Analysis and Policy in Economics

Master Thesis

Optimal Inheritance Taxation with Uncertain Lifespan

PARIS SCHOOL OF ECONOMICS
ECOLE D'ECONOMIE DE PARIS

Qinkun Yao

Supervisor: Prof. Bertrand Wiginolle
Co-supervisor: Prof. Thomas Piketty
Referee: Prof. Stéphane Gauthier

Keywords: Optimal Taxation, Wealth Distribution,
Altruism, Intergenerational Transfers, Overlapping Generation

JEL code: H2, D31, D64, J18

Abstract

We incorporate the death risk into a classic two-period overlapping generations model to investigate its effects on the optimal inheritance taxation. The death risk in absence of actuarially fair annuities causes accidental bequest as an extra source of inequality, thus we distinguish different bequest motives. At first we study the competitive equilibrium in a closed economy and see the effect of death risk on the capital accumulation, and decentralize the first-best social optimum by tax instruments including inheritance taxation. In the second-best analysis, we adopt a small open economy and study the optimal inheritance taxation that maximizes the social welfare in the long run, where the distribution of inheritance converges to a stationary one. We obtain the result that the optimal inheritance tax is positively related to the variance of shocks resulted from the death risk in the accumulative process of inheritance. Besides, for a Rawlsian social welfare criterion, the optimal tax rate is about 60% for logarithmic utility functions, while it equals 30% if the intertemporal elasticity of substitution in the CIES utility function is 2, when the surviving rate to second period of life is 0.8.

Acknowledgements

I would like to thank Professor Bertrand Wigniolle for his excellent comments and patient instruction in the several meetings during the redaction of this master thesis. I also would like to thank Professor Thomas Piketty who gave me insightful and challenging suggestions. I am also grateful to Professor Stéphane Gauthier for accepting to referee this thesis. Last but not least, I would like to thank my parents who have always given me the strongest support.

Contents

1 Introduction

The taxation on inheritance has been a remarkably controversial topic with a long history. Historically, inheritance taxation dates back to the days of the Roman Empire, when Augustus introduced in A.D. 6 the *vicesima hereditatium*, a 5% death duty on estates above a certain value provided that they were not inherited from close relatives, in order to support veterans. The supporters of inheritance taxation argue that it plays an essential role in redistributing the wealth and reducing inequality, since inheritance accounts for a large share of aggregate wealth, for instance, the estimation in Kotlikoff and Summers(1981) showed that 80% of total wealth in the U.S. are inherited[1], and its distribution is highly skewed (see Gale and Slemord(2001) and Piketty(2010)). The opponents of the inheritance taxation not only emphasize its efficiency cost, such as its effect in capital accumulation because of changed incentive to save and invest (the sign of this effect depends on the intertemporal elasticity of substitution), but also point out several moral reasons: it damages the interest of the parents who love their children and may go against the will of the deceased to bequeath. This controversial debate is reflected in the work of Mill: *Principles and Political Economy*(1848), in which he pointed out that inheritance runs counter to a fair and free competition by creating a different starting point for different individuals which is not related to their own efforts. Meanwhile, Mill acknowledged that one should respect the deceased's will concerning the bequest left to their children. His attitude can be considered as a balance between the moral reason for respecting the existence of inheritance and the consideration of equality and fairness that contradicts inheritance. As a consequence, Mill proposed a limit on the inheritance level to achieve a compromise.

Unlike concerns on inequality and pure moral reasons, the efficiency cost of inheritance tax depends largely on the bequest motives. In the literature, there are generally four categories of bequest motives (see Cremer and Pestieau(2006)): pure altruistic bequest or Barro-type altruism where parents care about the lifetime utility of children and consequently the utilities of all future generations; paternalistic bequest, where parents derive utility from bequeathing only through the joy of giving, or "warm glow" giving; strategic bequest where children offer some "service" or "attention" to their parents in exchange for inheritance as a remuneration, which consists of a strategic game between parents and children; and bequest without any motive, which is the accidental bequest or unplanned bequest, resulting from the fact that wealth of the parents is held on a bequeathable form and in case of premature death or accidental death with imperfect annuity market, their precautionary saving will be left to their children. Note that bequest in real life tends to result from some combination of these motives and cannot be easily distinguished. The three first bequest motives come from the

[1]This result has been dismissed by Modigliani, who estimates that the life cycle saving constitutes 80% of the total wealth, while inheritance constitutes only 20%. According to Davies and Shorrocks(1999), a more reasonable estimated result should be that inheritance makes a contribution of $35\% - 45\%$ of the aggregate wealth.

voluntary willingness of the parents, and bring utility to them. Thus, inheritance taxation on these intended bequests will have a distortionary effect on their behavior and damage economic efficiency. The accidental bequest, however, constitutes a special case where, according to most of the literature[2], inheritance taxation does not cause any distortion since the agents cannot react to the inheritance tax. Besides, based on the thought of Mill, the accidental bequest is not written in the will of the deceased. Thus, there is no moral obstacle for implementing a confiscatory 100% accidental inheritance tax. The optimal inheritance taxation, which takes into account its implications in terms of equality, efficiency, as well as simplicity and compliance, is thus largely determined by the relative importance of intended bequest and accidental bequest. According to Pestieau and Poterba(2001), there is always controversy about the importance of altruistic as opposed to accidental bequest, leading to uncertainty in the magnitude of efficiency cost caused by inheritance taxation. But at least qualitatively, when other things remain constant, a higher share of intended bequest motive would induce lower inheritance taxation.

In this master thesis, we will analyze the relationship between the optimal inheritance taxation and the death risk of the individuals, which can be considered as a proxy of the relative importance of the different bequest motives: a higher death risk provokes more premature death and more deceased leaving bequest without explicit will. We will see that the optimal inheritance taxation is affected by the death risk through a comprehensive set of channels with implications for inequality and efficiency. Apart from the study on taxation, another interesting part consists of analyzing the theoretical evolution of the inheritance flow from period to period and the long-term stationary distribution of inheritance. Basic methodological settings are discussed in the next section, with comparisons in existing literature. The third section is devoted to the competitive equilibrium without any governmental intervention. The first-best social optimum and its decentralization will be discussed in the fourth section, while the fifth section is reserved for discussing the second-best analysis, including the study on the long-term distribution of normalized inheritance and the optimal inheritance taxation in a more realistic world. The last section concludes.

2 Methodology

We will address the optimal inheritance taxation problem in a classical two-period overlapping generations model, in which individuals work and consume in the first period, while they retire, consume and bequeath in the second, as in the remarkable textbook of De la Croix and Michel(2002). The parent who survives to the end of the second period leaves intended bequest at the end of the first period of the kid's life. If the parent lives only one period,

[2]Some objections against this commonly held opinion will be briefly presented in Appendix (7.1).

her first-period saving is left as the total raw bequest at the very beginning of the kid's life. As for demographic setting, we assume that the population grows at a constant rate $n > -1$ and inheritance is equally shared among siblings within a family. As we shall see, a higher population growth rate is also a factor that mitigates the wealth concentration. There are two similar ways in the associated literature to incorporate the death risk in the two-period OLG model: firstly, as in Blumkin and Sadka(2004), Pestieau and Sato(2008),and Fleurbaey et al.(2017), individual lives for sure until the end of the first period, and there exists a probability $p \in (0, 1)$ to survive for the entire second period. Thus, the individual's lifespan is a binary variable whose realization is about living for either one or two periods. Secondly, each individual lives with certainty the entire first period but lives for a share of $\theta_i \in [0, 1]$ in the second period. That is, the maximum lifespan is two entire periods while the minimum is one entire period, and the individual's realized lifespan can take any value between the two extreme cases. This approach was only used in Michel and Pestieau(2002). To be consistent with the majority of the authors and to be simple with regard to the distribution of lifespans, we will adopt the first approach in this paper but compare some of the results with those in Michel and Pestieau(2002).

Michel and Pestieau(2002) analyze this question in a dynamic OLG model with a closed economy. They obtain the results of competitive equilibrium but do not provide a formula of the optimal taxation, as they believe it to be "not analytically tractable even within the simple model". Blumkin and Sadka(2004) obtain an explicit and reasonable optimal bequest tax formula in a particular case where labor supply is fixed. In contrast, they adopt a static OLG model where there are only two generations for each dynasty, since they focus on the "short run" that they believe to be the most relevant with practical policy designs. In the present master thesis, we will study the optimal taxation in a dynamic framework, where there are an infinite number of generations for every dynasty, and there are a large number of dynasties such that the law of large number is applied in terms of the distribution of lifespan and productivity. To make the model analytically tractable, we will switch from a closed economy where the factor prices are endogenously determined by the domestic capital accumulation to a small open economy where the wage and interest rate remain the same for each period, as in Piketty and Saez(2012), in the second-best analysis. As for the competitive equilibrium and the first-best decentralization, as in Michel and Pestieau(2002), it is totally tractable with endogenous factor prices, thus we maintain this setting for the first-best part.

In order to concentrate on the inheritance taxation instead of labor income taxation, we assume that labor supply of the young generation is inelastic, and the inequality in labor income comes from the heterogeneous productivity which we assume to be drawn i.i.d. from a given distribution for each individual of each generation. Thus it is not an optimal income taxation question as in Mirrless(1971), where agents value leisure and choose the labor supply by themselves, and the government cannot observe their productivity because they cannot observe their “true” labor supply. The labor income taxation part is rather similar to that in Piketty and Saez(2012), where the labor supply is given while productivity is heterogeneous.

In terms of the bequest motives, unlike a Barro-type altruism form (see Barro(1974)) where the parent takes into account the discounted sum of the life cycle utilities of all her descendants, we will adopt a joy of giving bequest motives to represent the intended bequest motive. The reason for not adopting the Barro-type dynamic altruism is that this type of altruism turns the OLG model equivalent to an infinite-horizon model, where all generations are linked by altruism. Consequently, a dynasty composed with infinite generations simply behaves like an individual with an infinite lifespan, which guarantees the validity of *Ricardian equivalence* where parents compensate all the governmental intergenerational redistribution such as public borrowing and pay-as-you-go pension system by smoothing their consumption through bequests. According to Chamley(1986), the long run optimal inheritance taxation should be zero, since there will be a much larger distortion on the future consumption than the benefit from redistribution across heterogeneous individuals. The Barro-type altruism is unrealistic in the sense that individuals in real life tend to have imperfect altruism without caring about the lifetime utility of the children, leaving room for a positive optimal inheritance tax rate.

In the first-best analysis, we will assume an omnipotent government who knows every private information of the individual, and most importantly, has the capability to distinguish the bequest motives. When switching to the more realistic second-best analysis, the government is no longer able to distinguish accidental bequest from joy of giving, and thus applies a unique inheritance tax for both types of bequest. In this master thesis, we will only focus on the flat rate taxation that is applied to all periods and generations, both for labor income and inheritance. The final optimal inheritance tax formula requires some numerical simulations, which we present in Appendix 2.

3 A laissez-faire model and competitive equilibrium

3.1 Consumers

The preference is assumed to be homogeneous and presented by a separable, homothetic, and strictly concave utility function. The utility of individual i of generation t writes:

$$U_{ti} = u(c_{ti}) + p\beta u(d_{t+1i}) + \gamma u(x_{t+1i}) \tag{3.1.1}$$

where c_{ti} is the consumption in the first period of live, d_{t+1i} the consumption in the second period (if the individual survives), and x_{t+1i} the intended bequest that she plans to leave at the end of the second period. $u(\cdot)$ is a constant intertemporal elasticity of substitution

(CIES) utility function. $p \in (0,1)$ is the probability of surviving at the end of the first period which is the same for all individuals. $\beta \in (0,1)$ is the discount factor and $\gamma > 0$ a parameter measuring the degree of altruism, or more precisely, of the joy of giving. The utility derived from bequeathing comes from solely the joy of leaving intended bequest, which corresponds to the opinion of Mill that the deceased does not value the accidental bequest for which there is no testimony of the deceased's will. However, this paternalistic altruism is not adopted by Fleurbaey et al.(2018) where they showed that we should even subsidize the accidental bequest under ex-post egalitarian criterion. Here we adopt the point of view of Mill, to make the model more tractable and comparable with existing literature. Besides, the inclusion of accidental bequest in the utility function can make accidental bequest taxation distortionary, even though the social planner is capable to perfectly distinguish different bequest motives(see Appendix (7.1)). There is a coefficient p for the second period consumption as it is expected to be realized with probability p. Note that the intended bequest x_{t+1i} does not depend on whether the individual survives: if she survives, she will leave purely intended bequest x_{t+1i} for each of her children at the end of her life; otherwise she leaves s_{ti} which is however not purely accidental, since the first period saving includes also the intended bequest which she wished to leave. This setting is similar with the utility function in Michel and Pestieau(2002).

The inheritor knows her first-period resources when she is about to make the decision of consumption, saving, and bequeathing. Indeed, the possible accidental death of her parent takes place when the inheritor is born. Thus she is in one of the two cases, instead of an "expected" inheriting scenario. This setting is also adopted by Blumkin and Sadka(2004) in their two generations static model, where the inheritor faces one of the two cases.

The heterogeneity in lifespan leads to two kinds of inheritors in terms of the structure of their inheritance. In the first case where the parent survives to the end of the second period, the kid receives only the intended bequest. In the second case where the parent dies at the end of the first period, the orphan will receive the parent's first-period saving whose capitalized value includes both the intended bequest (which is to be left in both cases), and the planned second-period consumption of the parent (which is not realized because of the sudden death). Importantly, the structure of the inheritance does not necessarily imply the level of the inheritance. The level of inheritance depends not only on the lifespan of the parent, but also on the entire lifespan history of all the ancestors, let alone the productivity shocks' history that makes it possible for some individuals to receive an intended bequest which is already higher than the sum of intended and accidental bequest received by others. [3]

[3] This complexity of dynamic model has been eliminated in Fleurbaey et al.(2018), where they assumed a quasi-linear utility function U with $u(c) = c$, such that the marginal utility in first-period consumption is 1. Consequently, the intended bequest and second-period consumption does not depend on the first-period resource, and thus are independent of the lifespan of ancestors. For an inheritor, her inheritance only depends on the characteristics of her parent, instead of the whole history of lifespan and productivity of the dynasty. We will adopt a more general and realistic preference and handle this complexity.

The individual i of generation t with productivity h_{ti} 's maximization program writes:

$$\max_{c_{ti},d_{t+1i},x_{t+1i}} U_{ti} = u(c_{ti}) + p\beta u(d_{t+1i}) + \gamma u(x_{t+1i})$$

subject to:

$$\begin{cases} c_{ti} + s_{ti} = b_{ti} + w_t h_{ti} \\ d_{t+1i} + (n+1)x_{t+1i} = s_{ti}(1+r_{t+1}) \end{cases} \tag{3.1.2}$$

where $w_t h_{ti}$ is the labor income of the individual with productivity h_{ti}, meaning that she has h_{ti} units of efficient labor. To be simple, assume that the average productivity of the young population is normalized to one. w_t and r_{t+1} are factor prices common for everyone. b_{ti} is the total inheritance received by this individual i, whose distribution depends on the entire productivity and lifespan history of all ancestors, as well as its initial distribution. As b_{ti} is exogenous for individual i to make her consumption, saving, and bequeathing decision, we will rather focus on the transition function and the long run evolution of inheritance distribution in the next sections.

By plugging the budget constraints into the utility function, and maximize w.r.t. s_{ti} and x_{t+1i}, the FOCs give:

$$\begin{cases} p\beta u'(d_{t+1i})(1+r_{t+1}) = u'(c_{ti}) \\ \gamma u'(x_{t+1i}) = p\beta u'(d_{t+1i})(n+1) \end{cases} \tag{3.1.3}$$

The first equation is the intertemporal trade-off between consuming in the first and second period. A higher surviving probability leads to a higher second-period consumption. The second one is the intratemporal trade-off between consuming and leaving intended bequest in the second period. A higher surviving probability gives individuals the incentive to devote more saving for the second-period consumption and less for intended bequest, since it becomes more likely to enjoy a realized second-period consumption.

In the following, assume that $u(Z) = \ln(Z)$(it is an assumption for simplicity, where income effect and substitution effect offset each other completely), for $Z \in \{c_{ti}, d_{t+1i}, x_{t+1i}\}$.Then the solution of the program writes:

$$\begin{cases} c_{ti} = \frac{1}{1+p\beta+\gamma}(w_t h_{ti} + b_{ti}) \\ s_{ti} = \frac{p\beta+\gamma}{1+p\beta+\gamma}(w_t h_{ti} + b_{ti}) \\ d_{t+1i} = s_{ti}(1+r_{t+1})\frac{p\beta}{p\beta+\gamma} \\ x_{t+1i}(1+n) = s_{ti}(1+r_{t+1})\frac{\gamma}{p\beta+\gamma} \end{cases} \tag{3.1.4}$$

To distinguish the two possible structures of the total inheritance received, we can write b_{ti} in terms of the planned second-period consumption d_{ti} and intended bequest x_{ti} of the parent:

$$b_{ti} = \begin{cases} x_{ti} & \textit{with probability p} \\ x_{ti} + \dfrac{d_{ti}}{1+n} & \textit{with probability 1-p} \end{cases} \tag{3.1.5}$$

All inheritors have the same saving rate $\tilde{s} = \frac{p\beta+\gamma}{1+p\beta+\gamma}$, the share of savings devoted to the second period consumption $\frac{p\beta}{\gamma+p\beta}$, and the share devoted to the intended bequest $\frac{\gamma}{\gamma+p\beta}$. Thus the difference among them lays only on the labor productivity and the level of total bequest received. The binary expression of b_{ti} implies that it is determined by the level of x_{ti} and d_{ti} depending on the lifespan path history until that of the grandparent (generation $t-2$), as well as its structure (whether it includes an accidental part) depending on the lifespan of parent (generation $t-1$). This distinction of the parent's lifespan from other ancestors' comes from the fact that her intended bequest and second-period consumption decisions do not hinge on her own lifespan. Thus, the intended inheritance received by an individual is independent of her parent's lifespan. However, the level of the intended inheritance depends on the whole lifespan history until her grandparent, since it affects the total inheritance received by the parent, so as the level of the parent's intended bequest and second-period consumption. The parent's lifespan plays a different role which determines the structure of the total inheritance: whether or not there is an accidental bequest. Appendix (7.2) provides a more clear and formalized explanation, and shows how the intended inheritance is related to the lifespan history until the grandparent.

Another observation is that an increase in the surviving rate p leads to a higher $\tilde{s}$, which increases the planned second-period consumption. As a higher p also means that there is less accidental bequest inheritors in the economy, there will be less inheritors of accidental bequest but the level of accidental bequest they receive is higher. This explanation for the concentration of wealth by increasing life expectancy is also mentioned by Fleurbaey et al.(2018).

3.2 Production

The production is standard: There is a constant return to scale production function $Y_t = F(K_t, H_t L_t)$, where L_t is the aggregate labor supply, which is equal to the population of the young generation at t. H_t is the aggregate level of productivity. As mentioned before, H_t is assumed to be the same in every period and normalized to 1. Thus, the per (young) capita variable is equivalent to the per efficient labor unit variable. K_t is the aggregate capital stock at t.

At the per capita level, we have:

$$\begin{cases} y_t = f(k_t) \\ f'(k_t) = 1 + r_t \\ w_t \equiv w(k_t) = f(k_t) - f'(k_t)k_t \end{cases} \tag{3.2.1}$$

w_t is the average wage level in the economy. With productivity heterogeneity, the individual's labor income is equal to the individual marginal productivity $w_t h_{ti}$, and the average labor income is w_t. During the production process, the capital stock is assumed to depreciate at rate $\delta = 1$, which is convenient for a length of period measuring $30 \sim 40$ years. The representative firm produces using capital and efficient labor, and collects savings from the young population of the current period to build up the capital stock of the next period[4]: it receives the "deposits" I_t from the young population which is equal to the young generation's savings S_t at t and turns it into the capital K_{t+1} for production at $t+1$:

$$I_t = S_t = K_{t+1} - K_t(1 - \delta) = K_{t+1} \tag{3.2.2}$$

which constitutes the capital market equilibrium.

3.3 Competitive equilibrium

The individuals of the generation (-1) are the initial old embedded with an initial exogenous distribution of savings s_{-1} with a mean of $\bar{s_{-1}}$. The capital market equilibrium (3.2.2) is equivalent to $k_{t+1}(n+1) = \bar{s}_t$ at a per young inhabitant level, thus $\bar{s_{-1}}$ is such that $k_0 = (n+1)s_{-1}$. The competitive equilibrium in this closed economy is given by the sequences of macro-variables $(K_t, L_t, Y_t)_t$, the sequences of factor prices $(w_t, r_t)_t$, the sequences of shocks $(h_t, \lambda_t)_t$, as well as the sequences of micro-variables $(c_t, s_t, d_t, x_t, b_t)_t$ that are chosen by individuals facing the factor prices of the respective periods.

3.3.1 Micro-variables

The competitive equilibrium of micro-variables can be written as follows according to the solution set (3.1.4): Denote the distributions of state variables for generation t as s_{t-1}, λ_{t-1}, h_t, which stand for the distribution of generation $t-1$'s saving, the distribution of generation $t-1$'s lifespan indicator (which is Bernoulli), the distribution of generation t's productivity, respectively. Given an initial distribution of saving s_{-1}, which is the saving of the initial old generation, it determines directly x_0 and d_0. The initial old generation's lifespan follows

[4] It is equivalent to having a producing firm who produces from inputs and an investing firm who collects savings to build capital stock. See De la Croix and Michel(2002), page 10.

the distribution λ_{-1}, which determines the total bequest distribution b_0, the first-period consumption distribution c_0 and saving distribution s_0 of generation 0. The distribution of the planned second-period consumption d_1 and that of the intended bequest x_1 are determined directly by s_0. So on and so forth, all micro variable distributions are given by the state variables $\{s_{t-1}, \lambda_{t-1}, h_t\}$.

