Abstract

We characterize the distribution of incomes and expected inheritances in the steady-state of a dynamic economy, when the transition from parent ability to child ability is described by a Markov chain. We show that the Atkinson-Stiglitz result on the redundancy of indirect taxes does not hold in this framework, though in the long run the differences in abilities represent the only source of heterogeneity. In particular, a bequest tax allows redistribution above the level attained by an optimal labor income tax alone. While this tax also has an adverse effect by distorting the bequest decision, a uniform tax on consumption and bequests has an unambiguously positive welfare effect. For both results the empirically validated condition that a more able individual on average has more able parents than a less able individual has to be fulfilled.

Keywords: Inheritance tax, expenditure tax, wealth transfers

JEL classification: H21; H24
1 Introduction

The tax on bequests or inheritances is still a contested issue in tax policy. Some industrialized countries have abolished the tax completely, while it is still imposed in others, though typically with large exemptions. The variety of administrative regulations indicates that the fundamental economic question of how to judge the efficiency and equity effects of the bequest tax against those of other taxes is far from being settled.

In this paper we provide further insights into the nature of the bequest tax by analyzing its long-run consequences. We follow an optimal-taxation approach and formulate a dynamic Mirrlees-type model of an economy with individuals differing in labor productivities. We ask how a shift from labor-income taxation to a bequest tax affects social welfare in the steady-state of this economy. The starting point is the well-known Atkinson-Stiglitz result derived in a static version of this model: if labor income is taxed optimally then other - indirect - taxes are redundant, given weak separability of preferences in consumption and labor. As a consequence, if leaving bequests is just seen as one form of consumption, there is no role for a tax on estates; neither is any form of capital taxation useful. In particular, an estate tax does not allow more redistribution than what is possible through an income tax alone. On the contrary, one can also take the view that even a subsidy for bequests increases welfare. The reason is the positive external effect associated with bequests, because in addition to the value they provide for the donor, they also contribute to utility of the recipient.

The situation changes if one takes into account that, as the very consequence of wealth transmission from parents to children, individuals of some generation differ not only in abilities but also in inherited wealth. In Brunner and Pech (2012ab) we have shown that in such an extended model a bequest tax indeed has a redistributive effect if there is a positive relation between inherited wealth and labor productivity; taxing bequests left by

\footnote{For instance, in Germany inheritances from spouses below 500 000 Euro are untaxed. In the US, $ 5m are exempt from the estate tax. Inheritances are completely free of tax in Sweden and Austria.}

\footnote{Unless bequests are a complement to leisure, see e.g., Gale and Slemrod (2001), Kaplow (2001).}

\footnote{See Farhi and Werning (2010), among others. Recently, Kopczuk 2013a has presented a formula for the optimal nonlinear tax on bequests which accounts for the positive externality as well as for the negative income effect of received inheritances on labor supply of children.}

1
a parent is a surrogate instrument for taxing inherited wealth received by this parent. To find the overall effect, the equalizing effect has to be weighed against the reduction of transmitted wealth and, thus, of the positive external effect.\footnote{In a related approach, Boadway et al. (2000) and Cremer et al. (2001, 2003) show that indirect taxes as well as a tax on capital income make sense if individuals differ in initial wealth, which they model as being unobservable.}

In Brunner and Pech (2012ab) we considered (unequal) inheritances as being exogenously given in some generation, and we studied the optimal tax system in the following generations. In contrast, in the present paper we analyze the welfare consequences of taxes in the steady-state equilibrium of an economy, when inheritances are completely endogenous. More precisely, we formulate a dynamic model where in each period a new generation consisting of \( n \) different ability-groups of individuals exists. Each individual is the parent of one child, who again belongs to one of the \( n \) groups in the next generation. The transition probabilities from parent ability to child ability are constant over time and can be described by a Markov chain. Assuming ergodicity of this stochastic process we consider the steady-state distribution of abilities.

Each individual uses her labor income and received inheritances for own consumption and for leaving bequests to her single child of the next generation. Bequests are motivated by joy of giving.\footnote{The alternative would be an altruistic motive resulting from dynastic preferences. We discuss the appropriateness of the bequest motive in Section 6.} Depending on the transition probabilities, the differences in bequest shares determine the distribution of inheritances over groups in the next generation. The same transition occurs in later periods and eventually leads to a steady-state distribution of inheritances, within and across ability groups. This distribution is rather diverse, as it depends on all bequest decisions of all prior generations. Still, under the assumption of affine-linear Engel curves for consumption and bequests in each period, we are able to characterize expected inheritances of each ability group in the steady-state. We then study the optimal linear income tax as the solution to maximizing expected social welfare of a generation. Whereas in the static model the optimal marginal tax rate is unambiguously positive (see e.g., Hellwig 1986, Brunner 1989), because of its redistributive effect on income, this is no longer the case in the present steady-state model. Here the redistributive
effect is even more pronounced, because it extends to bequests financed by labor income, but at the same time a negative effect on the amount of inheritances arises, due to the distortion of labor supply. Altogether, the desirability of a positive marginal tax rate is guaranteed if the labor supply elasticity is not too large.

Next, we turn to the bequest tax and ask whether its introduction increases steady-state social welfare, given an optimal income tax. The answer to this question is unclear a-priori, because on the one hand we know that the bequest tax is desirable if individuals differ in a second characteristic, namely inherited wealth (in addition to labor productivities). On the other hand, however, in the steady-state the differences in (expected) inheritances are in fact the consequence of the differences in labor productivities. Therefore, as the latter are the only (exogenous) source of heterogeneity one might suppose that again the Atkinson-Stiglitz result applies and no other tax than the (optimal) labor income tax is desirable.

Our analysis shows that this conjecture is wrong. Assuming that a positive marginal income tax rate is optimal, we also find a specific redistributive role of the bequest tax in the steady-state of the economy, if the condition that more productive individuals also receive higher expected inheritances is fulfilled. However, compared to the model with exogenous differences in initial wealth, the effect is now smaller; its size depends, among other factors, on how much inequality remains after the imposition of the optimal labor income tax, and how this is valued in the social welfare function, that is, on the degree of inequality aversion.

As a further instrument of tax policy we finally investigate the effect of a common tax on all expenditures of an individual - for consumption as well as for bequests. In a model with exogenous inheritances we have shown earlier (Brunner and Pech 2012b) that such a tax is equivalent to a specific tax on received inheritances, hence it is a preferable lump-sum instrument for redistribution. In the present steady-state model it is not lump-sum any more, because it distorts the labor supply decision and, as a consequence, the accumulation of bequests. But this distortion turns out to be smaller than the one caused by a bequest tax alone; and indeed our analysis shows that the expenditure tax has an
unambiguously positive effect on social welfare if introduced in addition to an optimal linear income tax.

For the latter result again the condition that in the steady-state more able individuals receive higher expected inheritances is crucial. This condition in fact requires that a more able individual on average has more able parents than a less able individual. That this is indeed the case can be concluded from a number of empirical studies on the correlation between parent and child income. The pioneering work by Solon (1992) and Zimmerman (1992) found an intergenerational correlation in the long-run income (between fathers and sons) of around 0.4 for the United States (suggesting less social mobility than previously believed). By now, there is a large set of estimates for a range of different countries (for a recent overview see Black and Devereux 2011 and Leigh 2007). Although these studies use different estimation methods, variable definitions and sample selections (limiting their comparability), all of them estimated a positive intergenerational correlation in earnings, ranging from 0.1 up 0.4, which corroborates the above condition.6

Our study establishes a role for taxing bequests as an instrument for increasing equality of opportunity. It is related to a recent work by Piketty and Saez (2012) who also provide a theory of optimal inheritance taxation in the steady-state of an economy. Their intention is to derive a formula for the optimal bequest tax rate in terms of estimable parameter such as the elasticity of bequests with respect to the tax. However, their approach differs substantially from the standard optimal-taxation framework employed in our paper. In particular, Piketty and Saez take the amount of government transfers as given and discuss to which extent these should be financed by a tax on bequests as opposed to a tax on labor income. Thus they do not analyze the optimum scale of redistribution which is the main objective of the work in the tradition of Mirrlees (1971). Moreover, by considering a general dynamic stochastic model they do not explicitly formulate the process of wealth transmission across ability groups. As a consequence, they do not analyze the condition

---

6In particular, these studies suggest that the correlation in the Nordic countries (Sweden, Finland, Denmark, Norway), Canada and Australia is about 0.1 - 0.2 and thus lower than in the U.S. and U.K. Interestingly, Jäntti et al. (2006) find that the greatest cross-country differences arise at the tails of distribution. Moreover, these studies estimate higher correlations between son and father than between daughter and farther, while the pattern across countries is similar.
for a positive relation between average received inheritances and abilities which is crucial for our results.

