Age and Health Related Inheritance Taxation

Marie-Louise Leroux et Pierre Pestieau

Juin 2021
Age and Health Related Inheritance Taxation

Marie-Louise Leroux,
Université du Québec à Montréal, CORE et CESifo

Pierre Pestieau,
Université de Liège, CORE et Toulouse School of Economics

Document de travail No. 2021-04
Juin 2021

Département des Sciences Économiques
Université du Québec à Montréal
Case postale 8888,
Succ. Centre-Ville
Montréal, (Québec), H3C 3P8, Canada
Courriel: brisson.lorraine@uqam.ca
Site web: http://economie.esg.uqam.ca

Les documents de travail contiennent des travaux souvent préliminaires et/ou partiels. Ils sont publiés pour encourager et stimuler les discussions. Toute référence à ces documents devrait tenir compte de leur caractère provisoire. Les opinions exprimées dans les documents de travail sont celles de leurs auteurs et elles ne reflètent pas nécessairement celles du Département des sciences économiques ou de l'ESG.

De courts extraits de texte peuvent être cités et reproduits sans permission explicite des auteurs à condition de faire référence au document de travail de manière appropriée.
Age and health related inheritance taxation.

Marie-Louise Leroux*  Pierre Pestieau†

April 5, 2021

Abstract

This paper studies the design of an optimal non linear inheritance taxation when individuals differ in wage as well as in their risks of both mortality and old-age dependance. We assume that the government cannot distinguish between bequests motives, that is whether bequests result from precautionary reasons or from pure joy of giving reasons. Instead, we assume that it only observes whether bequests are made early in life or late in life, and in the latter case, whether the donor is autonomous or not.

The main result is that, under asymmetric information, in addition to labour income taxation, early bequests of the low-productivity agent should be distorted downward, that is, they should be taxed so as to relax incentive constraints.

Keywords: Bequest taxation; Long term care; Utilitarianism; Old-age dependency; Non linear taxation.


*Corresponding author. Département des Sciences Economiques, ESG-UQAM; CORE (UcL, Belgium); CESifo (Munich, Germany). E-mail: leroux.marie-louise@uqam.ca Financial support from FQRSC and SSHRC is gratefully acknowledged.

†University of Liege, CORE, and Toulouse School of Economics. Financial support from the Chaire “Marché des risques et création de valeur” of the FdR/SCOR is gratefully acknowledged.
1 Introduction

“I see nothing objectionable in fixing a limit to what anyone may acquire by mere favor of others, without any exercise of his faculties, and in requiring that if he desires any further accession of fortune, he shall work for it.” John Stuart Mill (1848)’s argument in favor of inheritance taxation has convinced a lot of thinkers over the years and yet, this form of taxation has never been as unpopular as today. Half of OECD members countries have abolished it. Among them, one finds social democratic Sweden and Norway, Canada and Austria. In OECD countries, the proportion of total government revenues raised by such taxes has fallen since the 1960s from over 1 per cent to less than 0.5 per cent (half of Europe’s billionaires have inherited their wealth). There is a puzzle over why inheritance taxes are unpopular relative to other taxes, since they are progressive and, assuming they are spent wisely on welfare goods, more people should gain than lose through inheritance taxation. One of the main reasons why inheritance tax might be unpopular is its design. This tax is plagued by too many loopholes and cases of horizontal inequity. The purpose of this paper is to address one issue concerning the design of the inheritance tax, namely whether the tax rates should vary with the age of the deceased. This question has been dealt with by Vickrey (1945), who was concerned by the fact that the tax burden was decreasing with the spacing between the occurrences of inheritance. He thus proposed to relate the tax to the number of years during which donors hold their wealth.