Imagine the case where productivity is homogeneous, and all individuals live over two entire periods. Then for the dynasties with a same initial wealth, they will have the same sequence of all variables. But with different lifespan of the parent, some inherit a larger amount of total bequest, probably because of accidental death taking place for several generations consecutively, making the average level of savings higher. It gives an intuition that a shorter overall lifespan causes higher capital accumulation under some conditions, which we will investigate afterwards.

3.3.2 Capital Accumulation

In order to see how an uncertain lifespan affects the process of capital accumulation and the steady state capital stock in the long run, we have the following proposition on the transition function of capital:

Proposition 3.3.2.1: The transition function of capital per efficient labor $k_t = \frac{K_t}{H_t L_t}$ writes:

$$k_{t+1} = \frac{1}{(1+p\beta+\gamma)(n+1)}[(p\beta+\gamma)f(k_t) - p^2\beta f'(k_t)k_t] \tag{3.3.2.1}$$

with $t \geq 0$, and $k_0 = \frac{\bar{s_{-1}}}{n+1}$. *Proof*: See Appendix (7.3).

It is obvious in our case that for each k_t, there exists one unique k_{t+1} such that (3.3.2.1) holds. A direct observation from equation (3.3.2.1) is that if $p = \gamma = 0$, we have $k_t = 0$, $\forall t \geq 1$, and $k_0 = \frac{\bar{s_{-1}}}{n+1}$. It means that when everyone lives for only one period and there is no motive to leave intended bequest, there will be no capital accumulation since there is no room for life-cycle saving.

3.3.3 Steady state capital stock

The competitive stationary state of capital is k^* such that $(n+1)k^* = \frac{1}{1+p\beta+\gamma}[(p\beta+\gamma)f(k^*) - p^2\beta f'(k^*)k^*]$. And the steady state of average intended bequest stock is $x^* = \frac{\gamma}{\gamma+p\beta}f'(k^*)k^*$. The steady state of the average planned second-period consumption satisfies $d^* = \frac{p\beta}{\gamma+p\beta}k^*f'(k^*)(1+$

n). As a result, the steady state total inheritance per inheritor is:

$$b^* = x^* + (1-p)\frac{d^*}{1+n} = \frac{\gamma + p\beta(1-p)}{\gamma + p\beta} f'(k^*)k^* \tag{3.3.3.1}$$

From now on, assume that $f(k_t) = k_t^\alpha$, $\alpha \in (0,1)$.

Proposition 3.3.3.1 The unique steady state value of the capital per efficient labor k_t writes:

$$k^* = \left(\frac{(n+1)(1+p\beta+\gamma)}{p\beta(1-p\alpha)+\gamma}\right)^{\frac{1}{\alpha-1}} \tag{3.3.3.2}$$

Proof: In equation (3.3.2.1), denote $\mu(k) = (p\beta+\gamma)\frac{f(k)}{k} - (p\beta)f'(k)$, and the condition satisfied by the steady state k is thus $(n+1)(1+p\beta+\gamma) = \mu(k)$. With $f(k) = k^\alpha$, we have $\mu(k) = (p\beta+\gamma-\alpha p^2\beta)k^{\alpha-1}$, where $p\beta+\gamma-\alpha p^2\beta = p\beta(1-\alpha p)+\gamma$ is obviously positive. Thus it gives $\lim_{k\to 0}\mu(k) = \infty$, and $\lim_{k\to\infty}\mu(k) = 0$. Furthermore, $\mu'(k) = (p\beta+\gamma-\alpha p^2\beta)(\alpha-1)k^{\alpha-2} < 0$, which ensures that k^* is the unique steady state value of k_t. QED.

We can thus study the effect of surviving probability on the steady state capital stock with this expression of k^*:

Proposition 3.3.3.2 When parameters α, β, γ are such that $\alpha(\beta+2(1+\gamma)) > 1$, the steady state value k^* is first increasing and then decreasing in the surviving rate p, with a threshold value $p^* \in (0,1)$.

Proof: Denote $h(p) = \frac{(n+1)(1+p\beta+\gamma)}{p\beta(1-p\alpha)+\gamma}$ which affects k^* negatively, then:

$$h'(p) = \frac{\beta(n+1)}{(p\beta(1-p\alpha)+\gamma)^2}(p\alpha(p\beta+2(1+\gamma))-1)$$

Its sign depends on that of $m(p) = p\alpha(p\beta+2(1+\gamma))-1$. Set $m(p) = 0$, it gives a positive root $p^* = \frac{\sqrt{\alpha^2(1+\gamma)^2+\alpha\beta}-\alpha(1+\gamma)}{\alpha\beta}$ and another negative root that are symmetric around $p = -\frac{1+\gamma}{\beta} < 0$. Since $m(0) = -1$, it is straightforward to see that $p^* < 1$ if and only if $m(1) > 0 \iff \alpha(\beta+2(1+\gamma)) > 1$. In this case $m(p)$ turns from a negative value to become positive, and $h(p)$ decreases then increases, making k^* at first increasing and then decreasing in $p \in (0,1)$.QED.

We can interpret ***Proposition 3.3.3.2*** in the following way:

- When $p < p^*$, k^* is increasing in p. As mentioned above, when $p = \gamma = 0$, $k_t = 0, \forall t \geq 1$. When p becomes positive, individuals know that they are more likely to realize a second period of consumption, which encourages them to save in the first period. It can be seen from the formula of the saving rate $\tilde{s} = \frac{p\beta+\gamma}{1+p\beta+\gamma}$ that agents tend to have a higher saving rate when the surviving probability increases. Also, one can see from the solution (3.1.4) that the planned second-period consumption is increasing in p, which makes the relative level of accidental bequest compared with intended bequest to be higher for those who receive accidental bequest, even though there are fewer individuals leaving an accidental bequest.

- When $p > p^*$, k^* is decreasing in p. Less accidental death leads to less accidental bequest because the increase in the level of accidental bequest per death is dominated by the drop in the number of deceased who leave accidental bequest, i.e., at aggregate level, more lifetime resources are consumed instead of being left accidentally. Besides, the intended bequest decreases according to solution (3.1.4). As a result, the total resources in the first period becomes lower, making both first-period consumption and savings lower, as is the steady state capital stock.

Hence there is a trade-off between higher saving rate and the ratio of accidental bequest over intended bequest for accidental bequest receivers, versus less accidental bequest units and lower intended bequest, when the surviving probability rises. When $p < p^*$, the effect of the first two channels dominates, otherwise that of the last two dominates. Imagine a social planner who is capable to control surviving rate p, e.g., through policies in the health system, p^* is thus the surviving rate which maximizes the capital stock at steady state, when other things remain equal.

4 The first-best analysis

4.1 Social planner's optimum

In the first-best study, the government will not take into account the paternalistic altruism or joy of giving, because it is a pure transfer between individuals: it does not make sense to value a pure transfer, since the joy of giving of the parent is also included in the utility of the kid, otherwise it would be redundant. Besides, if valuing a pure transfer within individuals, then people can just transfer resources to each other to increase the social welfare, which seems absurd. [5]

[5] The question concerning whether or not the social planner ignores the part of joy of giving in the utility function of individuals when designing the social welfare function has been a controversial issue. Some argue that including the paternalistic altruism in the SWF causes double counting and we should "exclude all external preferences, even benevolent ones, from our social utility function" (Harsanyi (1995)), Hammond (1988) is also in favor of this opinion. The advocates of including joy of giving in the SWF claim that the social planner cannot modify the individual's preferences in a paternalistic way. We exclude the joy of giving in the first-best analysis, also because otherwise it will generate an optimal bequest without upper bound (See Appendix (7.4)).

The social planner will not care about the lifespan of the ancestors because it cares only about the own characteristics of the individual, which are their own lifespan and productivity type. Also, the social planner in the first best question is omnipotent, in the sense that it can observe all personal characteristics and distinguish perfectly different bequest motives in order to make optimal decisions in allocating consumption of each life period according to the characteristics of the individual.

To simplify the presentation, we constrain the labor productivity to be binary ($h_{ti} \in \{h_{low}, h_{high}\}$ such that $E(h_{ti}) = 1, \forall t \geq 0$), so that there are four types of individuals in total, which are the combinations of long and short lifespan, along with high and low productivity:

productivity / lifespan	$High(q)$	$Low(1-q)$
$Long(p)$	LH	LL
$Short(1-p)$	SH	SL

Table 1: Individual characteristics' combinations

where $q \in (0, 1)$ is the share of high productivity individuals. The first letter of the indicators of characteristic combinations represents the lifespan and the second represents the productivity.

The utilitarian social welfare function at the steady state writes:

$$\begin{aligned} SWF = &\, pqu(c_{LH}) + p(1-q)u(c_{LL}) + \beta pqu(d_{LH}) + \beta p(1-q)u(d_{LL}) \\ &+ (1-p)qu(c_{SH}) + (1-p)(1-q)u(c_{SL}) \end{aligned} \tag{4.1.1}$$

subject to the resource constraint:

$$f(k) = pqc_{LH} + p(1-q)c_{LL} + (1-p)qc_{SH} + (1-p)(1-q)c_{SL} + \frac{pqd_{LH}}{n+1} + \frac{p(1-q)d_{LL}}{n+1} + (n+1)k \tag{4.1.2}$$

Note that one can interpret the social welfare function (4.1.1) twofold: firstly, it is the average utility from consumption of all individuals living at a given period in the long run, weighted by the respective share of the four characteristic combinations in the living population. Secondly, it is the expected utility of an unborn individual living in the long run, with the respective probabilities of realizing different characteristic combinations during her coming life. These two interpretations are equivalent with the law of large number. Beside, equation (4.1.2) corresponds to the fact that output is used either for consuming or for investing, which gives no room for bequeathing.

The FOCs give the optimal conditions:

$$u'(c_{LL}) = u'(c_{LH}) = u'(c_{SH}) = u'(c_{SL}) \Longleftrightarrow c_{LL} = c_{LH} = c_{SH} = c_{SL} = c \tag{4.1.3}$$

$$u'(d_{LH}) = u'(d_{LL}) \Longleftrightarrow d_{LH} = d_{LL} = d_L \tag{4.1.4}$$

$$\beta(1+r)u'(d_L) = u'(c) \tag{4.1.5}$$

$$f'(k) = n+1 \tag{4.1.6}$$

Thus, according to conditions (4.1.3) and (4.1.4), all inequalities are eliminated in the first-best solution: neither the productivity nor the lifespan of the agents plays a role in determining the optimal consumption. Equation (4.1.5) is the intertemporal condition for individuals who live for two periods, and (4.1.6) is the "Golden rule" condition which determines the optimal level of capital accumulation that maximizes the utility from consumption.

4.2 First-best decentralization

Comparing the social planner optimum and the competitive equilibrium, it can be seen that there is a difference in the intertemporal allocation of consumption (4.1.5) and (3.1.3). The social planner optimum gives $\beta(1+r)u'(d_L) = u'(c)$, while the FOCs from the competitive equilibrium gives $u'(c) = p\beta(1+r)u'(d)$. The ratio of consumption of the first period over that of the second period therefore tends to be lower in the social planner optimum than in the competitive equilibrium. This difference of intertemporal allocation of consumption is determined by the fact that the social planner faces a certain objective function as the death risk does not lead to aggregate risk, and the lower number of old individuals cannot legitimate a lower second-period consumption for those survivors. However, for each of the individuals, she faces an uncertain world where she may die at the end of the first period. Thus she tends to consume more in the first period to hedge against the risk of not surviving to the second period. Besides, a lower surviving probability makes the competitive intertemporal consumption allocation further away from the social optimum. This gap can be filled by a commodity tax imposed on, for example, the second-period consumption.

The second difference consists of the intratemporal allocation of consumption. In the competitive equilibrium, individuals whose parents live for one period enjoy higher consumption in both periods than those whose parents live for two periods, as long as they receive the same amount of intended bequest. However, the social planner do not care about the lifespan of their parents, and implies the same consumption level for all types of individuals in the same life period.[6] From the competitive equilibrium result, it implies that a 100% taxation on the accidental bequest is necessary to achieve the social planner optimum. Similarly, the higher productivity should be totally neutralized by a labor income taxation specialized for each type.

[6] It is easy to show that, even if the social planner considers the parent's lifespan as a characteristic together with productivity and the individual's own lifespan, the first order conditions still eliminate all inequalities and suggest a same consumption level for parented individuals and orphans. In fact, we obtain always the result such that all inequalities are eliminated in the first-best solution.

In terms of the inheritance taxation, as the social planner does not take into account the joy of giving, it is used indirectly to achieve the golden rule level of capital accumulation, via the saving of the individuals. Thus, to achieve the first-best solution, the social planner should apply a commodity tax on the second-period consumption τ_d, a set of labor income tax (τ_L), and an intended inheritance tax τ_x. Of course, as accidental bequest tax is non-distortionary in our model, i.e., individuals cannot preview the accidental death and change their behavior, it is optimal to set it to be 100%. Indeed, the 100% accidental bequest taxation allows to eliminate the inequalities coming from the lifespan of all ancestors.

Proposition 4.2.1 The first-best optimum can be decentralized by the set of tax instruments: $\{\tau_L, \tau_d, \tau_x, \tau_A\}$, with $\tau_L = \{\tau_{Lhigh}, \tau_{Llow}\}$ such that $h_{high}(1-\tau_{Lhigh}) = h_{low}(1-\tau_{Llow}) = H = 1$ that eliminate productivity heterogeneity, $1+\tau_d = p$ that eliminates the effect of death risk on intertemporal consumption decision, τ_x such that $1-\tau_x = \frac{\alpha(1+p\beta+\gamma)-(1-\alpha)(p\beta+\gamma)}{\alpha\gamma}$ that ensures the "Golden rule" level of capital accumulation, as well as $\tau_A = 100\%$ that eliminates all heterogeneity coming from lifespan of ancestors.

Proof: As mentioned above, it is optimal to apply $\tau_A = 100\%$ in our framework (See more detail in Appendix (7.1)). After taxing all accidental bequest, the maximization program of individual i at period t writes:

$$\max_{c_{ti}, d_{t+1i}, x_{t+1i}} U_{ti} = u(c_{ti}) + p\beta u(d_{t+1i}) + \gamma u(x_{t+1i}(1-\tau_x))$$

subject to:

$$\begin{cases} c_{ti} + s_{ti} = x_{ti}(1-\tau_x) + w_t h_{ti}(1-\tau_{Li}) \\ d_{t+1i} + (n+1)x_{t+1i} = s_{ti}(1+r_{t+1}) \end{cases} \tag{4.2.1}$$

By assuming a logarithmic utility function[7], the optimal decisions made by an individual in a competitive equilibrium writes:

$$\begin{cases} c_{ti} = \frac{1}{p\beta+\gamma+1}(x_{ti}(1-\tau_x) + w_t h_{ti}(1-\tau_{Li})) \\ s_{ti} = \frac{p\beta+\gamma}{p\beta+\gamma+1}(x_{ti}(1-\tau_x) + w_t h_{ti}(1-\tau_{Li})) \\ (1+\tau_d)d_{t+1i} = \frac{p\beta}{p\beta+\gamma} s_{ti}(1+r_{t+1}) \\ x_{t+1i}(n+1) = \frac{\gamma}{p\beta+\gamma} s_{ti}(1+r_{t+1}) \end{cases} \tag{4.2.2}$$

The labor income tax τ_{Li} is productivity-specified to keep the overall productivity $H = 1$. So the omnipotent government can impose a labor income tax $\tau_{Lhigh} > 0$ to the higher type, and $\tau_{Llow} < 0$ to the lower type such that: $h_{high}(1-\tau_{Lhigh}) = h_{low}(1-\tau_{Llow}) = H = 1$.

[7]This simplification is for obtaining a tax formula in a more tractable way, but it makes the behaviors of agents unaffected by the intended bequest taxation, since the substitution effect and income effect cancel out each other. One may consider a more general CIES utility function as in the second-best analysis to see the distortionary effect on the individual's bequeathing behaviors. See Michel and Pestieau(2002).

The expressions for c_{ti} and d_{t+1i} give at steady state:

$$d = c(1+r)\beta p \frac{1}{1+\tau_d} \tag{4.2.3}$$

The commodity tax τ_d is such that:

$$1+\tau_d = p \tag{4.2.4}$$

in order to realize the social planner's intertemporal allocation of consumption (4.1.5) by decentralization. When p decreases, the individuals face a higher death rate, and they tend to reduce their consumption in the second-period. With a negative commodity tax τ_d (a subsidy for the second-period consumption), individuals will consume more in the second period: the commodity tax can restore the social optimum for those who live two periods.

To deduce the optimal tax on the intended bequest, it is useful to write the capital accumulation function, when all the available tax instruments have already been applied. It consists of a classical capital accumulation process with a representative agent. We have the saving equation:

$$\bar{s_t} = \frac{p\beta+\gamma}{1+p\beta+\gamma}(w_t + \frac{\bar{s_{t-1}}(1+r_t)}{n+1}\frac{\gamma}{\gamma+p\beta}(1-\tau_x)) \tag{4.2.5}$$

With the capital market equilibrium (3.3.1), it can be written as:

$$(n+1)k_{t+1} = \frac{p\beta+\gamma}{1+p\beta+\gamma}(f(k_t) - f'(k_t)k_t) + k_t f'(k_t)(1-\tau_x)\frac{\gamma}{1+p\beta+\gamma} \tag{4.2.6}$$

Take $f(k_t) = k_t^{\alpha}$:

$$(n+1)k_{t+1} = \frac{p\beta+\gamma}{1+p\beta+\gamma}(1-\alpha)k_t^{\alpha} + \frac{\alpha\gamma}{1+p\beta+\gamma}(1-\tau_x)k_t^{\alpha} \tag{4.2.7}$$

This relation is obviously monotonic, and the steady state capital per capita k satisfies:

$$(n+1)k^{1-\alpha} = \frac{(p\beta+\gamma)(1-\alpha)+\alpha\gamma(1-\tau_x)}{1+p\beta+\gamma} \tag{4.2.8}$$

It can be seen that a higher intended inheritance tax rate will reduce the steady state capital level, since individuals have less first-period resources to save. Knowing that the golden rule (4.1.6) gives the optimal per capita capital level $\bar{k}$ which satisfies:

$$(n+1)\bar{k}^{1-\alpha} = \alpha \tag{4.2.9}$$

When $k = \bar{k}$, the optimal tax on the intended inheritance satisfies:

$$1-\tau_x = \frac{\alpha(1+p\beta+\gamma) - (1-\alpha)(p\beta+\gamma)}{\alpha\gamma} \qquad QED.(4.2.10)$$

This formula is computed in order to avoid over-accumulation or under-accumulation of capital, which do not maximize the overall consumption in the economy.

It would be interesting to study how the optimal intended tax rate is affected by the surviving rate p. The numerator of the RHS of (4.2.10) can be rearranged as:

$$(2\alpha - 1)p\beta + \alpha(1 + 2\gamma) - \gamma$$

If $2\alpha - 1 < 0 \iff \alpha < \frac{1}{2}$, when p increases, $1 - \tau_x$ decreases, and τ_x increases. In contrast, when $\alpha > \frac{1}{2}$, a higher p means a lower optimal tax on intended bequest. To see the mechanism behind this result, we can rewrite the transition function of the saving in the decentralized first-best solution:

$$\bar{s_t} = \frac{p\beta + \gamma}{p\beta + \gamma + 1} w_t + \frac{\gamma}{p\beta + \gamma + 1} \bar{s_{t-1}} \frac{1 + r_t}{n + 1} (1 - \tau_x) \tag{4.2.5'}$$

- The first term of RHS is the saving from labor income, which consists of a new resource contributing to the savings' accumulation. When the surviving rate p increases, this part increases because individuals expect a higher probability to enjoy the second-period consumption, and they save from the labor income.

- The second term of RHS, however, is the savings from the inheritance, which comes from the saving of the parent. This part, as a share of the parent's savings, is decreasing in p: the savings of the parent was used to finance more second-period consumption related to bequest.

Hence, the overall relationship between p and savings depends on the relative importance of the two parts. Knowing that, $1 + r_t = f'(k_t) = \alpha k_t^{\alpha - 1}$, and $w_t = f(k_t) - f'(k_t)k_t = (1 - \alpha)k_t^{\alpha}$. So we have $\frac{1+r_t}{w_t} = \frac{\alpha}{(1-\alpha)k_t}$, which is increasing in α. Thus, when α is relatively low, the first part of $\bar{s_t}$ coming from labor income dominates, and since this part increases in p, the social planner needs to reduce savings by increasing the tax τ_x. When α is relatively high, the second part from inheritance dominates, and since it is decreasing in p, a higher p tends to under-accumulation of capital, thus the social planner will decrease the tax τ_x to increase savings. τ_x can also be a subsidy in some cases, if there is a large risk of under-accumulation of capital.

However, this first-best study is not for real world policy recommendations. The government in the first-case is somewhat too omnipotent: it knows everything about the characteristics of individuals: productivity, lifespan, and bequest motives. In real world, the real productivity of individuals can be cached by them, in order to masquerade and benefice a more favorable tax policy. Besides, the government has no way to know the lifespan of an individual when allocating a consumption to her, and actually even this person herself has no idea about a possible premature death. Moreover, it is difficult for the government to know the bequest motives of individuals, they often observe the total bequest without knowing its share of different bequest motives.

In the following, we will turn to the second-best case, where the government will be much less powerful, for example, it can only apply one single inheritance tax, on both intended and accidental bequest, which means that it cannot eliminate all inequalities. To study the optimal overall inheritance taxation, it is firstly useful to analyze the transition of the overall bequest b_t in this economy and its steady state.

