{c2}(\theta) \geq 0 & \textit{if } \left\{\frac{\dot{P}(\theta)}{1-P(\theta)} - \frac{\dot{\beta}(\theta)}{\beta(\theta)}\right\} + \frac{V''_c(c^*_{c2}(\theta))}{V'_c(c^*_{c2}(\theta))}\dot{m}(\theta) \geq 0 \\ \dot{c}^*_{c2}(\theta) \leq 0 & \textit{otherwise.} \end{cases}$$

Lemma 2. *If an allocation $(c_1, c_2, c^*_{c2}, c^*_{c3}, y)$ satisfies (4) and the monotonicity condition, then the allocation is incentive compatible.*[4]

Proof We denote the utility of a θ-type agent who reports $\hat{\theta}$ by

$$U(\hat{\theta}, \theta) = u(c_1(\hat{\theta})) - \psi(\frac{y(\hat{\theta})}{\theta}) + P(\theta)\beta(\theta)\left[u(c_2(\hat{\theta})) + \beta(\theta)V_c(c^*_{c3}(\hat{\theta}))\right] + (1-P(\theta))\beta(\theta)V_c(c^*_{c2}(\hat{\theta})).$$

Then the first order condition $\frac{\partial U(\hat{\theta},\theta)}{\partial \hat{\theta}} = 0$ implies the envelope condition (4). Using the first order condition, we obtain $\frac{\partial^2 U(\hat{\theta},\theta)}{\partial \theta \partial \hat{\theta}} + \frac{\partial^2 U(\hat{\theta},\theta)}{\partial \hat{\theta}^2} = 0$, and thus the second order condition of the agent's problem ($\frac{\partial^2 U(\hat{\theta},\theta)}{\partial \hat{\theta}^2} \leq 0$) is equivalent to $\frac{\partial^2 U(\hat{\theta},\theta)}{\partial \theta \partial \hat{\theta}} \geq 0$, where

$$\begin{aligned}
\frac{\partial^2 U(\hat{\theta},\theta)}{\partial \theta \partial \hat{\theta}} \quad = \quad & \left\{\frac{1}{\theta^2}\psi'(\frac{y(\theta)}{\theta}) + \frac{y(\theta)}{\theta^3}\psi''(\frac{y(\theta)}{\theta})\right\}\dot{y}(\theta) \\
& + \dot{P}(\theta)\beta(\theta)\left[u'(c_2(\theta))\dot{c}_2(\theta) + \beta(\theta)V'_c(c^*_{c3}(\theta))\dot{c}^*_{c3}(\theta)\right] \\
& - P(\theta)\beta(\theta)V''_c(c^*_{c3}(\theta))\dot{m}(\theta)\dot{c}^*_{c3}(\theta) - (1-P(\theta))\beta(\theta)V''_c(c^*_{c2}(\theta))\dot{m}(\theta)\dot{c}^*_{c2}(\theta) \\
& + P(\theta)\dot{\beta}(\theta)\left[u'(c_2(\theta))\dot{c}_2(\theta) + 2\beta(\theta)V'_c(c^*_{c3}(\theta))\dot{c}_{c3^*}(\theta)\right] + (1-P(\theta))\dot{\beta}(\theta)V'_c(c^*_{c2}(\theta))\dot{c}_{c2^*}(\theta).
\end{aligned}$$

Then, the monotonicity condition proves the second order condition, and thus (4) is sufficient. ■

Note that the envelope condition (4) and the monotonicity conditions are sufficient, but the monotonicity conditions are not necessary, which is different from the standard Mirrlees optimal taxation analysis.[5] For example, even if $c^*_{c3}(\theta)$ violates the monotonicity condition **(MON)**, if the absolute value of the $\left\{\frac{\dot{P}(\theta)}{1-P(\theta)} - \frac{\dot{\beta}(\theta)}{\beta(\theta)}\right\} + \frac{V''_c(c^*_{c2}(\theta))}{V'_c(c^*_{c2}(\theta))}\dot{m}(\theta)$ is relatively small, then the second order condition is still satisfied. Thus, for most of the theoretical analysis, we focus on the solution of the relaxed problem — the planner's problem without the monotonicity conditions, and then we provide some more investigation when the bunching occurs — when the monotonicity conditions are binding. In the quantitative analysis, we plan to solve the relaxed problem and to check the monotonicity conditions ex post.

[4] If assumption **??** does not hold for all θ, then for those θ-interval, the monotonicity condition changes the sign : $\dot{c}_c(\theta) \geq 0$

[5] In Mirrlees optimal taxation (Mirrlees (1971)), the envelope condition and the monotonicity condition are both necessary and sufficient.

We now recast the planning problem as the problem whose incentive constraints (4) are replaced by the envelope condition (4) and the monotonicity conditions (**MON**).

$$\textbf{(P)} \quad \max_{c_1,c_2,c_c,y,v} \quad \int \lambda(\theta)v(\theta)f(\theta)d\theta \tag{5}$$

$$s.t. \quad v(\theta) = u(c_1(\theta)) - \psi\left(\frac{y(\theta)}{\theta}\right) + P(\theta)u(c_2(\theta)) + (1-P(\theta))\gamma(\theta)V_c(c_c(\theta)) \tag{6}$$

$$(1), (4) \text{ and } (\textbf{MON}).$$

We call the planning problem **(P)** *without* the monotonicity constraint (**MON**) the relaxed problem **(RP)**. Note that the solution of the problem **(P)** and the relaxed problem are the same except for the productivity types whose monotonicity conditions are binding.

We first characterize the optimal allocation — the solution to the relaxed planning problem **(RP)** — by analyzing the wedges implied by the solution. Then, we also discuss how to implement the optimal allocation using a nonlinear tax system.

3 Main Result: Role of Medical Expenses, Mortality, and Discount Factor

3.1 Optimal Wedges

To analyze the implications for the optimal inheritance taxation, we now analyze the wedges implied by the optimal allocation. To analyze the role of the heterogeneous medical expenses, mortality, and the discount factor on the optimal wedges, we solve the planning problem by incorporating these factors one by one. By following these steps, we will be able to clearly understand the role of the each factor on the implicit inheritance taxation.

Before getting into the inheritance tax analysis, we start with the intratemporal wedge, which shows the standard results as in canonical Mirrleesian tax theory.

3.1.1 Intratemporal wedge

Given an allocation $(c_1, c_2, c_{c2}^*, c_{c3}^*, y)$ and a type θ, we define the intratemporal wedges

$$\tau_{intra}(\theta) = 1 - \frac{\psi'(\frac{y(\theta)}{\theta})}{\theta u'(c_1(\theta))}.$$

The intratemporal wedge is the labor wedge, and the optimal labor wedge is characterized as usual.

Proposition 3. *Suppose that* (c_1, c_2, c_c, y) *solves the relaxed planning problem. Then the intratemporal wedge is given by*

$$\tau_{intra}(\theta) = -\frac{\mu(\theta)}{\theta^2 \eta f(\theta)} \left[\psi''(\frac{y(\theta)}{\theta})\frac{y(\theta)}{\theta} + \psi'(\frac{y(\theta)}{\theta}) \right],$$

where $\mu(\theta)$ *and* η *are the multipliers on the envelope condition (4) and the resource constraint (1), respectively.*

If we assume that Pareto weight $\lambda(\theta)$ *is nonincreasing in* θ*, then* $\mu(\theta) < 0$*, which implies that when there is no bunching,*[6]

$$\tau_{intra}(\theta) > 0.$$

Proof See the appendix. ■

Note that the sign of $\mu(\theta)$ determines the sign of the labor wedge, and the sign of $\mu(\theta)$ does depend on the Pareto weight function $\lambda(\theta)$. Proposition 3 shows that $\mu(\theta)$ is negative if $\lambda(\theta)$ is nonincreasing, including the utilitarian case, which is commonly used in the literature. We also want to remark that $\mu(\theta)$ can be negative even if $\lambda(\theta)$ is increasing as long as the slope of $\lambda(\theta)$ is not too high — as long as the social preferences for redistribution across earning ability is not too weak. To formalize this argument, we consider the special parametric functional form assumption on the Pareto weight $\lambda(\theta)$ as in Heathcote and Tsujiyama (2017):

$$\lambda(\theta) = \frac{exp(-\alpha\theta)}{\int exp(-\alpha\theta) f(\theta) d\theta} \tag{7}$$

for some constant α which reflects the planner's preferences for redistribution. A negative α implies that the planner puts higher weight on more productive parents ($\dot{\lambda}(\theta) > 0$), while a positive α implies the planner prefers less productive parents ($\dot{\lambda}(\theta) < 0$). Under this special Pareto weight function, we have the following proposition. We will discuss this in more detail when the sign of $\mu(\theta)$ is crucial for the savings wedges and the bequest wedges.