In the following Section we present the basic model of individual decisions on labor supply, consumption and bequests. Section 3 provides an analysis of the dynamics of wealth transmission over generations, whose steady-state is studied in Section 4. In Section 5 the optimal tax on labor income is characterized, and the results on the desirability of additional taxes on bequests and total expenditures, respectively, are derived. Section 6 concludes.

2 The Model

In each period $t$ a population of mass one exists. It is split into $n$ groups with different abilities $w_1 < w_2 < \ldots < w_n$, whose shares are $f_1, \ldots, f_n$. That is, $f_1, \ldots, f_n$ are the probabilities of an individual to belong to the respective ability group. Individuals of a generation live for one period; they have common preferences over consumption $c_i$, labor time $l_i$, and leaving bequests $b_i$. Preferences are described by the utility function $u(c_i, b_i, l_i) = \varphi(c_i, b_i) - g(l_i)$, where $\varphi$ is a concave function, increasing in both arguments, and $g$ is strictly concave and increasing. This formulation indicates that we assume bequests to be motivated by joy-of-giving. Consumption and bequests are financed out of labor income $w_i l_i$ and received inheritances $e_i$, $i = 1, \ldots, n$.

There exists a linear tax on labor income, given by $-\alpha + \sigma w_i l_i$ with $\alpha$ as a uniform demogrant and $\sigma$ as the marginal tax rate on gross labor income. Moreover, the tax system consists of a proportional tax $\tau$ on all expenditures or a proportional tax $\tau_b$ on bequests alone. For given taxes the maximization problem of an individual $i$ reads

$$\max \varphi(c_i, b_i) - g(l_i),$$

s.t. $(1 + \tau)(c + (1 + \tau_b)b_i) \leq e_i + \alpha + (1 - \sigma)w_i l_i$, \hspace{1cm} (2)

$$c_i, b_i, l_i \geq 0,$$ \hspace{1cm} (3)
where we typically assume that \( \tau, \tau_b \) do not exist both. The first-order conditions with \( \lambda \) as the Lagrangian variable associated with (2) read as

\[
\frac{\partial \varphi}{\partial c_i} - \lambda(1 + \tau) = 0, \\
\frac{\partial \varphi}{\partial b_i} - \lambda(1 + \tau)(1 + \tau_b) = 0, \\
-g'(l_i) + \lambda(1 - \sigma)w_i = 0.
\]

From these we obtain demand functions for \( c_i \) and \( b_i \). We assume that demand can be described by linear Engel curves which need not pass through the origin. That is, \( c_i \) and \( b_i \) can be written as shares \( 1 - \gamma \) and \( \gamma/(1 + \tau_b) \), with \( \gamma \in (0, 1) \), of the available budget, given a constant \( c_0 \) representing minimum consumption:

\[
c_i = (1 - \gamma)((e_i + x_i)/(1 + \tau) - c_0) + c_0, \\
b_i = \gamma_b((e_i + x_i)/(1 + \tau) - c_0).
\]

Here \( x_i \) denotes net income of the household, \( x_i \equiv \alpha + (1 - \sigma)w_il_i \), and \( \gamma_b \equiv \gamma/(1 + \tau_b) \) describes the share of bequests in the "real" budget after deducting minimum consumption. Such a type of demand functions arises if \( \varphi \) is homogeneous in \( c_i - c_0 \) and \( b_i \). In addition, we assume that \( \varphi \) is homogeneous of degree 1. Note that \( \gamma \) itself depends on \( \tau_b \), except in case that \( \varphi \) is a "generalized" Cobb-Douglas function \( (c_i - c_0)^{1-\gamma}b_i^\gamma \), then \( \gamma \) is the constant exponent. We generally assume that the expenditures for \( c_0 \) are smaller than net income, that is, \( x_i/(1 + \tau) - c_0 > 0 \).

Indirect utility of an individual \( i \), denoted by \( V_i \), is given by evaluating \( \varphi(c_i, b_i) - g(l_i) \) at the values (7) and (8) and with \( l_i \) determined by (6). Moreover, application of the Euler Theorem to \( \varphi \) at optimal \( c_i - c_0 \) and \( b_i \) gives

\[
\varphi((1 - \gamma)(c_i + x_i)/(1 + \tau) - c_0), \gamma_b^{(c_i + x_i)/(1 + \tau) - c_0}) = \frac{\partial \varphi(\cdot)}{\partial c_i}(1 - \gamma)(c_i + x_i)/(1 + \tau) - c_0) \\
+ \frac{\partial \varphi(\cdot)}{\partial b_i}\gamma_b(c_i + x_i)/(1 + \tau) - c_0).
\]
By use of (4) and (5) and \( \gamma_b = \gamma/(1 + \tau_b) \), the RHS of (9) can be reduced to \( \lambda (e_i + x_i - (1 + \tau)c_0) \), where \( \lambda \) is the marginal utility of income, which is independent of \( w_i \) but depends on \( \tau \) and \( \tau_b \) as can be seen from (4) and (5). Indirect utility \( V_i \) is therefore given by

\[
V_i(e_i; \alpha, \sigma, \tau, \tau_b) = \lambda (e_i + x_i - (1 + \tau)c_0) - g(l_i)
\]  

(10)

where \( l_i \) is determined by (6) as \( l_i = (g')^{-1}(\lambda(1 - \sigma)w_i) \), depending on the after-tax wage rate \( (1 - \sigma)w_i \) and the marginal utility of income \( \lambda \). Note that from (6) the relation \( x_1 < x_2 < ... < x_n \) follows, because \( \lambda \) is independent of \( w_i \).

In a static context our formulation of the model - with weakly separable preferences - implies the well-known Atkinson-Stiglitz result for the optimal structure of the tax system: if initial endowments \( e_i \) are zero (or nonzero, but identical, see Brunner and Pech 2012b), then the labor income tax alone is a sufficient instrument in order to maximize social welfare; a tax on consumption or, in particular, on bequests does not increase welfare (and the same applies for a tax on expenditures). A role for the latter taxes only arises if inheritances \( e_i \) exist and vary across individuals, that is, if individuals differ in a second characteristic, in addition to abilities. As we have shown in Brunner and Pech (2012ab), these taxes have a positive welfare effect, when imposed in addition to an optimal labor income tax, if inheritances are positively correlated with abilities.\(^7\) As already mentioned in the Introduction, to derive this result inheritances were taken as exogenously given in some generation, being the result of bequests of earlier generations, without specifying this relation precisely. Now we consider the transmission of wealth explicitly and study inheritances - which arise endogenously - in the steady-state of this process. In the next Section we describe the transmission process.

3 Dynamics of wealth transmission and abilities

We assume that abilities \( w_1 < w_2 < ... < w_n \) remain constant over time (that is, over generations). Each individual has a single descendant to whom she leaves all her bequests.\(^7\) To be precise, this result was shown for an optimal nonlinear tax on labor income. It is, however, straightforward to extend the result to the case of an optimal linear tax.
There is a constant transition probability \( p_{ij} \) that the descendant of an individual with ability \( w_i \) has ability \( w_j \), with \( \sum_{j=1}^{n} p_{ij} = 1 \) for \( i = 1, ..., n \). With constant transition probabilities the dynamics of the shares of the ability groups over time represents a Markov chain. We assume the Markov chain to be ergodic and to converge to a steady-state distribution \( \pi = (\pi_1, ..., \pi_n) \). \( \pi \) is unique and independent of the initial distribution \( f \equiv (f_1, ..., f_n) \); it is determined by the equation \( \pi = \pi P \) together with the normalization \( \sum_{i=1}^{n} \pi_i = 1 \), where \( P \) is the \( n \times n \) transition matrix with entries \( p_{ij} \). As is well-known, \( \pi \) can be found by considering the limit

\[
W \equiv \lim_{t \to \infty} P^t.
\]

All \( n \) rows of \( W \) are identical and equal to \( \pi \). We assume in the following that the distribution of abilities is already in the steady-state, that is \( f_i = \pi_i, i = 1, ..., n \). Thus, \( f \) has the property

\[
f = fP. \tag{11}
\]

Next we study the transmission of bequests. For the following considerations we neglect the tax \( \tau \) on total expenditures; its effect will be studied in Section 4. Let \( \beta_{it} = \gamma_b(x_i - c_0) \) be the bequest out of own net labor income (minus minimum consumption expenditures), left by an individual with ability \( w_i \) in some period \( t \). Remember that we generally assume \( x_i > c_0 \). With a linear Engel curve for bequests, we can think of total bequests \( b_{it} \) left by some individual in period \( t \) as being split into bequest \( \beta_{it} \) originating from own net labor income, and bequests \( \gamma_b \epsilon_{it} \) originating from received inheritances. As gross and net incomes increase with the ability level, the relation \( \beta_{it} < ... < \beta_{nt} \) holds. We study the diffusion of these particular bequests over the population in the course of time.