5 The second-best analysis

Several features of the first-best solution and its model setting are quite questionable in a more realistic framework, where we have tax instruments limitations, endogenous and probably heterogeneous preferences, and imperfect observations of individual characteristics by the social planner. In this section we will begin with a study on the transition function of individual inheritance which depends on random processes, as well as its long term distribution in normalized value. As said in section 2, it would be analytically unfeasible without some compromises in the model setting to obtain a proper and tractable tax formula in the second-best analysis. Thus we will switch to a small open economy whose factor prices are given exogenously. After studying the dynamics of inheritance, we will focus on the second-best optimal tax formula.

5.1 Total inheritance dynamics

For an individual i of generation $t + 1$, the intended bequest she receives is given by:

$$x_{t+1i} = \frac{\gamma(1+r)}{(1+p\beta+\gamma)(n+1)}(b_{ti} + wh_{ti})$$

which comes directly from the solutions (3.1.4) except that we have now exogenous factor prices w and r. The total inheritance b_{t+1i} includes this intended inheritance x_{t+1i} for sure, along with a possible accidental part which equals to the second-period consumption of her parent of generation t with probability $(1-p)$. Thanks to the intertemporal condition $d_{ti} = x_{ti}(n+1)\frac{p\beta}{\gamma}$, the total inheritance she receives can be written as:

$$b_{t+1i} = x_{t+1i} + \mathbb{1}_{ti}\frac{d_{t+1i}}{n+1} = (1 + \mathbb{1}_{ti}\frac{p\beta}{\gamma})x_{t+1i}$$

with $\mathbb{1}_{ti}$ following a Bernoulli distribution: its realized value is 1 with probability $(1-p)$ and 0 with probability p. Thus, we have the transition function of b_{ti}:

$$b_{t+1i} = \underbrace{[\frac{\gamma(1+r)}{(1+p\beta+\gamma)(n+1)}b_{ti} + \frac{\gamma(1+r)}{\gamma+p\beta+1}\frac{wh_{ti}}{n+1}]}_{x_{t+1i}}(1+\mathbb{1}_{ti}\frac{p\beta}{\gamma}) \tag{5.1.1}$$

where b_{ti} can be written as $b_{ti} = (1+\mathbb{1}_{t-1i}\frac{p\beta}{\gamma})x_{ti}$. This implies that generation t receives an intended bequest x_{ti} from generation $t-1$ for sure, and an accidental bequest $\frac{p\beta}{\gamma}x_{ti}$ if the generation $t-1$ died young. The generation t uses this amount to bequeath for the generation $t+1$ voluntarily, as a result, the intended bequest x_{t+1i} does not depend on the lifespan of generation t, but depends on the lifespans of older ancestors.

We assume that the government has access to a flat rate labor income taxation τ_L unique for everyone and a flat rate inheritance taxation τ_B applied to all inheritance, regardless of their bequest motives. With these tax instruments the second-best competitive equilibrium satisfies:

$$\begin{cases} c_{ti} = \frac{1}{1+p\beta+\gamma}(wh_{ti}(1-\tau_L) + b_{ti}(1-\tau_B)) \\ s_{ti} = \frac{p\beta+\gamma}{1+p\beta+\gamma}(wh_{ti}(1-\tau_L) + b_{ti}(1-\tau_B)) \\ d_{t+1i} = s_{ti}(1+r)\frac{p\beta}{p\beta+\gamma} \\ x_{t+1i}(1+n) = s_{ti}(1+r)\frac{\gamma}{\gamma+p\beta} \end{cases} \tag{5.1.2}$$

Thus, the individual total bequest transition function with tax instruments writes:

$$b_{t+1i} = [\frac{\gamma(1+r)}{(1+p\beta+\gamma)(n+1)}b_{ti}(1-\tau_B) + \frac{\gamma(1+r)}{(\gamma+p\beta+1)(n+1)}wh_{ti}(1-\tau_L)](1+\mathbb{1}_{ti}\frac{p\beta}{\gamma}) \tag{5.1.1'}$$

It is useful to observe the sources of inequalities from the individual bequest transition equation, as in Piketty and Saez(2012). In our case, if an individual receives a different inheritance b_{t+1i} compared with her contemporaries, it is because:

- Her parent receives a different bequest b_{ti}.
- Her parent's labor productivity, or labor income is different. In our case the labor productivity is independent of generations and dynasties, i.e., every individual draws i.i.d. her productivity from a given distribution, as a result one's labor productivity is independent of the inheritance received. In contrast, if assuming a positive correlation between the productivity throughout generations for each dynasty, there will be a positive correlation between one's inheritance and her own productivity, since she tends to have ancestors with high productivity who were capable to bequeath more.
- The lifespan of her parent plays a role here. It determines the structure of the kid's total inheritance, namely, whether there is an accidental part. It is equivalent to consider it

as a binary random variable drawn i.i.d. from a given Bernoulli distribution by every individual.

Denote $b_{t+1} \equiv E(b_{t+1i})$ as the average inheritance received by individuals of generation $t+1$, $b_t \equiv E(b_{ti})$ as the average bequest received by individuals of generation t. With $E(h_{ti}) = 1$ and the fact that one's inheritance and productivity are both independent of her lifespan, the average inheritance's transition function writes:

$$b_{t+1} = [\frac{\gamma(1+r)}{(1+p\beta+\gamma)(n+1)} b_t(1-\tau_B) + \frac{\gamma(1+r)}{(\gamma+p\beta+1)(n+1)} w(1-\tau_L)](1+(1-p)\frac{p\beta}{\gamma}) \quad (5.1.3)$$

Equation (5.1.3) gives the steady state total inheritance b (if the interest rate is not too much high and the non-explosive steady state exists)[8]:

$$b = \frac{w(1-\tau_L)}{\frac{(1+p\beta+\gamma)(n+1)}{(1+r)(\gamma+p\beta(1-p))} - 1 + \tau_B} \quad (5.1.4)$$

Obviously, stationary per capita inheritance is linear and increasing in the wage level, as individuals have more resource for leaving bequest. It is decreasing in population growth: the wealth will be distributed into more parts if there are more kids. b increases in r since there will be a higher capitalized saving for both intended and accidental bequest. A higher τ_B clearly causes a lower b, as for the moment we do not introduce the relationship between τ_B and τ_L.

However, it makes more sense to study the relative importance of inheritance in comparison with the aggregate output, rather than simply focusing on the absolute level of inheritance. It allows for a historical comprehension of the inheritance share's evolution, as empirical data about the inheritance flow-output share exist and the so-called "sufficient statistics" approach can be adopted. Besides, by focusing on the inheritance-output ratio, the bequest transition function can be more comparable with the relevant study in Piketty-Saez(2012).

5.2 Inheritance-output ratio: r and g

We can apply the "r and g" narrative from Piketty and Saez(2012) in the present two period OLG model with uncertainty in lifespan. To do that, we should first introduce a driving force for economic growth at the per capita level: the population growth only amplify the total

[8] Alternatively, the steady state of inheritance b can be computed from the variables at steady state we obtained in section (3.3.3) according to which $b^* = \frac{\gamma+p\beta(1-p))}{\gamma+p\beta} k^* f'(k^*)$. Substituting $f'(k_t)$ with $1+r$ and k^* with steady state saving which can be further written as a function of b^*, we can obtain the same expression of (5.1.4) with taxation absent.

output but does not affect the per capita output. Thus, we assume in the following that the labor-augmenting productivity H_t grows at a constant rate g per generation.

Modified setting of production side in a small open economy: The production side is modified in the small open economy: the firm sets the capital input to achieve the exogenous interest rate: $F'(K) = 1 + r = R$. Assume that the production is Cobb-Douglas $F(K_t) = K_t^{\alpha}(H_tL_t)^{1-\alpha} = K_t^{\alpha}(H_0(1+g)^tL_0(1+n)^t)^{1-\alpha}$. The marginal product of capital $F'(K_t) = \alpha K_t^{\alpha-1}(H_tL_t)^{1-\alpha} = R$. Thus, we have $K_t = (\frac{R}{\alpha})^{\frac{1}{\alpha-1}}H_tL_t$, then the output writes: $Y_t = ((\frac{R}{\alpha})^{\frac{1}{\alpha-1}}H_tL_t)^{\alpha}(H_tL_t)^{1-\alpha} = (\frac{R}{\alpha})^{\frac{\alpha}{\alpha-1}}H_tL_t = (\frac{R}{\alpha})^{\frac{\alpha}{\alpha-1}}L_0(1+n)^tH_0(1+g)^t$. Since R, α, L_0, H_0 are exogenously given, the growth of the total output Y_t is completely driven by the population growth n and the overall productivity growth g. As a result, the per capita output $y_t = \frac{Y_t}{L_t} = (\frac{R}{\alpha})^{\frac{\alpha}{\alpha-1}}H_0(1+g)^t$'s growth is only driven by the productivity growth g. Besides, the wage per efficient labor writes $w = (1-\alpha)(\frac{R}{\alpha})^{\frac{\alpha}{\alpha-1}}$ and the labor income per capita writes $y_{Lt} = wH_t = wH_0(1+g)^t = (1-\alpha)y_t$. Note that the per capita variables b_t, y_t,y_{Lt} are computed on the young population, which makes sense because only the young population works and inherits. Therefore, the economy is on a balanced growth path where all per capita variables grow at rate g per generation. In this case, we have a modified version of (5.1.3):

$$b_{t+1} = [\frac{\gamma(1+r)}{(1+p\beta+\gamma)(n+1)}b_t(1-\tau_B) + \frac{\gamma(1+r)}{(\gamma+p\beta+1)(n+1)}wH_t(1-\tau_L)](1+(1-p)\frac{p\beta}{\gamma}) \tag{5.1.3'}$$

Denote $b_{yt} = \frac{b_t}{y_t}$ as the inheritance-output ratio at period t. Note that b_t is not exactly the average of purely raw bequest left at period t, it is instead a combination of raw intended bequest x_t and capitalized bequest $s_{t-1}(1+r)$ which comprises an intended part and an accidental part. By dividing both sides of (5.1.5′) by y_t, we have the transition function of this ratio:

$$(1+g)b_{yt+1} = [\frac{\gamma(1+r)(1-\tau_B)}{(1+p\beta+\gamma)(1+n)}b_{yt} + \frac{\gamma(1+r)(1-\tau_L)}{(1+p\beta+\gamma)(1+n)}(1-\alpha)](1+\frac{p\beta}{\gamma}(1-p)) \tag{5.2.1}$$

where $b_{yt+1} = \frac{b_{t+1}}{y_{t+1}}$. To ensure the existence of a non-explosive steady state, we need the coefficient of b_{yt} to be less than 1:

Assumption 5.2.1:

$$\frac{\gamma(1+r)}{1+p\beta+\gamma}(1-\tau_B)(1+\frac{p\beta}{\gamma}(1-p)) < (1+g)(1+n)$$

which means that, the after-tax marginal propensity to bequeath for all children out of the first-period resource, augmented by the magnitude of accidental bequest, must be lower than $(1+g)(1+n)$, which captures the relative importance of r and g: the relative importance between existing wealth and new income. For example, if the capital return r is sufficiently high that the marginal propensity to bequeath out of the first-period resource is higher than the growth rate of total output, then the inheritance-output ratio's path will be explosive.

If *Assumption 5.2.1* is satisfied, the non-explosive steady state of inheritance-output ratio writes:

$$b_y = \frac{(1-\tau_L)(1-\alpha)}{\frac{(1+g)(1+n)}{1+r}\frac{1+p\beta+\gamma}{\gamma+p\beta(1-p)} - 1 + \tau_B} \tag{5.2.2}$$

Observations on (5.2.2):

- A higher overall bequest tax τ_B lowers b_y when we do not consider the relationship between τ_B and τ_L.
- The inheritance-output ratio b_y is a decreasing function in the per capita growth rate and interest rate ratio $\frac{1+g}{1+r}$. When the importance of new income is relatively larger than that of the existing wealth, the inheritance-output ratio will decrease. It can be reflected by the reduction in the share of annual inheritance flow in national income during the world wars. With the large destruction of the existing wealth by the war, the new income is emphasized, and the inheritance flow share dropped from $20\% - 25\%$ around 1900-1910 to 10% in the 1920s-1930s, and fell to less than 5% in the 1950s (see Piketty (2010)).
- The term $\frac{1+p\beta+\gamma}{(\gamma+p\beta(1-p))(1+r)}$ associated with p can be rewritten as:

$$\nu(p) = \frac{\frac{1+p\beta+\gamma}{\gamma(1+r)}}{1+\frac{p\beta}{\gamma}(1-p)}$$

  where the numerator is the inverse of the marginal propensity to voluntary bequeathing $\frac{\gamma(1+r)}{1+p\beta+\gamma}$, which is also the share of first-period resource used for intended bequest. As for the denominator, $\frac{p\beta}{\gamma}(1-p)$ is the aggregate accidental bequest magnitude relative to the intended bequest. When p increases, the propensity for voluntary bequest is lower, and the share of accidental bequest units $(1-p)$ decreases as there is less accidental death, these two effects lead to a higher $\nu(p)$ and thus a lower b_y. However, a higher p also means a higher ratio of accidental bequest over intended bequest $\frac{p\beta}{\gamma}$ for those who still receive it, since the planned second-period consumption of their parents is higher, which has a positive impact on b_y.

Proposition 5.2.1 When $\gamma < \frac{1}{2}$, there exists a threshold value $p^* \in (0,1)$ such that b_y increases in p when $0 < p < p^*$ and decreases in p when $p^* < p < 1$. Otherwise, b_y is always decreasing in p when $p \in (0,1)$.

Proof: Denote $\upsilon(p) = \frac{1+p\beta+\gamma}{\gamma+p\beta(1-p)}$, then $\upsilon'(p) = \frac{\beta(\beta p^2+2p+2\gamma-1)}{(\gamma+p\beta(1-p))^2}$. Setting $\upsilon'(p^*) = 0$, when γ is high enough such that $\beta(2\gamma-1) > 1$, p^* does not exist, and $\upsilon'(p) > 0, \forall p \in (0,1)$, meaning that b_y is always decreasing in p. Otherwise, it gives a root $p^* = \frac{-1+\sqrt{1-\beta(2\gamma-1)}}{\beta}$ with another root which is surely negative. If $p^* < 0$, meaning that $2\gamma - 1 > 0 \Longleftrightarrow \gamma > \frac{1}{2}$, it also ends up with a b_y always decreasing in p. To ensure that $p^* < 1$, we need $\gamma > \frac{-\beta^2-\beta}{2\beta}$, which always holds. Thus, when $\gamma < \frac{1}{2}$, we have $p^* \in (0, 1)$. We see that $v'(0) < 0$ when $\gamma < \frac{1}{2}$, thus $v'(p) < 0$ when $p < p^*$ and $v'(p) > 0$ when $p > p^*$, and consequently b_y increases and then decreases in p, with p^* as the threshold. QED.

When $\gamma > \frac{1}{2}$, the monotonic and negative relationship between b_y and p comes from the fact that the negative effect of a higher surviving rate on the marginal propensity for bequeathing, together with a less number of parents who leave accidental bequest, overweighs its positive effect on the accidental-intended bequest ratio for those who still receive accidental bequest. When $\gamma < \frac{1}{2}$, the ratio $\frac{p\beta}{\gamma}$ becomes more important and the positive effect on the accidental bequest-intended bequest ratio of accidental bequest receivers can dominate on the two negative effects when p is relatively low, making b_y increasing in p. When p increases and surpasses p^*, this ratio's increment is again overshadowed by the decrease in bequeathing incentive and number of accidental bequest receivers, and b_y decreases in p.

5.3 A more general preference

The second-best taxation should be related to the elasticity of individual decision, e.g., the elasticity of leaving intended bequest with regard to the net-of-tax rate, which measures the sensibility of the individual's reaction to taxation and the ensuing distortion. It is useful to generalize the utility function of a certain individual i to incorporate the influence of tax instruments on its decision. In the previous study on the competitive equilibrium and first-best social planner, it was assumed that individuals were endowed with a logarithm utility function, which simplified the computations and allowed cleaner expressions but made the elasticity of bequeathing with regard to inheritance tax always equal to 0, since the income effect and substitution effect cancel out. The logarithmic utility function can be viewed as a particular case of a general constant intertemporal elasticity of substitution utility function:

$$u(Z) = \frac{Z^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}}$$

where $Z \in \{c, d, x\}$, and $\sigma > 0, \sigma \neq 1$. σ is the intertemporal elasticity of substitution. The special case where $\sigma = 1$ leads to the logarithmic utility function. Since empirical estimates such as Kopczuk and Slemrod(2001) suggest that the elasticity of inheritance-output ratio with regard to net-of-tax rate is slightly positive (0.2 using US data), meaning that the substitution effect dominates the income effect $\Longleftrightarrow \sigma > 1$.

For an individual i of generation t, her maximization program writes:

$$\max_{c_{ti},d_{t+1i},x_{t+1i}} u(c_{ti}) + p\beta u(d_{t+1i}) + \gamma(x_{t+1i}(1-\tau_B))$$

subject to:

$$\begin{cases} c_{ti} + s_{ti} = b_{ti}(1-\tau_B) + wh_{ti}(1-\tau_L) \\ d_{t+1i} + (n+1)x_{t+1i} = s_{ti}(1+r) \end{cases} \tag{5.3.1}$$

where b_{ti} is formed by the process (5.1.1) and the initial wealth of i's dynasty. Solving this program with the general constant inter temporal elasticity of substitution utility function, we obtain:

$$\begin{cases} c_{ti} = \tilde{c}(b_{ti}(1-\tau_B) + wh_{ti}(1-\tau_L)) \\ s_{ti} = \tilde{s}(b_{ti}(1-\tau_B) + wh_{ti}(1-\tau_L)) \\ d_{t+1i} = \tilde{d}(b_{ti}(1-\tau_B) + wh_{ti}(1-\tau_L)) \\ x_{t+1i} = \tilde{x}(b_{ti}(1-\tau_B) + wh_{ti}(1-\tau_L)) \end{cases} \tag{5.3.2}$$

where:

$$\begin{cases} \tilde{c} = \frac{1}{1+(1+r)^{\sigma-1}(\gamma(1-\tau_B)^{1-\frac{1}{\sigma}})^{\sigma}(1+n)^{1-\sigma}+(p\beta)^{\sigma}(1+r)^{\sigma-1}} \\ \tilde{s} = \frac{(1+r)^{\sigma-1}(\gamma(1-\tau_B)^{1-\frac{1}{\sigma}})^{\sigma}(1+n)^{1-\sigma}+(p\beta)^{\sigma}(1+r)^{\sigma-1}}{1+(1+r)^{\sigma-1}(\gamma(1-\tau_B)^{1-\frac{1}{\sigma}})^{\sigma}(1+n)^{1-\sigma}+(p\beta)^{\sigma}(1+r)^{\sigma-1}} \\ \tilde{d} = \frac{(p\beta)^{\sigma}(1+r)^{\sigma}}{1+(1+r)^{\sigma-1}(\gamma(1-\tau_B)^{1-\frac{1}{\sigma}})^{\sigma}(1+n)^{1-\sigma}+(p\beta)^{\sigma}(1+r)^{\sigma-1}} \\ \tilde{x} = \frac{(1+r)^{\sigma}(\gamma(1-\tau_B)^{1-\frac{1}{\sigma}})^{\sigma}(1+n)^{-\sigma}}{1+(1+r)^{\sigma-1}(\gamma(1-\tau_B)^{1-\frac{1}{\sigma}})^{\sigma}(1+n)^{1-\sigma}+(p\beta)^{\sigma}(1+r)^{\sigma-1}} \end{cases} \tag{5.3.3}$$

When $\sigma = 1$, these coefficients are:

$$\begin{cases} \tilde{c} = \frac{1}{1+\gamma+p\beta} \\ \tilde{s} = \frac{\gamma+p\beta}{1+\gamma+p\beta} \\ \tilde{d} = \frac{p\beta(1+r)}{1+\gamma+p\beta} \\ \tilde{x} = \frac{(1+r)\gamma}{(n+1)(1+\gamma+p\beta)} \end{cases}$$

which are what we obtained in the previous case with a logarithmic utility function. It can be seen from (5.3.3) that the share of the first-period consumption $\tilde{c}$ is increasing in τ_B while that of saving $\tilde{s}$ is decreasing in τ_B when $\sigma > 1$, which implies that individuals whose substitute effect dominates income effect prefer to consume the resource instead of smoothing them to the next period where she will have to pay extra tax when bequeathing. $\tilde{d}$ is increasing in τ_B while $\tilde{x}$ is decreasing in τ_B when $\sigma > 1$, since individuals are reluctant to bequeath when τ_B is higher. The ratio between the accidental bequest and intended bequest for an inheritor who receives the both, which captures the magnitude of the effects of premature death on the total bequest level, writes:

$$M \equiv \frac{d_{ti}}{(n+1)x_{ti}} = \frac{(n+1)^{\sigma-1}(p\beta)^{\sigma}}{(\gamma(1-\tau_B)^{1-\frac{1}{\sigma}})^{\sigma}} \tag{5.3.4}$$

Obviously, this ratio is equal to $\frac{p\beta}{\gamma}$ when $\sigma = 1$, corresponding to the logarithm case. M decreases in γ because individuals favor more intended bequest. The opposite is true for higher $p\beta$. When $\sigma > 1$, a higher τ_B leads to a higher M since individuals with higher substitution effect than income effect will give less bequest and plan more second-period consumption. Similarly, as in our model the parent only values the intended bequest given to one of her children, although there is no heterogeneity among the siblings, this valuation based on the per-inheritor rather than total intended bequest lets σ play a role: when $\sigma > 1$, having more children to share the bequest reduces the incentive of the parent to leave intended bequest, thus she switches to consume more in the second period of life, which increases the relative magnitude of accidental bequest in case of sudden death.