Proposition 4. *Suppose that the Pareto weight* $\lambda(\theta)$ *takes the form of (7). Then,*

1. *There exists a cutoff* $\hat{\alpha} < 0$ *such that* $\mu(\theta) \leq 0$ *for all* $\alpha \geq \hat{\alpha}$ *and* $\mu(\theta) > 0$ *for all* $\alpha < \hat{\alpha}$.

2. *If* $\alpha \geq \hat{\alpha}$*, then* $\tau_{intra}(\theta) > 0$.

[6] See the appendix for the investigation on the sign of the intratemporal wedge when there is bunching.

Proof See the appendix. ■

Proposition 4 shows that even if the Pareto weight $\lambda(\theta)$ is increasing with negative α, the characterization of the intertemporal wedge above still applies, as long as the planner's preferences for redistribution is not too weak ($\alpha > \hat{\alpha}$).

3.1.2 Atkinson and Stiglitz

As a benchmark, we show that if there is no heterogeneity in medical expenses, mortality, and discount factor, then it is optimal not to have any intertemporal taxes. Thus, the inheritance taxes and any taxes on saving are not useful.

Proposition 5. *Suppose that $\dot{m}(\theta) = \dot{P}(\theta) = \dot{\beta}(\theta) = 0$ for all θ and (c_1, c_{c2}^*, y) solves the planning problem (**P**). Then the following intertemporal conditions hold: for all θ,*

$$u'(c_1(\theta)) = R\beta u'(c_2(\theta)), \quad u'(c_1(\theta)) = R\beta V_c'(c_{c2}^*(\theta)), \quad u'(c_2(\theta)) = R\beta V_c'(c_{c3}^*(\theta)).$$

Proof See the appendix. ■

3.1.3 Role of Medical Expenses

We now start analyzing the role of each factor on the optimal inheritance tax. We begin with the effects of heterogeneous medical expenses. The empirical evidences show that the average medical expenses near death can be very large. Mariacristina, Eric, Bailey, and Jeremy (2016) find that medical expenses more than double between ages 70 and 90. A long stay in nursing home near death can easily exhaust the wealth even for the upper half of the wealth distribution (See Lockwood (2014) and the references there in). French, De Nardi, Jones, Baker, and Doctor (2006) show that the increase in medical spending before death — including the burial expenses — can explain about 24 percent to 37 percent of decline in assets for the elderly households.

Another important feature of the medical expenses is that, as we discussed above, the average medical expenses are increasing as age increases especially at older ages, and combined with the longer lifespan of the higher productivity, higher skilled workers have higher medical expenditures near death. We now show that this positive correlation between the medical expenses and the skill can generate a rationale for the positive inheritance taxation.

To focus on the role of medical expenses, we assume that $P(\theta) = 0$ and $\beta(\theta) = 1$ for all θ. Then the indirect utility for parents is simply

$$v(\theta) = u(c_1(\theta)) - \psi(\frac{y(\theta)}{\theta}) + V_c(c^*_{c2}(\theta)).$$

That is, every parent lives for 1 period dies after the first period, and thus there is no heterogeneity in mortality risk. Then, the only relevant intertemporal distortion we can consider is the inheritance wedge:

$$t_b(\theta) = 1 - \frac{u'(c_1(\theta))}{RV'_c(c^*_{c2}(\theta))}.$$

Next proposition clearly shows the role of medical expenses on the inheritance wedge.

Proposition 6. *Suppose that $P(\theta) = 0$ and $\beta(\theta) = 1$ for all θ. Suppose that $\lambda(\theta)$ is nonincreasing and (c_1, c^*_{c2}, y) solves the relaxed problem (**RP**). Then the optimal inheritance wedge is*

$$t_b(\theta) = \frac{\mu(\theta)}{\eta f(\theta)} \frac{V''_c(c^*_{c2}(\theta))}{V'_c(c^*_{c2}(\theta))} u'(c_1(\theta))\dot{m}(\theta),$$

which implies that $t_b(\theta) > 0$ if $\dot{m}(\theta) > 0$ ($t_b(\theta) < 0$ if $\dot{m}(\theta) < 0$).

Proof See the appendix. ■

If the parent with type θ saves $A(\theta)$ for the medical expenses and bequest, the inherited wealth to children is $R \cdot A(\theta) - m(\theta)$. Thus, higher medical expenses drop the asset near the death of parents and reduces the inheritance, which increases the marginal value of the bequest. Since the average medical expenses are increasing as the θ increases, the parents with higher productivity have a relatively higher valuation for child's consumption. Thus, reducing the children's consumption by taxing inherited wealth can relax the cost of redistribution, and thus positive inheritance is optimal.

As we briefly discussed above, the key correlation for the rationale of the positive bequest tax is not the correlation between the income and the medical expenses, but the correlation between the skill and the medical expenses. The higher medical expenses for the higher income is endogenous result driven by the government medical program — means-tested programs such as Medicaid (see De Nardi, French, and Jones (2016) and Mariacristina, Eric, Bailey, and Jeremy (2016).) The effects of skill on the medical expenses through this endogenous income effects do not provide reason for taxing the bequest, because when high skilled individuals mimick the low skilled individual by reducing the labor supply, the medical expenses are also reduced as the income decreases, which decreases the marginal value of the bequest. Thus, taxing the bequest does not relax the incentive constraints. On the other hand, higher medical expenses for the high skilled agents through the age effects does not depend on the labor supply, and thus taxing the bequest can relax the incentive constraints because high skilled agents value the bequest even more when they are shirking.

If the agents have myopic perspective in their labor decision, however, the income effects — higher medical expenditure of higher skilled agents through higher income — can also generate rationales for taxing the bequest. If the myopic young agents do not fully consider the effects of the future government medical policy when they are supplying labor, this can generate another rationales for taxing the bequest. In this case, the government medical program and the inheritance taxes should be considered together. More redistributive medical program provides a rationale for the positive inheritance taxation.

3.1.4 Role of Mortality

We now introduce heterogeneous mortality to analyze the effects of correlation between the mortality and the income on the inheritance taxes. Empirical evidences show that the uncertainty of the lifespan is huge. When there is uncertainty after the retirement, parents save for both the precautionary motive and the bequest motive. Despite many empirical evidences of showing the importance of the bequest motive, most of the survey on the elderly saving reveals that the most important reasons of savings are life-cycle or precautionary considerations. Dynan, Skinner, and Zeldes (2002) pointed out that this seemingly contradictory survey evidences could be explained by the overlapping motive of the savings. That is, one unit of saving can serve for two purposes, but the first purpose is to insure against the uncertainties (such as living a long time or high health expenditure risk), and the second purpose — bequest motive — becomes operative in the event the parents die early with positive wealth.[7]

Another important feature of the mortality risk is that the rich tend to live long. Many empirical studies show that rich people live much longer than the poor (For example, see Cristia (2009), Waldron (2013), De Nardi, French, and Jones (2016)). The income-generating ability θ in our model is a comprehensive productivity including the health status, and thus the positive correlation between the probability of living long and the productivities are partially reflects the fact that health people live longer. However, the empirical evidences show that even when we control for the health status, the rich tend to live longer.

[7] Recently, Lockwood (2014) also showed that this operative bequest motive is crucial to explain the retiree's saving behavior — why many retirees self-insure late-life risk.

We show that the mortality risk generate the positive intertemporal wedge, but the operative bequest motive and the negative correlation between the mortality has implication toward the positive taxes on the return of the retirement savings and the negative taxes on the inheritance taxes.

To show the role of mortality on the optimal wedge, we now assume that the probability of living long is higher for more productive parents: $\dot{P}(\theta) > 0$ for all θ. To focus on the role of mortality, we assume $m(\theta) = 0$ — no medical expenses at the end of life — and $\beta(\theta) = 1$ for all θ.

Ex ante Intertemporal Wedge

In the presence of life-span uncertainty, we can defined the intertemporal distortions for both ex ante and ex post. The ex ante intertemporal distortion is instructive for the optimal level of saving. However, the intertemporal wedge is the combination of the bequest distortion and the asset distortion, and thus the sign of the intertemporal wedge does not tell us the sign of each distortion. We first investigate the ex ante intertemporal wedge and then study each distortion. We can define the the ex ante intertemporal wedge as usual:

$$\tau_{inter}(\theta) = 1 - \frac{u'(c_1(\theta))}{R[P(\theta)u'(c_2(\theta)) + (1 - P(\theta))V_c'(c_{c2}(\theta))]}.$$

Next proposition shows that the well known inverse Euler equation holds when there is no bunching.