Let \( \epsilon_{i,t+1} \) denote inheritances received in period \( t+1 \) by an individual in ability group \( i \) out of bequest initiated by the bequests \( \beta_{jt} \) of the proceeding period. Clearly, each \( \epsilon_{i,t+1} \) has possible realizations \( \beta_{it} < ... < \beta_{nt} \). Let further \( \mu_{ij} \) denote the probability that in \( t+1 \) an individual in group \( i \) has a parent in group \( j \), that is, receives inheritance \( \beta_{jt} \). Note that the possible realizations \( \beta_{jt} \) are the same for all ability groups (are independent of \( i \),
but the associated probabilities \( \mu_{ij} \) differ across groups. To determine \( \mu_{ij} \) we observe that the probability of a descendant to belong to ability group \( i \) is \( \sum_{j=1}^{n} f_{j} p_{ji} \), which is equal to \( f_{i} \) in the steady-state population. Thus, the conditional probability of a descendant in group \( i \) to have a parent of group \( j \) is

\[
\mu_{ij} = \text{Prob}(\text{parent} = j \mid \text{descendant} = i) = \frac{f_{j} p_{ji}}{f_{i}}. \tag{12}
\]

Obviously, \( \mu_{ij} \) is independent of \( t \), given the steady-state distribution of abilities. In matrix notation, where \( M \) denotes the \( n \times n \) matrix with elements \( \mu_{ij} \), we have

\[
M = \begin{pmatrix}
  p_{11} & \frac{f_{2} p_{11}}{f_{1}} & \cdots & \frac{f_{n} p_{11}}{f_{1}} \\
  \frac{f_{1} p_{12}}{f_{2}} & p_{22} & \cdots & \frac{f_{n} p_{12}}{f_{2}} \\
  \vdots & \vdots & \ddots & \vdots \\
  \frac{f_{1} p_{1n}}{f_{n}} & \frac{f_{2} p_{1n}}{f_{n}} & \cdots & p_{nn}
\end{pmatrix}. \tag{13}
\]

By definition, \( M \) is also a stochastic matrix, that is, all elements are nonnegative and \( \sum_{j=1}^{n} \mu_{ij} = 1 \), for all \( i = 1, \ldots, n \). Expected inheritances of a member of group \( i \) in period \( t + 1 \) out of bequests \( \beta_{jt} \) are then computed as

\[
E[t^{t+1}i] = \sum_{j=1}^{n} \mu_{ij} \beta_{jt} = \sum_{j=1}^{n} \beta_{jt} f_{j} p_{ji} / f_{i}. \tag{14}
\]

In matrix notation with \( E^{t,t+1} \) denoting the column vector of expected inheritances out of the \( \beta_{jt} \) and \( \beta_{t} = (\beta_{1t}, \ldots, \beta_{nt})^T \), this is written as\(^8\)

\[
E^{t,t+1} = M \beta_{t}. \tag{15}
\]

Next, with homothetic preferences, an individual in generation \( t + 1 \) who has received \( \beta_{jt} \) as inheritance out of labor income in period \( t \), again leaves some share \( \gamma_{b} \beta_{jt} \) as own bequests, and the same happens in all further periods \( s > t \). The important observation is that what is left in period \( s \) to the descendants in period \( s + 1 \) is equal to one of the

\(^8\)The upper index \( T \) denotes the transposed vector, hence \( \beta_{t} \) is a column vector.
values $\gamma_b^{s-t-1}\beta_{1t}, \ldots, \gamma_b^{s-t-1}\beta_{nt}$, irrespective of the ability groups of the individuals who were involved in the transmission process between $t$ and $s$. This is a consequence of our assumption of homothetic preferences. Altogether, bequests $\beta_{jt}, j = 1, \ldots, n$ out of labor income in generation $t$ initiate what we call a bequest series $\gamma_b^{s-t-1}\beta_{jt}$, for all periods $s > t$. That is, in some period $s$ the possible realizations $\ell_i^{t,s}$ of received inheritances of an individual in ability group $i$ are $\gamma_b^{s-t-1}\beta_{jt}, j = 1, \ldots, n$, independent of $i$. Let $\kappa_{ij}^{t,s}$ denote the associated probabilities.

We study the relation between $\kappa_{ij}^{t,s}$ and $\kappa_{ij}^{t,s+1}$. In period $s + 1$ an individual in ability group $i$ inherits $\gamma_b^{s-t}\beta_{jt}$ from all ability groups of the previous generation, thus

$$
\kappa_{ij}^{t,s+1} = \frac{f_1 \kappa_{ij}^{t,s} p_{1i} + \ldots + f_n \kappa_{nj}^{t,s} p_{ni}}{f_i}.
$$

In the first term of the numerator, $f_1$ describes the probability that a parent belongs to group 1, $\kappa_{ij}^{t,s}$ describes the probability that a parent in group 1 has received inheritance $\gamma_b^{s-t-1}\beta_{jt}$, and $p_{1i}$ represents the transition probability from group 1 in period $s$ to group $i$ in period $s + 1$. The same reasoning is behind the other terms in the numerator. The denominator results from the normalization that the sum of the $\kappa_{ij}^{t,s+1}$ over $j$ must be equal to 1. Thus, the denominator reads as

$$
\sum_{j=1}^{n} (f_1 \kappa_{ij}^{t,s} p_{1i} + \ldots + f_n \kappa_{nj}^{t,s} p_{ni}) = f_1 p_{1i} \sum_{j=1}^{n} \kappa_{ij}^{t,s} + \ldots + f_n p_{ni} \sum_{j=1}^{n} \kappa_{nj}^{t,s} = f_1 p_{1i} + \ldots + f_n p_{ni} = f_i,
$$

where the last equality comes from the steady-state property of the ability distribution.
Written in matrix notation with $K^{t,s}$ denoting the matrix with elements $\kappa_{ij}^{t,s}$:

$$
K^{t,s+1} = \begin{pmatrix}
\frac{1}{f_1}(f_1 p_{11}, \ldots, f_n p_{1n}) & \kappa_{11}^{t,s} & \cdots & \kappa_{1n}^{t,s} \\
\frac{1}{f_1}(f_1 p_{11}, \ldots, f_n p_{1n}) & \kappa_{11}^{t,s} & \cdots & \kappa_{1n}^{t,s} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{1}{f_n}(f_1 p_{n1}, \ldots, f_n p_{nn}) & \kappa_{n1}^{t,s} & \cdots & \kappa_{nn}^{t,s}
\end{pmatrix}
= MK^{t,s},
$$

where $M$ is defined in (13) above.

As a consequence, we find for the vector $E^{t,s+1}$ of expected inheritances $E^{t,s+1}$ in period $s+1$, stemming from initial bequests $\beta_{jt}$:

$$
E^{t,s+1} = K^{t,s+1}(\gamma_b^{s-t} \beta_{1t}, \ldots, \gamma_b^{s-t} \beta_{nt})^T = \gamma_b M E^{t,s}.
$$

This formula tells us that when considering a bequest series $\gamma_b^{s-t-1} \beta_{jt}$, the same matrix $M$, found above to determine expected inheritances in the first descendant generation, also determines the dynamics of expected inheritances over all later periods. In fact, applying the recursive formula (16) $r$ times in turn gives us

$$
E^{t,s+1} = \gamma_b^r M^r E^{t,s+1-r},
$$

and setting $r = s - t$ we arrive at

$$
E^{t,s+1} = \gamma_b^{s-t} M^{s-t} E^{t,t+1} = \gamma_b^{s-t-1} M^{s-t-1} \beta_{1t},
$$

where the latter equality follows from (15).

So far we have characterized the evolution of a bequest series $\gamma_b^{s-t-1} \beta_{jt}$, for all periods
$s > t$, initiated by labor income of an individual $j$ in some period $t$. Considering total inheritances occurring in some period $s$, these obviously can be seen as the result of all bequest series initiated in earlier periods. That is, all possible realizations of inheritances $e_{is}$, received by an individual $i$ in period $s$, are given by

$$e_{is} = \sum_{r=1}^{\infty} \gamma_{b}^{r-1} \beta_{j,s-r};$$

for all possible values $\beta_{j,s-r} = \gamma_{b} (\alpha + w_{j} (1 - \sigma) l_{j} - c_{0})$, where $l_{j}$ is optimal labor supply of group $j$. Labor supply is constant over time (independent of the received inheritances, as can be seen from (4), (5) and (6)), given that the tax rates $\sigma, \tau, \tau_{b}$ do not change. Observe that again the possible realizations are identical for all receiving ability groups $i$. The size of inheritances belonging to a specific bequest series diminishes over time by the factor $\gamma_{b}$; total inheritances are thus dominated by those originating from labor income of more recent generations. We need not specify the probabilities of the respective realizations, but we note that the expected value $E[e_{is}]$ is finite because it is smaller than the largest realization $\sum_{r=1}^{\infty} \gamma_{b}^{r-1} \beta_{n,s-r} = \gamma_{b} (x_{n} - c_{0})/(1 - \gamma_{b})$, a finite value.