Thus, with $\sigma > 1$, a higher τ_B increases M but decreases $\tilde{x}$. The effect of τ_B on the total bequest level is ambiguous and requires an investigation to determine its overall signal. A way to see the overall impact of τ_B on total bequest magnitude is to study its impact on the steady state inheritance-output ratio b_y.

5.4 Inheritance-output ratio and taxation

The total bequest transition function, similar as equation (5.1.5'), writes:

$$b_{t+1} = \tilde{x}(y_{Lt}(1-\tau_L) + b_t(1-\tau_B))(1+(1-p)M) \tag{5.4.1}$$

which can be transformed into the transition function of the inheritance-output ratio b_{yt}:

$$b_{yt+1} = \frac{\tilde{x}}{1+g}((1-\alpha)(1-\tau_L) + b_{yt}(1-\tau_B))(1+(1-p)M) \tag{5.4.2}$$

When the parameter $(1-\tau_B)(1+(1-p)M)\frac{\tilde{x}}{1+g} < 1$, the steady state b_y is:

$$b_y = \frac{\tilde{x}(1-\alpha)(1-\tau_L)(1+(1-p)M)}{1+g-\tilde{x}(1-\tau_B)(1+(1-p)M)} \tag{5.4.3}$$

This expression has two ambiguities when considering the impact of τ_B on b_y. In the first place, a higher τ_B may imply a softer tax burden for labor income taxation if the government uses the two taxations to finance a certain public expenditure; we see that a higher τ_B leads to a lower b_y, while a lower τ_L leads to a higher b_y. Additionally, $\tilde{x}$ and M also depends on τ_B in different directions, although b_y hinges positively on the product $\tilde{x}(1+(1-p)M)$.

One approach to incorporate the relationship between the two tax instruments is to consider the government's budget constraint. Assume that the government can only raise revenue from the two tax instruments, and it uses this revenue to finance a public expenditure which constitutes τ as a share of y_t at each period. Assume that taxes and expenditures are executed at the end of the period, when all the labor income and inheritance of the current generation are formed. The government's budget constraint writes:

$$b_t\tau_B + y_{Lt}\tau_L = \tau y_t \Longleftrightarrow b_{yt}\tau_B + (1-\alpha)\tau_L = \tau \tag{5.4.4}$$

Assumption 5.4.1: $\tau < 1-\alpha$, meaning that the government expenditure can be totally covered by a labor income tax $\tau_L < 1$.

Note that here τ_B not only applies to the raw inheritance received at middle age for those whose parent lived for two periods, but also to the capitalized bequest which comprises both an intended and an accidental part for those whose parent lived for only one period, since the tax base b_t at the end of period t is also a hybrid of raw bequest and capitalized bequest. When there is no risk on the return to capital, for example in our model with a constant and exogenous r faced by everyone, it is equivalent to tax the unlucky short-lived parent's first-period saving at the beginning of the kid's life or to tax its capitalized value at the end of the kid's first period of life. In other words, τ_B is a unique rate taxation imposed on raw bequest and on the capital return generated by the raw bequest. [9]

Plugging $\tau_B = \frac{\tau-(1-\alpha)\tau_L}{b_{yt}}$, the expression of b_y rewrites:

$$b_y = \frac{\tilde{x}(1+(1-p)M)(1-\alpha-\tau)}{1+g-\tilde{x}(1+(1-p)M)} \tag{5.4.5}$$

Obviously, $b_y > 0$ when *Assumption 5.4.1* is satisfied. It can be seen that when the combination of τ_B and τ_L is such that a fixed share of output τy_t as public expenditure is financed, the steady state of inheritance-output ratio does not depend explicitly on τ_B or on τ_L. The only channels through which the inheritance taxation can affect the inheritance-output ratio are its effect on the ratio $\tilde{x}$ and on the relative magnitude of accidental bequest compared with intended bequest M. As shown before, these two channels have opposite directions, and to see the overall effect of τ_B, we need to compute the product $\tilde{x}(1+(1-p)M)$:

$$\tilde{x}(1+(1-p)M) = \frac{1+r}{1+n}\left(1 - \frac{p(n+1)^{\sigma-1}(p\beta)^{\sigma} + (n+1)^{\sigma-1}(1+r)^{1-\sigma}}{(n+1)^{\sigma-1}(1+r)^{1-\sigma} + (\gamma(1-\tau_B)^{1-\frac{1}{\sigma}})^{\sigma} + (p\beta)^{\sigma}(n+1)^{\sigma-1}}\right) \tag{5.4.6}$$

[9] When there is no risk on the capital return, the government can divide the tax burden into a taxation on the raw bequest and another on the capital return on it, in a very flexible way. For instance, the government can choose not to tax raw bequest and only tax capital return at a higher rate.

Obviously, if $\sigma > 1$, a lower τ_B (a higher net-of-tax rate) leads to a higher $\tilde{x}(1+(1-p)M)$, which in turns results in a higher b_y, if everything else remains the same. As a consequence, when substitution effect dominates income effect, even if there is a reversed effect from accidental bequest, the overall effect is still dominated by the voluntary bequeathing rate $\tilde{x}$: a higher inheritance tax brings about a lower steady state inheritance-output ratio. And we obtain: when $\sigma > 1$, the elasticity of b_y with regard to the net-of-tax rate $1 - \tau_B$ is positive:

$$e_B = \frac{db_y}{b_y} \frac{1 - \tau_B}{d(1 - \tau_B)} > 0 \tag{5.4.7}$$

5.5 Long run distribution of normalized inheritance

Each individual's resource in our model is characterized by her labor productivity which is drawn i.i.d. from a given distribution, as well as the total inheritance she receives, which relies on more complex conditions: the realization of the lifespan and productivity of all her ancestors, together with the position in the initial old generation's wealth distribution. However, we can show that its distribution in the long run will converge to a stationary one which has implications for inequality. To begin with, we can rewrite the individual transition function of inheritance:

$$b_{t+1i} = \tilde{x}[b_{ti}(1 - \tau_B) + h_{ti} w_t (1 - \tau_L)]\xi_{ti} \tag{5.5.1}$$

where $\xi_{ti} = (1 + \mathbb{1}_{ti} M)$, meaning that generation $t + 1$ may receive an accidental part of the inheritance, depending on the lifespan of generation t. Besides, w_t is the average wage level at t: $w_t = (1 - \alpha) y_t$.

To analyze the transition function, we need to formalize some assumptions as follows:

- *Assumption 5.5.1*: The lifespan indicator (which indicates the existence of accidental bequest) ξ_{ti} is drawn i.i.d. from an exogenous distribution $\varphi_t(\xi) = \varphi(\xi), \forall t \geq 0$ where $\xi_{ti} = (1 + M)$ with probability $1 - p$, and $\xi_{ti} = 1$ with probability p.
- *Assumption 5.5.2*: The labor productivity h_{ti} is drawn i.i.d. from an exogenous distribution $\phi_t(h) = \phi(h), \forall t \geq 0$, with $E(h_{ti}) = 1$.
- *Assumption 5.5.3*: As in the previous section, to ensure the existence of a steady state of b_y, we need $(1 - \tau_B)(1 + (1 - p)M)\frac{\tilde{x}}{1+g} < 1$.

In order to concentrate on the relative position of individuals in the inheritance distribution, denote $z_{ti} = \frac{b_{ti}}{b_t}$ the normalized bequest received by individual i of generation t.[10] Then given an initial distribution of normalized inheritance $\theta_0(z)$, the random process for lifespan and productivity $\varphi(\xi), \phi(h)$, and the individual transition equation of bequest, entirely determine the law of motion of the distribution of normalized inheritance $\theta_t(z)$, and the joint distribution of normalized inheritance and productivity $\Phi_t(z,h) = \theta_t(z)\phi(h)$, because inheritance and productivity are assumed to be independent for a given individual.

Proposition 5.5.1: Under assumptions 5.5.1 $\sim$ 5.5.3, there is a unique steady state for the average inheritance-output ratio b_y, the normalized inheritance distribution $\theta(z)$, the joint inheritance-productivity distribution $\Phi(z,h)$. For any initial conditions, as $t \to \infty$, $b_{yt} \to b_y, \theta_t \to \theta$, and $\Phi_t \to \Phi$.

Proof: The three dimensional discrete time stochastic process $S_{ti} = (z_{ti}, \xi_{ti}, h_{ti})$ is a Markovian process with a state variable b_{yt}. In my model, the lifespan indicator ξ_{ti} and labor productivity h_{ti} are drawn i.i.d. from the exogenous distributions $\varphi(\xi)$ and $\phi(hi)$ for every generation, so that we can concentrate on the convergence of the Markovian process of z_{ti}.

From equation (5.5.1), we can deduce the individual transition function of z_{ti}:

$$z_{t+1i} = \frac{\tilde{x}\xi_{ti}[(1-\tau_B)z_{ti}b_{yt} + (1-\alpha)h_{ti}(1-\tau_L)]}{(1+g)b_{yt+1}} \tag{5.5.2}$$

Plugging (5.4.3) for b_y, the transition function of z_{ti} when $t \to \infty$ can be written as:

$$z_{t+1i} = \frac{\xi_{ti}}{1+g}[(1-\tau_B)\tilde{x}z_{ti} + \frac{h_{ti}(1+g-\tilde{x}(1-\tau_B)(1+(1-p)M))}{1+(1-p)M}] \tag{5.5.3}$$

Denote $\frac{(1-\tau_B)\tilde{x}}{1+g} = Q$, assume that $0 < Q < \frac{1}{1+(1-p)M}$[11]. The transition function (5.5.3) rewrites:

$$z_{t+1i} = \xi_{ti}[Qz_{ti} + (\frac{1}{1+(1-p)M} - Q)h_{ti}] \tag{5.5.3$'$}$$

Knowing that the minimum labor productivity is $h_0 < 1$, and the maximum is $h_1 > 1$, the minimum outcome of steady state z_0 satisfies $z_0 = Qz_0 + (\frac{1}{1+(1-p)A} - Q)h_0$, meaning that all the ancestors had lived two periods and never left an accidental bequest, and they all had the lowest labor productivity, thus we have:

$$z_0 = \frac{1-Q(1+(1-p)M)}{(1-Q)(1+(1-p)M)}h_0 \tag{5.5.4}$$

Similarly, the highest normalized inheritance at steady state (when the coefficient in this case

[10] It is easy to transform the normalized inheritance distribution into the distribution of absolute levels of inheritance in the long run: $b_{ti} = z_{ti}b_t = z_{ti}b_y y_t$, with b_y the constant steady state value, y_t grows at a constant rate g for each period.

[11] It is a natural assumption since a positive productivity shock of the parent has a positive impact on the kid's normalized inheritance position, other things remaining equal.

$(1+M)Q < 1)$, when all ancestors had the highest productivity but only lived for one period, is:

$$z_1 = \frac{(1+M)(1-(1+(1-p)M)Q)}{(1-(1+M)Q)(1+(1-p)M)} h_1 \tag{5.5.5}$$

It can be easily shown that $z_0 < 1$ whilst $z_1 > 1$. In addition, when $(1+M)Q > 1$, the highest normalized inheritance tends toward infinity $z_1 \to \infty$, since a non-explosive steady state does not exist.

Thanks to *assumption 5.5.1* and *assumption 5.5.2*, the Markovian process for z_{ti} verifies the following property over the interval $[z_0, z_1]$. For any relative inheritance positions $z_0 \leq z < z' < z'' \leq z_1$, there exists $T \geq 1$ and $\epsilon > 0$ such that $proba(z_{t+Ti} > z' | z_{ti} = z) > \epsilon$, and $proba(z_{t+Ti} < z' | z_{ti} = z'') \geq \epsilon$. In words, for a given relative position of inheritance, it is always possible to reach another relative position of inheritance in some finite number of generations. For example, to reach a relatively higher position, one needs to have consecutively several periods of high productivity and shorter lifespan, while for a lower position of inheritance, one needs to have consecutively several periods' low productivity and longer lifespan. Besides, the transition function is monotonic, in the sense that $z_{t+1i}(z_{ti})$ has a first-order stochastic dominance over $z_{t+1i}(z'_{ti})$ if $z_{ti} > z'_{ti}$: For all outcome values z of z_{t+1i}, the probability of having a higher outcome value than z when the current outcome is z_{ti} is at least equal to the probability when the current outcome is $z'_{ti} < z_{ti}$, and for some outcome values z, the probability of having a higher outcome value than z when the current outcome is z_{ti} is strictly higher than when it is z'_{ti}. Hence, according to the standard ergodic convergence theorem, there exists a unique stationary distribution $\theta(z)$ towards which $\theta_t(z)$ converges, independently of the initial distribution $\theta_0(z)$. QED.

The transition function (5.5.3') is in fact a so-called Kesten process which is a stochastic process with a multiplicative shock and an additive shock (See Kesten(1973) and Fournier(2015)). According to Piketty and Zucman(2015), one can show that all accumulation processes with multiplicative random shocks engender distributions with Pareto tails. The transition function of this kind writes as: $m_{t+1i} = \omega_{ti} m_{ti} + \epsilon_{ti}$, where ω_{ti} is an i.i.d. multiplicative shock with mean $\omega = E(\omega_{ti}) < 1$, and ϵ_{ti} an additive random shock. Our transition function of normalized bequest (5.5.3') verifies these conditions, with particularly $E(\xi_{ti} Q) = Q(1+M)(1-p) + Qp = Q(1+M(1-p)) < 1$ because we have $Q < \frac{1}{1+(1-p)M}$(see footnote 11). More importantly, the stationary distribution of m_i has a Pareto upper tail with a Pareto coefficient a which solves the Champernowne's equation: $E(\omega_{ti}^a) = 1$.[12] When $w_{ti} > 1$ with positive probability, there exists a unique $a > 1$ satisfying $E(\omega_{ti}^a) = 1$. In such case, it would be interesting to study the relationship between the inheritance taxation and the inverted Pareto coefficient $b = \frac{a}{a-1}$ which measures the inequality of distribution $\theta(z)$ of our model[13].

[12] See Appendix (7.5.1) for a brief demonstration.

[13] It can be shown that, for a certain level of normalized inheritance z, the average normalized inheritance

5.6 Optimal inheritance tax

In this section we will take the optimal inheritance taxation study from Piketty and Saez(2012) as a reference. Define the general social welfare function as follows:

$$SWF = \int\int_{z,h} \omega_{z,h} \frac{V_{zh}^{1-\Gamma}}{1-\Gamma} d\Phi(z,h) \tag{5.6.1}$$

where $V_{zh} = E(V_i|z_i = z, h_i = h)$ is the average steady state utility of individuals with a normalized inheritance z and productivity h. $\omega_{z,h}$ is the (subjective) social welfare weight associated with the utility of these individuals of productivity h and normalized inheritance z. Besides, Γ measures the concavity of the social welfare function: when $\Gamma = 0$, the social welfare function is linear: the government does not have much intention to redistribute income from high productivity individuals to low productivity individuals, which consists of a "meritocratic SWF" since the government considers that individuals are responsible for their own productivity. If $\Gamma \to \infty$, the concavity of the SWF is high, and the government considers that the productivity is by sheer luck, and it consists in a "radical" version of SWF. As we know, the productivity and normalized inheritance are independently distributed, which makes the joint distribution $\Phi(z,h)$ two dimensional: for any z, the distribution of h is the same. In the following, we assume that the government will take the "meritocratic approach" by setting $\Gamma = 0$, as V_{zh} itself is concave, there should still be some redistribution in productivity. Take the most general case where the social planner gives every citizen an equal social welfare weight, then the social welfare function can be simplified as:

$$SWF = \int\int_{z,h} V_{zh} d\Phi(z,h) \tag{5.6.1$'$}$$

It is direct to modify (5.6.1') into a "Rawlsian" one, by eliminating all other individuals than those who receive z_0: $SWF_{z_0} = \int_h V_{z_0 h} d\phi(h)$.

Proposition 5.6.1 When we consider a linear social welfare function ($\Gamma = 0$) with welfare weight $\omega_{z,h} = 1$ for each individual regardless of her position in the stationary joint distribution $\Phi(z,h)$, the optimal inheritance tax rate τ_B^* written in statistics that can be observed or estimated is:

$$\tau_B^* = 1 - \frac{\frac{\tilde{x}(n+1)}{1+r}\frac{1-\alpha-\tau}{b_y} + e_B + \frac{\tilde{x}(n+1)}{1+r}}{(1 + e_B + \frac{\tilde{x}(n+1)}{1+r})(1 - Z_h)} \tag{5.6.2}$$

and the corresponding labor income tax rate is $\tau_L^* = \frac{\tau - \tau_B b_y}{1-\alpha}$, where $Z_h = \frac{E(V_{ci} z_i)}{E(V_{ci} h_i)}$ is the ratio between average normalized inheritance and average productivity, weighted by the marginal

of those who receive at least z is $E(z_i|z_i \geq z) = bz$. For instance, when we focus on the average normalized inheritance of top inheritors who receive $z_i \geq 10$, if $b = 2$, then these top inheritors have an average of 20 for their normalized inheritance. A brief proof of this key property of Pareto upper tail is provided in Appendix(7.5.2).

utility of first-period consumption.

Proof: Imagine that the government implements a tax reform in which $d\tau_B > 0$ and the budget balance $\tau_B b_{yt} + (1-\alpha)\tau_L = \tau$ still holds. The total differentiation of the budget constraint writes (with regard to (τ_B, τ_L)):

$$0 = (1-\alpha)d\tau_L + (b_y + \tau_B \frac{db_y}{d\tau_B})d\tau_B$$

Given that $e_B = \frac{1-\tau_B}{b_y}\frac{db_y}{d(1-\tau_B)}$, we have:

$$d\tau_L = -\frac{d\tau_B b_y}{1-\alpha}(1 - \frac{\tau_B e_B}{1-\tau_B})$$

We see from this expression of $d\tau_L$ that labor income tax rate will not necessarily decrease when the inheritance tax rate increases, namely, $d\tau_L$ is not always negative when $d\tau_B > 0$. There are two possible cases where $d\tau_L$ can be positive: firstly, e_B is large enough, meaning that an increase in τ_B causes a huge loss in the tax base, namely, the inheritance that can be taxed, and the government has to resort to labor income taxation to re-balance the budget. Secondly, the existent inheritance tax rate τ_B is already very high such that a little erosion in the tax base of inheritance tax makes the loss in tax revenue $\frac{d_{b_y}}{d\tau_B}\tau_B$ considerable.

The utility function of individual i who receives normalized inheritance $z_i = z$ and possesses productivity $h_i = h$ in the stationary joint distribution $\Phi(z,h)$, expressed in the budget constraint, writes:

$$V_i = u(y_{Lti}(1-\tau_L)+b_{ti}(1-\tau_B)-s_{ti})+p\beta u(s_{ti}(1+r)-x_{t+1i}(n+1))+\gamma u(x_{t+1i}(1-\tau_B)) \quad (5.6.3)$$

where $b_{ti} = z_i b_t = z b_t = z b_y y_t$ when $t \to \infty$. $y_{Lti} = wH_t h_i = (1-\alpha)y_t h$. To calculate the total differentiation of V_i with respect to τ_B and τ_L, it is necessary to take into account the fact that the received inheritance b_{ti} is also a function of τ_B in the sense that, to maintain a constant $z = \frac{b_{ti}}{b_t} = \frac{b_{ti}}{b_y y_t}$, as b_y changes with τ_B, b_{ti} should also change.

$$dV_i = -V_{ci}y_{Lti}d\tau_L - V_{ci}b_{ti}d\tau_B - V_{xi}x_{t+1i}d\tau_B + V_{ci}(1-\tau_B)\frac{db_{ti}}{d\tau_B}d\tau_B \quad (5.6.4)$$

where $V_{vi} = u'(c_{ti})$, $V_{di} = p\beta u'(d_{t+1i})$, and $V_{xi} = \gamma u'(x_{t+1i}(1-\tau_B))$. Note that x_{t+1i} and s_{ti} are also functions in τ_B and in τ_L, but their derivatives w.r.t. τ_B and τ_L have been eliminated from dV_i thanks to the Envelope theorem, because they allow us to maximize V_i and $\frac{dV_i}{dx_{t+1i}} = \frac{dV_i}{ds_{ti}} = 0$.

