Taxing bequests and consumption in the steady state

by

Johann Brunner and Susanne Pech

Taxing bequests and consumption in the steady state

Johann K. Brunner[*] and Susanne Pech[†]

University of Linz[‡]

October 21, 2013

Abstract

We study the optimal tax system in a dynamic model where differences in wages induce differences in inheritances, and the transition from parent ability to child ability is described by a Markov chain. We characterize expected inheritances in the steady state and show that the Atkinson-Stiglitz result on the redundancy of indirect taxes does not hold in this framework. In particular, given an optimal income tax, a bequest tax as well as a consumption tax are potential instruments for additional redistribution. For the bequest tax the sign of the overall welfare effect depends on the reaction of bequests and on inequality aversion, while for the consumption tax the sign is always positive because the distorting effect is outweighed by the induced increase in wealth accumulation. A necessary condition for a positive welfare effect is the empirically validated relation that more able individuals on average have more able parents than less able individuals.

---

[*] *Email:* johann.brunner@jku.at.

[†] *Email:* susanne.pech@jku.at.

[‡] *both:* Department of Economics, Altenberger Straße 69, 4040 Linz, Austria.

1 Introduction

In developed economies each year a substantial amount of wealth is transmitted from one generation to the next (see Piketty 2011, Wolff 2002). Hence, inheritances represent an important source of individual wealth, but their variation across individuals also creates additional inequality in a society, on top of differences in income. In this paper we study the implications of this fact for the design of the tax system. In particular, we ask whether the existence of inheritances requires specific tax instruments such as an estate tax to mitigate inequality. As is well-know, inheritance taxation is one of the most controversial issues in tax policy.

To provide an answer to this question we formulate a dynamic optimal-taxation model of an economy with individuals differing in labor productivity and leaving bequests to their descendants. We ask how a shift from labor-income taxation to bequest or consumption taxation affects welfare in the steady state of this economy. The starting point is the well-known result by Atkinson and Stiglitz (1976) derived in a model without bequests: if labor income is taxed optimally then any other - indirect - tax is redundant, given weak separability of preferences in consumption and labor.[1] As a consequence, if leaving bequests is just seen as one form of consumption, there is no role for a tax on estates;[2] neither is any kind of consumption or capital taxation useful. In particular, an estate tax does not allow more redistribution than what is possible through an income tax alone.[3]

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. Brunner and Pech (2012ab) have shown that with exogenously given inherited wealth a tax on wealth transfers indeed has a redistributive effect, given that more productive individuals receive more inheritances.[4]

---

[1] This holds in case of an optimal nonlinear income tax. If it is restricted to be linear the condition that preferences over consumption are homothetic is needed (see Deaton 1979).

[2] Unless bequests are a complement to leisure, see e.g., Gale and Slemrod (2001), Kaplow (2001).

[3] 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 (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.

[4] In a related approach, Boadway et al. (2000) and Cremer et al. (2001, 2003) show that indirect

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; the stochastic process can be described by a Markov chain and 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. Depending on the transition probabilities, the budget share devoted to bequests in each group determines 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.

As a result we arrive at a Mirrlees-type model with differing labor productivities, extended by the existence of endogenously determined inheritances. We then apply this model to study the optimal design of a tax system. First we analyze the optimal linear income tax, determined by maximization of expected social welfare of a generation. It turns out that in the extended model the redistributive effect of this tax is more pronounced, because it refers not only to labor income as such, but also indirectly to steady-state inheritances financed by labor income. On the other hand, the distortion of labor supply causes an indirect negative effect on inheritances. Altogether, the desirability of a positive marginal tax rate is guaranteed if the labor supply elasticity is not too large.[5]

Next, we address the main question, namely whether the introduction of indirect taxes

taxes as well as a tax on capital income make sense if individuals differ in inherited wealth. They assume, however, that bequests and inheritances are unobservable and, thus, cannot be taxed. In Saez (2002) exogenous differences in the tastes for consumption goods make indirect taxes useful.