Proposition 7. *Suppose that $\dot{P}(\theta) > 0$, $m(\theta) = 0$, and $\beta(\theta) = 1$ for all θ. Suppose that (c_1, c_2, c_{c2}, y) solves the relaxed planning problem. Then the inverse Euler equation holds:*

$$\frac{P(\theta)}{Ru'(c_2(\theta))} + \frac{1 - P(\theta)}{RV_c'(c_{c2}(\theta))} = \frac{1}{u'(c_1(\theta))}, \quad \forall \theta \in \Theta. \tag{8}$$

Proof See the appendix. ∎

By applying the Jensen's inequality to (8) those parents, we get the following inequality :

$$u'(c_1(\theta)) \leq P(\theta)Ru'(c_2(\theta)) + (1 - P(\theta))RV_c'(c_c(\theta)), \quad \forall \theta,$$

which implies the positive intertemporal wedges :

$$\tau_{inter}(\theta) \geq 0, \quad \forall \theta.$$

Note that the inequality holds for any Pareto weight function $\lambda(\theta)$. We also note that the strict inequality holds as long as $u'(c_2(\theta)) \neq V_c'(c_{c2}(\theta))$. That is, the intertemporal wedge is positive, as long as the the constrained efficient allocation does not achieve the perfect insurance and there is no bunching. As long as the Pareto weight $\lambda(\theta)$ is set to guarantee that $\mu(\theta) \neq 0$, perfect insurance cannot be achieved .

As in the Dynamic Mirrleesian tax literature, the positive intertemporal wedge result is to provide better insurance which is the first order benefit at the second-order cost of reducing the consumption smoothing. The insurance benefit in this economy, however is not the insurance against the future productivity uncertainty, but against the life-cycle uncertainty — the mortality risk.

In general, a positive intertemporal distortion implies that there is need to tax the return of savings. As we discussed above, however, it does not mean that both the bequest distortion and the retirement savings distortion should be positive. To see the implication of each distortion, we now investigate the ex post wedges.

Ex post Wedges

We now define ex post intertemporal distortions, the retirement savings wedge and the bequest wedges. Given an allocation $(c_1, c_2, c_{c2}, c_{c3}, y)$ and a type θ, we define the ex post retirement savings wedge

$$t_a(\theta) = 1 - \frac{u'(c_1)}{Ru'(c_2)}$$

and the ex post bequest wedges

$$t_{b2}(\theta) = 1 - \frac{u'(c_1)}{RV_c'(c_{c2})}, \qquad t_{b3}(\theta) = 1 - \frac{u'(c_2)}{RV_c'(c_{c3})}.$$

These ex post wedges can be understood as the implicit tax on the return to the retirement savings when surviving and the the implicit tax on the inheritance when parents are dying.

Next proposition shows that in the absence of heterogeneity in altruism, the retirement saving wedge and inheritance wedge have the opposite sings.

Proposition 8. *Suppose that $\dot{P}(\theta) > 0$, $m(\theta) = 0$, and $\beta(\theta) = 1$ for all θ. Suppose that $\lambda(\theta)$ is nonincreasing and $(c_1, c_2, c_{c2}, c_{c3}, y)$ solves the relaxed planning problem. Then,*

$$t_a(\theta) > 0, \qquad t_{b2}(\theta) < 0, \quad \forall\theta.$$

Proof See the appendix. ■

This proposition shows that in the presence of mortality risk which is negatively correlated with productivity, the planner who has preferences for redistribution (with nonincreasing $\lambda(\theta)$) would tax the return of the retirement savings and subsidize the inherited wealth. This is because the parents who are likely to live longer values the consumption after the retirement relatively more than the children's consumption. Thus, the planner who wants to redistribute from the high productive parents to low productive parents can achieve this redistribution at the lower efficiency cost by taxing the asset income after the retirement and subsidizing the inherited wealth. In other words, in the presence of a negative correlation between the mortality and productivity, the planner can accomplish distinction among different productivities beyond what can be done by labor income taxes.

Thus, the planner wants to subsidize bequest and tax retirement savings so that the planner can redistribute from the more productive parents who are likely to live longer to the less productive parents whose bequest motive becomes operative with high probability. Note that this redistribution is possible, because the government can observe the realization of the survival and death, and can tax the returns of the saving with contingency. The taxes on the retirement savings are the taxes contingent on the survival, while the inheritance subsidy is the subsidy contingent on the death.

Once again, we want to note that proposition 8 includes the utilitarian case and it is also applied to the case with increasing $\lambda(\theta)$, as long as the preferences for redistribution are not too weak. With the special parametric Pareto weight function (7) we discussed above, proposition 8 holds as long as $\alpha > \hat{\alpha}$.

Note that in this two period model, the ex post bequest wedge is zero for the parents who live long: $t_{b3} = 0$. However, this is because of the simplification assumptions that parents die for sure after living two periods. If we extend the analysis to the economy with multi-periods, then this zero bequest wedge result will only apply to the last periods with no probability of survive.

Another important take-away message is that in the world where the savings can serve for both the precautionary life-cycle function and the bequest function, the asset taxes and the inheritance taxes should be considered together. Proposition 8 shows that the implication for these two distortions are the opposite in this economy with contingent bequest motive.

Retirement Saving Wedge and Bequest Wedge and their Decomposition

The ex post wedges defined above can be interpreted as the tax imposed on the return of state contingent assets. These ex post wedges are helpful for understanding the mechanisms of the intertemporal distortion, but these wedges are hypothetical, since the incomplete market does not allow state contingent asset holdings against the uncertainty of the life cycle. To interpret the retirement saving wedges and the bequest wedges as the taxes imposed on the existing asset in this economy — state non-contingent bond, we now define new wedges.

Given an allocation $c_1, c_2, c_{c2}, c_{c3}, y)$ and a type θ, we define the noncontingnet retirement savings wedge $\tau_a(\theta)$ and the noncontingent bequest wedge $\tau_b(\theta)$ as follows:

$$\tau_a(\theta) = \left(1 - \frac{1}{P(\theta)}\right) + \frac{1}{P(\theta)} \underbrace{\left\{1 - \frac{u'(c_1(\theta))}{Ru'(c_2(\theta))}\right\}}_{=t_a(\theta)} + (1-\tau_b(\theta))\frac{1-P(\theta)}{P(\theta)}\frac{V_c'(c_{c2}(\theta))}{u'(c_2(\theta))}, \quad (9)$$

$$\tau_b(\theta) = \left(1 - \frac{1}{1-P(\theta)}\right) + \frac{1}{1-P(\theta)} \underbrace{\left\{1 - \frac{u'(c_1(\theta))}{RV_c'(c_{c2}(\theta))}\right\}}_{=t_b(\theta)} + (1-\tau_a(\theta))\frac{P(\theta)}{1-P(\theta)}\frac{u'(c_2(\theta))}{V_c'(c_{c2}(\theta))}. \quad (10)$$

We can understand the retirement savings wedge $\tau_a(\theta)$ as the tax imposed on the return of the non-contingent saving if the parents live long, while we can understand the bequest wedge $\tau_b(\theta)$ as the tax imposed on the return of the same non-contingent saving if the parents die early.

Note that the retirement saving wedge (9) is composed of the three components. We want to remark that this decomposition of the wedge is based on the definition of the wedge given any allocation, not necessarily the optimal allocation. The first term $1 - \frac{1}{P(\theta)}$, which is negative, is a component arising due to non-existence of annuity market. The second term $\frac{1}{P(\theta)}t_a(\theta)$ is the normalized ex post retirement saving wedge. The third term $(1-\tau_b(\theta))\frac{1-P(\theta)}{P(\theta)}\frac{V_c'(c_{c2}(\theta))}{u'(c_2(\theta))}$ captures the interaction of the asset and the inheritance due to overlapping functions of saving.

We can better understand the third term by rewriting the third term:

$$(1-\tau_b(\theta))\frac{1-P(\theta)}{P(\theta)}\frac{V_c'(c_{c2}(\theta))}{u'(c_2(\theta))} = -\left(1 - \frac{1}{P(\theta)}\right) + \frac{1-P(\theta)}{P(\theta)}\left(\frac{V_c'(c_{c2}(\theta))}{u'(c_2(\theta))} - 1\right) - \tau_B(\theta)\left(\frac{1-P(\theta)}{P(\theta)}\right)\frac{V_c'(c_{c2}(\theta))}{u'(c_2(\theta))}.$$

The third term is again decomposed into the three component: (i) undoing the subsidy, (ii) adjusting the risk — which arises due to no state-contingent asset, (iii) undoing the indirect distortion of the inheritance tax $\tau_b(\theta)$ on the retirement saving.

Notice that the net cost/benefit of the no contingent market is $\frac{1-P(\theta)}{P(\theta)}\left(\frac{V_c'(c_c(\theta))}{u'(c_2(\theta))} - 1\right)$, which is the component (ii) in the third term. For example, if the marginal utility the children in the event of the parent's death is smaller larger than the marginal utility of the parent in the event of their surviving, then we can adjust this risk by increasing tax on the parent's asset when they survive.

It is instructive to compare the retirement savings wedge in this economy with that of an economy without the bequest motive (with $V_c(\cdot) = 0$), which is typically considered in the literature on the retirement financing and the annuity. In an economy without the bequest motive, the saving wedge is only composed of the first two terms in (9) because there is no interaction between the retirement saving and the inheritance. Hosseini and Shourideh (2016) argue that there should be a huge asset subsidy using the model without bequest motive, because the first component — subsidy to complete the annuity market — dominates the second component in their quantitative analysis. Our results show that when there is an overlapping motive of the retirement saving, the cost of the no annuity market is lower, and thus there is less need for the subsidy.