Since $e_B = \frac{db_y}{b_y}\frac{1-\tau_B}{d(1-\tau_B)}$, we have $db_y = -e_B b_y \frac{d\tau_B}{1-\tau_B}$, and:

$$db_{ti} = db_y z_i y_t = -e_B b_y \frac{d\tau_B}{1-\tau_B} z_i y_t = -e_B \frac{d\tau_B}{1-\tau_B} b_{ti}$$

Thus:

$$dV_i = -V_{ci} y_{Lti} d\tau_L - V_{xi} x_{t+1i} d\tau_B - V_{ci}(1+e_B) b_{ti} d\tau_B \tag{5.6.5}$$

The FOCs of V_i w.r.t x_{t+1i} and s_{ti} write:

$$\begin{cases} u'(c_{ti}) = p\beta u'(d_{t+1i})(1+r) \\ \gamma u'((x_{t+1i}(1-\tau_B))(1-\tau_B) = p\beta u'(d_{t+1i})(n+1) \end{cases}$$

which is equivalent to:

$$\begin{cases} V_{di}(1+r) = V_{ci} \\ V_{xi} = \frac{n+1}{1-\tau_B} V_{di} \end{cases}$$

Thus we obtain the relationship between V_{ci} and V_{xi}:

$$V_{xi} = \frac{(n+1)V_{ci}}{(1+r)(1-\tau_B)}$$

Replacing V_{xi} by V_{ci} using this relationship, equation (5.6.5) can be written as:

$$dV_i = -V_{ci} y_{Lti} d\tau_L - V_{ci} \frac{x_{t+1i}(n+1)}{(1+r)(1-\tau_B)} d\tau_B - V_{ci}(1+e_B) b_{ti} d\tau_B$$

Plugging $d\tau_L = -\frac{d\tau_B b_y}{1-\alpha}(1 - \frac{\tau_B e_B}{1-\tau_B})$, we have:

$$dV_i = V_{ci} d\tau_B [y_{Lti} \frac{b_y}{1-\alpha} \frac{1-\tau(1+e_B)}{1-\tau_B} - \frac{x_{t+1i}(n+1)}{(1+r)(1-\tau_B)} - (1+e_B) b_{ti}]$$

Then it is useful to write b_{ti} and x_{t+1i} as functions of y_{Lti}. By definition:

$$b_{ti} = b_y y_t z_i = b_y z_i \frac{y_{Lti}}{(1-\alpha)h_i}$$

And for x_{t+1i}:

$$x_{t+1i} = \tilde{x}(y_{Lti}(1-\tau_L) + b_{ti}(1-\tau_B)) = \tilde{x}(y_{Lti}(1-\tau_L) + b_y \frac{y_{Lti}}{(1-\alpha)h_i} z_i (1-\tau_B))$$

Thus, we obtain:

$$dV_i = V_{ci} y_{Lti} d\tau_B [\frac{b_y}{1-\alpha} \frac{1-\tau_B(1+e_B)}{1-\tau_B} - \frac{\tilde{x}(n+1)}{(1+r)(1-\tau_B)} (1-\tau_L + b_y \frac{z_i(1-\tau_B)}{(1-\alpha)h_i}) - (1+e_B) b_y \frac{z_i}{(1-\alpha)h_i}]$$

Since $y_{Lti} = y_{Lt}h_i$, it is equivalent to write:

$$dVi = V_{ci}y_{Lt}d\tau_B[\frac{b_y}{1-\alpha}\frac{1-\tau_B(1+e_B)}{1-\tau_B}h_i - \frac{\tilde{x}(n+1)}{1+r}(\frac{1-\tau_L}{1-\tau_B}h_i + \frac{b_y z_i}{1-\alpha}) - (1+e_B)\frac{b_y z_i}{(1-\alpha)}]$$

Replacing τ_L by its expression in τ_B by keeping the budget balance $1-\tau_L = \frac{1-\alpha-\tau+\tau_B b_y}{1-\alpha}$ and using $y_{Lt} = y_t(1-\alpha)$, we obtain:

$$dV_i = \frac{V_{ci}y_t d\tau_B}{(1-\tau_B)}[b_y(1-\tau_B(1+e_B))h_i - \frac{\tilde{x}(n+1)}{1+r}(1-\alpha-\tau+\tau_B b_y)h_i - (\frac{\tilde{x}(n+1)}{1+r}+1+e_B)b_y z_i(1-\tau_B)] \tag{5.6.6}$$

Summing up (5.6.6) over the joint distribution $\Phi(z,h)$ to obtain the marginal change in the social welfare of the tax reform:

$$\begin{aligned} dSWF = \frac{y_t d\tau_B}{1-\tau_B}&[b_y(1-\tau_B(1+e_B))E(h_iV_{ci}) - \frac{\tilde{x}(n+1)}{1+r}(1-\alpha-\tau+\tau_B b_y)E(h_iV_{ci}) \\ &-(\frac{\tilde{x}(n+1)}{1+r}+1+e_B)b_y(1-\tau_B)E(z_iV_{ci})] \end{aligned} \tag{5.6.7}$$

Remind that we aim to obtain the optimal tax rate τ_B^* as a function of observable variable b_y and other variables that one can estimate from empirical data: $\tilde{x}, e_B, E(h_iV_{ci}), E(z_iV_{ci})$, thus we consider SWF as a function of τ_B only through explicit channels, meaning that we consider these mentioned variables as given by empirical statistics. Therefore, the *second order condition* can be verified:

From (5.6.6), the first order derivative of individual indirect utility w.r.t. τ_B writes:

$$\frac{dV_i}{d\tau_B} = \frac{V_{ci}y_t}{(1-\tau_B)}[b_y(1-\tau_B(1+e_B))h_i - \frac{\tilde{x}(n+1)}{1+r}(1-\alpha-\tau+\tau_B b_y)h_i - (\frac{\tilde{x}(n+1)}{1+r}+1+e_B)b_y z_i(1-\tau_B)] \tag{5.6.6'}$$

And the second order derivative can be rearranged as:

$$\frac{d^2V_i}{d\tau_B^2} = \frac{V_{ci}y_t}{(1-\tau_B)^2}[-b_y e_B h_i - \frac{\tilde{x}(1+n)}{(1+r)}(1-\alpha-\tau+\tau_B b_y)h_i - \frac{\tilde{x}(1+n)}{(1+r)}b_y h_i(1-\tau_B)] \tag{5.6.8}$$

The government budget constraint $\tau_B b_y + (1-\alpha)\tau_L = \tau$, along with *Assumption 5.4.1*, gives $1-\alpha-\tau+\tau_B b_y > 0$. Knowing that V_{ci} and e_B are both positive, the second order derivative w.r.t. τ_B is thus negative, meaning that V_i is concave in τ_B, so is SWF since it is the sum of concave functions. Thus τ_B such that $\frac{dSWF}{d\tau_B} = 0$ is the tax rate that maximizes the social welfare instead of minimizing it.

When $\frac{dSWF}{d\tau_B} = 0$, the optimal inheritance tax rate writes:

$$\tau_B^* = \frac{1 - \frac{\tilde{x}(n+1)}{1+r}\frac{1-\alpha-\tau}{b_y} - (\frac{\tilde{x}(n+1)}{1+r}+1+e_B)\frac{E(z_iV_{ci})}{E(h_iV_{ci})}}{(1+e_B+\frac{\tilde{x}(n+1)}{1+r})(1-\frac{E(z_iV_{ci})}{E(h_iV_{ci})})} \tag{5.6.9}$$

which can be rearranged to equation (5.6.2). QED.

Observations:

- τ_B^* is increasing in b_y: a higher steady state inheritance-output ratio leads to a higher inheritance tax. A higher b_y implies that the wealth from the past is relatively more important compared to the current labor income, possibly as a result of low growth rate g, which gives a reason for taxing more inheritance to mitigate the inequality.

- τ_B^* is decreasing in $\frac{\tilde{x}(n+1)}{1+r}$ which measures the share of first-period resources devoted to intended bequest for all children and thus reflects the willingness of leaving bequest. It means that the inheritance tax should decrease when parents have a higher preference for leaving bequest, particularly since $\frac{\tilde{x}(n+1)}{1+r}$ is positively correlated with the altruism degree γ. Otherwise it harms too much their utility derived from joy of giving.

- τ_B^* is decreasing in the elasticity of inheritance-output ratio with regard to the net-of-tax rate e_B when $\tau > 0$. It is not a good idea to tax inheritance when the elasticity of inheritance w.r.t. taxation is high in absolute value, since it will be highly distortionary and difficult to raise tax revenues. As it is possible for τ_B to be negative (an inheritance subsidy), we see that when $e_B \to \infty$, $\tau_B^* = -\frac{Z_h}{1-Z_h}$.

- The formula of τ_B^* is not defined in the case where $Z_h = \frac{E(z_i V_{ci})}{E(h_i V_{ci})} \geq 1$, which is in fact consistent with the basic setting of our model. Recall from the maximization program (5.3.1) that the individual only values the net-of-tax bequest she leaves to her child, and the argument of the CIES utility function (logarithmic when $\sigma = 1$) must be positive, as a result we exclude all possibility of obtaining a meaningful τ_B^* which is higher than or equal to 100%. From (5.6.2), when $\frac{E(z_i V_{ci})}{E(h_i V_{ci})} > 1$ we have clearly $\tau_B^* > 1$, which goes against the basic setting of exclusion of more than confiscatory tax.

- For $Z_h < 1$, we can see clearly from (5.6.2) that τ_B^* decreases in this ratio, and for a certain value of it τ_B^* turns negative: the government should subsidize the inheritance. The driving force of change in Z_h can be complicated, and hard to illustrate analytically. In particular, we have an intuition that a higher inequality in the productivity distribution will induce a higher labor income tax rate and thus a lower inheritance tax rate, by keeping the government budget balanced. This effect goes through Z_h: we use numerical method to show that a higher standard deviation in labor productivity S_h is correlated with a higher Z_h and thus a lower τ_B^*. See Appendix (8.3.1) for a numerical example.

Again, it is worth noting that (5.6.2) is not a proper formula of optimal inheritance tax rate, in the sense that they are not a function of purely primitive parameters. Some variables on the RHS also depend on τ_B, such as b_y, $\tilde{x}$, the ratio Z_h and even e_B. A higher τ_B reduces

b_y, which will make τ_B to be lower to reach the optimal rate; the higher τ_B also reduces e_B, which implies an even higher τ_B; it also reduces $\tilde{x}$, meaning τ_B should still increase. As for Z_h, our numerical simulation suggests that it is increasing in τ_B which needs τ_B to decrease to be really optimal (See Appendix (8.3.1)). These effects in different directions imply that there may exist an equilibrium where τ_B^{**} is the solution to:

$$\tau_B = \frac{1 - \frac{\tilde{x}(\tau_B)(n+1)}{1+r}\frac{1-\alpha-\tau}{b_y(\tau_B)} - (\frac{\tilde{x}(\tau_B)(n+1)}{1+r} + 1 + e_B(\tau_B))Z_h(\tau_B)}{(1 + e_B(\tau_B) + \frac{\tilde{x}(\tau_B)(n+1)}{1+r})(1 - Z_h(\tau_B))} \tag{5.6.10}$$

For numerical solutions of τ_B^{**} with Rawlsian social welfare criterion and general CIES utility function where $\sigma > 1$, see Appendix (8.3.4) where we show that a higher τ_B^{**} is highly associated with higher variance of shocks coming from uncertain lifespan and accidental bequest.

Nevertheless, the advantage of writing τ_B^* as these endogenous variables is that they can be more or less observed or empirically estimated from real data, for example there exist empirical data of b_y and e_B has been estimated to be around $0.1 - 0.2$ by Kopczuk and Slemrod(2001) using US data, therefore they are easier to calibrate. In the following we present a special case where this issue of endogeneity does not come into play. After that we switch to alternative social welfare criteria using a numerical method, but we shall take into account the endogeneity to some extent, meaning that variables such as $b_y, \tilde{x}, Z_h, e_B$ are determined by an existing tax rate τ_B, and to see the relationship between the new optimal τ_B^* and different welfare criteria.

A particular case: Logarithmic utility function with Rawlsian social welfare criterion: Let us come back to the special case of logarithmic utility function in order to have a look at the effects of surviving rate p on the optimal tax rate. From the steady state $b_y = \frac{\tilde{x}(1+(1-p)M)(1-\alpha-\tau)}{1+g-\tilde{x}(1+(1-p)M}$ we see that $e_B = 0$, since $\tilde{x}, M$ are both independent of τ_B when $\sigma = 1$. In this specific case, the optimal inheritance tax rate writes:

$$\tau_B^\star = \frac{1 - \frac{\gamma}{p\beta+\gamma+1}(\frac{1-\alpha-\tau}{b_y} + Z_h) - Z_h}{(1 + \frac{\gamma}{p\beta+\gamma+1})(1 - Z_h)} \tag{5.6.11}$$

In addition, when the social planner adopts a Rawlsian social welfare criterion, meaning that only the welfare weight of those who receive the minimum value of z_i in the long run distribution z_0 is 1, and that of all others is 0. The Rawlsian optimal inheritance tax rate is obtained by the same procedure, except that we only aggregate over h_i and set $z_i = z_0$ from (5.6.6) to (5.6.7). From equation (5.5.3'), for an individual i who have a path $(\xi_{ti})_t$ where $\xi_{ti} = 1, \forall t$, and $(h_{ti})_t$ where $h_{ti} = h_0$. Assume that the lowest productivity is $h_0 = 0$, thus (5.5.3') is modified to $z_{t+1i} = Qz_{ti}$, with $Q < 1$. Therefore, $z_0 \to 0$.[14] Consequently,

[14]The expression of z_0 by equation (5.5.4) gives $z_0 = 0$ when $h_0 = 0$, because it was calculated by setting $z_{t+1i} = z_{ti}$. In practical, our model does not give rise to $z_0 = 0$, it can only approach to 0.

we see that $Z_h = \frac{E(V_{ci}z_i|z_i=z_0)}{E(V_{ci}h_i|z_i=z_0)} \to 0$ under some conditions. [15] Hence the Rawlsian optimal inheritance tax rate with logarithm utility function writes:

$$\tau_B^\star = \frac{1 - \frac{\gamma}{p\beta+\gamma+1}(\frac{1-\alpha-\tau}{b_y})}{(1 + \frac{\gamma}{p\beta+\gamma+1})} \tag{5.6.12}$$

There are four channels for surviving rate p to influence τ_B that we can see analytically:

- The ratio $\frac{\gamma}{p\beta+\gamma+1}$, which reflects the share of first-period resources devoted to intended bequest, decreases when p is higher, because they have a higher preference for the second-period consumption than leaving bequest. This relative decrease in the taste for bequeathing makes the optimal inheritance tax rate higher. Hence a higher surviving rate leads to a higher inheritance tax. This result also allows us to mitigate the concentration of accidental bequest shown at the end of section (3.1).

- A higher p leads to a lower bequeathing rate, where in b_y tends to be lower. Thus a higher p also leads to a lower τ_B.

- A higher p also means that, for an individual receiving an accidental bequest, the ratio of accidental bequest over intended bequest she receives is larger, so b_y tends to be higher, so as τ_B.

- A higher surviving rate p means that there are less individuals who receive an accidental inheritance, so b_y tends to decrease, so does τ_B.

Plugging b_y which is independent with τ_B, we obtain its formula in terms of pure primitive parameters:

$$\tau_B^\star = 1 - \frac{(1+g)(1+n)}{1+r}\frac{\gamma(p\beta+\gamma+1)}{(\gamma+p\beta(1-p))(p\beta+2\gamma+1)} \tag{5.6.13}$$

It can be seen that $\tau_B^\star$ is a decreasing function in g and n, while an increasing function in r. The effects of these three parameters all go through b_y. As we saw in section (5.2), b_y decreases in n, g while increases in r, as the current income's growth reduces b_y and return to existing wealth accentuates it. In particular, a higher population growth rate means that there are more young individuals who work and contribute to output compared with old individuals who live on their savings at each period, which makes current income more important than existing wealth, thus $\tau_B^\star$ decreases. Alternatively, a higher n means that the given existing wealth is

[15] When $z_0 \to 0$, $V_{ci}z_i|_{z_i=z_0} = \frac{(1+\gamma+p\beta)z_0}{z_0 b_y y_t(1-\tau_B)+y_{Lti}(1-\tau_L)} = \frac{1+\gamma+p\beta}{b_y y_t(1-\tau_B)+(\frac{y_{Lt}h_i(1-\tau_L)}{z_0})} \to 0$, as long as h_i of those who receive z_0 cannot be 0. Then $E(V_{ci}z_i|z_i = z_0) \to 0$, which gives $Z_h \to 0$ since $E(V_{ci}h_i|z_i = z_0)$ is positive when $h_i > 0$ for those who receive z_0. This assumption seems to be strong since it is not likely to have $h_i > 0$ for all nearly zero bequest receivers, considering that all their ancestors had $h_0 = 0$ as productivity(which is already a very rare case), but it seems to be the sole case where Z_h becomes exogenous.

divided up by more inheritors while the per capita output is not affected by a higher number of workers. It can be shown that, $\tau_B^\star$ is first increasing and then decreasing in p, which means that the first and third channels dominate the second and fourth when p is relatively low, and then they become dominated when p becomes higher. We can also show that $\tau_B^\star$ is at first decreasing and then increasing in γ, with $\tau_B^\star \to 100\%$ when $\gamma \to 0$. However, a higher γ does not always lead to a lower $\tau_B^\star$ since it can result in too much inheritance in the economy, and the benefit from joy of giving is less than that of an enhanced redistribution of inheritance. For reasonable value of primitive parameters, the Rawlsian optimal inheritance tax rate with $\sigma = 1$ and $e_B = 0$ is approximately 60%. See Appendix (8.3.2) for the numerical example of p and γ's effect on $\tau_B^\star$.

Alternative welfare criteria: This Rawlsian social welfare criterion is nevertheless not highly suitable for our model, since the features of shocks associated with lifespan and productivity does not lead to a majority of inheritors who receive negligible inheritance. The social criterion focusing on the inheritors receiving the average normalized inheritance $E(z) = 1$ is not advisable either, since the top inheritors raises the overall level while the majority of inheritors receive less than the average level of inheritance, which leads to an enormous inheritance subsidy[16]. Intuitively, it is due to the fact that top inheritors tend to have a huge joy of giving (there is no taste shock in our model, so the amount of bequest they leave is highly related to their inheritance received), and subsiding inheritance heavily allows to increase the overall social welfare. Thus, we calculate the optimal inheritance tax rate by focusing on inheritors of different positions in the long run distribution $\theta(z)$. Note that the existing inheritance tax rate τ_B plays a role in the level of $b_y, \tilde{x}, e_B$, as well as the transition function of z_{ti} to affect $\theta(z)$, implying different level of $Z_h|z_i = z$ for different existing tax rate τ_B. We show that when the rank of inheritors that the social planner cares about increases (from bottom to top inheritors), the optimal inheritance tax rate τ_B^* decreases. Besides, a higher existing τ_B means higher optimal τ_B^* to implement. For example, with an existing $\tau_B = 40\%$, if the social planner gives a positive welfare weight to those ranked at 5% of the normalized inheritance distribution while gives zero weight to all others, the optimal inheritance tax rate τ_B^* to implement is approximately 20%. See Appendix(8.3.3).

6 Conclusion

In this master thesis we used a classic two-period overlapping generations model and incorporated the death risk in order to study the effect of uncertain lifespan on optimal inheritance taxation. We illustrated the relationship between the degree of death risk and macro variables such as

[16]As we show in section (8.3.1), Z_h in this case is close to 1 and induces a too large subsidy which breaks up the budget constraint since τ_L can be higher than 100% and lead to negative inheritance for some individuals. We exclude these extreme scenarios since inheritance must be positive in our model.

capital stock and inheritance-output ratio. We find that the effect of death risk is likely to be non monotonic since it involves different driving forces that go in opposite directions. We studied the first-best case where all accidental inheritance is eliminated and the intended inheritance taxation is a simple tool that induces the golden rule level of capital stock. The more realistic second-best analysis allowed us to look into the tax problem when the government can no longer distinguish between intended and accidental bequest. We find that the twofold shocks in lifespan and productivity give rise to a transition function of inheritance that corresponds to a Kesten process converging to a limit distribution of inheritance with a Pareto upper tail. With reasonable parameter values, we showed that a higher inheritance tax rate can reduce the inverted Pareto coefficient and mitigate the inequality caused by the top inheritors, when the maximum steady state level of inheritance tends to infinity. The second-best optimal inheritance taxation we obtained is written in variables whose levels can be either observed or estimated using empirical data. This gives an advantage in determining the optimal tax rate for practical policies since the tax formula is clear and tractable for policy makers. Nevertheless, it gives rise to an issue of endogeneity because some statistics hinge on the inheritance tax rate. To deal with this endogeneity, we first used exogenous existing inheritance tax rate to fix the level of these statistics, and they compute the new optimal tax that is supposed to be implemented. Furthermore, we tried to compute the optimal inheritance tax rate that incorporates all endogeneity and reaches an equilibrium where the existing tax rate equals the optimal tax rate. The effects of death risk on the optimal inheritance tax rate goes through the variance of the shock resulted from it. We found that higher variance is related to a higher tax rate, as they are both induced by the evolution in surviving rate.

Many papers in the existing literature obviated the complexity from the accumulated effect of lifespan uncertainty in a dynamic framework, either because of difficulty in obtaining analytical solution or from preference for short term policy design. To make our model tractable, we adopt a small open economy setting in the second-best study and focus on the long run distribution of inheritance to maximize social welfare function. We obtained an optimal inheritance taxation formula analytically but we need to use numerical methods to study its properties in most of the case. The contribution of this study is that it incorporates death risk in a two-period OLG model that allows for accumulative effect of shocks and brings implications in terms of inequality, which can serve as a continuation for purely analytical papers such as Michel and Pestieau(2002), even though we changed some model settings to make the problem tractable.

Nevertheless, the limits in this study are also noticeable. For concentrating on the lifespan shock, we did not take into account the shock in bequeathing tastes, i.e., the heterogeneity in altruism degree γ. As γ in that case can be zero, it will engender a higher heterogeneity in the distribution of inheritance at the limit, with a large number of zero bequest receivers. As we can see, our model does not allow zero bequest receivers, we only have inheritors with negligible inheritance and they are not the majority in the long run distribution. This compromise in terms of shock types leads to the fact that the Rawlsian social welfare criterion may not be the most convenient in our case. Therefore, for alternative welfare criteria where the planner focuses on subgroup of inheritors with relatively higher inheritance, the optimal tax rate tends to be lower or even negative. Another reason for obtaining relatively lower inheritance tax is that the elasticity of inheritance with regard to tax is endogenous and takes substantial values in our model, especially if the tax rate is not too high, implying more distortion from inheritance tax increase. For future study, it would be useful to incorporate more shocks to make it more realistic and comprehensive, in order to achieve a long run distribution that corresponds better to the real data. Also, we need to consider how to improve the model in order to obtain a value of elasticity which matches empirical estimates that are much lower.