[5]This is in contrast to the static, one-period model where the optimal marginal tax rate is unambiguously positive (see e.g., Hellwig 1986, Brunner 1989).

on bequests or consumption increases steady-state social welfare, given an optimal linear income tax. The answer to this question is unclear a priori, because on the one hand we know for the bequest tax that it is desirable if individuals differ in a second exogenously given 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 represent the only (exogenous) source of heterogeneity, one might suppose that again the Atkinson-Stiglitz result applies and no other (indirect) tax than the optimal labor income tax is desirable.

Our analysis shows that this conjecture is wrong. We first 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. On the other hand, the tax distorts the bequest decision and induces, thus, lower accumulated wealth in the steady state. The overall effect is positive in case that the substitution elasticity between bequests and consumption is small or that the degree of inequality aversion in the social welfare function is high.

Further, we find for a consumption tax, introduced in addition to an optimally chosen tax on labor income, that in our model it has a redistributive potential similar to the bequest tax. A closer analysis shows that both taxes, though being targeted on some form of spending of the individuals' budgets (for bequests versus own consumption), in fact are instruments for an equalization of received (inherited) wealth. The difference is, however, that the distortion caused by the consumption tax works in favor of bequests and induces, thus, a higher accumulation of wealth. For small tax rates, this accumulation effect is large enough to outweigh the deadweight loss so that the overall effect of a consumption tax is unambiguously positive. This result can be seen as a justification of a specific tax on consumption which in fact exists in many countries, like the VAT in the European Union.

It should be stressed that in standard optimal-taxation models such a tax has no particular role: if the individuals' budget consists only of labor income which is spent for the consumption of different commodities, a uniform tax on the latter is equivalent to an income tax. In our model the analogous instrument is a uniform tax on bequests and

consumption which we call an expenditure tax. We show that such a tax unambiguously increases social welfare effect, because it combines the redistributive effect of both separate taxes and causes no distortion.

For these results again the condition that in the steady-state more able individuals receive higher expected inheritances is crucial. This condition actually requires that a more able individual on average has more able parents than a less able individual. That this conforms to reality 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 used 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 indirect taxation as an instrument for increasing equality of opportunity. In a related work Piketty and Saez (2013) derive a formula for the optimal bequest tax rate in terms of estimable parameters such as the elasticity of bequests with respect to the tax. Taking the amount of government transfers as given, this formula states to which extent they should be financed by a tax on bequests as opposed to a tax on labor income, depending on the parameter constellation. In contrast, our focus is on whether a welfare gain by additional redistribution through indirect taxes is possible, given an optimal linear income tax. In particular, we also characterize the specific effects of a consumption or expenditure tax arising from differences in inheritances across individuals.

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

---

[6]In 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.

5 the optimal tax on labor income is characterized, and the results on the desirability of additional taxes on bequests and consumption, respectively, are derived. Section 6 provides an extensive discussion of these results.

2 The Model

In each period t a population of mass one exists. It is split into n groups with different abilities $w_1 < w_2 < .... < w_n$, whose shares are $f_1, ..., f_n$. That is, $f_1, ..., f_n$ 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 φ 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. Pre-tax labor income is $w_i l_i$ and the individual receives inheritances $e_i, i = 1, ..., n$.

There exists a linear tax on labor income, given by $-\alpha + \sigma w_i l_i$ with $-\alpha$ as a uniform negative tax and σ as the marginal tax rate on gross labor income. Moreover, the tax system consists of a proportional tax τ_b on bequests and of a proportional tax τ_c on consumption. For given taxes the maximization problem of an individual i is

$$\max \varphi(c_i, b_i) - g(l_i), \tag{1}$$

$$\text{s.t.} \quad (1 + \tau_c)c_i + (1 + \tau_b)b_i \leq e_i + \alpha + (1 - \sigma)w_i l_i, \tag{2}$$

$$c_i, b_i, l_i \geq 0. \tag{3}$$

The first-order conditions for an interior solution with λ as the Lagrangian variable associated with (2) read as

$$\frac{\partial \varphi}{\partial c_i} - \lambda(1 + \tau_c) = 0, \tag{4}$$

$$\frac{\partial \varphi}{\partial b_i} - \lambda(1 + \tau_b) = 0, \tag{5}$$

$$-g'(l_i) + \lambda(1 - \sigma)w_i = 0. \tag{6}$$

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 $(1 - \gamma)/(1 + \tau_c)$ and $\gamma/(1 + \tau_b)$, with $\gamma \in (0, 1)$, of the available budget, given a constant c_a representing minimum consumption:

$$c_i = \gamma_c(e_i + x_i - (1 + \tau_c)c_a) + c_a, \tag{7}$$

$$b_i = \gamma_b(e_i + x_i - (1 + \tau_c)c_a). \tag{8}$$

Here x_i denotes net income of the household, $x_i \equiv \alpha + (1 - \sigma)w_i l_i$, and $\gamma_b \equiv \gamma/(1 + \tau_b)$ describes the share of bequests in the "real" budget after deducting minimum consumption, while $\gamma_c \equiv (1 - \gamma)/(1 + \tau_c)$ describes the corresponding share of consumption above minimum consumption. We generally assume that the expenditures for c_a are not larger than net income, that is, $x_i - c_a(1 + \tau_c) \geqslant 0$.

Such a type of demand functions, as described by (7) and (8), arises if φ can be written as a function $\widetilde{\varphi}(c_i - c_a, b_i)$, homogeneous in $c_i - c_a$ and b_i. In addition, we assume that $\widetilde{\varphi}$ is homogeneous of degree 1. Note that in general γ itself depends on τ_b and τ_c, such as in case of a properly adjusted CES utility function with the functional form $\widetilde{\varphi}(c_i - c_a, b_i) = ((1 - \delta)^{1-\eta}(c_i - c_a)^\eta + \delta^{1-\eta}b_i^\eta)^{\frac{1}{\eta}}$ where the parameters δ and $(1 - \delta)$, respectively, express the weights on b_i and $c_i - c_a$, while η determines the constant elasticity of substitution $1/(1 - \eta)$.[7]

Indirect utility of an individual 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). Linear homogeneity of $\widetilde{\varphi}$ implies that $\widetilde{\varphi}(\gamma_c(e_i + x_i - (1 + \tau_c)c_a), \gamma_b(e_i + x_i - (1 + \tau_c)c_a) = \lambda(e_i + x_i - (1 + \tau_c)c_a)$, where $\lambda = \widetilde{\varphi}(\gamma_c, \gamma_b)$ is the marginal utility of income which is independent of w_i, e_i and x_i but depends on τ_c and τ_b. Hence, indirect utility V_i is given by

$$V_i(e_i; \alpha, \sigma, \tau_c, \tau_b) = \lambda(e_i + x_i - c_a(1 + \tau_c)) - g(l_i) \tag{9}$$

---

[7] Only in case of a Stone-Geary utility function $(c_i - c_a)^{1-\delta}b_i^\delta$ (η converges to zero), γ is the constant weight δ of bequests.

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 on the marginal utility of income λ. Note that from (6) the relation $x_1 < x_2 < ... < x_n$ follows, because λ is independent of w_i.

3 Dynamics of wealth transmission and abilities

Having specified the model of the economy in any given period, including the individual decision to bequeath part of the available budget to the descendants, we now turn to the dynamics. We consider the transmission of wealth in detail and study the distribution of inheritances in the steady-state of this process. 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. There is a constant transition probability p_{ij} that an individual with ability w_i has a descendant with ability w_j, where $\sum_{j=1}^{n} p_{ij} = 1$ 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 $\pi = (\pi_1, ..., \pi_n)$. π 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, π 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 π. 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

$$f = fP. \tag{10}$$

Next we study the transmission of bequests. Tax parameters α, σ and τ_b, τ_c are assumed to remain constant over time, therefore labor supply and net income of each type is also constant over time. Let b_{jt} be the bequests left by an individual with ability w_j in some period t. We study the diffusion of these bequests to the individuals of the next generation $t+1$.