The three components of the bequest wedge (10) can be also interpreted in the similar way. The first term $1 - \frac{1}{1-P(\theta)}$ is to complete the incomplete market, and the second term $\frac{1}{1-P(\theta)} t_b(\theta)$ is the normalized ex post bequest wedge. The third term $(1-\tau_a(\theta))\frac{P(\theta)}{1-P(\theta)}\frac{u'(c_2(\theta))}{V_c'(c_c(\theta))}$ captures the interaction of the asset and the inheritance — adjustment of the subsidy due to less costly incomplete market and the compensation for the indirect distortion caused by the asset tax $\tau_a(\theta)$.

As we mentioned above, the definition of $\tau_a(\theta)$ and $\tau_b(\theta)$ is given for any allocation, and there are many combinations of $\tau_a(\theta)$ and $\tau_b(\theta)$ that satisfy (9) and (10), since (9) and (10) are essentially the same Euler equations. To guarantee the optimal ex post wedges $t_a(\theta)$ and $t_b(\theta)$, however, the retirement savings wedge $\tau_a(\theta)$ and the bequest wedge $\tau_b(\theta)$ should guarantee that the after distortion marginal utility should be equalized:

$$(1 - \tau_a(\theta))Ru'(c_2(\theta)) = (1 - \tau_b(\theta))RV_c'(c_c(\theta)). \tag{11}$$

Then the wedges that can satisfy both the Euler equation (9) (or equivalently (10)) and (11) are :

$$\tau_a(\theta) = t_a(\theta), \qquad \tau_b(\theta) = t_b(\theta), \quad \forall\theta.$$

3.1.5 Role of Discount Factor

Finally, we now briefly analyze the role of the heterogeneous discount factor. Empirical evidences show that rich people tend to save more with higher saving propensity (e.g., Dynan, Skinner, and Zeldes (2004)), and one of the reason for this higher saving could be higher discount factor for the rich. Saez (2002) pointed out that the heterogeneous discount factor which is increasing in productivity can be a rationale for the positive taxes on saving. Here, we show that the same argument can be applied to the rationale for the positive taxes on inheritance. We can easily see that by assuming $\dot{\beta}(\theta) > 0$ for all θ. To focus on the role of discount factor, we assume that every parent lives for two period for sure and there is no medical expenses.

Proposition 9. *Suppose that $P(\theta) = 1$, $m(\theta) = 0$, and $\dot{\beta}(\theta) > 0$ for all θ. Suppose that $\lambda(\theta)$ is nonincreasing and (c_1, c_2, c_{c3}, y) solves the relaxed planning problem. Then,*

$$t_a(\theta) > 0, \qquad t_{b3}(\theta) > 0, \quad \forall\theta.$$

Proposition 9 is the straightforward extension of the result in Saez (2002). The positive ex post retirement saving wedge ($t_a(\theta) > 0$) is exactly because of the same reason in Saez (2002). Parents with higher productivity has higher propensity to save because of higher discount factor, then taxation of saving can relax the efficiency cost of redistribution across θ, because higher saving can be an indirect evidence about who has higher productivity. The same argument can be applied to the taxation of inherited wealth if the bequest motive is based on pure altruism with the exactly same discount factor $\beta(\theta)$ for both parent's own future utility and child's utility.

There can be other reasons of why the rich save more and bequest more such as uncertainties about future earnings and medical expenses, bequest motive when the bequest is luxury good, and the asset tested program of the government (See Dynan, Skinner, and Zeldes (2004)). Thus, higher propensity to save for the rich cannot be the direct evidence for the increasing discount factor, and we need to investigate this more seriously in the quantitative analysis.

3.1.6 Summary and Discussion

We now sum up the analyses so far by putting all factors together. Suppose that $\dot{m}(\theta) > 0$, $\dot{P(\theta)} > 0$, and $\dot{\beta}(\theta) > 0$. We also assume that the preferences for redistribution are not too weak so that $\mu(\theta) < 0$. Under this assumption, the ex post wedges can be expressed by

$$\begin{aligned}
t_a(\theta) &= -\frac{\mu(\theta)}{\eta f(\theta)} u'(c_1(\theta)) \left[\frac{\dot{P}(\theta)}{P(\theta)} + \frac{\dot{\beta}(\theta)}{\beta(\theta)} \right] > 0 \\
t_{b2}(\theta) &= \underbrace{-\frac{\mu(\theta)}{\eta f(\theta)} u'(c_1(\theta))}_{+} \Bigg[\underbrace{-\frac{V_c''(c_{c2}^*(\theta))}{V_c'(c_{c2}^*(\theta))} \dot{m}(\theta)}_{\equiv t_{b2,m} > 0} \underbrace{- \frac{\dot{P}(\theta)}{1 - P(\theta)}}_{\equiv t_{b2,P} < 0} \underbrace{+ \frac{\dot{\beta}(\theta)}{\beta(\theta)}}_{\equiv t_{b2,\beta} > 0} \Bigg] \\
t_{b3}(\theta) &= -\frac{\mu(\theta)}{\eta f(\theta)} R\beta(\theta) u'(c_2(\theta)) \left[-\frac{V_c''(c_{c3}^*(\theta))}{V_c'(c_{c3}^*(\theta))} \dot{m}(\theta) + \frac{\dot{\beta}(\theta)}{\beta(\theta)} \right] > 0.
\end{aligned}$$

This expression of the wedge shows that the ex post intertemporal wedges are decomposed into the three components, where each of them reflects the effects of each channel highlighted above — the medical expense channel, the mortality channel, and the (discounting) preference channel.

The implicit tax on the return of retirement savings $t_a(\theta)$ is positive because more productive parents tend to save more as they tend to live longer and be more patient. Thus taxing the retirement saving makes the redistributive tax system more efficient. Taxation of savings at lower income prevents productive parents from under save.

On the other hand, the implicit tax on the inherited wealth $t_b(\theta)$ can be either positive or negative because there are counter effects from the mortality channel which has the opposite implications from those of the medical expense channel and the β-preference channel. Both increasing medical expenses and increasing discount factor make the more productive parents value the bequest more, implying that the inheritance taxation is useful for redistribution. However, uncertainty in the lifespan has more subtle implication. The mortality risk makes the ex ante intertemporal wedge positive, which implies that the return of saving should be taxed in general. However, the negative correlation between the mortality and productivity has the opposite implications on the taxation of saving and the taxation of inheritance — the planner wants to tax the return to retirement savings but want to subsidize the bequest. If this mortality channel is strong enough, then it can be optimal to subsidize the inheritance for the efficient redistribution.

Previous studies found the economic reasons for either taxing or subsidizing the bequest from the preferences of the planner — meritocratic preferences, direct Pareto weight on the future generation, or the planner's direct preferences for altruism.[8]. We show that even if the planner does not have direct preferences over the bequest or altruism — the planner's preferences for redistribution is mainly based on the heterogeneity in earning, the correlation between the earning ability and the key factors which determines the bequest motive could be another reason for either taxing or subsidizing bequest.

There can be other factors that can affect the propensity to bequest for rich people. For example, the rich might have higher altruism than the poor, or the bequest is luxury good. What are the impact of these factors on the optimal inheritance taxation? Heterogeneity in altruism is another

[8]When the government puts a direct Pareto weight for children in addition to indirect valuation through altruism there is positive externality of the bequest — which is rationale for subsidy (Farhi and Werning (2010)) On the other hand, the heterogeneity in altruism and the planner's direct preferences across the altruism can generate rationales for both subsidy and tax (Farhi and Werning (2013a))

type of preference heterogeneity which is very similar to the discount factor heterogeneity. Thus, if the degree of altruism is increasing in income, then the same argument as in the β-preference channel applies : the positive correlation between the altruism and the income can provide a rationale for the positive inheritance taxation. However, we do not have clear empirical evidences on this correlation, and thus the heterogeneity in altruism might not be the first order concern for the inheritance tax.

Many previous studies have pointed out that bequests are luxury good, and this is very important to explain why a large fraction of poor parents do not leave bequest at all (or leave only insignificant amount). Typical way of modeling this feature is assuming that there is some consumption floor in the utility from bequests : $V_c(c_c) = U(\underline{c} + c_c)$, where $\underline{c} > 0$ is the threshold consumption level under which people do not leave bequest. It is important to have this factor to match the bequest behavior in the data, but this is not important for the optimal inheritance tax unless the consumption floor $\underline{c}$ has some correlation with the income-generating abilities. This is rather straightforward implication of uniform commodity taxation results of Atkinson and Stiglitz (1976).