In addition, the tax formula is not defined for $Z_h > 1$, which is associated with a high productivity heterogeneity and high existing inheritance tax rate. For overcoming this issue, we may imagine incorporate more tax instruments to reduce the burden of inheritance taxation and have a more realistic assumption on the distribution of productivity. Moreover, we could try to implement specific rates for different categories of inheritors, which would give implications in terms of progressivity of inheritance taxation. Finally, the basic two-period OLG model facilitates the work but it is not really capable to produce a meaningful tax formula that can be directly used with real numbers and give concrete policy recommendations. To do that, it is necessary to incorporate a realistic age structure, ideally using a continuous-time OLG model, such as in Piketty(2010).

In conclusion, this master thesis illustrates how an uncertain lifespan may have comprehensive effects on the optimal inheritance taxation, by distinguishing different bequest motives in the first-best analysis and by focusing on the accumulative process of total inheritance and its long run distribution in the second-best study. In the end we find the reasonable result that a higher death risk does affect the optimal inheritance taxation, with τ_B^* highly related to the variances of shock coming from the uncertain lifespan.

7 Appendix 1: Remarks and proofs

7.1 Confiscatory taxation on accidental inheritance

Intuitively, individuals cannot react to the consequence brought about by a sudden death, which leads to a “common sense” that tax on the accidental bequest is not distortionary: it causes no efficiency cost and enhances the equality. However, it depends on the preference of the individuals. If the individuals take accidental bequest in their utility obtained from joy of giving, the result might change. Firstly, it is straightforward to show that an accidental bequest is capable to change the behavior of individuals when their intertemporal elasticity $\sigma \neq 1$, in other words, when their income effect and substitution effect do not cancel each other entirely, even though the social planner distinguishes perfectly accidental bequest from intended bequest. We can formulate their preference as follows:

$$U_{it} = u(c_{ti}) + p\beta u(d_{t+1i}) + \gamma u(x_{t+1i}(1-\tau_x)) + (1-p)\gamma u(\frac{d_{t+1i}}{n+1}(1-\tau_A)) \tag{7.1.1}$$

where τ_A is the specific tax rate applied to accidental bequest. The last term implies that individual i has $(1-p)$ to leave an accidental bequest and she values it with the same degree of altruism γ as with intended bequest. The budget constraints write as usual:

$$\begin{cases} c_{ti} + s_{ti} = b_{ti} + w_t h_{ti}(1-\tau_L) \\ d_{t+1i} + (n+1)x_{t+1i} = s_{ti}(1+r_{t+1}) \end{cases} \tag{7.1.2}$$

where b_{ti} is the net-of-tax inheritance which can be written as $b_{ti} = x_{ti}(1-\tau_x) + \mathbb{1}_{t-1i}\frac{d_{ti}}{n+1}(1-\tau_A)$ with $\mathbb{1}_{t-1i}$ indicates the lifespan of her parent. we obtain the following FOCs:

$$\begin{cases} u'(c_{ti}) = p\beta u'(d_{t+1i})(1+r_{t+1}) + (1-p)\gamma u'(\frac{1-\tau_A}{n+1}d_{t+1i})\frac{1-\tau_A}{n+1}(1+r_{t+1}) \\ \gamma u'(x_{t+1i}(1-\tau_x))(1-\tau_x) = p\beta u'(d_{t+1i})(n+1) + (1-p)\gamma u'(\frac{(1-\tau_A)}{n+1}d_{t+1i})\frac{1-\tau_A}{n+1}(n+1) \end{cases} \tag{7.1.3}$$

As we can see, for general CIES utility function, τ_A plays a role in the individual’s decision. For instance, a higher τ_A implies a relatively lower d_{t+1i} when the substitution effect dominates the income effect, and vice versa. Particularly, with a logarithm utility function for accidental bequest, τ_A cancel out and it does not affect the agent’s behaviors. Thus, when the agent values accidental bequest, for joy of giving, or as a compensation for not enjoying the second-period consumption, an accidental bequest taxation is distortionary except for a logarithm utility function. Nevertheless, valuing accidental bequest is not a necessary condition for its efficiency cost. According to Blumkin and Sadka(2004), when agents do not value accidental bequest but their labor supply is endogenous, a 100% accidental bequest taxation is not always optimal. The redistribution from a confiscatory accidental bequest will affect the labor supply decision of the agents. In other words, even if there is no substitution effect, there is an income effect when agents' income is changed by the redistribution. Also, the extent of redistribution from other tax instruments will be reduced when the redistribution of inheritance is enhance. Thus, the confiscatory accidental bequest taxation should be considered in a more complex way with its indirect effects, and to warrant it we need its marginal benefits to outweigh its costs.

7.2 Inheritance and lifespan history in section 3.1

We can write the transition function of the intended bequest for a certain dynasty i:

$$x_{t+1i}(n+1) = \frac{p\beta+\gamma}{1+p\beta+\gamma}(w_t h_{ti} + b_{ti})\frac{\gamma}{\gamma+p\beta}(1+r_{t+1}) \tag{7.2.1}$$

where b_{ti} can be written as $b_{ti} = x_{ti} + \mathbb{1}_{t-1i}\frac{d_{ti}}{n+1}, \forall t \geq 0$. $\mathbb{1}_{t-1i}$ indicates the lifespan of generation $t-1$ following a Bernoulli distribution: it is equal to 1 with probability $(1-p)$ and 0 with probability p. Plugging the intratemporal condition $d_{ti} = x_{ti}(n+1)\frac{p\beta}{\gamma}$, the transition function (7.2.1) can be written as:

$$x_{t+1i}(n+1) = \frac{\gamma}{1+p\beta+\gamma}(w_t h_{ti} + (1+\mathbb{1}_{t-1i}\frac{p\beta}{\gamma})x_{ti})(1+r_{t+1}) \tag{7.2.2}$$

And the intended inheritance x_{t+1i} received by generation $t+1$ writes:

$$x_{t+1i} = \frac{\gamma(1+r_{t+1})}{(1+p\beta+\gamma)(n+1)}(1+\mathbb{1}_{t-1i}\frac{p\beta}{\gamma})x_{ti} + \frac{\gamma(1+r_{t+1})}{(1+p\beta+\gamma)(n+1)}w_t h_{ti} \tag{7.2.3}$$

Equation (7.2.3) shows that grandchild's intended inheritance relies on the lifespan of grandparent explicitly (and recursively on the lifespan of all other older ancestors): to concentrate on the random process of lifespan, assume that $r_{t+1} = r, w_t = w, \forall t$, and $h_{ti} = 1, \forall t, i$, equation (7.2.3) can be rewritten as:

$$x_{t+1i} = (A + \mathbb{1}_{t-1i}B)x_{ti} + C$$

where $A = \frac{\gamma(1+r)}{(1+p\beta+\gamma)(n+1)}$, $B = \frac{p\beta(1+r)}{(1+p\beta+\gamma)(n+1)}$, and $C = \frac{\gamma(1+r)w}{(1+p\beta+\gamma)(n+1)}$. Obviously, A, B, C are all positive. Denote $\Omega_{t-1i} = A + \mathbb{1}_{t-1i}B$, which can take randomly two values, depending on the lifespan of the generation $t-1$:

$$\Omega_{t-1i} = \begin{cases} A+B & \text{with probability (1-p)} \\ A & \text{with probability (p)} \end{cases}$$

Hence, we have the expression of x_{t+1i} relating to the lifespan history until that of generation $t-1$:

$$x_{t+1i} = x_{0i}(\Omega_{t-1i}\Omega_{t-2i}\Omega_{t-3i}\dots\Omega_{-1i})+C(1+\Omega_{t-1i}+\Omega_{t-1i}\Omega_{t-2i}+\dots+\Omega_{t-1i}\Omega_{t-2i}\Omega_{t-3i}\dots\Omega_{-1i}) \tag{7.2.4}$$

It can be seen that, if $t \to +\infty$, by the law of large number, the coefficient of x_{0i} converges to a certain value for every dynasty:

$$\Omega_{t-1i}\Omega_{t-2i}\Omega_{t-3i}\dots\Omega_{-1i}|_{t\to+\infty} = A^{p(t+1)}(A+B)^{(1-p)(t+1)}$$

However, the coefficient of C shows that, even though after an unlimited number of generations, this coefficient does not converge.

7.3 Transition function of capital in section 3.2

From equation (3.2.2), the capital market equilibrium writes:

$$(1+n)k_{t+1} = \bar{s}_t \tag{7.3.1}$$

where $\bar{s}_t$ is the average saving in period t. This can be written as follows:

$$\bar{s}_t = pE(s_{ti}|P) + (1-p)E(s_{ti}|O) \tag{7.3.2}$$

where $E(s_{ti}|P)$ is the average of saving of individuals whose parent lived for two periods (P for "parented"); $E(s_{ti}|O)$ is the average of saving of individuals whose parent lived for only one period (O for "orphan"). The first term can be further written as:

$$\begin{aligned} E(s_{ti}|P) &= E((w_t h_{ti} + x_{ti})\tfrac{p\beta+\gamma}{1+p\beta+\gamma}|P) \\ &= \tfrac{p\beta+\gamma}{1+p\beta+\gamma}(w_t E(h_{ti}|P) + E(x_{ti}|P)) \\ &= \tfrac{p\beta+\gamma}{1+p\beta+\gamma}(w_t + \bar{x}_t) \end{aligned} \tag{7.3.3}$$

w_t and $\bar{x}_t$ are average wage and intended inheritance for generation t respectively. They are independent of whether their parent of generation $t-1$ survives: neither the productivity nor the intended inheritance depends on the parent's lifespan. As we showed before, x_{ti} depends on the lifespan history until that of the grandparent from generation $t-2$.

Similarly, the second term can be developed as:

$$\begin{aligned} E(s_{ti}|O) &= E((w_t h_{ti} + x_{ti} + \tfrac{d_{ti}}{n+1})\tfrac{p\beta+\gamma}{1+p\beta+\gamma}|O) \\ &= \tfrac{p\beta+\gamma}{1+p\beta+\gamma}(w_t E(h_{ti}|O) + E(x_{ti}|O) + E(\tfrac{d_{ti}}{n+1}|O)) \\ &= \tfrac{p\beta+\gamma}{1+p\beta+\gamma}(w_t + \bar{x}_t + \tfrac{\bar{d}_t}{n+1}) \end{aligned} \tag{7.3.4}$$

where $\bar{d}_t$ is the average second-period consumption of generation $t-1$. Note that d_{ti} is the planned second-period consumption, realized or not, its amount does not depend on whether this person survives: if she survives, then it is realized in her second life-period; otherwise it becomes the accidental bequest.

Hence we have for the average saving at period t:

$$\bar{s_t} = \frac{p\beta+\gamma}{1+p\beta+\gamma}(w_t + \bar{x_t}) + (1-p)\frac{p\beta+\gamma}{1+p\beta+\gamma}\frac{\bar{d_t}}{n+1} \tag{7.3.5}$$

In case "P" where generation t do not receive accidental inheritance, the relationship between x_{t+1i} and x_{ti} is:

$$x_{t+1i} = (w_t h_{ti} + x_{ti})\frac{\gamma}{1+p\beta+\gamma}(1+r_{t+1})\frac{1}{n+1}$$

Aggregating on all parented individuals at t:

$$E(x_{t+1i}|P) = (w_t + \bar{x_t})\frac{\gamma}{1+p\beta+\gamma}(1+r_{t+1})\frac{1}{n+1} \tag{7.3.6}$$

In case "O" where generation t receives an accidental part of inheritance, the relationship is:

$$x_{t+1i} = (w_t h_{ti} + x_{ti} + \frac{d_{ti}}{1+n})\frac{\gamma}{1+p\beta+\gamma}(1+r_{t+1})\frac{1}{n+1}$$

Aggregating on all orphans at t:

$$E(x_{t+1i}|O) = (w_t + \bar{x_t} + \frac{\bar{d_t}}{1+n})\frac{\gamma}{1+p\beta+\gamma}(1+r_{t+1})\frac{1}{n+1} \tag{7.3.7}$$

Then the average intended bequest writes:

$$\begin{aligned} \bar{x_{t+1}} &= pE(x_{t+1i}|P) + (1-p)E(x_{t+1i}|O) \\ &= (1+r_{t+1})\tfrac{\gamma}{\gamma+p\beta}\tfrac{1}{n+1}\underbrace{(\frac{p\beta+\gamma}{1+p\beta+\gamma}(w_t + \bar{x}_t + (1-p)\frac{\bar{d_t}}{n+1}))}_{\bar{s_t}} \\ &= (1+r_{t+1})\tfrac{\gamma}{\gamma+p\beta}k_{t+1} \end{aligned}$$

Thus we have:

$$\bar{x_t} = (1 + r_t)\frac{\gamma}{\gamma + p\beta}k_t = \frac{\gamma}{\gamma + p\beta}f'(k_t)k_t \tag{7.3.8}$$

The same applies to d_t. In case "P'" when the grandparent lived two periods:

$$E(d_{ti}|P') = (\bar{x_{t-1}} + w_{t-1})(1 + r_t)\frac{p\beta}{1 + p\beta + \gamma} \tag{7.3.9}$$

In case "O'" when the grandparent lived one period:

$$E(d_{ti}|O') = (\bar{x_{t-1}} + w_{t-1} + \frac{\bar{d_{t-1}}}{1 + n})(1 + r_t)\frac{p\beta}{1 + p\beta + \gamma} \tag{7.3.10}$$

because the intended inheritance x_{t-1i} and productivity h_{t-1i} of the parent does not depend on the lifespan of the grandparent, and the planned second-period consumption d_{t-1i} of the grandparent do not rely on the lifespan of the grandparent. Then the average planned second-period consumption $\bar{d_t}$ of generation $t-1$ is:

$$\begin{aligned} \bar{d_t} &= pE(d_{ti}|P') + (1-p)E(d_{ti}|O') \\ &= (1 + r_t)\tfrac{p\beta}{\gamma + p\beta}\underbrace{(\frac{p\beta + \gamma}{1 + p\beta + \gamma}(w_{t-1} + \bar{x_{t-1}} + (1-p)\frac{\bar{d_{t-1}}}{n+1}))}_{\bar{s_{t-1}}} \\ &= \tfrac{p\beta}{\gamma + p\beta}f'(k_t)k_t(n+1) \end{aligned} \tag{7.3.11}$$

Thus the transition function of k_t writes:

$$\begin{aligned} (n+1)k_{t+1} &= \bar{s_t} = \tfrac{p\beta+\gamma}{1+p\beta+\gamma}(w_t + \bar{x_t}) + (1-p)\tfrac{p\beta+\gamma}{1+p\beta+\gamma}\tfrac{\bar{d_t}}{n+1} \\ &= \tfrac{p\beta+\gamma}{1+p\beta+\gamma}(f(k_t) - f'(k_t)k_t + \tfrac{\gamma}{\gamma+p\beta}f'(k_t)k_t) + (1-p)\tfrac{p\beta+\gamma}{1+p\beta+\gamma}f'(k_t)k_t\tfrac{p\beta}{\gamma+p\beta} \\ &= \tfrac{1}{1+p\beta+\gamma}[(p\beta+\gamma)f(k_t) - p^2\beta f'(k_t)k_t] \end{aligned} \tag{7.3.12}$$

which is equivalent to equation (3.3.2.1).QED.

We can make a comparison with the result of Michel and Pestieau(2002), section 2.3, where the authors modeled the death risk in an alternative way: individuals live for a random fraction θ_i of the second period of life. It can be shown that their transition function $(1+n)k_{t+1} = \tilde{s}(f'(k_{t+1}))[f(k_t) - \delta\bar{\theta}k_t f'(k_t)]$ is in fact a general expression of (3.3.2.1), with $\bar{\theta}$ as the expected second-period lifespan which plays the same role as the surviving rate p in our model, and δ the share of saving devoted for the second-period consumption which is equivalent to $\frac{p\beta}{p\beta+\gamma}$ with our logarithm utility function.

7.4 Exclusion of bequest in section 4.1

If the social planner includes bequest in the first-best social welfare function, the SWF writes:

$$
\begin{aligned}
SWF = {} & pqu(c_{LH}) + p(1-q)u(c_{LL}) + \beta pqu(d_{LH}) + \beta p(1-q)u(d_{LL}) \\
& + (1-p)qu(c_{SH}) + (1-p)(1-q)u(c_{SL}) + \gamma pqu(x_{LH}) + \gamma p(1-q)u(x_{LL}) \\
& + \gamma(1-p)qu(s_{SH}) + \gamma(1-p)(1-q)u(x_{SL})
\end{aligned} \tag{7.4.1}
$$

Subject to the same resource constraint (4.1.2), because it is formed by the fact that output is equal to consumption and investment in this model. Bequest is a pure transfer without any real counterpart. The FOCs w.r.t. intended bequest give that $u'(x) = 0 \iff x \to \infty$, $\forall x \in \{x_{LH}, x_{LL}, x_{SL}, x_{SH}\}$. It means that the individuals can just transfer wealth to each other and increase the social welfare, which is not a reasonable result. In virtue of the debate concerning bequest in the SWF, Michel and Pestieau(2004) defined the utility from leaving bequest as $v(x)$ and they had a term for bequest in the SWF $\epsilon v(x)$, with $\epsilon \in [0,1]$ measuring the degree of laundering out of joy of giving in the social welfare function: $\epsilon = 1$ means no laundering out while $\epsilon = 0$ implies full laundering out. They assumed that $v(x)$ satisfies $v'(x^*) = 0$ with x^* the optimal level of intended bequest, which averted the infinite bequest in our logarithm preference.

As for the second-best, since we focus on the indirect utility of the agents, we naturally need to incorporate the joy of giving in the social welfare function. Nevertheless, this incorporation may bring about a lower bequest tax rate because of double counting of the bequest in utility of parent and child.

7.5 Pareto upper tail in section 5.5

7.5.1 Proof for Champernowne's equation

For the normalized inheritance z_{ti} whose stationary distribution $\theta(z)$ has a upper tail following a Pareto distribution, we have $Pr(z_{ti} > z) = 1 - F(z) = C_0 z^{-a}$ for a large z, where C_0 is a parameter and a the Pareto coefficient. Thus for z_{t+1i} we have:

$$
\begin{aligned}
Pr(z_{t+1i} > z) &= Pr(\xi_{ti} Q z_{ti} + \xi_{ti}(\tfrac{1}{1+(1-p)M} - Q) h_{ti} > z) \\
&= Pr(z_{ti} > \frac{z - \xi_{ti}(\frac{1}{(1+(1-p)M} - Q) h_{ti}}{\xi_{ti} Q}) \\
&= C_o E((\frac{z - \xi_{ti} h_{ti}(\frac{1}{1+(1-p)M} - Q)}{\xi_{ti} Q})^{-a})
\end{aligned} \tag{7.5.1.1}
$$

When z is very large, the term $\xi_{ti}h_{ti}(\frac{1}{(1+(1-p)M} - Q)$ becomes negligible and:

$$\begin{aligned} Pr(z_{t+1i} > z) &= C_o E((\frac{z}{\xi_{ti}Q})^{-a}) \\ &= C_0 E(z^{-a}(\xi_{ti}Q)^a) \\ &= C_0 z^{-a} E((\xi_{ti}Q)^a) \end{aligned} \tag{7.5.1.2}$$

Therefore, the only way that the distribution $\theta(z)$ is stationary is that $E((\xi_{ti}Q)^a) = 1$. QED.

7.5.2 Proof for Van der Wijk's law

As long as the biggest value of the random coefficient of z_{ti} is superior to 1, ($(1 + M)Q > 1$ in our case), the Champernowne's equation determines a unique $a > 1$, which can be used to calculate $b = \frac{a}{a-1}$. The average inheritance of those whose inheritance is at least z times the average inheritance level of the economy $E(z_i) = 1$ writes:

$$\begin{aligned} E(z_{ti}|z_{ti} \geq z) &= \frac{\int_{z_{ti}\geq z} z_{ti} f(z_{ti}) dz_{ti}}{\int_{z_{ti}\geq z} f(z_{ti}) dz_{ti}} \\ &= \frac{\int_{z_{ti}>z} z_{ti}^{-a} dz_{ti}}{\int_{z_{ti}>z} z_{ti}^{-a-1} dz_{ti}} \\ &= \frac{\frac{1}{1-a}[z_{ti}^{1-a}]_z^\infty}{\frac{-1}{a}[z_{ti}^{-a}]_z^\infty} = \frac{a}{a-1} z = bz \end{aligned} \tag{7.5.2.1}$$

where $f(z_{ti}) = F'(z_{ti}) = C_0 a z_{ti}^{-a-1}$. Thus, for any given z at the Pareto upper tail, the average normalized inheritance equals to the subgroup's minimum level z multiplied by the inverted Pareto coefficient b. This is called Van der Wijk's law. (See Dickinson(1940), Van der Wijk(1939)).

8 Appendix 2: Numerical examples

8.1 Basic relationships

Calibration for primitives: To begin with, note that the interest rate r, per capita output growth rate g, discount factor β, population growth rate n are per generation rates. By setting the annual interest rate at 4%, annual per capita output growth rate at 2%, yearly discount factor at 0.98, annual population growth rate at 0.5%, and length of a generation equal to 35 years, it is straightforward to obtain the generational value of these parameters written in table 2. The value of γ, β are calibrated according to Nishiyama(2000) when $\sigma = 2$. Besides,

Primitive	Value
r	3.0552
g	1.0138
β	0.4931
n	0.1912
α	0.36
γ	0.50
σ	2.00
τ	0.30

Table 2: Primitives' values

the surviving rate p is 0.8 to fix the idea when we concentrate on the effect of other variables, such as τ_B. Similarly, τ_B is set as given values when we deal with the effect of surviving rate.