4 The steady-state

We have assumed from the beginning that the population is constant and the distribution of abilities is in the steady-state, that is, the shares $f_{i}$ of all groups remain constant over time. The overall steady-state of this economy is defined by the condition that expected inheritances of each ability group also remain constant over time. Note that because of our assumption of a population of mass one, split into $n$ ability groups, expected inheritances are equal to realized inheritances; there is no aggregate uncertainty. With $E[b_{i}]$ and $E[e_{i}]$ denoting expected bequests and inheritances, respectively, of group $i$ in the steady-state, we have

$$E[e_{i}] = \frac{f_{1} p_{1i} E[b_{1}] + \ldots + f_{n} p_{ni} E[b_{n}]}{f_{i}}, \quad i = 1, \ldots, n, \quad (17)$$
with (we still neglect the expenditure tax \( \tau \))

\[
E[b_j] = \gamma_b(E[e_j] + x_j - c_0).
\]  

(18)

We rewrite (17) in matrix notation with \( E[e] \) denoting the column vector of steady-state expected inheritances, \( E[e] \equiv (E[e_1], ..., E[e_n])^T \):

\[
E[e] = \gamma_bM \begin{pmatrix} E[e_1] + x_1 - c_0 \\ \vdots \\ E[e_n] + x_n - c_0 \end{pmatrix}
\]

and introducing the column vector \( x \) of steady-state net incomes \( x_j = \alpha + w_j(1 - \sigma)l_j \) as well as \( c_m \equiv (c_0, ..., c_0)^T \) we arrive at the steady-state condition

\[
E[e] = \gamma_bM(E[e] + x - c_m)
\]  

(19)

or

\[
E[e] = \gamma_b(I - \gamma_bM)^{-1}M(x - c_m),
\]  

(20)

where \( I \) denotes the unit matrix. Note that \( (I - \gamma_bM)^{-1} = \sum_{t=0}^{\infty}(\gamma_bM)^t \), thus (20) can be written as

\[
E[e] = \sum_{t=1}^{\infty} \gamma_bM^t(x - c_m),
\]  

(21)

which suits well to our earlier result that inheritances in some period consist of bequests out of labor income in all past periods.

Before we turn to the optimal tax problem we discuss a condition on \( M \) which will turn out to be important in the next section:

**Definition 1** A \( n \times n \) matrix \( Z \) fulfills condition (C), if for any column vector \( y \in \mathbb{R}^n \) with increasing components, \( y_1 < y_2 < ... < y_n \), the column vector \( Zy \) also has increasing components.

For an interpretation of this condition in our framework let some vector \( x \) of steady-
state net incomes of the \( n \) groups in the descendant generation be given and consider the parents of each descendant group. By definition, \( Mx \) describes mean parent net incomes for each group and condition (C) states that, given increasing incomes in each generation, mean parent income is increasing as well. As it is known from empirical studies that parent and descendant income is positively correlated, it seems indeed very plausible that condition (C) is fulfilled in reality. On observes immediately from (21) that whether \( M \) fulfills these condition is also essential for the question of how incomes and inheritances are related. Let in the following \( A \) denote the matrix \( A \equiv \gamma_b(I - \gamma_bM)^{-1}M. \)

**Lemma 1** \( A \) has nonnegative elements and the components of each of its row vectors sum up to \( \gamma_b/(1 - \gamma_b) \). Moreover, if \( M \) has property (C), also \( A \) has this property; then expected inheritances are increasing with abilities.

**Proof.** We know that \( A \) can be written as \( \sum_{t=1}^{\infty}(\gamma_bM)^t \). Elementary properties of matrix multiplication show that in each term of the infinite series the row vectors have components summing up to \( \gamma_b^t \) (remember that components of the rows of \( M \) add up to 1), which gives \( \sum_{t=1}^{\infty}\gamma_b^t = \gamma_b/(1 - \gamma_b) \). Multiplying by \( y \), each term in the infinite series consists of the product \( \gamma_b^tM \cdot M \cdot M \cdot \ldots \cdot My \) (where \( M \) occurs \( t \)-times). Clearly then, if \( My \) has increasing components, the \( M^ty \) has increasing components as well, and this remains true after multiplication with \( \gamma_b^t \) and summing up the terms. \( \blacksquare \)

A sufficient condition on \( M \) to fulfill (C) can be given:

**Lemma 2** Let the \( \mu_{ij} \), the elements of \( M \) have the following properties:

\[
\mu_{ii} > \mu_{ji} \quad \text{for any } i, j = 1, \ldots, n, i \neq j \quad (22)
\]

\[
\mu_{ij} \geq \mu_{i+1,j} \quad \text{for } j < i \leq n - 1 \quad (23)
\]

\[
\mu_{ij} \leq \mu_{i+1,j} \quad \text{for } 1 \leq i \leq j - 2 \quad (24)
\]

Then \( M \) fulfills condition (C).

**Proof.** Consider two subsequent rows of \( M \), \( M_i \) and \( M_{i+1} \). Construct \( \tilde{M}_i \) from \( M_i \) by reducing the first \( i \) components of \( M_i \) to those of \( M_{i+1} \) and increasing the component
appropriately so that the sum of all components remains 1. Because of increasing $y_i$ this gives $\tilde{M}_i y > M_i y$. Note that in case that all the remaining components $i + 2, \ldots, n$ of $\tilde{M}_i$ are equal to those of $M_{i+1}$, we have $\tilde{M}_i = M_{i+1}$. Otherwise, in a second step, if one or more of the remaining components $i + 2, \ldots, n$ of $\tilde{M}_i$ are larger than that of $M_{i+1}$, increase the these components to those of $M_{i+1}$ and decrease the $i + 1$-component so that the sum of all components remains 1. The resulting vector is $M_{i+1}$ and because of increasing $y_i$ we have $M_{i+1} y > \tilde{M}_i y > M_i y$. Altogether, we have $M_{i+1} y \geq \tilde{M}_i y > M_i y$. ■

In fact, the property (C) is a condition on the transition matrix $P$, from which $M$ is derived (see 13). A general condition on $P$ which implies (C) for $M$ is not easily found. Instead, we present two special cases of $P$ such that $M$ is equal to $P$.

Lemma 3 Let $P$ have one of the two properties:

(i) $p_{ij} = 0$ for $|i - j| \geq 2$

(ii) $p_{ij} = c_i$ for $j \neq i$. Then the probability matrix $M$ is equal to the transition matrix $P$.

Proof. (i) Assume that $P$ is a tridiagonal matrix, with nonzero elements only in the main diagonal and in the two neighboring ones. By the steady-state condition (11) on $f$ we have

$$f_1(1 - p_{12}) + f_2 p_{21} = f_1, \text{ or } f_2 / f_1 = p_{12} / p_{21},$$

which implies (see (13)) $\mu_{12} = f_2 p_{21} / f_1 = p_{12}$ as well as $\mu_{21} = p_{21}$. Similarly, $f_1 p_{12} + f_2 (1 - (p_{21} + p_{23})) + f_3 p_{32} = f_2$ gives, after substituting for $f_1$, $f_3 / f_2 = p_{23} / p_{32}$, which in turn implies $\mu_{23} = p_{23}$ and $\mu_{32} = p_{32}$, and so on.

(ii) Next assume that in each row $i$ all elements outside off the main diagonal are equal to a constant $c_i$. Then the steady-state condition (11) reads as $f_j (1 - (n - 1)c_j) + \sum_{i \neq j} f_i c_i = f_j$, or $f_j = (\sum_{i=1}^n f_i c_i) / (nc_j)$ for any $j = 1, \ldots, n$. This implies $f_i / f_j = c_j / c_i$ and further $\mu_{ij} = p_{ij}, i, j = 1, \ldots, n$. ■

Intuitively, these are cases where (i) only switches to neighboring ability groups are possible or where (ii) switches to other ability groups are equally likely. Finally we note for later use that $M$ has the same property as $P$: 15
Lemma 4 Let $f$ be the steady-state distribution of abilities and the stochastic matrix $M$ defined as in (13). Then

$$f = fM.$$ \hspace{1cm} (25)

Proof. One observes immediately that $fM = (f_1(p_{11} + p_{12} + \ldots + p_{1n}), \ldots, f_n(p_{11} + p_{12} + \ldots + p_{1n})) = f$. ■

5 Optimal Taxes

Formula (20) characterizes steady-state inheritances in terms of labor supply $l_j, j = 1, \ldots, n$ and the share $\gamma_b$ of bequests which, in turn, depend on the tax parameters $\alpha, \sigma, \tau, \tau_b$, fixed by the government. In this section we study how the government should choose the tax parameters in order to maximize social welfare in the steady-state. In particular, we are interested in the question of whether a tax on bequests and a tax on expenditures are appropriate instruments. It was already mentioned that both are redundant in a static context with identical inheritances, as a consequence of the Atkinson-Stiglitz result. However, introducing differing exogenous inheritances as a second characteristic, in addition to varying abilities, reverses this outcome. Now, in the steady-state inheritances are again not exogenous but are determined by labor supply of the different groups. It is therefore a-priori unclear whether taxes on bequests or expenditures represent a desirable additional instrument or are redundant as in the Atkinson-Stiglitz case.