3.2 Nonlinear Taxation : Implementation

to be filled out

4 Extensions

4.1 Intergenerational Redistribution: Pareto-Efficient Allocation

So far, we have analyzed the constrained efficient allocation when the planner only cares about the utility of the parents. We now extend our analysis to the planner who also cares about children's directly.

[to be filled out...]

4.2 Infinite Horizon Model with Productivity Shock

We now extend our anaysis on the infinite horizon economy with overlapping generations. The economy lasts for the infinite periods: $t = 1, \cdots, \infty$. In each period t, a continuum of new generation is born, and we assume that each individual in any generation can live for at most

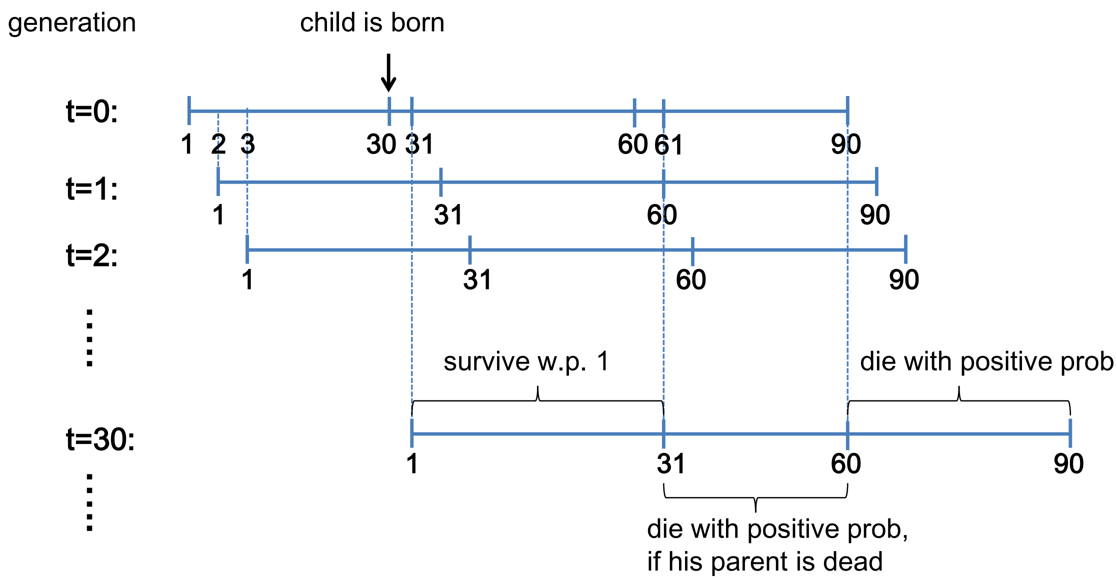

Figure 1: Timeline

90 years, and we also assume that each individual gives a birth to his child at age 30. Thus, the age gap between parent and child is 30 years.

In the two period model, we have maintained the assumption that only parents have working productivity and the parent's productivity is perfectly inherited to the consumption of the child. We now consider the case where there is a productivity shock for each generation. That is, an individual in generation t is born with type θ_t, where θ_t is drawn from the distribution $F(\theta)$, which is i.i.d. across people within generation and across generations. In this economy, the parent's income-generating ability is not perfectly inherited to the consumption of the children, and there will be some mean reversion of consumption across generation.

The productivity of the individual with θ_t at age j is $\varphi_j(\theta_t)$. In addition, $\varphi_j(\theta_t)$ is correlated with the factors of the bequest motive. We assume perfect correlation, by assuming that the medical expenses, mortality, and the discount factors are functions of θ_t. The survival probability of individual with θ_t at age j conditional on the survival by the age $j-1$ is denoted by $p_j(\theta_t|S_{j+29}^{t-30})$, where S_j^t is the indicator function of survival — $S_j^t = 1$, if t-th generation with θ_t is alive at age j. For tractability, we assume that an individual dies with positive probability only after their parents die: $p_j(\theta_t|S_{j+29}^{t-30} = 1) = 1$ and $p_j(\theta_t|S_{j+29}^{t-30} = 0) < 1$, and we denote the survival probability conditional on the parent's death by $p_{0,j}(\theta_t) \equiv p_j(\theta_t|S_{j+29}^{t-30} = 0)$. The lifespan of the generation t is $l_t = \sum_{j=1}^{90} S_j^t$, and the unconditional survival probability given the parent's lifespan is denoted by

$P_j(\theta_t|l_{t-30}) \equiv \prod_{k=1}^{j} p_j(\theta_t|S_{j+29}^{t-30}) = \prod_{k=l_{t-30}-28}^{j} p_{0,k}(\theta_t)$.

Note that each dynasty has a member of each generation in every 30 years, and thus t-generation belongs to the $m(t)$-dynasty, where $m(t) \in \{0, 1, \cdots, 29\}$ is the initial generation in the dynasty where t-generation belongs to. We denote the k-th generation in the $m(t)$-dynasty by $\tau(t,k) = 30 * k + m(t)$, where $t = \tau(t, d(t)) = 30 * d(t) + m(t)$. We denote the history of the productivity and the lifespan in the $m(t)$-dynasty by θ^t and l^{t-30}, separately, and the allocations are function of the history $\{c_j(\theta^t, l^{t-30}, y_j(\theta^t, l^{t-30}))\}$. The continuation utility of the t-generation

We assume that the current utility takes the separable form: $u(c_j, y_j, \varphi_j) = u(c_j) - v\left(\frac{y_j}{\varphi_j}\right)$. The continuation utility of the t-generation in the $m(t)$-dynasty will be

$$\begin{aligned} U(\{c,y\}|\theta^t, l^{t-30}) &= \sum_{j=1}^{l_{t-30}-30} \beta(\theta_t)^j u\left(c_j(\theta^t, l^{t-30}), y_j(\theta^t, l^{t-30}), \varphi_j(\theta_t)\right) \\ &+ \sum_{k=d(t)}^{\infty} \sum_{\theta^{\tau(t,k)}, l^{\tau(t,k-1)}} Pr(\theta^{\tau(t,k)}, l^{\tau(t,k-1)}|\theta^t, l^{t-30}) \times \\ &\sum_{j=l_{\tau(t,k-1)}-29}^{90} P_j(\theta_{\tau(t,k)}|l^{\tau(t,k-1)}) \beta(\theta^{\tau(t,k)}, l^{\tau(t,k-1)}, j|t) u\left(c_j(\theta^{\tau(t,k)}, l^{\tau(t,k-1)}), y_j(\theta^{\tau(t,k)}, l^{\tau(t,k-1)}), \varphi_j(\theta_{\tau(t,k)})\right), \end{aligned}$$

where

$$\beta(\theta^{\tau(t,k)}, l^{\tau(t,k-1)}, j|t) = \begin{cases} \beta(\theta_t) \times \prod_{n=d(t)+1}^{k-1} \beta(\theta_{\tau(t,n)})^{l_{\tau(t,n)} - l_{\tau(t,n-1)}+30} \times \beta(\theta_{\tau(t,k)})^{j - l_{\tau(t,k-1)}+29}, & k > d(t) \\ \beta(\theta_t)^j & k = d(t) \end{cases}$$

is the discount factor of t-generation for the utility of $\tau(t,k)$-generation in the same dynasty.

By the revelation principle, we can focus on direct mechanism, where agents make report regarding θ_t. Let Σ denote the set of all reporting strategies $\sigma = \{\sigma_t(\theta^t)\}$.

Consider an allocation $\{c, y\}$. Then the continuation utility of t-generation after his/her parent's death can be written recursively:

$$\begin{aligned} w_{ap}^{\sigma}(\theta^t, l^{t-30}) &= \sum_{j=l_{t-30}-29}^{90} P_j(\theta_t|l_{t-30}) \beta(\theta_t)^{j-l_{t-30}+29} u\left(c_j(\theta^t, l^{t-30}), y_j(\theta^t, l^{t-30}), \varphi_j(\theta_t)\right) \\ &+ \sum_{j=l_{t-30}-29}^{90} \left(P_j(\theta_t|l_{t-30}) - P_{j+1}(\theta_t|l_{t-30})\right) \beta(\theta_t)^{j-l_{t-30}+30} \int w_{ap}^{\sigma}(\theta^t, \theta_{t+30}, l^{t-30}, j) f(\theta_{t+30}) d\theta_{t+30}. \end{aligned}$$

Denote the continuation utility of t-generation by

$$w^{\sigma}(\theta^t, l^{t-30}) = \sum_{j=1}^{l_{t-30}-30} \beta(\theta_t)^j u\left(c_j(\theta^t, l^{t-30}), y_j(\theta^t, l^{t-30}), \varphi_j(\theta_t)\right) + \beta(\theta_t)^{l_{t-30}-29} w_{ap}^{\sigma}(\theta^t, l^{t-30}),$$

and the continuation utility associated with the truth-telling strategy by $w(\theta^t, l^{t-30})$. We say that $\{c, y\}$ is incentive compatible if and only if

$$w^{\sigma}(\theta^t, l^{t-30}) \geq w^{\sigma}(\theta^t, l^{t-30}), \quad \forall \sigma \in \Sigma, \quad \forall (\theta^t, l^{t-30}),$$

and IC denotes the set of all incentive compatible allocation $\{c, y\}$.