Our paper studies whether inheritance taxation should depend on the age of the deceased and on his health state at death, i.e. whether he lived long under autonomy or under dependency. In case of early death, inheritance comprises both a planned and an unplanned component, whereas in case of late death, it only comprises the planned component. Also, when the deceased had to go through long term care (LTC) expenditures, his estate is likely to be lower than that of someone who remained autonomous till the end. If we distinguish between three types of bequests, depending or whether they are left early, late under autonomy or late under dependency, there are opposite arguments to taxing more heavily early bequests. On the one hand, early bequests comprise unintended bequests that are known to be inelastic to taxation. But, on the other hand, they concern individuals who unluckily suffer from a short life. As to taxing late bequests of dependent people, there is a good equity argument to tax them less heavily than those who remained healthy. Those are intuitive arguments that explain why those three taxes may differ.

In this paper, we want to tackle this problem using the non linear tax approach. We consider individuals who differ in their productivity, in their survival probability as well as their probability to become dependent. Following Atkinson and Stiglitz (1980), we know that, under asymmetric

---

1OECD (2018).
2For a survey of the literature, see Cremer and Pestieau (2006).
3On accidental bequests, see Blumkin and Sadka (2003) and Cremer et al. (2012).
4This point is developed by Pestieau and Ponthiere (2021) and by Fleurbaey et al. (2019).
5Leroux and Pestieau (2021) deals with the same problem using a linear taxation approach.
information, there is usually no need for any additional tax besides optimal income taxation. This will not be the case in our setting since individuals differ not only in their productivity but, also in their risk of early death and their risk of becoming dependent. An additional tax (or a subsidy) on bequests could then be desirable depending on the relation between these demographic risks and earning. On this, the key evidence we rely on comes from Lefebvre et al. (2018), who show that the correlation between income and survival probability is positive, whereas the correlation between income and dependency probability is negative. We are then able to show that, besides income taxation, a tax on the early bequests of the low-productivity individual is desirable. This is a way to relax self-selection constraints.

The rest of the paper is organized as follows. Section 2 analyses the problem of the individuals. Section 3 deals with the first-best utilitarian problem and section 4 with the second best. A final section concludes.

2 Individuals’ problem

Our society comprises individuals with different types, indexed by $i = \{1, ..., N\}$, and each group is composed of $n_i$ individuals. They all live at most two periods, the first period with certainty but the second is uncertain. Further, this second period can be healthy or not. Each type is characterized by a wage $w_i$, a survival probability $\pi_i$ and a dependency risk $p_i$. Each individual $i$ supplies an amount of labor $l_i$, which creates some disutility $h(l_i)$. Out of his wage earnings, $y_i = w_i l_i$, each individual finances first-period consumption $c_i$, planned bequest $b_i$ and saving $s_i$ for second-period consumption. In case of autonomy, second-period consumption is denoted by $d_i$, while in case of dependency, it is denoted by $m_i$ (and it includes LTC expenses). When agents are in good health, consumption utility is denoted by $u(\cdot)$ and, it is increasing and concave. When dependent, consumption utility is denoted by $H(\cdot)$. It is increasing and concave and, such that $H(z) < u(z) \forall z$. Finally, the joy of giving utility, or equivalently, the utility obtained from leaving bequests, is denoted by $v(b^j_i)$ where $j = \{E, L, D\}$ stand for the type of bequests (Early, Late in good health, Dependent). It is increasing and concave in its argument.

Throughout the paper, we assume that the rate of interest is zero. Assuming away any market for annuities and LTC insurance, the problem of each individual $i$ amounts to maximizing the following expression:

$$EU_i = u(w_i l_i - s_i - b_i) - h(l_i) + \pi_i p_i [H(s_i) + v(b_i)] + (1 - \pi_i) v(b_i + s_i)$$

$$+ \pi_i (1 - p_i) [u(s_i - x_i) + v(x_i + b_i)]$$

(1)

where $x_i$ is the (additional) amount of saving that the healthy individual bequeathes to his heir.
(we come back on this point below). This lifetime utility can similarly be written as

$$EU_i = u(c_i) + \pi_i p_i \left[ H(m_i) + v(b^D_i) \right] + \pi_i (1 - p_i) \left[ u(d_i) + v(b^L_i) \right] + (1 - \pi_i) v(b^E_i)$$

where $b^E_i = (b_i + s_i)$, $b^L_i = (b_i + x_i)$ and $b^D_i = b_i$ denote the three different types of bequests.