Let $e_{i,t+1}$ denote inheritances received in period $t+1$ by an individual in ability group i out of bequests of the preceding period. Clearly, each $e_{i,t+1}$ has possible realizations

7

$b_{1t}, ..., b_{nt}$. Let further μ_{ij} denote the probability that in $t+1$ an individual in group i has a parent in group j, that is, receives inheritance b_{jt}. Note that the possible realizations b_{jt} are the same for all ability groups (are independent of i) in $t+1$, but the associated probabilities μ_{ij} differ across groups. To determine μ_{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

$$\mu_{ij} = \text{Prob}(\text{parent} = j \mid \text{descendant} = i) = \frac{f_j p_{ji}}{f_i}. \tag{11}$$

Obviously, μ_{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 μ_{ij}, we have

$$M = \begin{pmatrix} p_{11} & \frac{f_2 p_{21}}{f_1} & ...... & \frac{f_n p_{n1}}{f_1} \\ \frac{f_1 p_{12}}{f_2} & p_{22} & ...... & \frac{f_n p_{n2}}{f_2} \\ ...... & ...... & ...... & ...... \\ \frac{f_1 p_{1n}}{f_n} & \frac{f_2 p_{2n}}{f_n} & ...... & p_{nn} \end{pmatrix}. \tag{12}$$

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

$$E[e_{i,t+1}] = \sum_{j=1}^{n} \mu_{ij} b_{jt} = \sum_{j=1}^{n} b_{jt} f_j p_{ji}/f_i. \tag{13}$$

In matrix notation with $E[e_{t+1} \mid b_t]$ denoting the column vector of expected inheritances out of $b_t \equiv (b_{1t}, ..., b_{nt})^T$, this is written as[8]

$$E[e_{t+1} \mid b_t] = M b_t. \tag{14}$$

Next, consider the initial generation $t = 0$ and assume that in this generation no received inheritances exist, $e_{j0} = 0, j = 1, ..., n$. Hence, individuals of this generation

---

[8]The upper index T denotes the transposed vector, hence b_t is a column vector.

leave bequests out of labor income only, $b_{j0} = \gamma_b(x_{j0} - c_a(1 + \tau_c))$, and these values are the possible realizations of inheritances e_{i1} received by an individual i in generation 1.[9] Furthermore, with homothetic preferences, bequests left by an individual i in generation 1 can be separated into those out of received inheritances and those out of own labor income, that is, $b_{i1} = \gamma_b e_{i1} + \gamma_b(x_{i1} - c_a(1 + \tau_c))$. As a consequence, we get $b_{i1} = \gamma_b^2(x_{j0} - c_a(1 + \tau_c)) + \gamma_b(x_{i1} - c_a(1 + \tau_c))$, for some $j \in \{1, ..., n\}$. By the same reasoning, bequests left by an individual k in generation 2 can be expressed as $b_{k2} = \gamma_b^3(x_{j0} - c_a(1 + \tau_c)) + \gamma_b^2(x_{i1} - c_a(1 + \tau_c)) + \gamma_b(x_{k2} - c_a(1 + \tau_c))$, for some $i, j \in \{1, ..., n\}$, which are the possible realizations of inheritances e_{m3} for any individual m (independent of the type) in generation 3.

Generalizing this idea we find that in any generation t received inheritances of an individual $r \in \{1, ..., n\}$ can assume one of the values

$$e_{rt} = \sum_{s=0}^{t-1} \gamma_b^{t-s}(x_{j_s s} - c_a(1 + \tau_c)), j_s \in \{1, ..., n\}, \tag{15}$$

which again are the same for any r. What is inherited by some individual in t only depends on the ability group of the individual who originally leaves bequests out of own labor income; but is independent of the ability types of individuals further involved in the transition process by just transmitting a share γ_b of received inheritances to the next generation. Obviously, because of this decline by the factor γ_b, total inheritances in each generation are dominated by bequests originating from labor income of more recent generations.

One observes that the number of possible realizations increases by the factor n in each generation, when received inheritances are combined with n values of labor income. Fortunately, we need not describe the distribution in detail but confine our analysis to the evolution of expectations. Clearly, the above insights also remain true in the long run, for t going to infinity. In particular, given our assumption of constant abilities, the values $x_{j_s} - c_a(1 + \tau_c)$ in (15) are obviously bounded by $x_n - c_a(1 + \tau_c)$ for all s,[10] thus

---

[9] For the moment, to make the time structure more transparent, we add a time index to net income (which in fact is constant over time for each type, given constant tax rates).