In order to set up the maximization problem of the government we note first an implication of linear homogeneity of $\varphi$: expected utility \[ \sum_{h \in H} \kappa_{ih} \varphi((1 - \gamma)((e_h + x_i)/(1 + \tau) - c_0), \gamma_b((e_h + x_i)/(1 + \tau) - c_0)) \] where $e_h$ denotes all possible realizations of $e$ with associated probabilities $\kappa_{ih}$, is equal to utility from expected inheritances \[ \varphi((1 - \gamma)(E[e_i] + x_i)/(1 + \tau) - c_0), \gamma_b((E[e_i] + x_i)/(1 + \tau) - c_0)) \]. This allows us to formulate the problem of maximizing steady-state social welfare in the following way: let $\rho$ denote the parameter

\[ \text{9}\footnote{Remember that the realizations of $e$ are the same for each ability group, but the associated probabilities differ.} \]
of inequality aversion, $\rho > 0$, then the task is

$$\max \sum_{i=1}^{n} f_i \varphi((1 - \gamma)\left(\frac{E[e_i] + x_i}{1 + \tau} - c_0\right), \gamma_b\left(\frac{E[e_i] + x_i}{1 + \tau} - c_0\right) - g(l_i)\right)^{1-\rho}/(1 - \rho),$$

(26)

subject to

$$\sigma \sum_{i=1}^{n} f_i w_i l_i + \gamma_b \sum_{i=1}^{n} f_i \gamma_b\left(\frac{E[e_i] + x_i}{1 + \tau} - c_0\right) + \tau \sum_{i=1}^{n} f_i\left(\frac{E[e_i] + x_i}{1 + \tau}\right) - \alpha \geq 0,$$

(27)

$$E[e_i] = A_i(x/(1 + \tau) - c_m),$$

(28)

where $A_i$ denotes the $i$-th row vector of the matrix $A = \gamma_b(I - \gamma_b/(1 + \tau)M)^{-1}M$ and $x$ is the column vector of net incomes as defined above, with components $x_i = \alpha + (1 - \sigma)w_i l_i$.

Note that from now on we take the expenditure tax $\tau$ into account; the formula for $A$ is derived from the steady-state condition (19), which now reads $E[e] = \gamma_b M((E[e] + x)/(1 + \tau) - c_m)$ or $E[e] = \gamma_b(I - \gamma_b/(1 + \tau)M)^{-1}M(x/(1 + \tau) - c_m)$. However, we still assume that $\tau_b$ and $\tau$ do not exist simultaneously. We start our analysis with the welfare effect of a linear income for some given tax rates $\tau_b, \tau \geq 0$. Remember that $V_i$ is indirect utility (10) and $\lambda = \partial V_i/\partial x_i$ is the marginal utility of income. We note first

**Lemma 5** If $M$ has property (C), then $V_1 < V_2 < \cdots < V_n$ and $V_1^{-\rho} > V_2^{-\rho} > \cdots > V_n^{-\rho}$.

**Proof.** By Lemma 1, Property (C) implies that inheritances are increasing with abilities, as is net income. As a consequence, social marginal utility $V_i^{-\rho}$ decreases if $\rho > 0$. ■

**Proposition 2** Let $\tau_b = \tau = 0$. The welfare effect of an introduction of a linear income tax is

$$\lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(\overline{w}l - w_i l_i) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} A_i \begin{pmatrix} \overline{w}l - w_1 l_1 \\ \vdots \\ \overline{w}l - w_n l_n \end{pmatrix} + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} A_i \begin{pmatrix} \lambda_1 \frac{\partial \lambda_1}{\partial \sigma} \\ \vdots \\ \lambda_n \frac{\partial \lambda_n}{\partial \sigma} \end{pmatrix},$$

(29)

with $\overline{w}l = \sum_{i=1}^{n} f_i w_i l_i$ as the average gross income. The first and second term are positive, given that $M$ fulfills property (C). The third term is negative.

**Proof.** We substitute (28) into (26) - (27) and get the first-order condition with respect
\( \alpha \), with \( \mathcal{L} \) as the Lagrangian and \( v \) as the Lagrange multiplier to (27):

\[
\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{n} f_i V_i^{-\rho} \left[ \frac{\partial \varphi(.)}{\partial c_i} (1 - \gamma)(\frac{\partial E[e_i]}{\partial \alpha} + 1 + (1 - \sigma)w_i \frac{\partial l_i}{\partial \alpha}) 
+ \frac{\partial \varphi(.)}{\partial b_i} \gamma_b \left( \frac{\partial E[e_i]}{\partial \sigma} + g \frac{\partial l_i}{\partial \sigma} \right) + v \right] + \sum_{i=1}^{n} f_i w_i \frac{\partial l_i}{\partial \alpha} - 1 = 0. 
\]

Using the first-order conditions (4) - (6) of the individual optimization problem, together with \( b_i = \gamma_b / (1 + \tau_b) \), and \( \partial l_i / \partial \alpha = 0 \), (30) can be transformed to

\[
\frac{\partial \mathcal{L}}{\partial \alpha} : \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left( \frac{\partial E[e_i]}{\partial \alpha} + 1 \right) - v = 0. 
\]

Similarly, the derivative of the Lagrangian with respect to \( \sigma \) is

\[
\frac{\partial \mathcal{L}}{\partial \sigma} = \sum_{i=1}^{n} f_i V_i^{-\rho} \left[ \frac{\partial \varphi(.)}{\partial c_i} (1 - \gamma)(\frac{\partial E[e_i]}{\partial \sigma} - w_i l_i + (1 - \sigma)w_i \frac{\partial l_i}{\partial \sigma}) 
+ \frac{\partial \varphi(.)}{\partial b_i} \gamma_b \left( \frac{\partial E[e_i]}{\partial \sigma} - w_i l_i + (1 - \sigma)w_i \frac{\partial l_i}{\partial \sigma} \right) - g \frac{\partial l_i}{\partial \sigma} \right] + \sum_{i=1}^{n} f_i (w_i l_i + \sigma w_i \frac{\partial l_i}{\partial \sigma}), 
\]

which can be written as

\[
\frac{\partial \mathcal{L}}{\partial \sigma} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left( \frac{\partial E[e_i]}{\partial \sigma} - w_i l_i \right) + v \sum_{i=1}^{n} f_i (w_i l_i + \sigma w_i \frac{\partial l_i}{\partial \sigma}). 
\]

Substituting for \( v \) from (31) in (33) gives after rearrangement

\[
\frac{\partial \mathcal{L}}{\partial \sigma} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (\overline{w} - w_i l_i) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left( \frac{\partial E[e_i]}{\partial \alpha} \frac{\overline{w} l_i}{\overline{w}} + \frac{\partial E[e_i]}{\partial \sigma} \right) + \\
\lambda \sigma \left( \sum_{i=1}^{n} f_i w_i \frac{\partial l_i}{\partial \sigma} \right) \sum_{i=1}^{n} f_i V_i^{-\rho} \left( \frac{\partial E[e_i]}{\partial \alpha} + 1 \right) 
\]

and, thus, we have at \( \sigma = 0 \)

\[
\frac{\partial \mathcal{L}}{\partial \sigma}_{\sigma=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (\overline{w} - w_i l_i) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left( \frac{\partial E[e_i]}{\partial \alpha} \frac{\overline{w} l_i}{\overline{w}} + \frac{\partial E[e_i]}{\partial \sigma} \right). 
\]

In view of \( \sum_{i=1}^{n} f_i (\overline{w} - w_i l_i) = \overline{w} - \overline{w} = 0 \), and \( \overline{w} - w_1 l_1 > \overline{w} - w_2 l_2 > \cdots > \overline{w} - w_n l_n \),
the first term on the RHS of (35) is positive if $M$ has the property (C), because then $V_1^{-\rho} > V_2^{-\rho} > \cdots > V_n^{-\rho}$ by Lemma 10. From the steady-state equation (28) for $E[e_i]$ we find

$$
\frac{\partial E[e_i]}{\partial \alpha} = A_i \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},
$$