Two words of clarification on this specification are in order. First, we purposely assume that there is no LTC insurance, nor annuity market. As a consequence, in case of early death, parents leave an amount $s_i$ of unplanned bequests besides the planned bequest $b_i$ and, individuals choose a saving level higher than what would be needed if $p_i = 0$. As a consequence, in case of good health in the second period, they optimally choose to leave an additional transfer $x_i$ to their children. If actuarially fair annuities and LTC insurance were available, there would be no accidental bequest $s_i$, nor any additional transfer $x_i$. In other words, within such an hypothetical setting, the only type of bequests is the intentional one denoted $b$ and the issue of differential inheritance taxes vanishes. In the real world, we only witness a partial annuitisation of retirement saving through defined benefits public or private schemes. The extent of LTC insurance, both public and private, is limited.

This implies that there is room for unintended bequests. We could introduce those partial schemes in the analysis and as long as they are taken as given, the results would be qualitatively unchanged. In any case, as long as insurance is partial, we are left with different levels of bequests in the three states of the world and, any optimal policy will try to reduce the gap between those bequests.

Second, we deliberately opted for a particular type of intended bequests, namely that, which arises from a joy of giving motivation. The modelling of the joy of giving utility is similar, for instance, to Fleurbaey et al. (2019), Glomm and Ravikumar (1992), Kopczuk and Lupton (2007) and Piketty and Saez (2013) who also study bequest taxation. Like in these papers, we will assume that the deceased cares about the amount of bequests received by his heirs, i.e. net of taxation, and not about the gross amount (i.e. before taxation). There exist alternative motivations for intended bequest: perfect and imperfect altruism and exchange (including strategic bequests). Empirically, it is not clear to assess which motivation is the most relevant. For the problem at hand, the exchange motivation does not apply and the altruistic one would imply a dynamic setting that we prefer to avoid in order to keep the analysis tractable.

We assume that the only variables that can be observed are the three different types of bequests, gross earnings $y_i = w_i l_i$ and consumption $c_i$. We will therefore express the individual’s utility in terms of these variables:

$$EU_i = u(c_i) - h \left( \frac{y_i}{w_i} \right) + \pi_i p_i \left[ H(b^E_i - b^D_i) + v(b^D_i) \right] + \pi_i (1 - p_i) \left[ u(b^E_i - b^L_i) + v(b^L_i) \right] + (1 - \pi_i) v(b^E_i)$$

---

$^6$This is the so-called LTC insurance puzzle. See Pestieau and Ponthiere (2011).
where first-period consumption can be rewritten as follows: $c_i = y_i - b_i^E$.

For further use, let us introduce a system of non linear taxes $\theta(.)$ so that we rewrite the lifetime utility as:

$$U_i = u(y_i - \theta(y_i) - b_i^E - \theta(b_i^E)) - h\left(\frac{y_i}{w_i}\right) + \pi_ip_i \left[H(b_i^E - b_i^D - \theta(b_i^D)) + v(b_i^D)\right]$$

$$+ \pi_i(1 - p_i) \left[u(b_i^E - b_i^L - \theta(b_i^L)) + v(b_i^L)\right] + (1 - \pi_i)v(b_i^E).$$

From the FOCs, we obtain the relevant marginal rates of substitution and the marginal tax rates:

$$1 - \theta'(y_i) = \frac{1}{w_i}h'(\frac{y_i}{w_i})$$

$$1 + \theta'(b_i^E) = \frac{\pi_ip_iH'(m_i) + \pi_i(1 - p_i)w'(d_i) + (1 - \pi_i)v'(b_i^E)}{u'(c_i)}$$

$$1 + \theta'(b_i^L) = \frac{v'(b_i^L)}{u'(d_i)}$$

$$1 + \theta'(b_i^D) = \frac{v'(b_i^D)}{H'(m_i)}.$$ (5)

These trade-offs will be used in order to find the levels of marginal taxation at the second-best optimum.