[10] From now we drop the time index to net income again.

for $t \to \infty$ inheritances are not larger than $(x_n - c_a(1 + \tau_c))\gamma_b/(1 - \gamma_b)$. This implies that all realizations of inheritances are finite, then also expected inheritances of each group are finite for all t going to infinity.

For a characterization of expected inheritances we return to formula (14) which describes the relation between bequests and inheritances of succeeding generations. We know that in generation t bequests result as $b_{jt} = \gamma_b(e_{jt} + x_j - c_a(1 + \tau_c))$, with received inheritances e_{jt} being a random variable, $j = 1, ..., n$. In vector notation, with column vectors $e_t = (e_{1t}, ..., e_{nt})^T$, $x = (x_1, ..., x_n)^T$ and $c_a^v = (c_a, ..., c_a)^T$, (14) reads as

$$E[e_{t+1} \mid e_t] = \gamma_b M(e_t + x - c_a^v(1 + \tau_c)) \tag{16}$$

and computing the expectation over all possible realizations of e_t gives[11]

$$E[e_{t+1}] = \gamma_b M(E[e_t] + x - c_a^v(1 + \tau_c)). \tag{17}$$

That is, the matrix M governs the evolution of expected values over time.

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. 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. The overall steady-state of this economy is defined by the condition that expected inheritances of each ability group also remain constant over time, that is

$$E[e] = \gamma_b M(E[e] + x - c_a^v(1 + \tau_c)) \tag{18}$$

---

[11]Formally, let H denote the index set of all possible realizations e_{th} of the vector e_t, with associated probabiliies $\kappa_{th}, h \in H$ (thus $\kappa_{t1} + \kappa_{t2} + ... = 1$). Then $E[e_{t+1}] = \sum_{h \in H} \kappa_{th} E[e_{t+1} \mid e_{th}] = \sum_{h \in H} \kappa_{th} \gamma_b M(e_{th} + x - (1 + \tau_c)c_s^v) = \gamma_b M(E[e_t] + x - (1 + \tau_c)c_s^v)$.

or

$$E[e] = A(x - c_a^v(1 + \tau_c)), \tag{19}$$

with $A \equiv \gamma_b(I - \gamma_b M)^{-1}M$ and I denoting the unit matrix. In view of the formula for a geometric series, $(I - \gamma_b M)^{-1} = \sum_{t=0}^{\infty}(\gamma_b M)^t$, equation (19) can be written as

$$E[e] = \sum_{t=1}^{\infty} \gamma_b^t M^t(x - c_a^v(1 + \tau_c)), \tag{20}$$

which suits well to our earlier result (see (15)) 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, the column vector 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 mentioned in the introduction, there is substantial empirical evidence that parent and descendant income is positively correlated, therefore it seems indeed very plausible that condition (C) is fulfilled in reality. One observes immediately from (20) that whether M fulfills these condition is also essential for the question of how incomes and inheritances are related.

Lemma 1 *The matrix 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), then also A has this property; that is expected inheritances are increasing with abilities.*

Proof. We know that A can be written as $\sum_{t=1}^{\infty}(\gamma_b M)^t$. Elementary properties of matrix multiplication show that in each term of the infinite series the row vectors have components

adding up to γ_b^t (remember that components of the rows of M add up to 1 and the same is true for M^t, for any t), which gives the sum $\sum_{t=1}^{\infty} \gamma_b^t = \gamma_b/(1 - \gamma_b)$. Multiplying by y, each term in the infinite series consists of the product $\gamma_b^t \, M \cdot M \cdot M....My$ (where M occurs t-times). Clearly, if My has increasing components, then $M^t y$ has increasing components as well, and this remains true after multiplication with γ_b^t and summing up the terms. ∎

A sufficient condition on M to fulfill (C) is that the components of the column vectors of M are increasing above and decreasing below the diagonal, with the diagonal element being the largest:

Lemma 2 *Let the μ_{ij}, the elements of M, have the following properties:*

$$\mu_{ii} > \mu_{ji} \qquad \text{for any } i,j = 1,..,n, i \neq j \tag{21}$$

$$\mu_{ij} \geq \mu_{i+1,j} \qquad \text{for } j < i \leq n-1, \tag{22}$$

$$\mu_{ij} \leq \mu_{i+1,j} \qquad \text{for } 1 \leq i \leq j-2. \tag{23}$$

Then M fulfills condition (C).