(36)

$$
\frac{\partial E[e_i]}{\partial \sigma} = A_i \begin{pmatrix} -w_1 l_1 + (1 - \sigma)w_1 \frac{\partial l_1}{\partial \sigma} \\ \vdots \\ -w_n l_n + (1 - \sigma)w_n \frac{\partial l_n}{\partial \sigma} \end{pmatrix},
$$

(37)

thus at $\sigma = 0$ the second expression in (35) reads as

$$
\lambda \sum_{i=1}^{n} f_i V_i^{-\rho} A_i \begin{pmatrix} \bar{w}l - w_1 l_1 \\ \vdots \\ \bar{w}l - w_n l_n \end{pmatrix} + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} A_i \begin{pmatrix} w_1 \frac{\partial l_1}{\partial \sigma} \\ \vdots \\ w_n \frac{\partial l_n}{\partial \sigma} \end{pmatrix}.
$$

(38)

If $M$ has the property (C) then $A(w_1 l_1, \ldots, w_n l_n)^T$ has increasing components and the first term of (38) is positive, for the same reason as discussed above for the first term in (35). The second term in (38) is negative because $\partial l_i / \partial \sigma$ is negative. ■

In a static Mirrlees-model an optimal linear income tax with a positive marginal tax rate is always desirable (Hellwig (1986), Brunner (1989) among others); its introduction allows some redistribution, while its distorting effect is of second order and therefore zero at $\sigma = 0$. This result corresponds to the positive first term of (29) in Proposition 2. In addition, this Proposition shows that in the present model with endogenous inheritances the redistributive potential of the linear tax on labor income is even larger if $M$ fulfills condition (C). Then steady-state inheritances increase with abilities and their distribution is also made more equal as a consequence of the linear income tax (second term in (29)). However, in the extended model there arises now a first-order negative effect on the amount of bequests left in the steady-state (third term in (29)), resulting from the distortion of labor supply. The overall effect is undetermined and depends on the labor supply elasticity.
as well as on the ability distribution and the degree of inequality aversion. Clearly, with a low supply elasticity the direct redistributive effect (the first-term in (29)) dominates. 

Note that the first term involves gross incomes which are more unequally distributed than inheritances; the distribution of the latter is equalized by the transmission of wealth if there is some variation in ability groups over generations.

In the following we assume that a positive marginal tax rate is optimal. Its value is determined by the first-order condition (set (33) equal to zero)

\[
\lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left( \frac{\partial E[e_i]}{\partial \sigma} - w_i l_i \right) + v \sum_{i=1}^{n} f_i (w_i l_i + \sigma w_i \partial l_i / \partial \sigma) = 0.
\]

(39)

Observe that with \( \sigma > 0 \) the distorting effect on labor supply of the marginal tax rate (in a static context) becomes increasingly important. Together with the third term in (29) it eventually provides an upper bound for the tax rate. We now ask whether, in case of an optimal choice of the labor income tax, other taxes are desirable.

**Proposition 3** Let \( \tau = 0 \). The welfare effect of an introduction of a tax \( \tau_b \) on bequests, given that the linear income tax is chosen optimally, is

\[
\frac{\partial S}{\partial \tau_b} \bigg|_{\tau_b=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (\gamma \bar{e} + \frac{\partial \gamma_b}{\partial \tau_b} \frac{1}{\bar{e}} E[e_i])
\]

(40)

\[+ \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (\frac{\gamma^2}{1-\gamma} \bar{e} + \frac{\partial \gamma_b}{\partial \tau_b} [(I - \gamma M)^{-1} AM]_i (x - c_m)),\]

with \( \bar{e} = \sum_{i=1}^{n} f_i E[e_i] \) as the average inheritances received by a generation. In case of a generalized Cobb-Douglas utility function \( \varphi = (c_i - c_0)^{(1-\gamma)} b_i^{\gamma} \) with constant \( \gamma \), \( \partial \gamma_b / \partial \tau_b = -\gamma \) at \( \tau_b = 0 \). Then, given that \( A \) fulfills property (C) the first term in (40) is positive, if \( \gamma \) and/or \( \rho \) is large enough. The sign of the second term is ambiguous.

**Proof.** Using the Envelope Theorem, we get for the optimal value function \( S(\tau_b, \tau) \) of
the maximization problem (26) and (27), after inserting (28) in (26) and (27) 

\[
\frac{\partial S}{\partial \tau_b} \bigg|_{\tau_b=0} = \sum_{i=1}^{n} f_i V_i^{-\rho} \left[ \left(1 - \gamma\right) \frac{\partial E[e_i]}{\partial \tau_b} + (1 - \sigma) w_i \frac{\partial l_i}{\partial \tau_b} \right] 
- \gamma \left( E[e_i] + \alpha + (1 - \sigma) w_i l_i - c_0 \right) + \gamma \left( \frac{\partial E[e_i]}{\partial \tau_b} + (1 - \sigma) w_i \frac{\partial l_i}{\partial \tau_b} \right) 
+ v \sum_{i=1}^{n} f_i (\sigma w_i \frac{\partial l_i}{\partial \tau_b} + \gamma (E[e_i] + \alpha + (1 - \sigma) w_i l_i - c_0)) .
\]

By use of \(\frac{\partial \gamma_b}{\partial \tau_b}|_{\tau_b=0} = \frac{\partial \gamma}{\partial \tau_b} - \gamma\) (and \(\gamma_b = \gamma\)), and of the first-order conditions (4) - (6) of the individual optimization problem, (41) can be reduced to 

\[
\frac{\partial S}{\partial \tau_b} \bigg|_{\tau_b=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left[ \frac{\partial E[e_i]}{\partial \tau_b} - \gamma (E[e_i] + \alpha + (1 - \sigma) w_i l_i - c_0) \right] 
+ v \sum_{i=1}^{n} f_i (\sigma w_i \frac{\partial l_i}{\partial \tau_b} + \gamma (E[e_i] + \alpha + (1 - \sigma) w_i l_i - c_0)) ,
\]

and rearrangement gives us 

\[
\frac{\partial S}{\partial \tau_b} \bigg|_{\tau_b=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \frac{\partial E[e_i]}{\partial \tau_b} + \gamma (1 - \sigma) \left[ \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (w_i l_i) + v \sum_{i=1}^{n} f_i w_i l_i \right] 
- \lambda \gamma \sum_{i=1}^{n} f_i V_i^{-\rho} (E[e_i] + \alpha - c_0) + v \sigma \sum_{i=1}^{n} f_i w_i \frac{\partial l_i}{\partial \tau_b} + v \gamma (\tau + \alpha - c_0) .
\]

We use the condition (39) for the optimum linear income tax to substitute for the term in square brackets in (43), and we substitute for \(v\) from (31) in the last term to obtain further 

\[
\frac{\partial S}{\partial \tau_b} \bigg|_{\tau_b=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \frac{\partial E[e_i]}{\partial \tau_b} - \gamma (1 - \sigma) \left[ \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \frac{\partial E[e_i]}{\partial \sigma} + v \sigma \sum_{i=1}^{n} f_i w_i \frac{\partial l_i}{\partial \sigma} \right] 
- \lambda \gamma \sum_{i=1}^{n} f_i V_i^{-\rho} (E[e_i] + \alpha - c_0) + v \sigma \sum_{i=1}^{n} f_i w_i \frac{\partial l_i}{\partial \tau_b} 
+ \lambda \gamma \sum_{i=1}^{n} f_i V_i^{-\rho} \frac{\partial E[e_i]}{\partial \alpha} + \left( \tau + \alpha - c_0 \right) ,
\]
which we transform to

\[
\frac{\partial S}{\partial \tau_b} \bigg|_{\tau_b=0} = \varnothing \sum_{i=1}^{n} f_i w_i\left( \frac{\partial l_i}{\partial \tau_b} - \gamma(1 - \sigma)\frac{\partial l_i}{\partial \sigma} \right) + \lambda \gamma \sum_{i=1}^{n} f_i V_i^{-\rho}(\bar{\varnothing} - E[e_i]) \\
+ \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}\left[ \frac{\partial E[e_i]}{\partial \tau_b} - \gamma(1 - \sigma)\frac{\partial E[e_i]}{\partial \sigma} + \gamma(\bar{\varnothing} + \alpha - c_0)\frac{\partial E[e_i]}{\partial \alpha} \right].
\]

By implicit differentiation of the first-order conditions of the individual maximization problem (see the external appendix) we find that \( \frac{\partial l_i}{\partial \tau_b} - \gamma(1 - \sigma)\frac{\partial l_i}{\partial \tau_b} = 0 \). Thus, the first term in (44) cancels out.