### 3 First-best optimum

We can now express the optimality problem of an utilitarian government as the maximization of the following Lagrangian expression in terms of the observable variables, namely the three different types of bequest, gross earnings $y_i = w_il_i$ and consumption $c_i$.

$$\mathcal{L} = \sum_i n_i \{u(c_i) - h\left(\frac{y_i}{w_i}\right) + \pi_ip_i \left[H(b_i^E - b_i^D) + v(b_i^D)\right]$$

$$+ \pi_i(1 - p_i) \left[u(b_i^E - b_i^L) + v(b_i^L)\right] + (1 - \pi_i)v(b_i^E) - \mu (c_i + b_i^E - y_i)\}$$

where $\mu$ is the multiplier associated with the resource constraint that amounts to equalizing aggregate earnings to consumption and early bequest $b_i^E = (b_i + s_i)$. 
The FOCs of this problem are:

\[ u'(c_i) = \mu \quad (6) \]
\[ h'(l_i) = \mu w_i \quad (7) \]
\[ \pi_i p_i H'(b_i^E - b_i^D) + \pi_i (1 - p_i) u'(b_i^E - b_i^L) + (1 - \pi_i) v(b_i^E) = \mu \quad (8) \]
\[ -H'(b_i^E - b_i^D) + v'(b_i^D) = 0 \quad (9) \]
\[ -u'(b_i^E - b_i^L) + v'(b_i^L) = 0. \quad (10) \]

Before interpreting these conditions, it is important to note that this first best is somehow constrained by the absence of insurance mechanisms that would cover the risk of longevity and that of dependency. With such devices, both unplanned bequests, \( s_i \), in case of early death and \( x_i \), in case of long healthy life would disappear and we would have: \( b_i^E = b_i^L = b_i^D \). Without these insurance devices, we have condition (8) that establishes an equality between the marginal utility of first period consumption \( u'(c_i) = \mu \) and the weighted average of the marginal utilities of bequests. Replacing for (9) and (10) in (8) yields

\[ b_i^E > c_i > b_i^D, \]
\[ c_i \leq b_i^L. \]

Equations (9) and (10) indicate that there is no distortion in the choice of \( b_i^L \) and \( b_i^D \).

As to the implementation of the first best, it is clear from above that only interpersonal lump sum transfers would suffice and that \( \theta'(y_i) = \theta'(b_i^L) \forall i \forall j \).

4 Second-best optimum and non linear taxation.

We now turn to the second best problem that arises from the fact that the first best solution could lead the more productive individual to mimic the less productive one \(^7\). To keep the presentation simple we restrict the analysis to a two-type model, \( N = 2 \), with \( w_2 > w_1 \). Thus, individual 2 has a higher survival probability and a lower dependence probability than individual 1.

\(^7\)Indeed, more productive agents work more than less productive individuals but obtain the same consumption.
second-best problem can be expressed by the following Lagrangian:

\[
\mathcal{L} = \sum_{i=\{1,2\}} n_i \left\{ u(c_i) - h\left(\frac{y_i}{w_i}\right) + \pi_i p_i \left[ H(b_i^E - b_i^D) + v(b_i^D) \right] \\
+ \pi_i (1 - p_i) \left[ u(b_i^E - b_i^L) + v(b_i^L) \right] + (1 - \pi_i) v(b_i^E) - \mu (c_i + b_i^E - y_i) \right\} \\
+ \lambda \left\{ u(c_2) - h\left(\frac{y_2}{w_2}\right) + \pi_2 p_2 \left[ H(b_2^E - b_2^D) + v(b_2^D) \right] + \pi_2 (1 - p_2) \left[ u(b_2^E - b_2^L) + v(b_2^L) \right] + (1 - \pi_2) v(b_2^E) \\
- \left[ u(c_1) - h\left(\frac{y_1}{w_2}\right) + \pi_2 p_2 \left[ H(b_1^E - b_1^D) + v(b_1^D) \right] + \pi_2 (1 - p_2) \left[ u(b_1^E - b_1^L) + v(b_1^L) \right] + (1 - \pi_2) v(b_1^E) \right] \right\}
\]