In this section, we will present the simulated results concerning the endogeneity of b_y, e_B, as well as the relationship between b_y and p.

8.1.1 Inheritance-output ratio and inheritance tax rate

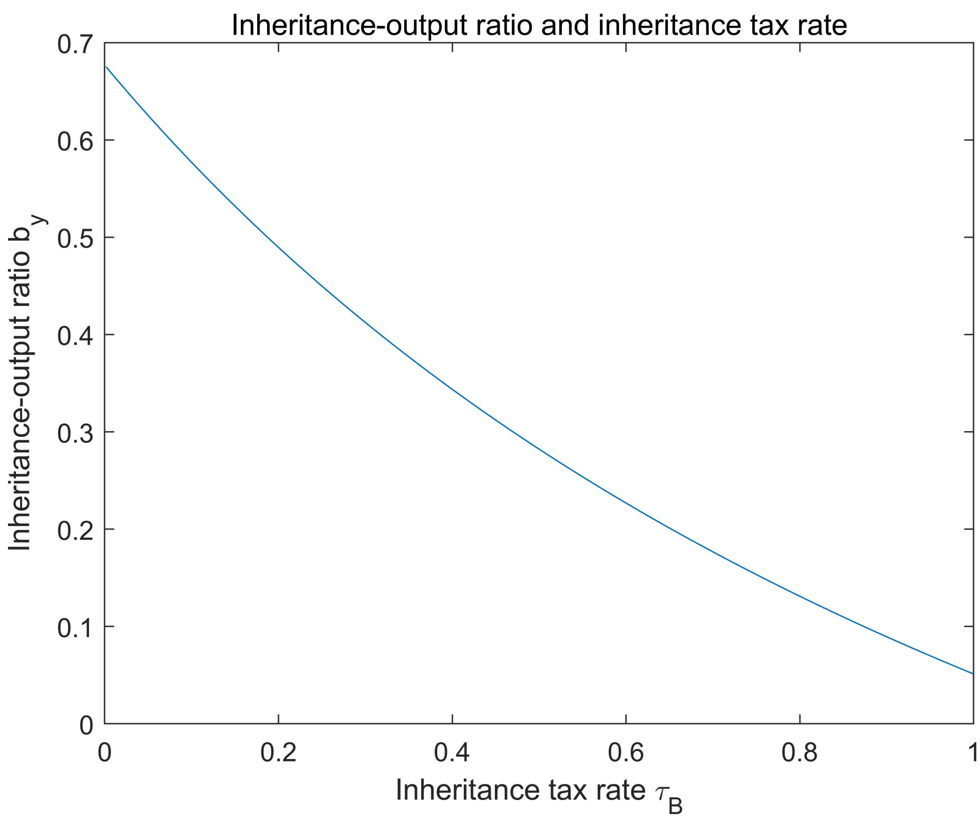

Figure 8.1.1.1: Relationship between b_y and τ_B when $\tau_B \in [0, 1)$.

$b_y(\tau_B)$ is computed using (5.3.3), (5.3.4) and (5.4.5). Note that τ_B must be strictly inferior to one, otherwise the ratio between accidental inheritance and intended inheritance for accidental bequest receivers M is not defined and b_y cannot be computed. We see clearly that b_y is decreasing in τ_B, and with a quasi confiscatory inheritance tax, b_y can be reduced to a very low level of 5%. The reason for not having $b_y \to 0$ is that $M \to \infty$ when $\tau_B \to 1$, which avoids the disappearance of b_y, but in this case the inheritance distribution will be highly concentrated.

8.1.2 Elasticity and inheritance tax rate

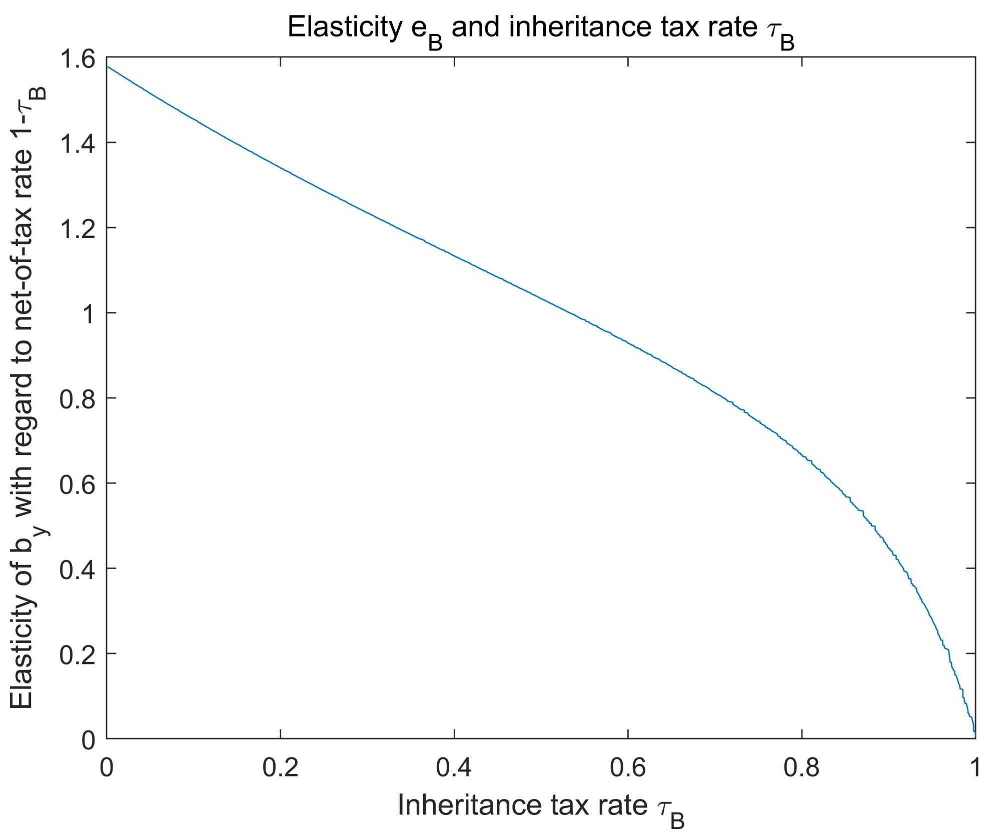

Figure 8.1.2.1: Relationship between e_B and τ_B.

The elasticity of inheritance-output ratio b_y w.r.t. the net-of-tax rate $1-\tau_B$ is always positive since $\sigma > 1$. Higher τ_B induces lower elasticity. Intuitively, a higher τ_B induces lower net-of-tax inheritance in the economy, since individuals' marginal utility from joy of giving is high in this case, they will not reduce a lot their bequest. The point here is that there is no reason to have an $e_B \to \infty$, and thus it is possible to have an optimal inheritance tax which is positive.

8.1.3 Inheritance-output ratio and surviving rate

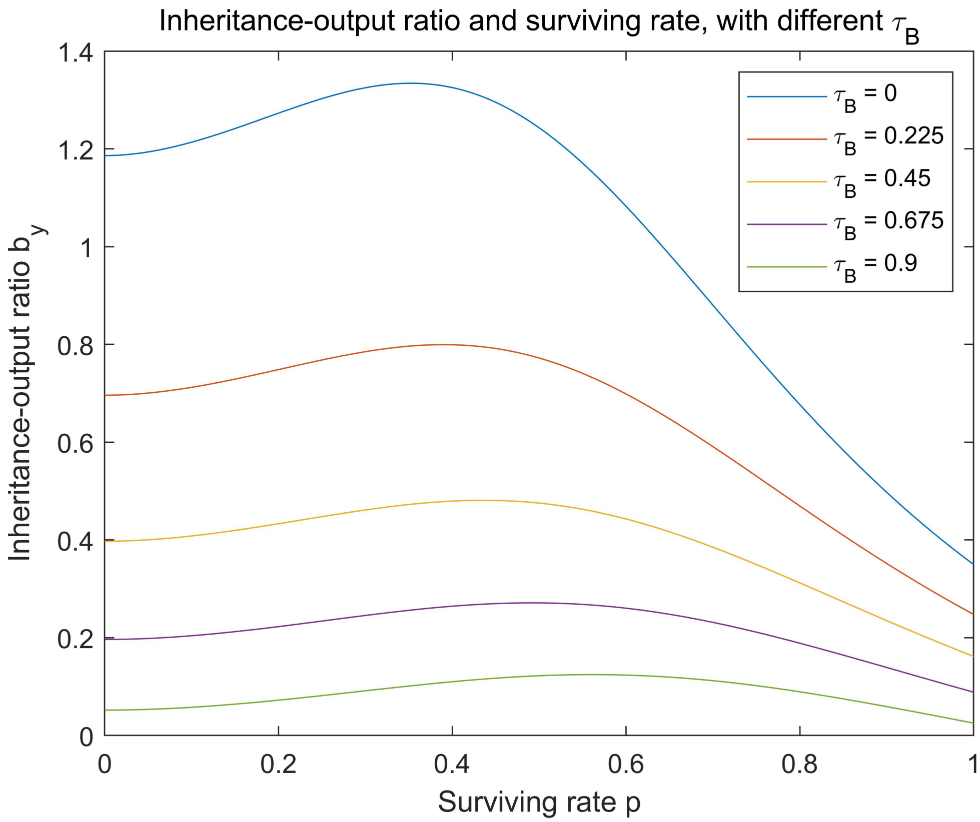

Figure 8.1.3.1: Relationship between b_y and p, for different τ_B.

We see again that higher τ_B leads to lower b_y. The effect of p on b_y is not monotonic: it increases and then decreases according to a threshold p^* which is specific to the given inheritance tax rate. Also, higher τ_B induces higher p^*. Intuitively, a higher τ_B enhances the positive effect on b_y of a higher p via a higher M. See ***Proposition 5.2.1*** that illustrates different driving forces with a logarithm utility function.

8.2 Dynamic of normalized inheritance

8.2.1 Convergence of the distribution

We set the initial distribution $\theta_0(z)$ to be uniform, with $0 < z_{0i} < 2$, $N = 100000$ as the number of dynasties, and $T = 1000$ as the number of generations. The productivity values are taken from a normal distribution $\phi(h)$ with mean $E(h) = 1$ and standard deviation S_h equal to 0.4 for the moment: $h \sim \mathcal{N}(1, 0.4)$. We set the maximum and minimum of productivity to be $h_0 = 0$ and $h_1 = 2$ to avoid negative values and keep the average at 1(In the strict sense this operation makes that h_i's no longer follow a normal distribution). To focus on the dynamic of normalized inheritance, we firstly assume that $\tau_B = 0$ which makes sure that the bigger coefficient $(1 + M)Q > 1$ and there is no upper bound for the maximum steady state value z_1.

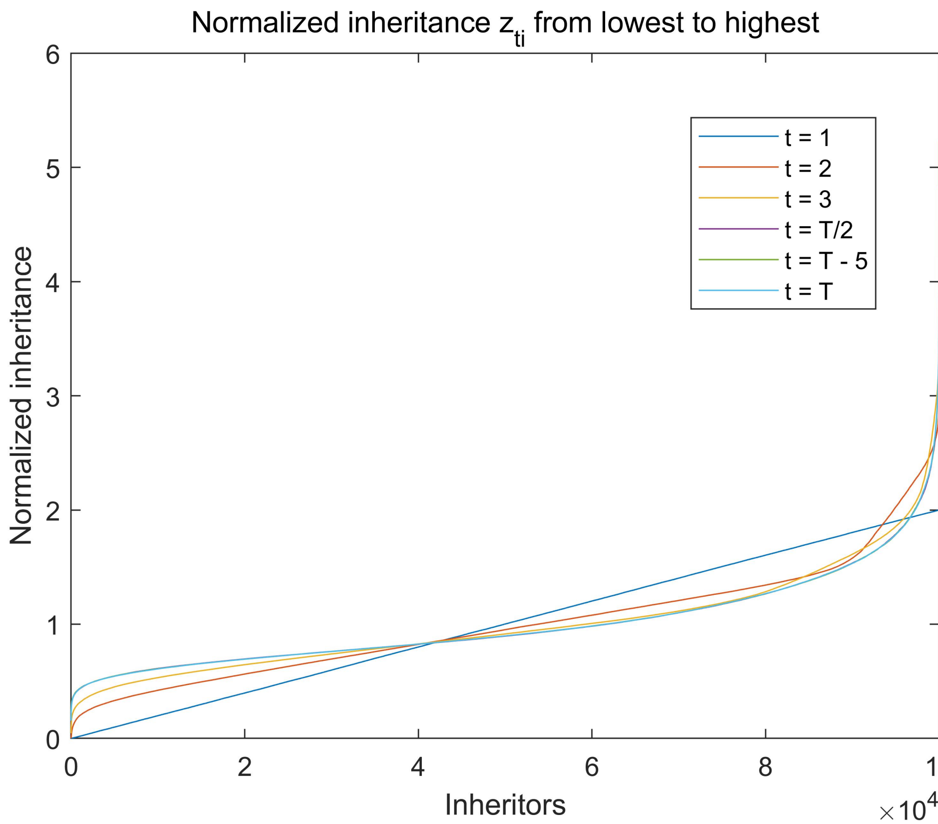

Figure 8.2.1.1: Inheritors ranked from bottom to top according to their inheritance.

The initial uniform distribution evolves very quickly to a stationary distribution, we can see that the first three periods' distributions have been moving towards a thick tail: the inheritance received by the top inheritors have increased enormously. The last three curves coincide with each other, and they display a very high normalized inheritance for a very tiny group of inheritors. However, the normalized inheritance of the bottom have also increased through the transition path. It seems that the biggest losers are those in the upper-middle "class" of inheritors. The convergence of distribution $\theta_t(z) \to \theta(z)$ is also shown by Figure 8.2.2.

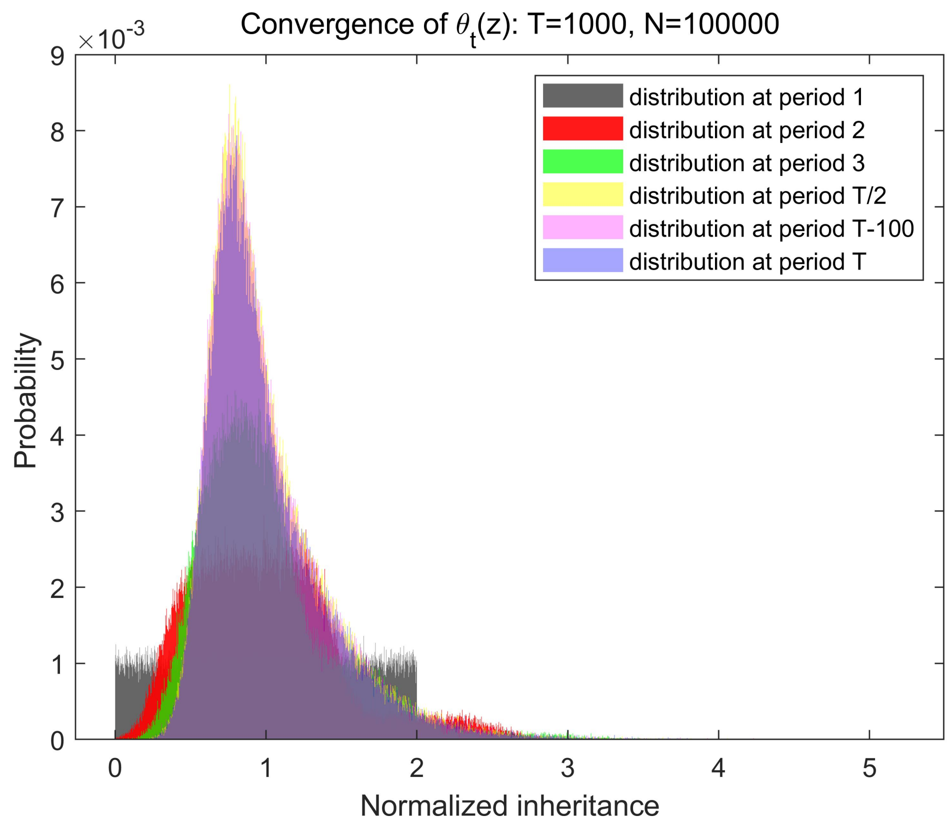

Figure 8.2.1.2: Convergence of $\theta_t(z)$ through generations.

It can be seen that more and more individuals receive an inheritance lower than the average $E(z) = 1$, and the number of top inheritors is negligible in the long run (but their inheritance is enormous). The distributions at $T/2, T - 100, T$ coincide, thus the convergence has been accomplished much before T.

8.2.2 On the Pareto upper tail

Our model does not always ensure a $Q(1 + M)$ superior to 1: from reasonable calibration, only for tiny inheritance tax rate $\tau_B \leq 0.7\%$ leads to $Q(1 + M) > 1$ and uniqueness of a that satisfies the Champernowne's equation:

$$(Q(1 + M))^a(1 - p) + Q^a p = 1 \tag{8.2.2.1}$$

For different τ_B such that $Q(1 + M) > 1$, we can have a look at their effect on the Pareto coefficient and inverted Pareto coefficient.

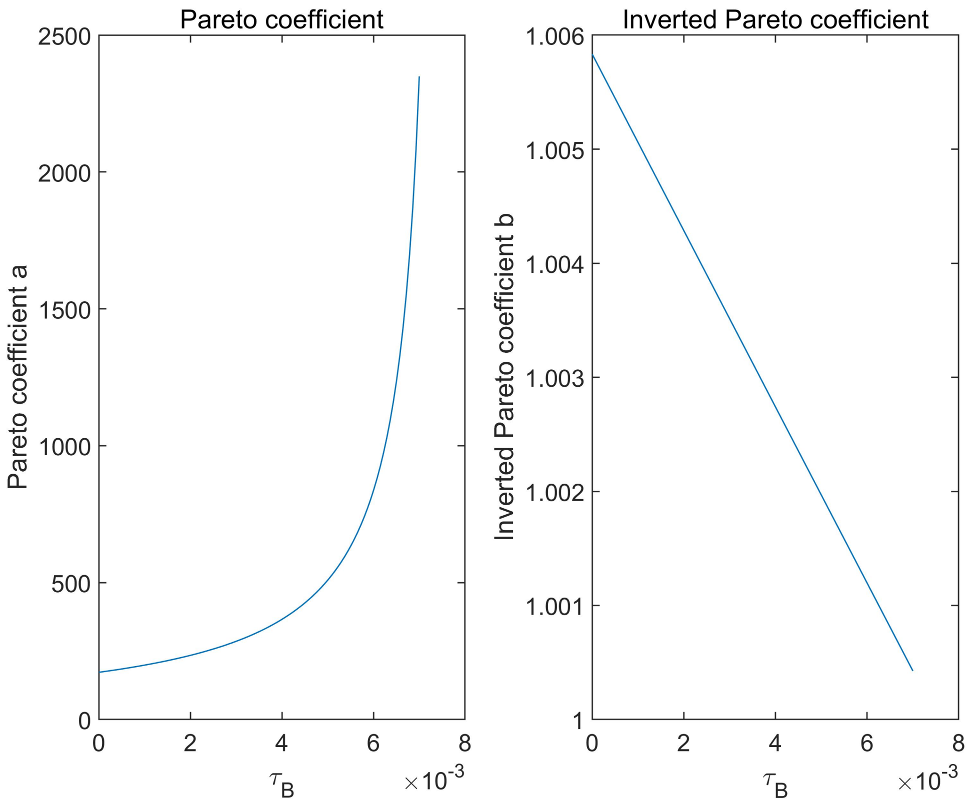

Figure 8.2.2.1: Inheritance tax rate and inheritance concentration.

Clearly, a higher inheritance tax rate induces a lower inverted Pareto coefficient, and confirms the role of inheritance taxation in mitigating inequality. Nevertheless, since a slightly higher τ_B decreases the maximum value that the shock can take, meaning that the solution of (8.2.2.1) cannot uniquely determine the Pareto coefficient, this result cannot be verified for larger τ_B(see Nirei(2009)).

8.3 Optimal inheritance tax rate computation

8.3.1 Productivity heterogeneity and optimal tax rate

As said in section (5.6), a higher Z_h induces a lower τ_B^*, and it is useful to investigate the relationship between Z_h and the dispersion in labor productivity, since intuitively a higher inequality in labor productivity should bring about a higher τ_L^* and a lower τ_B^*. Since Z_h in fact hinges on τ_B, we can see the relationship between productivity's standard deviation S_h and Z_h to verify this intuition, for different existing inheritance tax rate.