From the steady-state equation (28) for \( E[e_i] \) we find

\[
\frac{\partial E[e_i]}{\partial \tau_b} = A_i \left( (1 - \sigma)w_1 \frac{\partial l_1}{\partial \tau_b} \right) + \left[ \frac{\partial A}{\partial \tau_b} \right] (x - c_m),
\]

where the matrix \( \frac{\partial A}{\partial \tau_b} \) is derived from \( A \) by differentiating each component with respect to \( \tau_b \) and the index \( i \) denotes its \( i \)-th row. Thus, the third term of (44) reads as - by use of (37) and (36) -

\[
\lambda \sum_{i=1}^{n} f_i V_i^{-\rho}\left[ A_i(1 - \sigma) \left( \frac{w_1(\frac{\partial l_1}{\partial \tau_b} - \gamma(1 - \sigma)\frac{\partial l_1}{\partial \sigma})}{\gamma(1 - \sigma)\frac{\partial l_1}{\partial \sigma}} \right) \right]
+ A_i \gamma \left( (1 - \sigma)w_1 l_1 + \alpha - c_0 + \bar{\varnothing} \right) + \left[ \frac{\partial A}{\partial \tau_b} \right] (x - c_m),
\]

Observe that each component in the first column vector of (46) is zero (see above). In view of this and the above considerations, (44) can be rewritten as

\[
\frac{\partial S}{\partial \tau_b} \bigg|_{\tau_b=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(\bar{\varnothing} + \frac{\gamma^2}{1 - \gamma} + \frac{\partial A}{\partial \tau_b} (x - c_m)),
\]

where we have used \( x_i = \alpha + (1 - \sigma)w_i l_i \) and \( A_i(x - c_m) = E[e_i] \), as well as \( A_i(\bar{\varnothing}, \varnothing, \ldots, \bar{\varnothing})^T = \)

\[
22
\]
\( \bar{\gamma}/(1-\gamma) \), which follows from the fact that the components of \( A_i \) add up to \( \gamma/(1-\gamma) \), see Lemma 1.

Next we study the matrix \( \partial A/\partial \tau_b \). From the definition of \( A = \gamma_b(I - \gamma_b M)^{-1} M \) we get (\( M \) is independent of \( \tau_b \))

\[
\frac{\partial A}{\partial \tau_b} = \frac{\partial \gamma_b}{\partial \tau_b} (I - \gamma_b M)^{-1} M + \gamma_b \frac{\partial (I - \gamma_b M)^{-1}}{\partial \tau_b} M, \tag{48}
\]

where the matrix \( \partial (I - \gamma_b M)^{-1}/\partial \tau_b \) is derived from \( (I - \gamma_b M)^{-1} \) by differentiating each element with respect to \( \tau_b \). We know that \( (I - \gamma_b M)^{-1} = \sum_{t=0}^{\infty} (\gamma_b M)^t \), thus

\[
\frac{\partial (I - \gamma_b M)^{-1}}{\partial \tau_b} = \sum_{t=1}^{\infty} t \gamma_b^{t-1} M^{t-1} \frac{\partial \gamma_b}{\partial \tau_b} = \frac{\partial \gamma_b}{\partial \tau_b} [(I - \gamma_b M)^{-1}]^2 M, \tag{49}
\]

where the equality \( \sum_{t=1}^{\infty} t (\gamma_b M)^{t-1} = (\sum_{t=0}^{\infty} (\gamma_b M)^t)(\sum_{t=0}^{\infty} (\gamma_b M)^t) = [(I - \gamma_b M)^{-1}]^2 \) follows immediately from direct multiplication. Using (49) in (48) together with \( (I - \gamma_b M)^{-1} M(x-c_m) = (1/\gamma_b)E[e] \) gives us

\[
\frac{\partial A}{\partial \tau_b} (x-c_m) = \frac{\partial \gamma_b}{\partial \tau_b} 1 / \gamma_b E[e] + \frac{\partial \gamma_b}{\partial \tau_b} (I - \gamma M)^{-1} AM(x-c_m). \tag{50}
\]

Substituting (50) in (47) and rearranging terms leads to

\[
\frac{\partial S}{\partial \tau_b} \bigg|_{\tau_b=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (\gamma \bar{e} + \frac{\partial \gamma_b}{\partial \tau_b} 1 / \gamma_b E[e_i] + \frac{\gamma^2}{1-\gamma} \bar{e} + \frac{\partial \gamma_b}{\partial \tau_b} [(I - \gamma M)^{-1} AM]_i (x-c_m)), \tag{51}
\]

with the index \( i \) denoting the \( i \)-th row of the matrix. In case of a generalized Cobb-Douglas function \( \varphi \), when \( \gamma \) is independent of \( \tau_b \), we find \( \partial \gamma_b/\partial \tau_b = (\gamma/(1+\gamma_b))/\partial \tau_b = -\gamma \). Then the first two terms of (51) can be combined to

\[
\lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (\gamma \bar{e} - E[e_i]). \tag{52}
\]

Expected inheritances are increasing with \( i \) if \( M \) fulfills condition (C). Then, for large enough \( \gamma \), the expression is positive, because the \( \gamma \bar{e} - E[e_i] \) are decreasing with \( i \), \( \bar{e} = \sum_{i=1}^{n} f_i E[e_i] \), and the \( V_i^{-\rho} \) are larger for positive values of \( \gamma \bar{e} - E[e_i] \). The latter holds
even more, the larger is $\rho$. ■

The last formula (52) expresses the main redistributive effect of a tax on bequests as a supplement to an optimal labor income tax: if individuals of some generation differ not only in abilities (incomes) but also in received inheritances, then a tax on the bequests they leave to their descendants is an (indirect) instrument in order to reduce the differences in received initial endowments (the tax revenues are returned to the individuals via the lump-sum demogrant). Clearly, this instrument redistributes in the desired way (from high-ability to low-ability groups) only if there is a positive relation between abilities and received inheritances, that is, if $M$ and, thus, $A$ fulfills condition (C). Its negative effect on labor supply is smaller than that of the income tax, but it distorts the bequest decision. For the latter reason, compared to a model with exogenous inherited wealth of some generation (see Brunner and Pech 2012ab), the redistributive effect of tax on bequests of this generation is now weaker ($\gamma \bar{e}$ instead of $e$) occurs.

The magnitude of the first two terms (51) depends on the variation in inheritances. The more unequally they are distributed, the larger is the redistributive effect of the bequest tax. The variation is determined by the matrix $M$ which describes the wealth transmission, but also by the differences in bequests left by individuals of different ability groups. These differences increase with $c_0$, the fraction of the budget needed for minimum consumption.

In order to interpret the last two terms of (51) we turn to the special case that $M$ (or $P$) is the unit matrix, that is, there is no mobility across ability groups; all bequests go to the same ability group. Then $I - \gamma M$ is a diagonal matrix with elements $1 - \gamma$, its inverse has elements $1/(1 - \gamma)$, while $A$ is a diagonal matrix with elements $\gamma/(1 - \gamma)$. As a consequence, $A_i(x - c_m) = \gamma/(1 - \gamma)(x_i - c_0) = E[e_i]$ and for a generalized Cobb-Douglas function the last two terms of (51) read as

$$\lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (\gamma \bar{e} - E[e_i]) \gamma/(1 - \gamma),$$

which is again positive for sufficiently large $\gamma$ and/or $\rho$. 24
Note finally that in case of complete mobility, when all $E[e_i]$ are equal to $\bar{e}$, thus there is no case for redistribution of inheritances, the welfare effect of the tax on bequests is negative, due to the distortion of the bequest decision.

Next we turn to the effect of tax $\tau$ which is imposed on all expenditures of the individuals as opposed to the tax on bequests alone.

**Proposition 4** Let $\tau_b = 0$. The welfare effect of an introduction of a tax $\tau$ on expenditures for consumption and bequests, given that the linear income tax is chosen optimally, is

$$\frac{\partial S}{\partial \tau} \bigg|_{\tau=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left( \bar{e} - E[e_i] \right) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \frac{\gamma \bar{e}}{1 - \gamma} - \gamma (I - \gamma M)^{-1} AM(x - c_m).$$

This first expression is positive, given that $M$ fulfills condition (C).