Proof. Assume that M has the properties (21) - (23). Consider two subsequent rows M_i and M_{i+1} of M. Construct $\widetilde{M_i}$ from M_i by reducing the first i components of M_i to those of M_{i+1} and increasing the component $\mu_{i,i+1}$ appropriately so that the sum of all components remains 1. For any vector y with increasing components y_i this gives $\widetilde{M_i}y > M_i y$. If $\widetilde{M_i}$ is equal to M_{i+1}, we are ready. Otherwise, in a second step, if some of the remaining components $i + 2, ..., n$ of $\widetilde{M_i}$ (namely, of M_i) are smaller than the corresponding ones of M_{i+1}, increase the former to the latter and decrease the $i+1$-component of $\widetilde{M_i}$ so that the sum of all components remains 1. The resulting vector is M_{i+1} and because of increasing y_i we altogether have $M_{i+1}y \geq \widetilde{M_i}y > M_i y$. ∎

In fact, the property (C) is a condition on the transition matrix P, from which M is derived (see 12). 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. Then, clearly, if P fulfills (21) - (23), then M fulfills condition (C). Intuitively, the two cases are when (i) only

switches to neighboring ability groups are possible or when (ii) switches to other ability groups are equally likely:

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

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

(ii) $p_{ij} = d_i$ for $j \neq i$.

Then the probability matrix M is equal to the transition matrix P.

Proof. See Appendix. ∎

Finally we note for later use that M has the same property as P:

Lemma 4 *Let f be the steady-state distribution of abilities and the stochastic matrix M defined as in (12). Then*

$$f = fM. \tag{24}$$

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

5 Optimal Taxes

Formula (19) characterizes steady-state inheritances in terms of labor supply $l_j, j = 1, ..., n$ and the share γ_b of bequests which, in turn, depend on the tax parameters $\alpha, \sigma, \tau_b, \tau_c$, 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 taxes on bequests or consumption (or on both) are desirable instruments. It was already mentioned that, given an optimal tax on labor income, both are redundant in a static one-period context of the model when individuals receive no (or identical) inheritances, as a consequence of the Atkinson-Stiglitz result (note that we have assumed weakly separable preferences). However, previous work (Brunner and Pech 2012ab) has shown that this result does not apply in a model with (exogenously given) differing inheritances, which represent a second characteristic of

individuals in addition to differing abilities.[12] Now, in the model of this paper, inheritances in the steady-state are not exogenous, but are ultimately determined by labor income of the different groups. It is therefore a priori unclear whether taxes on bequests or consumption represent desirable additional instruments 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 $\widetilde{\varphi}(c_i - c_a, b_i)$: expected utility $\sum_{h \in H} \kappa_{ih} \widetilde{\varphi}(\gamma_c(e_h + x_i - c_a(1 + \tau_c)), \gamma_b(e_h + x_i - c_a(1 + \tau_c)))$, where e_h denotes all possible realizations of e with associated probabilities κ_{ih}, is equal to utility from expected inheritances $\widetilde{\varphi}(\gamma_c(E[e_i] + x_i - c_a(1 + \tau_c)), \gamma_b(E[e_i] + x_i - c_a(1 + \tau_c)))$.[13] This allows us to formulate the problem of maximizing steady-state social welfare in the following way: let $\rho > 0$ denote the parameter of inequality aversion and remember that $\varphi(c_i, b_i) = \widetilde{\varphi}(c_i - c_a, b_i)$, then the task is to

$$\max \sum_{i=1}^{n} \frac{f_i}{1 - \rho} [\varphi(\gamma_c(E[e_i] + x_i - c_a(1 + \tau_c)) + c_a, \gamma_b(E[e_i] + x_i - c_a(1 + \tau_c))) - g(l_i)]^{1-\rho},$$
(25)