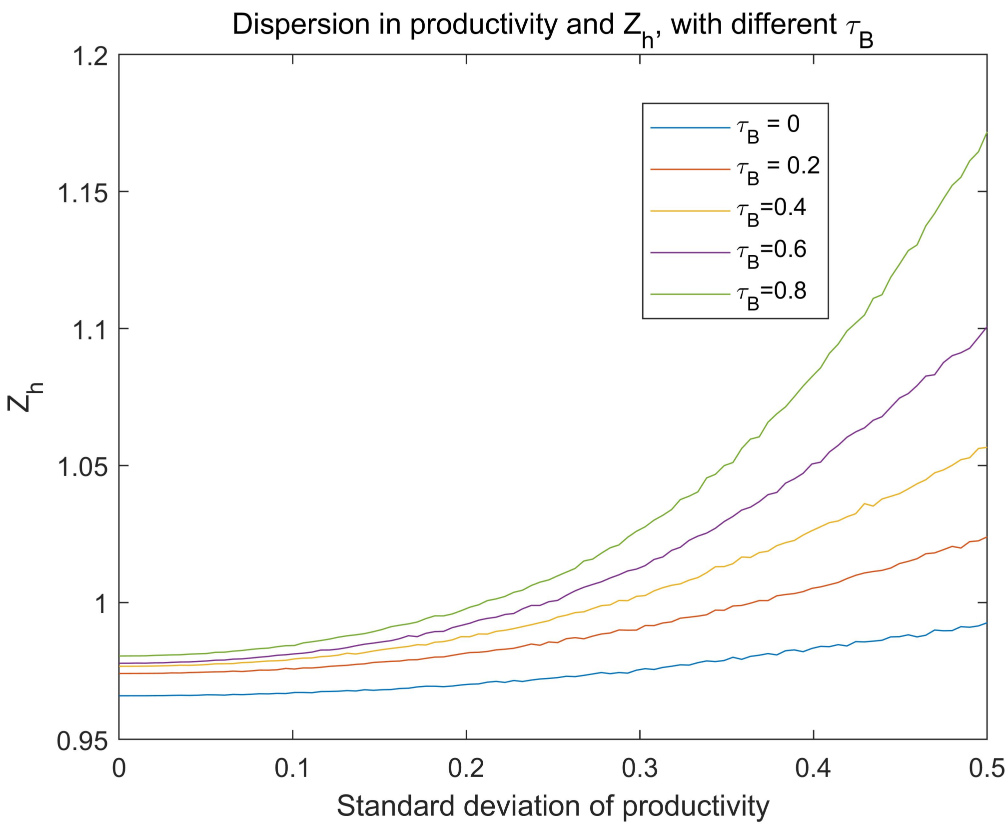

Figure 8.3.1.1: The relationship between Z_h and productivity heterogeneity.

Thus, the simulated result verifies the intuition that a higher productivity dispersion leads to a lower τ_B^* since τ_B^* is decreasing in Z_h. Note that this simulation takes into account that b_y, M,$\tilde{x}$ are all functions in the existing tax rate τ_B. Nevertheless, it is worth stressing that for high existing tax rate, it is likely to have $Z_h \geq 1$ when the standard deviation of productivity increases, which leads to $\tau_B^* > 1$ which is not defined in our model. We see that when the social planner gives a same welfare weight for every individual,i.e., $Z_h = \frac{E(V_{ci}z_i)}{E(V_{ci}h_i)}$, its value is in the neighborhood of 1, which makes τ_B^* likely to be a subsidy and increases the risk of having $Z_h \geq 1$. When the social planner focus on a certain subgroup of inheritors, for example in the Rawlsian welfare criterion where $z_i = z_0$, the value of $Z_h|_{z_i=z}$ will considerably decrease.

8.3.2 Optimal tax rate with Rawlsian welfare criterion

Firstly, when we assume logarithm utility function, $e_B = 0$ and $Z_h = 0$(if those receive $z_0 \to 0$ do not have a productivity $h_i = 0$). With the calibration for primitives in section (8.1) except for $\sigma = 1$ for this particular case, we obtain $\tau_B^\star = 0.5958$ from the formula (5.6.13). The effect of p and γ are visualized in figure (8.3.2.1) and (8.3.2.2).

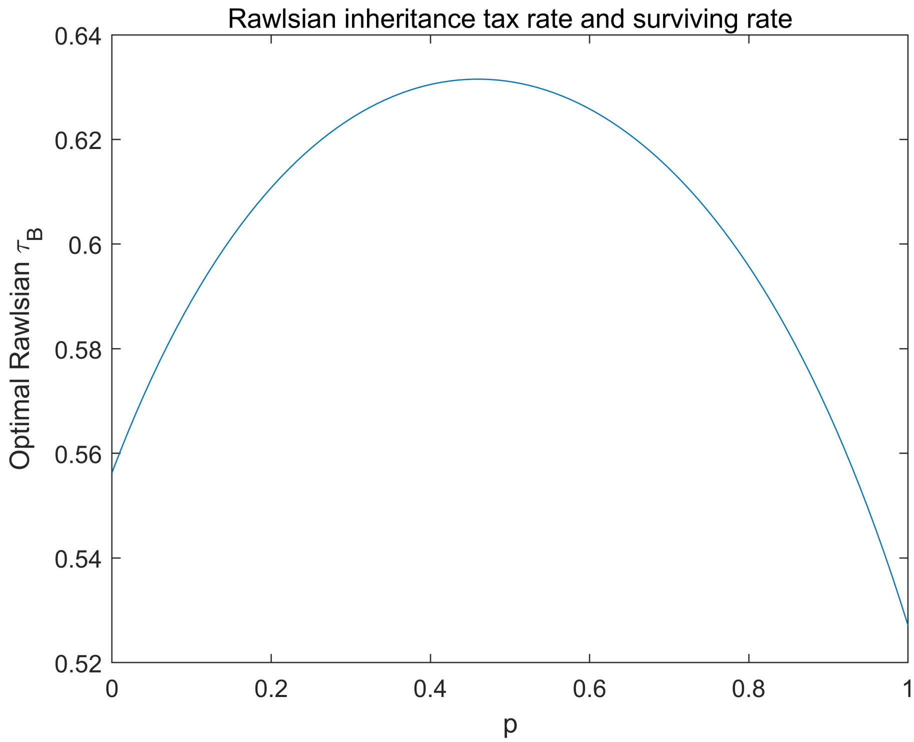

Figure 8.3.2.1: Rawlsian $\tau_B^\star$ and surviving rate p with $\sigma = 1$.

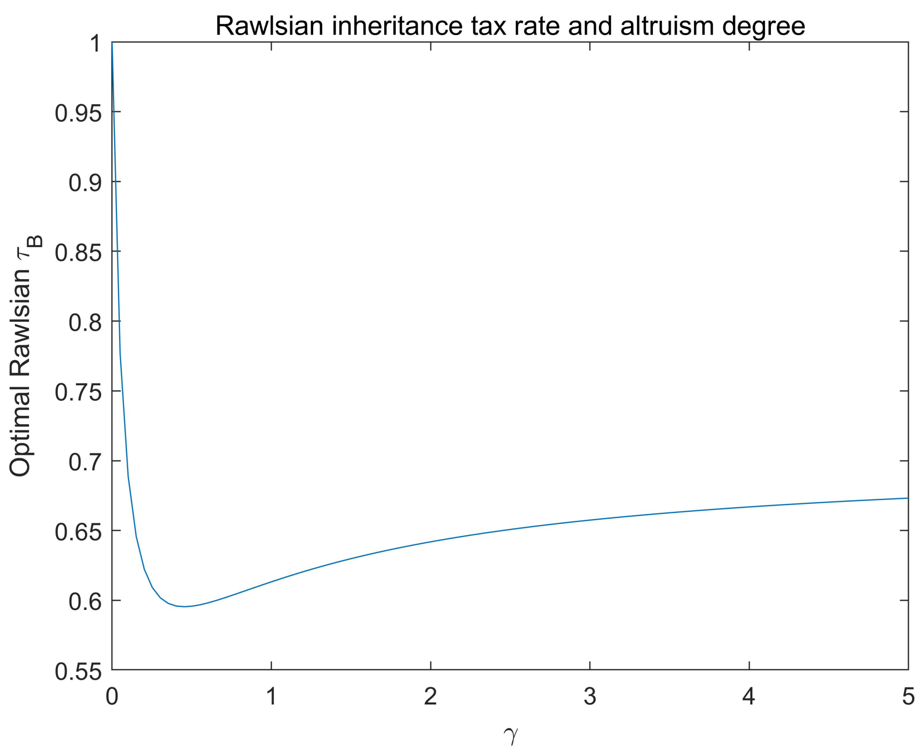

Figure 8.3.2.2: Rawlsian $\tau_B^\star$ and altruism γ with $\sigma = 1$.

8.3.3 Optimal inheritance with alternative welfare criteria

Using the population equal to 10000, standard deviation of productivity $S_h = 0.2$, we simulate the optimal inheritance tax rate τ_B^* for alternative social welfare criteria that focus on different subgroups of inheritors. Since $b_y, \tilde{x}, e_B, M, Z_h|_{z_i=z}$ are all functions in the level of τ_B that already exists in the economy, even if we do not consider their change in response to the new tax rate τ_B^*, we should use the existing tax rate to calculate their corresponding level and then determine the new tax rate τ_B^*.

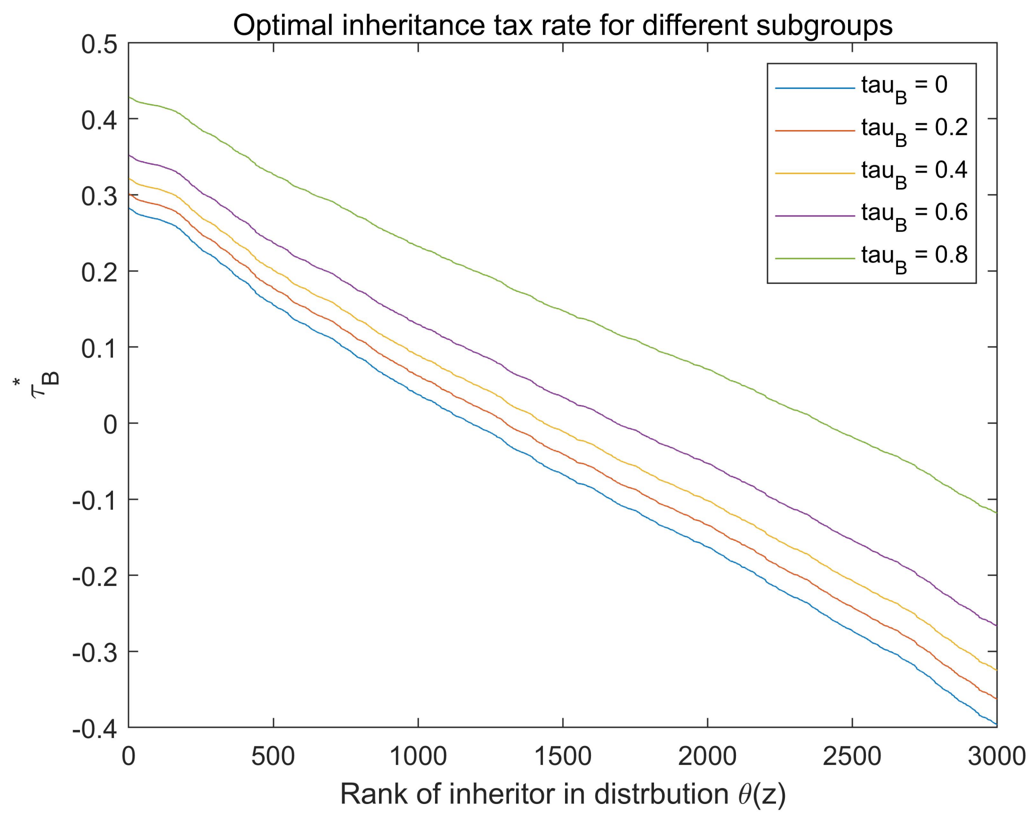

Figure 8.3.3.1: Optimal inheritance tax rate to implement for alternative welfare criteria, with different existing inheritance tax rate. The rank of inheritors is from bottom to top.

We see that a higher existing τ_B is associated with higher optimal rate τ_B^* to implement. As we saw in section (8.1.2), higher tax is related to lower elasticity, thus it creates more opportunity for a high τ_B^* to apply without bringing about too much distortion. When the social planner switches gradually her attention from inheritors at the bottom towards those in the middle class, τ_B^* decreases and it turns negative when she gives positive welfare weight for inheritors ranked at about 15% from the bottom, when b_y, e_B and other variables' level is determined by an existing $\tau_B = 40\%$. The apparently low level of τ_B^* and large likelihood to have inheritance subsidies is due to the fact that the elasticity e_B in our model is endogenous and quite large when τ_B is not too high.

8.3.4 An example of optimal tax rate at equilibrium

We show in this section an example of τ_B^{**} that satisfies equation (5.6.10). Note that we do not necessarily have meaningful solution for (5.6.10), meaning that we must have $\tau_B^{**} < 1$ and $\tau_L^* < 1$. Since the inheritance subsidies in many welfare criteria implies a too high labor income tax, we need to exclude those cases. Also, the effect of τ_B on the endogenous variables are in different directions, thus it is possible to see a higher existing τ_B increases the optimal tax rate τ_B^*, and τ_B^* changes the endogenous variables in such a way that they induce an even higher τ_B^{**}, and so on, which makes that τ_B^{**} does not exist. Nonetheless, we find that τ_B^{**} exists for a Rawlsian social welfare criterion, and it gives reasonable result.

The method for solving (5.6.10) is to calculate optimal inheritance tax rate τ_B^* for a large number of existing tax rate τ_B. Then we choose τ_B that has the slightest difference with its corresponding τ_B^{**}. The result gives $\tau_B^{**} = 31.32\%$. Note that it is only the half of the Rawlsian optimal tax rate when we use logarithm utility function ($\sigma = 1$). The reason is that logarithm utility function made the elasticity $e_B = 0$, so the inheritance tax rate could be much higher than in the case of $\sigma = 2$. This result can also be seen in Figure 8.3.3.1, when the existing τ_B is between 0.2 and 0.4, the optimal τ_B^* is also in this interval. To show the relationship between p and τ_B^{**}, we do the same for different levels of p: see Figure 8.3.4.1.

It is straightforward to write the standard deviation of the shock in inheritance associated death risk as $MQ(p(1-p))^{\frac{1}{2}}$(which comes from the transition function (5.5.3')). We can see the relationship between p and this standard deviation in Figure 8.3.4.2. By comparing the two figures, we see that the dispersion in the shocks induced by death risk has a similar relationship with p compared with the optimal inheritance tax rate. Thus, when the surviving rate leads to higher variance of shock, the associated optimal inheritance tax rate should increase, and vice versa.

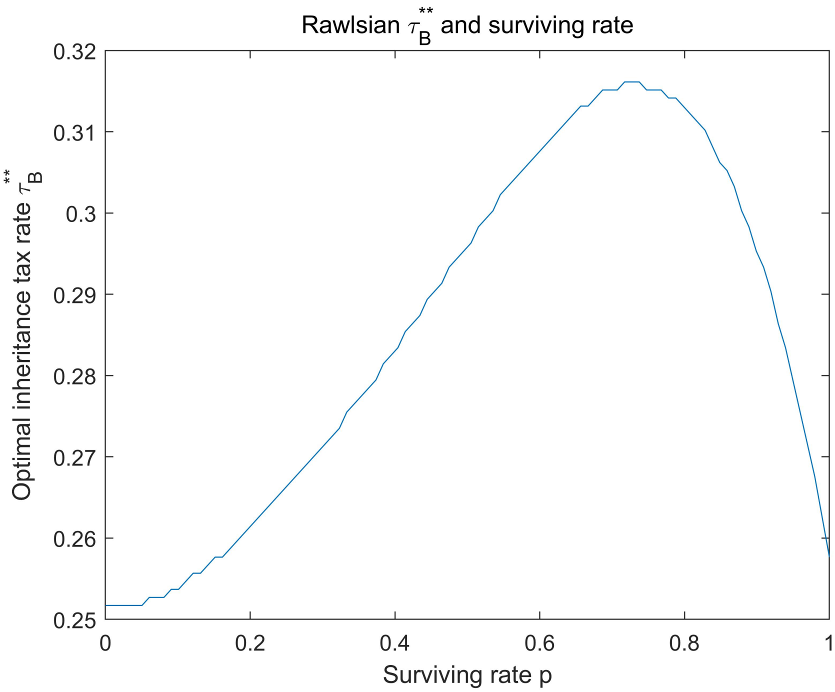

Figure 8.3.4.1: The relationship between surviving rate and the Rawlsian optimal inheritance tax rate, when taking into account all endogeneities with $\sigma = 2$.

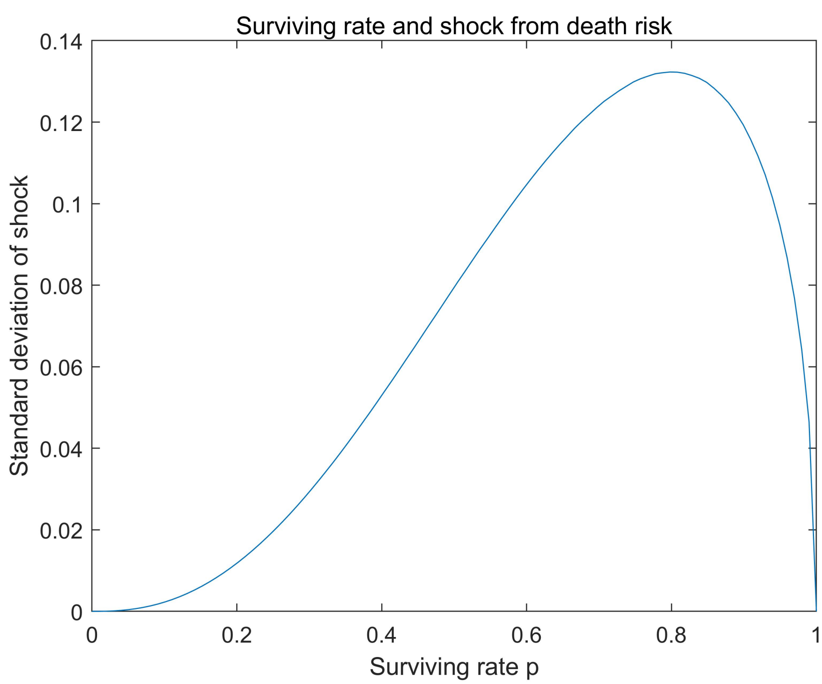

Figure 8.3.4.2: The relationship between surviving rate and the standard deviation of the shock, when taking into account all endogeneities with $\sigma = 2$.

References

[1] James Frank Gilliam. “The Minimum Subject to the vicesima hereditatium”. In: *The American Journal of Philology* 73.4 (1952), pp. 397–405.

[2] Laurence J Kotlikoff and Lawrence H Summers. “The role of intergenerational transfers in aggregate capital accumulation”. In: *Journal of political economy* 89.4 (1981), pp. 706–732.

[3] AF Shorrocks and J Davies. “The distribution of wealth”. In: *Conference on Intergenerational Transfers, Taxes and the Distribution of Wealth, Uppsala, Sweden, June*. 1999.

[4] William Gale and Joel Slemrod. *Rethinking the estate and gift tax: overview*. 2001.

[5] Thomas Piketty. “On the long-run evolution of inheritance: France 1820–2050”. In: *The quarterly journal of economics* 126.3 (2011), pp. 1071–1131.

[6] John Stuart Mill. *Principles of political economy*. D. Appleton, 1884.

[7] Helmuth Cremer and Pierre Pestieau. “Wealth transfer taxation: a survey of the theoretical literature”. In: *Handbook of the economics of giving, altruism and reciprocity* 2 (2006), pp. 1107–1134.

[8] David De La Croix, Philippe Michel, et al. *A theory of economic growth: dynamics and policy in overlapping generations*. Cambridge University Press, 2002.

[9] Tomer Blumkin and Efraim Sadka. “Estate taxation with intended and accidental bequests”. In: *Journal of Public Economics* 88.1-2 (2004), pp. 1–21.

[10] Pierre Pestieau and Motohiro Sato. “Estate taxation with both accidental and planned bequests”. In: *Asia-Pacific Journal of Accounting & Economics* 15.3 (2008), pp. 223–240.

[11] Marc Fleurbaey, M-L Leroux, Pierre Pestieau, Gregory Ponthiere, and Stephane Zuber. “Premature deaths, accidental bequests and fairness”. In: (2018).

[12] Philippe Michel, Pierre Pestieau, et al. *Wealth transfer taxation with both accidental and planned bequests*. Tech. rep. Universite catholique de Louvain, 2002.

[13] James A Mirrlees. “An exploration in the theory of optimum income taxation”. In: *The review of economic studies* 38.2 (1971), pp. 175–208.

[14] Thomas Piketty and Emmanuel Saez. *A theory of optimal capital taxation*. Tech. rep. National Bureau of Economic Research, 2012.

[15] John C Harsanyi. “A theory of prudential values and a rule utilitarian theory of morality”. In: *Social Choice and Welfare* 12.4 (1995), pp. 319–333.

[16] Peter Hammond. *Altruism in the New Palgrave: a Dictionary of Economics, J. Eatwell, M. Milgate and P. Newman, ed.* 1988.

[17] Wojciech Kopczuk and Joel Slemrod. *Dying to save taxes: evidence from estate tax returns on the death elasticity*. 2001.

[18] Harry Kesten. "Random difference equations and renewal theory for products of random matrices". In: *Acta Mathematatica* 131 (1973), pp. 207–248.

[19] Juliette Fournier. "Generalized Pareto curves: Theory and application using income and inheritance tabulations for France 1901–2012". In: *Paris School of Economics Masters thesis* (2015).

[20] Thomas Piketty and Gabriel Zucman. "Wealth and inheritance in the long run". In: *Handbook of income distribution*. Vol. 2. Elsevier, 2015, pp. 1303–1368.

[21] Philippe Michel and Pierre Pestieau. "Fiscal Policy in an Overlapping Generations Model with Bequest-as-Consumption". In: *Journal of Public Economic Theory* 6.3 (2004), pp. 397–407.

[22] H. D. Dickinson. In: *The Economic Journal* 50.198/199 (1940), pp. 274–276. ISSN: 00130133, 14680297. URL: `http://www.jstor.org/stable/2226114`.

[23] J Van der Wijk. *Inkomens-en vermogens-verdeling*. 26. De erven F. Bohn nv, 1939.

[24] Shinichi Nishiyama. *Measuring time preference and parental altruism*. Congressional Budget Office, 2000.

[25] Makoto Nirei. "Pareto distributions in economic growth models". In: (2009).