**Proof.** We get for the optimal value function (note that $\gamma_b = \gamma$, as $\tau_b$ is set to zero)

$$\frac{\partial S}{\partial \tau} = \sum_{i=1}^{n} f_i V_i^{-\rho} \left\{ \frac{1 - \gamma}{(1 + \tau)^2} (E[e_i] + x_i) + \frac{1 - \gamma}{1 + \tau} \left( \frac{\partial E[e_i]}{\partial \tau} + w_i (1 - \sigma) \frac{\partial l_i}{\partial \tau} \right) \right\} + \frac{\partial \varphi(.)}{\partial b_i} \left\{ - \frac{\gamma}{(1 + \tau)^2} (E[e_i] + x_i) + \frac{\gamma}{1 + \tau} \left( \frac{\partial E[e_i]}{\partial \tau} + w_i (1 - \sigma) \frac{\partial l_i}{\partial \tau} \right) - g'(l_i) \frac{\partial l_i}{\partial \tau} \right\} + \nu \sigma \sum_{i=1}^{n} f_i w_i \frac{\partial l_i}{\partial \tau} + \frac{1}{(1 + \tau)^2} \sum_{i=1}^{n} f_i (E[e_i] + x_i) + \frac{\tau}{1 + \tau} \sum_{i=1}^{n} f_i \left( \frac{\partial E[e_i] + x_i}{\partial \tau} \right).$$

Use of the individual first-order conditions (4) - (6) and evaluation at $\tau = 0$ allows a simplification to

$$\frac{\partial S}{\partial \tau} \bigg|_{\tau=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left( -(E[e_i] + x_i) + \frac{\partial E[e_i]}{\partial \tau} \right) + \nu \sigma \sum_{i=1}^{n} f_i w_i \frac{\partial l_i}{\partial \tau} + \nu \sum_{i=1}^{n} f_i (E[e_i] + x_i),$$

which, by writing $x_i = \alpha + (1 - \sigma) w_i l_i$ and substituting for $\nu$ from (31) in $\nu \sum_{i=1}^{n} f_i (E[e_i] + \alpha) = \nu(\bar{e} + \alpha)$, can be rearranged to

$$\frac{\partial S}{\partial \tau} \bigg|_{\tau=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left( \bar{e} - E[e_i] \right) + (1 - \sigma) \left[ \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (-w_i l_i) + \nu \sum_{i=1}^{n} f_i w_i l_i \right] (53) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \frac{\partial E[e_i]}{\partial \tau} + \nu \sigma \sum_{i=1}^{n} f_i w_i \frac{\partial l_i}{\partial \tau} + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \frac{\partial E[e_i]}{\partial \alpha}(\bar{e} + \alpha).$$
We substitute for the term in square brackets in (53) by using the condition (39) for the optimal linear income tax rate to obtain further

\[
\frac{\partial S}{\partial \tau} \bigg|_{\tau=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(\bar{\tau} - E[e_i]) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left( \frac{\partial E[e_i]}{\partial \tau} - (1 - \sigma) \frac{\partial E[e_i]}{\partial \sigma} \right) (54) \\
+ \nu \sigma \sum_{i=1}^{n} f_i w_i \left( \frac{\partial l_i}{\partial \tau} - (1 - \sigma) \frac{\partial l_i}{\partial \sigma} \right) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \frac{\partial E[e_i]}{\partial \alpha} (\bar{\alpha} + \alpha).
\]

The third term is zero (see the external Appendix). From the steady-state equation (28) for \(E[e_i]\) we find at \(\tau = 0\)

\[
\frac{\partial E[e_i]}{\partial \tau} \bigg|_{\tau=0} = A_i \begin{pmatrix} (1 - \sigma)w_1 \frac{\partial l_1}{\partial \tau} \\ \vdots \\ (1 - \sigma)w_n \frac{\partial l_n}{\partial \tau} \end{pmatrix} = 0
\]

Using this together with (36) and (37) in (54), we get

\[
\frac{\partial S}{\partial \tau} \bigg|_{\tau=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(\bar{\tau} - E[e_i]) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left( \frac{\partial E[e_i]}{\partial \tau} - (1 - \sigma) \frac{\partial E[e_i]}{\partial \sigma} \right) \\
+ \nu \sigma \sum_{i=1}^{n} f_i w_i \left( \frac{\partial l_i}{\partial \tau} - (1 - \sigma) \frac{\partial l_i}{\partial \sigma} \right) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \frac{\partial E[e_i]}{\partial \alpha} (\bar{\alpha} + \alpha),
\]

and further, as \(A_i(\bar{e}, ..., \bar{e})^T = \bar{e} \gamma/(1 - \gamma)\) and \(\partial l_i/\partial \tau = (1 - \sigma)\partial l_i/\partial \sigma = 0\)

\[
\frac{\partial S}{\partial \tau} \bigg|_{\tau=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(\bar{\tau} - E[e_i]) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left( \frac{\gamma \bar{e}}{1 - \gamma} + \left[ \frac{\partial A}{\partial \tau} \right]_i (x - c_m) \right).
\]

Finally, from the definition of \(A = \gamma(I - \gamma/(1 + \tau)M)^{-1}M\) we find, making use of \((I - \gamma/(1 + \tau)M)^{-1} = \sum_{t=0}^{\infty} (\gamma/(1 + \tau)M)^t\), similar to the procedure in the proof of Proposition 3

\[
\frac{\partial A}{\partial \tau} \bigg|_{\tau=0} = -\gamma^2[(I - \gamma M)^{-1}]^2 M^2 = -\gamma(I - \gamma M)^{-1} AM.
\]
which completes the proof. The first term in (56) is positive if \( M \) fulfills condition (C), because then \( V_1^{-\rho} > V_2^{-\rho} > \cdots > V_n^{-\rho} \) (Lemma 10), and we have \( \tau - E[e_1] > \tau - E[e_2] > \cdots > \bar{c} - E[e_n] \), in addition to \( \sum_{i=1}^{n} f_i(\bar{c} - E[e_i]) = \bar{c} - \bar{c} = 0 \). \( \blacksquare \)

For a better understanding of the last term, we turn to the special case when \( M = P \) is the identity matrix. Then we know that \( (I - \gamma M)^{-1} = 1/(1 - \gamma)I \) and \( (\partial A/\partial \tau)_i(x - c_m) = -E[e_i] \gamma/(1 - \gamma) \). Then the last term in the round brackets in (56) can be written as \( \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(\bar{\bar{c}} - E[e_i]) \gamma/(1 - \gamma) \), which is again positive if \( M \) fulfills condition (C). Moreover, this expression is zero if there is complete mobility, that is, \( \bar{\bar{c}} - E[e_i] \) for all \( i \). We can therefore conclude that a common tax on consumption as well as on bequests, whose revenues are redistributed uniformly to the individuals, has an unambiguously positive effect on steady-state welfare. This result conforms to the finding in a model with exogenous initial wealth. The reason is that, when combined with an optimal income tax, the expenditure tax has a less distorting effect on labor supply, compared to the income tax, and also compared to the bequest tax (observe the similarity of the formulas: in Proposition 3 \( \gamma \bar{\bar{c}} \) occurs instead of \( \bar{c} \)). The expenditure tax is essentially a tax on inherited wealth, and allows additional redistribution. We have discussed this fact in more detail in Brunner and Pech (2012ab) in a model with a nonlinear tax on labor income, but the finding also applies in case of an optimal linear income tax. Again, this redistribution has the desired effect only if inheritances and abilities are positively related.

6 Conclusion

The Atkinson-Stiglitz result that other taxes are redundant, if labor income is taxed optimally and preferences are weakly separable between consumption and labor, holds as long as individuals differ only in a single characteristic, namely labor productivity. A case for indirect taxes or capital taxes arises in a model where individuals also differ in inherited wealth as a second characteristic and when the correlation between these two characteristics is positive. The essential problem when formulating such a model is how to treat inherited wealth. In the short run it obviously makes sense to regard it as exogenously given, as we did in previous work (Brunner and Pech 2012ab).
However in the longer run inherited wealth - bequests left from earlier generations - is clearly a consequence of labor income and is therefore endogenous. Taking this fact into account was the intention of the present paper. We have modeled the evolution of abilities as a Markov chain, with an associated distribution of inheritances, and we have studied the steady-state of this process. In particular, we were able to show that, though in effect differing labor productivities are the only source of heterogeneity in this model, the Atkinson-Stiglitz result does not apply and bequest taxation may raise welfare through increased redistribution, above the level attained by an income tax alone. Moreover, a tax on total expenditures (on consumption and bequests) was shown to have an unambiguously positive welfare effect.

It should be emphasized that our results are derived within a model where leaving bequests is motivated by joy of giving. That is, we considered social welfare of each generation separately and asked how it is affected by redistributive taxes such as the income tax or the bequest tax. With the alternative approach of an altruistic motive, redistribution refers to whole dynasties and essentially affects the first generation which anticipates taxes and transfers of the descendents. It is difficult to see how in such a perfect-foresight framework the idea of wealth transmission across ability groups could be modeled. Furthermore, empirical studies do not suggest that the altruistic model is more in accordance with actual behavior.\[10\]

In our model individuals live for one period only. Therefore, the taxation of bequests or inheritances is equivalent to a tax on wealth. Moreover, as a consequence of our assumption of unproductive capital, a tax on income from capital is not included in this study. A discussion of the specific roles of these taxes and their relation to the bequest tax requires a more elaborated model and represents a task for future research.

\[10\] See Kopczuk 2013b, among others. Piketty and Saez (2013) also develop their main formula in a model with a joy-of-giving motive motive for leaving bequests.
References


Solon, Gary (1992), Intergenerational Income Mobility in the United States, American Economic Review 82, 393-408.