Thus, we solve the following planner's problem: given R and $\underline{V}$,

$$\begin{aligned}
\min \quad & \int \sum_{t=0}^{\infty} \frac{1}{R^t} \sum_{\theta^t, l^{t-30}} Pr(\theta^t, l^{t-30}) \sum_{j=1}^{90} \frac{1}{R^j} P_j(\theta_t | l_{t-30}) \left[e_j^v(\theta^t, l^{t-30}) - y_j^v(\theta^t, l^{t-30}) \right] d\psi(v) \\
s.t. \quad & v = \sum Pr(\theta^t, l^{t-30}) U(\{c^v, y^v\} | \theta^t, l^{t-30}), \quad \forall v, \quad t = 0, \cdots, 29, \qquad \textbf{(PK)} \\
& \int \sum_{\theta^t, l^{t-30}} Pr(\theta^t, l^{t-30}) U(\{c^v, y^v\} | \theta^t, l^{t-30}) d\psi(v) \geq \underline{V}, \quad t \geq 30, \qquad \textbf{(AC)} \\
& \{c^v, y^v\} \in IC, \quad \forall v,
\end{aligned}$$

where $e_j^v(\theta^t, l^{t-30}) = c_j^v(\theta^t, l^{t-30}) + 1_{j=l_{t-30}-29} \times \dot{m}(\theta_{t-30})$. The constraints **(PK)** represent the promise keeping constraints for each $m(t)$-dynasty with promise utility v, and the constraints **(AC)** are admissibility constraints for the future generation, which is equivalent to putting direct Pareto weights for the future generation. When $\underline{V} = -\infty$, the admissibility constraints are slack and the planner cares for the future generation indirectly only through the altruism of the parents in each dynasty.

By putting the multipliers $\alpha_t \beta^t d\psi(v)$ and $\alpha_t \beta^t$ on the constraints **(PK)** and **(AC)** respectively and using the sum of Lagrange multipliers, the planning problem is simplified to

$$\begin{aligned}
\min \quad & \int \sum_{m(t)=0}^{29} \mathscr{L}_{m(t)}^v d\psi(v) \\
s.t. \quad & \{c^v, y^v\} \in IC, \quad \forall v,
\end{aligned}$$

where each dynasty's Lagrangian $\mathscr{L}_{m(t)}^v$ is written as follows using the notation $t(n) \equiv 30n + m(t)$:

$$\begin{aligned}
\mathscr{L}_{m(t)}^v \;=\; & \sum_{n=0}^{\infty} \sum_{\theta^{t(n)}, l^{t(n-1)}} Pr(\theta^{t(n)}, l^{t(n-1)}) \sum_{j=1}^{90} \frac{1}{R^{t(n)+j}} P_j(\theta_{t(n)} | l_{t(n-1)}) \left[e_j^v(\theta^{t(n)}, l^{t(n-1)}) - y_j^v(\theta^{t(n)}, l^{t(n-1)}) \right] \\
& - \sum_{n=0}^{\infty} \sum_{\theta^{t(n)}, l^{t(n-1)}} Pr(\theta^{t(n)}, l^{t(n-1)}) \left[\begin{array}{l} \sum_{j=l_{t(n-1)}-29}^{90} P_j(\theta_t(n) | l_{t(n-1)}) \left\{ \sum_{k=0}^{n} \alpha_{t(k)} \beta(\theta^{t(n)}, l^{t(n-1)}, j | t(k)) \right\} \\ \times \left(u(c_j^v(\theta^{t(n)}, l^{t(n-1)})) - v\left(\frac{y_j^v(\theta^{t(n)}, l^{t(n-1)})}{\varphi_j(\theta_{t(n)})} \right) \right) \\ + \alpha_{t(n)} \beta^{t(n)} \sum_{j=1}^{l_{t(n-1)}} \beta(\theta_{t(n)})^j \left(u(c_j^v(\theta^{t(n)}, l^{t(n-1)})) - v\left(\frac{y_j^v(\theta^{t(n)}, l^{t(n-1)})}{\varphi_j(\theta_{t(n)})} \right) \right) \end{array} \right].
\end{aligned}$$

Since each dynasty's minimization problem can be separated out, we are solving the component planning problem for each $m(t)$-dynasty with promise utility v.

As in Pavan, Segal, and Toikka (2014) and Farhi and Werning (2013b), we use the first order approach and solve the relaxed component planning problem. Incentive compatibility requires

$$w(\theta^t, l^{t-30}) = \max_r w^{\sigma^r}(\theta^t, l^{t-30}), \quad \forall \theta^t, l^{t-30},$$

where a deviation strategy σ^r is such that $\sigma_t^t(\theta^{t-30}, \theta_t) = r$ and $\sigma_t^r(\theta^{t-30}, \tilde{\theta}) = \tilde{\theta}$, for $\tilde{\theta} \neq \theta_t$. Then, any incentive compatible allocation satisfies the following envelope condition

$$\begin{aligned}
\frac{\partial}{\partial \theta_t} w(\theta^t, l^t) \;\; = & \sum_{j=1}^{90} P_j(\theta_t | l_{t-30}) \beta(\theta_t)^j \left[v'\left(\frac{y_j(\theta^t, l^{t-30})}{\varphi_j(\theta_t)} \right) \frac{y_j(\theta^t, l^{t-30})}{\varphi_j(\theta_t)^2} \varphi_j'(\theta_t) \right] \qquad (12) \\
& + \sum_{j=1}^{90} \left(\dot{P}(\theta_t) | l_{t-30}) \beta(\theta_t)^j + P(\theta_t) | l_{t-30}) \cdot j \cdot \beta(\theta_t)^{j-1} \dot{\beta}(\theta_t) \right) \times \left[u(c_j(\theta^t, l^{t-30})) - v\left(\frac{y_j(\theta^t, l^{t-30})}{\varphi_j(\theta_t)} \right) \right] \\
& + \sum_{j=l_{t-30}-29}^{90} \left[\dot{P}_j(\theta_t | l_{t-30}) - \dot{P}_{j+1}(\theta_t | l^{t-30}) \right] \beta(\theta_t)^{j+1} \int w_{ap}(\theta^t, \theta_{t+30}, l^{t-30}, j) f(\theta_{t+30}) d\theta_{t+30} \\
& + \sum_{j=l_{t-30}-29}^{90} \left(P_j(\theta_t | l_{t-30}) - P_{j+1}(\theta_t | l_{t-30}) \right) \times \left[\begin{array}{l} (j+1)\beta(\theta_t)^j \dot{\beta}(\theta_t) \int w_{ap}(\theta^t, \theta_{t+30}, l^{t-30}, j) f(\theta_{t+30}) d\theta_{t+30} \\ -\beta(\theta_t)^{j+1} \dot{m}(\theta_t) \int u'(c_{j-29}(\theta^t, \theta_{t+30}, l^{t-30}, j)) d\theta_{t+30} \end{array} \right],
\end{aligned}$$

where

$w_{ap}(\theta^{t+30}, l^t) = \frac{1}{\beta(\theta_{t+30})^{l_t - 29}} \left[w(\theta^{t+30}, l^t) - \sum_{j=1}^{l_t - 30} \beta(\theta_{t+30})^j \left\{ u(c_j(\theta^{t+30}, l^t)) - v\left(\frac{y_j(\theta^{t+30}, l^t)}{\varphi_j(\theta_{t+30})} \right) \right\} \right]$. That is, for each dynasty, the relaxed component planner's problem solves

$$\begin{aligned}
\min \quad & \mathscr{L}^v_{m(t)} \qquad (13) \\
s.t. \quad & (12), \text{ and} \\
& w(\theta^{t(n)}, l^{t(n-1)}) = \sum_{j=1}^{90} P_j(\theta_{t(n)} | l_{t(n-1)}) \beta(\theta_{t(n)})^j u(c_j^v(\theta^{t(n)}, l^{t(n-1)}), y_j^v(\theta^{t(n)}, l^{t(n-1)}), \varphi(\theta_{t(n)})) \\
& \quad + \sum_{j=l_{t(n-1)}-29}^{90} \left(P_j(\theta_{t(n)} | l_{t(n-1)}) - P_{j+1}(\theta_{t(n)} | l_{t(n-1)}) \right) \beta(\theta_{t(n)})^{j+1} \\
& \quad \times \int \frac{f(\theta_{t(n+1)})}{\beta(\theta_{t(n+1)})^{j-29}} \left[\begin{array}{l} w(\theta^{t(n)}, \theta_{t(n+1)}, l^{t(n-1)}, j) - \\ \sum_{i=1}^{j-30} \beta(\theta_{t(n+1)})^i u(c_i(\theta^{t(n)}, \theta_{t(n)}, l^{t(n-1)}, j), y_i(\theta^{t(n)}, \theta_{t(n)}, l^{t(n-1)}, j), \varphi_i(\theta_{t(n+1)})) \end{array} \right]
\end{aligned}$$

4.2.1 Optimal Wedges

We now show optimal wedges by solving the planner's problem. The analysis not only confirms the mechanism of the simple two-period model, but also provides additional characterization of the optimal taxes, such as the age-dependent properties.