$$\text{s.t. } \sigma \sum_{i=1}^{n} f_i w_i l_i - \alpha + \tau_b \sum_{i=1}^{n} f_i \gamma_b(E[e_i] + x_i - c_a(1 + \tau_c))$$

$$+ \tau_c \sum_{i=1}^{n} f_i(\gamma_c(E[e_i] + x_i - c_a(1 + \tau_c)) + c_a) \geq 0,$$
(26)

$$E[e_i] = A_i(x - c_a^v(1 + \tau_c)),$$
(27)

where A_i denotes the i-th row vector of the matrix $A = \gamma_b(I - \gamma_b M)^{-1}M$ and x is the column vector of net incomes with components $x_i = \alpha + (1 - \sigma)w_i l_i$. Remember that V_i is indirect utility (9) and $\lambda = \partial V_i / \partial x_i$ is the marginal utility of income, independent of w_i. 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. With no (or identical) inheritances indirect utility V_i obviously rises with i because of a higher net wage. Moreover, we know from Lemma (1) that, given property (C),

---

[12]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 linar tax, given linear homogeneity of $\widetilde{\varphi}$.

[13]Remember that the realizations of e are the same for each ability group, but the associated probabilities differ.

expected inheritances increase with abilities which indeed strengthens the fact that $V_i < V_{i+1}$. Hence, social marginal utility $V_i^{-\rho}$ decreases if $\rho > 0$. ∎

We begin our study of the optimal tax system by analyzing the welfare effect of the introduction of a linear income tax with parameters α and σ. That is, we determine the change of social welfare if σ is marginally increased from 0, as is α to satisfy the resource constraint.[14] Let $\overline{wl} = \sum_{i=1}^{n} f_i w_i l_i$ be the average gross income.

Proposition 1 *Let $\tau_b = \tau_c = 0$. The welfare effect of an introduction of a linear income tax is*

$$\lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(\overline{wl} - w_i l_i) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} A_i \begin{pmatrix} \overline{wl} - w_1 l_1 \\ \vdots \\ \overline{wl} - w_n l_n \end{pmatrix} + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} A_i \begin{pmatrix} w_1 \frac{\partial l_1}{\partial \sigma} \\ \vdots \\ w_n \frac{\partial l_n}{\partial \sigma} \end{pmatrix}. \quad (28)$$

The first term is positive, and the second term is positive as well if M fulfills property (C). The third term is negative. The overall effect is positive if the labor-supply reaction is not too large.

Proof. See Appendix. ∎

A linear income tax with a marginal tax rate $\sigma > 0$ collects higher revenues from more able individuals than from less able, but returns the same amount α to everyone, which means redistribution. In a static one-period Mirrlees-model such a redistribution through a positive marginal tax rate is always desirable to some extent (Hellwig 1986, Brunner 1989, among others), because its distorting effect is of second order and therefore zero at $\sigma = 0$. Formally, in the static model only the first term of (28) in Proposition 1 occurs, which is unambiguously positive because $\sum_{i=1}^{n} f_i(\overline{wl} - w_i l_i) = 0$ and the weights $V_i^{-\rho}$ are higher for small i (that is, for positive values of $\overline{wl} - w_i l_i$) than for large i (for negative values of $\overline{wl} - w_i l_i$), given increasing $w_i l_i$.

---

[14]Note that introducing some government requirement G to finance public expenditures into the resource constraint (26) would not change the results of the upcoming analysis. The only difference would be that with $\sigma = \tau_b = \tau_c = 0$ a negative lump-element $\alpha = -G$ is needed.

In our steady-state model, redistribution via the introduction of a linear tax on labor income also affects the amount of bequests left by each ability type, and the long-run consequences on inheritances can be seen from the remaining terms of (28) in Proposition 1. If M fulfills condition (C) then expected inheritances in the steady state (described by A_i times the vector $(x - c_a^v)$ where $x_i = w_i l_i$ at $\sigma = \alpha = 0$) increase with abilities and the second term shows that the introduction of the linear income tax - by redistributing income - also makes the distribution of inheritances more equal.