where \( \lambda \) is the multiplier associated with the self-selection constraint.

One easily check that there will be no distortion on the optimal trade-offs of type-2 individuals. In other words, the above first-best conditions \([11, 12]\) apply, and no distortoray taxation is needed for high-productivity individuals. As to individual 1, some of his choices will be distorted as we now see from the FOCs:

\[
u'(c_1)(1 - \lambda) = \mu \quad (11)
\]

\[-h'(\frac{y_1}{w_1}) \frac{1}{w_1} + \lambda h'(\frac{y_1}{w_2}) \frac{1}{w_2} = -\mu \quad (12)
\]

\[\pi_1 p_1 H'(b_1^E - b_1^D) + \pi_1 (1 - p_1) u'(b_1^E - b_1^L) + (1 - \pi_1) v(b_1^E) = 0 \quad (13)
\]

\[\pi_1 p_1 \left[ -H'(b_1^E - b_1^D) + v'(b_1^D) \right] - \lambda \pi_1 p_1 \left[ -H'(b_1^E - b_1^D) + v'(b_1^D) \right] = 0 \quad (14)
\]

\[\pi_1 (1 - p_1) \left[ -u'(b_1^E - b_1^L) + v'(b_1^L) \right] - \lambda \pi_2 p_2 \left[ -u'(b_1^E - b_1^L) + v'(b_1^L) \right] = 0 \quad (15)
\]

Equation \((11)\) and \((12)\) can be combined to obtain the standard taxation of labour earnings, \(\theta'(y_1) > 1\), like in Atkinson and Stiglitz (1980). Equations \((13)\) and \((14)\) imply that there should be no distortion in the choice of \(b_1^D\) and \(b_1^L\). In other words, no marginal taxation of \(b_1^D\) and \(b_1^L\) is required at the second-best optimum: \(\theta'(b_1^E) = 0\) and \(\theta'(b_1^D) = 0\).

Let us now turn to the second-best optimal choice of \(b_1^E\) and see whether it should be distorted at the second-best optimum. Using the following notation \(A = \pi_1 p_1 H'(b_1^E - b_1^D) + \pi_1 (1 - p_1) u'(b_1^E - b_1^L) + (1 - \pi_1) v(b_1^E)\) and \(B = [\pi_2 p_2 H'(b_1^E - b_1^D) + \pi_2 (1 - p_2) u(b_1^E - b_1^L) + (1 - \pi_2) v(b_1^E)]\), together with \((11)\), we can rewrite \((13)\) as follows:

\[A = u'(c_1)(1 - \lambda) + \lambda B. \quad (16)\]

Comparing it with eq. \((9)\), we obtain

\[1 + \theta'(b_1^E) = \frac{A}{u'(c_1)} = 1 - \lambda + \frac{B}{u'(c_1)} \quad (17)\]
where the first equality above shows that at the decentralized equilibrium, the marginal rate of substitution between $c_1$ and $b_1^E$ must be equal to $1 + \theta'(b_1^E)$. Depending on whether the last term in (17) is greater or smaller than 1, we will then obtain that this trade-off should be distorted upward or downward, and thus, whether early bequests should be subsidized or taxed at the second best.

Not surprisingly, if $\lambda = 0$ (that is if there was no incentive problem), there would be no need for taxation and, $\theta'(b_1^E) = 0$. Also, when the probabilities of survival and of dependency are the same for the two types of individuals, $A = B$ and $\theta'(b_1^E) = 0$. In that situation, only labour supply of the low-productivity agent should be taxed.