First of all, we characterize the properties of the optimal labor wedges in the following proposition. For notational simplicity, we denote the history by $s^t \equiv (\theta^t, l^{t-30})$, the lifespan of the t-generation's parents by $l_- \equiv l_{t-30}$, and the type of the t-generation's parents by $\theta_- = \theta_{t-30}$. We also drop v from the allocation $\{c^v, y^v\}$.

Proposition 10. *Suppose that $\{c, y\}$ solve (13). Then, the labor wedges satisfy*

(i) for $j \leq l_- - 30$, $\frac{\tau_j^l(s^t)}{1-\tau_j^l(s^t)} = \frac{B_j(s^t)}{A_j(s^t)}$, where

$$
\begin{aligned}
B_j(s^t) &= -\frac{\mu(s^t)}{f(\theta_t)}\frac{\varphi_j'(\theta_t)}{\varphi_j(\theta_t)}\left(1+\frac{v''(y_j(s^t))}{v'(y_j(s^t))}\frac{y_j(s^t)}{\varphi_j(\theta_t)}\right) \\
A_j(s^t) &= \frac{1}{R^{t+l_- -29}\beta(\theta_t)^{l_- -29}}\left[\frac{1}{u'(c_{l_- -29}(s^t))} - \frac{R\beta(\theta_-)}{u'(c_{l_-}(s^{t-30}))}\right] + \frac{\mu(s^t)}{f(\theta)}(j-(l_- - 29))\frac{\dot\beta(\theta_t)}{\beta(\theta_t)} \\
&\quad -\frac{\mu(s^{t-30})}{f(\theta_-)}\frac{\beta(\theta_-)^{l_- +1}}{\beta(\theta_t)^{l_- -29}}\left[-\frac{\dot p_{0,l_- +1}(\theta_-)}{1-p_{0,l_- +1}(\theta_-)} + \frac{\dot\beta(\theta_-)}{\beta(\theta_-)} - \frac{u''(c_{l_- -29}(s^t))}{u'('(c_{l_- -29}(s^t)))}\dot m(\theta_-)\right],
\end{aligned}
$$

(ii) for $j = l_- - 29$, $\frac{\tau_j^l(s^t)}{1-\tau_j^l(s^t)} = \frac{B_j(s^t)+F_j(s^t)}{E_j(s^t)}$, where

$$
\begin{aligned}
F_j(s^t) &= \frac{\mu(s^{t-30})}{f(\theta_-)}\frac{\beta(\theta_-)^{l_- +1}}{\beta(\theta_t)^{l_- -29}}\frac{u''(c_{l_- -29}(s^t))}{u'('(c_{l_- -29}(s^t)))}\dot m(\theta_-), \\
E_j(s^t) &= \frac{1}{R^{t+l_- -29}\beta(\theta_t)^{l_- -29}u'(c_{l_- -29}(s^t))},
\end{aligned}
$$

(iii) for $j \geq l_- - 28$, $\frac{\tau_j^l(s^t)}{1-\tau_j^l(s^t)} = \frac{B_j(s^t)}{D_j(s^t)}$, where

$$
D_j(s^t) = E_j(s^t) + F_j(s^t) + \frac{\mu(s^t)}{f(\theta_t)}\left[\frac{\dot P_j(\theta_t|l_-)}{P(\theta_t|l_-)} + (j-(l_- - 29))\frac{\dot\beta(\theta_t)}{\beta(\theta_t)}\right].
$$

Proof See the appendix. ■

Proposition 10 shows the role of the age-dependent labor income tax. As in Weinzierl (2011), even without heterogeneity in the factors related to savings and bequest, labor income tax does depend on age because productivity depends on the age, $\varphi_j(\theta)$. With the heterogeneity in the factors of savings and bequest, labor income tax should be age-dependent not only because of the life-cycle productivities but also because of the correlation between the skill and the factors of savings and the bequest. This is because the planner wants to tax the goods that are preferred by the high skilled agents to reduce the cost of redistribution. The correlations between the skill and the factors imply that the skilled agents have relatively stronger preferences for the future generation's consumption and the leisure after their the death of their parents.

The distortion of the future generation's labor supply at their earlier age before the parent's death ($j \leq l_- - 30$) can be used to reduce their consumption at a later age after the parent's death if this later consumption is more preferred by the skilled parents. Thus, $\dot m(\theta_-) > 0$ and $\dot\beta(\theta_-) > 0$

— which make the parents prefer child's consumption after the parent's death — increase the labor income tax of age $j \leq l_- - 30$, while $\dot{p}_{0,l_-}(\theta_-) > 0$ decreases τ_j^l for j. This incentive of distortion gets higher when $u'(c_{l_-}(s^{t-30})) < R\beta(\theta_-)u'(c_{l_-}(s^t))$ — when the bequest is constrained.

High skilled parents especially values the utility of children right after the parents die, because this is the time when the children's consumption ($c_{l_- -29}$) decreases the most due to higher medical expenses of the high skilled parents. Thus, distorting labor supply at age $l_- - 29$ by increasing labor income tax can reduce the cost of redistribution by relaxing the incentive constraints.

On the other hand, if the skilled child prefers the consumption at a later age because of $\dot{\beta}(\theta) > 0$ and $\dot{P}_j(\theta_t) > 0$, taxing the later consumption by backloading the labor distortion can also reduce the cost of the redistribution. Thus, these factors make the labor income tax increase with age (reflected by the associated terms in A_j and D_j).

Next, we characterize the properties of the optimal bequest wedges. The inherited wealth wedge of the t-generation for each history $s^t = (\theta^t, l^{t-30})$ is

$$\tau^b(s^t) = 1 - \frac{u'(c_{l_{t-30}}(s^{t-30}))}{R\beta(\theta_{t-30})u'(c_{l_{t-30}-29}(s^t))},$$

and as in Farhi and Werning (2013a), we define the average inheritance tax by

$$\bar{\tau}^b(\theta^{t-30}, l^{t-30}) = \int \tau^b(\theta^t, l^{t-30}) f(\theta_t) d\theta_t.$$

Next proposition shows the properties of the average inheritance tax.

Proposition 11. *Suppose that $\{c, y\}$ solve (13). Then, the average inheritance tax satisfies*

$$\bar{\tau}^b(\theta^{t-30}, l^{t-30}) = \frac{R_{t+l_- -30}u'(c_{l_-}(s^{t-30}))}{\beta(\theta_-)} \times \left[\begin{array}{l} -\alpha_t \beta^t \times \int \beta(\theta_t)^{l_- -29} f(\theta_t) d\theta_t \\ -\frac{\mu(s^{t-30})}{f(\theta_-)}\beta(\theta_-)^{l_- +1} \times \left\{ \begin{array}{l} -\frac{\dot{p}_{0,l_- +1}(\theta_-)}{1-p_{0,l_- +1}(\theta_-)} + \frac{\dot{\beta}(\theta_-)}{\beta(\theta_-)} \\ -\dot{m}(\theta_-) \int \frac{u''(c_{l_- -29}(s^t))}{u'('(c_{l_- -29}(s^t)))} f(\theta_t) d\theta_t \end{array} \right\} \end{array} \right],$$

for $t \geq 30$, $\forall(\theta^{t-30}, l^{t-30})$.

Note that without putting additional Pareto weights to the future generations ($\alpha_t = 0$, for $t \geq$30), the average bequest tax boils down to

$$\bar{\tau}^b(\theta^{t-30}, l^{t-30}) = -\frac{\mu(s^{t-30})}{f(\theta_-)}\beta(\theta_-)^{l_-} R_{t+l_- -30}u'(c_{l_-}(s^{t-30})) \left\{ \begin{array}{l} -\frac{\dot{p}_{0,l_- +1}(\theta_-)}{1-p_{0,l_- +1}(\theta_-)} + \frac{\dot{\beta}(\theta_-)}{\beta(\theta_-)} \\ -\dot{m}(\theta_-) \int \frac{u''(c_{l_- -29}(s^t))}{u'('(c_{l_- -29}(s^t)))} f(\theta_t) d\theta_t \end{array} \right\},$$

which is essentially equivalent to the two-period model result. The planner wants to tax the bequest if the bequest is relatively preferred by the skilled parents, and thus $\dot{\beta}(\theta_-)$ and $\dot{m}(\theta_-)$ makes the bequest tax positive, while $\dot{p}_{0,l_-+1}(\theta_-)$ makes the bequest tax negative. Higher medical expenses of the higher skilled agents and higher patience of the higher skilled agents imply the higher skilled parents wants to bequest more, while higher survival rate of the high skilled parents make them value bequest less.

Note that the first term in the average inheritance tax is negative if $\alpha_t > 0$. As in Farhi and Werning (2013a), with additional Pareto weights on the future generations ($\alpha_t > 0$), the planner wants to subsidize the bequest. Parents with altruistic motive for bequest does not take into account the additional Pareto weight for the future generation, and thus there is positive externalities for the bequest.

Lastly, we characterize the properties of the optimal saving wedges. The savings wedges are defined by

$$\tau^a_{j+1}(\theta^t, l^{t-30}) \equiv 1 - \frac{u'(c_j(\theta^t, l^{t-30}))}{R\beta(\theta_t)u'(c_{j+1}(\theta^t, l^{t-30}))}.$$

Proposition 12. *Suppose that $\{c, y\}$ solve (13). Then, the savings wedges satisfy*

(i) for $j+1 \leq l_- - 30$, $\frac{\tau^a_{j+1}(s^t)}{1-\tau^a_{a+1}(s^t)} = \frac{A_j(s^t)-A_{j+1}(s^t)}{A_{j+1}(s^t)}$, where

$$A_j(s^t) - A_{j+1}(s^t) = -\frac{\mu(s^t)}{f(\theta_t)}\frac{\dot{\beta}(\theta_t)}{\beta(\theta)} > 0,$$

(ii) for $j+1 = l_- - 29$, $\frac{\tau^a_{j+1}(s^t)}{1-\tau^a_{a+1}(s^t)} = \frac{A_j(s^t)-E_{j+1}(s^t)}{E_{j+1}(s^t)}$, where

$$\begin{aligned} A_j(s^t) - E_{j+1}(s^t) &= -\frac{\mu(s^t)}{f(\theta)}\frac{\dot{\beta}(\theta_t)}{\beta(\theta_t)} - \frac{\beta(\theta_-)}{R^{t+l_--30}\beta(\theta_t)^{l_--29}u'(c_{l_-}(s^{t-30}))} \\ &\quad -\frac{\mu(s^{t-30})}{f(\theta_-)}\frac{\beta(\theta_-)^{l_-+1}}{\beta(\theta_t)^{l_--29}}\left[-\frac{\dot{p}_{0,l_-+1}(\theta_-)}{1-p_{0,l_-+1}(\theta_-)} + \frac{\dot{\beta}(\theta_-)}{\beta(\theta_-)} - \frac{u''(c_{l_--29}(s^t))}{u'('(c_{l_--29}(s^t)))}\dot{m}(\theta_-)\right], \end{aligned}$$

(iii) for $j+1 = l_- - 28$, $\frac{\tau^a_{j+1}(s^t)}{1-\tau^a_{a+1}(s^t)} = \frac{E_j(s^t)-D_{j+1}(s^t)}{D_{j+1}(s^t)}$, where

$$E_j(s^t) - D_{j+1}(s^t) = -\frac{\mu(s^t)}{f(\theta_t)}\left[\frac{\dot{P}_{j+1}(\theta_t|l_-)}{P_{j+1}(\theta_t|l_-)} + \frac{\dot{\beta}(\theta_t)}{\beta(\theta_t)}\right] - \frac{\mu(s^{t-30})}{f(\theta_-)}\frac{\beta(\theta_-)^{l_-+1}}{\beta(\theta_t)^{l_--29}}\dot{m}(\theta_-)\frac{u''(c_{l_--29}(s^t))}{u'(c_{l_--29}(s^t))},$$

(iv) for $j+1 \geq l_- - 27$, $\frac{\tau^a_{j+1}(s^t)}{1-\tau^a_{a+1}(s^t)} = \frac{D_j(s^t)-D_{j+1}(s^t)}{D_{j+1}(s^t)}$, where

$$D_j(s^t) - D_{j+1}(s^t) = -\frac{\mu(s^t)}{f(\theta_t)}\left[\frac{\dot{p}_{0,j+1}(\theta_t)}{p_{0,j+1}(\theta_t)} + \frac{\dot{\beta}(\theta_t)}{\beta(\theta_t)}\right] > 0.$$

For age $j+1 \geq l_- - 27$, the optimal asset income tax result is essentially the same with the that of the two-period model. The planner wants to tax savings because the higher skilled agents wants to save more when $\dot{p}_{0,j}(\theta) > 0$ and $\dot{\beta}(\theta) > 0$.

On the other hand, higher medical expenditure of the higher skilled parents make the higher skilled parents value the children's consumption right after death $c_{l_- - 29}$ more, and thus the planner wants to increase asset income tax at $j+1 = l_- - 29$ and decrease asset income tax at $j+1 = l_- - 28$.

The taxes on the future generation's asset in come at their earlier age before the parent's death ($j \leq l_- - 30$) can be also used to reduce their consumption at a later age after the parent's death if this later consumption is more preferred by the skilled parents. Thus, as for the labor income taxation, $\dot{m}(\theta_-) > 0$ and $\dot{\beta}(\theta_-) > 0$ increase the asset income tax of age $j \leq l_- - 30$, while $\dot{p}_{0,l_-}(\theta_-) > 0$ decreases τ^a_{j+1} for $j \leq l_- - 30$.

5 Numerical Analysis

From the theoretical analysis with the simple two period model, we could see that in theory, various range of results are possible and the sign and the shape of the optimal tax system does depend on the calibration of the mortality and the altruism and the welfare criterion. We thus need to investigate these important determinant seriously to get the quantitative results.

To solve the problem numerically, we represent the problem recursively. See the appendix for the detail.

[to be continued...]

References

Atkinson, A. B., and J. E. Stiglitz (1976): "The design of tax structure: Direct versus indirect taxation," *Journal of Public Economics*, 6(1), 55–75.

Cristia, J. P. (2009): "Rising mortality and life expectancy differentials by lifetime earnings in the United States," *Journal of Health Economics*, 28(5), 984–995.

De Nardi, M., E. French, and J. Jones (2010): "Why Do the Elderly Save? The Role of Medical Expenses," *Journal of Political Economy*, 118(1), 39–75.

De Nardi, M., E. French, and J. B. Jones (2016): "Savings After Retirement: A Survey," *Annual Review of Economics*, 8(1), 177–204.

Dynan, K., J. Skinner, and S. Zeldes (2004): "Do the Rich Save More?," *Journal of Political Economy*, 112(2), 397–444.

Dynan, K. E., J. Skinner, and S. P. Zeldes (2002): "The Importance of Bequests and Life-Cycle Saving in Capital Accumulation: A New Answer," *The American Economic Review*, 92(2), 274–278.

Farhi, E., and I. Werning (2010): "Progressive Estate Taxation*," *The Quarterly Journal of Economics*, 125(2), 635–673.

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

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

Heathcote, J., and H. Tsujiyama (2017): "Optimal Income Taxation: Mirrlees Meets Ramsey," Discussion Paper 10380.

Hosseini, R., and A. Shourideh (2016): "Retirement Financing: An Optimal Reform Approach," Discussion paper.

Kopczuk, W. (2010): "Economics of estate taxation: a brief review of theory and evidence," *National Bureau of Economic Research Working Paper Series*, No. 15741.

Laitner, J., and F. T. Juster (1996): “New Evidence on Altruism: A Study of TIAA-CREF Retirees,” *The American Economic Review*, 86(4), 893–908.

Lockwood, L. M. (2014): “Incidental Bequests: Bequest Motives and the Choice to Self-Insure Late-Life Risks,” *National Bureau of Economic Research Working Paper Series*, No. 20745.

Mariacristina, N., F. Eric, J. J. Bailey, and M. Jeremy (2016): “Medical Spending of the US Elderly,” *Fiscal Studies*, 37(3-4), 717–747.

Mirrlees, J. A. (1971): “An Exploration in the Theory of Optimum Income Taxation,” *The Review of Economic Studies*, 38(2), 175–208.

Pavan, A., I. Segal, and J. Toikka (2014): “Dynamic Mechanism Design: A Myersonian Approach,” *Econometrica*, 82(2), 601–653.

Pavoni, N., and H. Yazici (2017): “Intergenerational Disagreement and Optimal Taxation of Parental Transfers,” *The Review of Economic Studies*, 84(3), 1264–1305.

Piketty, T., and E. Saez (2013): “A Theory of Optimal Inheritance Taxation,” *Econometrica*, 81(5), 1851–1886.

Saez, E. (2002): “The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes,” *Journal of Public Economics*, 83(2), 217–230.

Waldron, H. (2013): “Mortality Differentials by Lifetime Earnings Decile: Implications for Evaluations of Proposed Social Security Law Changes ,” .

Weinzierl, M. (2011): “The Surprising Power of Age-Dependent Taxes,” *The Review of Economic Studies*, 78(4), 1490–1518.

Appendix

to be filled out