We now consider the case when $A$ and $B$ are different. Let us denote the difference between those two terms, by $D$, namely $D = B - A$. Replacing for (11) in (17), and rearranging terms, we obtain the following expression for the marginal tax on bequests:

$$\theta'(b_1^E) = \frac{D\lambda}{A - B\lambda}$$

where from (11) and (16), the denominator of the RHS is shown to be equal to $\mu$ and is thus positive. The sign of the tax thus depends on the sign of $D$ that can be expressed as:

$$D = (\pi_2p_2 - \pi_1p_1)H'(b_1^L - b_1^D) + (\pi_2(1 - p_2) - \pi_1(1 - p_1)) u'(b_1^E - b_1^L) + (\pi_1 - \pi_2)v'(b_1^E)$$

or equivalently,

$$D = (\pi_2p_2 - \pi_1p_1) [u'(b_1^L) - u'(b_1^D)] + (\pi_1 - \pi_2) [v'(b_1^E) - v'(b_1^L)]$$

where we used the FOCs (14) and (15). In the above equality, the first term is negative (since $\pi_2p_2 - \pi_1p_1 < 0$ and, $u'(b_1^L) - u'(b_1^D) > 0$, while the second term is positive (since $\pi_1 - \pi_2 < 0$ and $v'(b_1^E) - v'(b_1^L) < 0$), so that the sign of $D$ is a priori ambiguous.

If the probability to become dependent were not correlated to income, then the first term above would disappear and $D > 0$. In that case, early bequests of the low-productivity agent should unambiguously be taxed.

However, in the general case where there is a correlation between productivity and dependency, Lefebvre et al. (2018) show that the gap in survival probability is larger than the gap in the probability to become dependent. We also expect that $|v'(b_1^E) - v'(b_1^L)| > v'(b_1^D) - v'(b_1^L)$.

This implies that, at the second-best optimum, early bequests of the low productivity agent should be taxed at the margin. Our results are summarized in the following proposition:

**Proposition 1.** At the second-best optimum, when individuals differ in terms productivity

---

8Lefebvre et al. (2018), indeed, show that the effect of higher wealth on the probability $\pi p$ to become dependent is negative but very low.

9Necessary conditions for this inequality to hold are that $x_1 < b_1$ and $s_1 >> x_1$. 

8
• If survival probabilities and probabilities to become dependent are identical across individuals, the standard Atkinson-Stiglitz (1980) result holds: optimality can be attained by only imposing a marginal tax on the income of the low-productivity individuals.

• If individuals also differ in their demographic characteristics, in addition to income taxation, it is now optimal to also tax early bequests of the low-productivity individuals.

This proposition shows that in order to relax incentive constraints (due to the unobservability of individuals’ productivity and demographic characteristics by the government), it is optimal to tax early bequests of the low-productivity individuals, when agents differ not only in productivity but also in their survival probability and probability to become dependent. If there was no other difference than differences in productivity, only taxation of labour would be required, as in the standard model of Atkinson-Stiglitz (1980).

5 Conclusion

In this paper, we have studied the design of an optimal non-linear inheritance taxation in a setting where individuals differ in wage as well as in their risks of both mortality and old-age dependance. Assuming that they exhibit a joy of giving bequest motive and that there is no perfect annuity and LTC insurance market, we end up with three types of bequests: early, late under autonomy and late under dependency. The government cannot distinguish between bequests motives, that is whether bequests result from precautionary reasons or from pure joy of giving reasons. Instead, it only observes whether bequests are made early in life or late in life and in the latter case, whether the donor is healthy or not. In that setting, we show that in a second-best framework where the government cannot observe productivity and demographic characteristics, in addition to labour income taxation, the early bequests of the low-productivity agent should be distorted downward, i.e. they should be taxed, so as to make the problem incentive compatible.

References