As to the third term, note that in the static model without inheritances the welfare loss caused by the introduction of the income tax (at $\sigma = 0$), via its negative impact on labor supply and income, is just compensated by the welfare gain through the associated increase in leisure (the deadweight loss is zero). In the steady-state model with inheritances an additional, indirect welfare loss, visible in the third - negative - term of (28), occurs because the reduction of labor supply and income transforms into lower long-run bequests.

The overall welfare effect of introducing a linear income tax is therefore positive if the labor supply elasticity is small. Moreover, a positive overall effect is the more likely, the larger the first two terms in (28), that is, the larger the spread of individual incomes and the degree of inequality aversion.

In the following we take as given that a positive marginal tax rate σ is optimal, which means that some redistribution of income is desirable. Obviously, with $\sigma > 0$ the distorting effect on labor supply of the marginal tax rate becomes increasingly important, and $\sigma < 1$ must hold in the optimum, otherwise labor supply would be zero, as follows from (6) with $g'(l_i) > 0$. We now study the welfare effects of taxes on bequests and on consumption and ask whether their introduction makes sense as a supplement to an optimally chosen labor income tax.

Let in the following $\bar{e} \equiv \sum_{i=1}^{n} f_i E[e_i]$ denote average inheritances received by a generation, and $\bar{e}^v = (\bar{e}, ..., \bar{e})^T$ denote the corresponding column vector. Moreover, for any matrix Z and tax parameter θ we use the notation that $[\partial Z / \partial \theta]_i$ is row i of the matrix derived from Z by differentiating each component with respect to θ. With $S(\tau_b, \tau_c)$ denoting the optimal value function of the maximization problem (25) - (27), we find for the bequest tax:

Proposition 2 *Let $\tau_c = 0$. The welfare effect of an introduction of a tax τ_b on bequests, given that the linear income tax is chosen optimally, is*

$$
\begin{aligned}
\left. \frac{\partial S}{\partial \tau_b} \right|_{\tau_b=0} = & \; \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \gamma (\bar{e} - E[e_i]) \\
& + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left[\frac{\partial A}{\partial \tau_b} \right]_i (x - c_a^v) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \gamma A_i (\bar{e}^v + x - c_a^v). \quad (29)
\end{aligned}
$$

The first term is positive, given that M fulfills property (C); the second term is negative, and the third term is positive.

Given that M fulfills property (C), the overall effect on social welfare is positive (i) if the elasticity of substitution between bequests and consumption is small, or (ii) if inequality aversion is large and expected inheritances of the lowest-income group are small.

Proof. See Appendix. ∎

Intuitively, the introduction of the bequest tax causes two separate effects: the welfare effect for given inheritances (that is, holding inheritances fixed), which is the term in the first line of the above formula, and the welfare effect due to changes in steady-state inheritances, which is described by the second line. The first effect shows the main role of the bequest tax: it allows additional redistribution of $\gamma E[e_i]$, which is that particular part of (expected) bequests financed out of (expected) received inheritances. If these are increasing with i (M and hence A fulfill condition (C), that is, more able individuals receive larger inheritances), then the term in the first line of (29) is positive, because the weights $V_i^{-\rho}$ are decreasing. The point is that the bequest tax is an instrument for equalization of received inheritances; it redistributes the latter indirectly via the share γ used for bequests. (Clearly, the bequest tax is also imposed on the other part of bequests, γx_i), which is financed out of net labor income. But for these - as the Atkinson-Stiglitz result tells us - no further redistribution is possible in addition to that performed by the optimal income tax.) The first term also occurs in the static model with exogenous inheritances (Brunner and Pech 2012b).

The second effect (second line of (29)) consists of two terms. The first (which is negative) describes how the introduction of τ_b affects the accumulation process as it reduces

17

the share γ_b of income devoted to bequests. Note from the definition of A_i that it is indeed only through the change of γ_b that τ_b influences A_i. The last term (which is positive) in (29) occurs because steady-state inheritances increase as a consequence of the fact that the revenues from the bequest tax are redistributed via a lower marginal income tax σ and a larger lump-sum element α.

In order to gain further insights, we determine the effect of τ_b on A_i (see the Appendix) which allows us to rewrite the second line of (29) as

$$\lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \gamma (A_i \bar{e}^v - \gamma B_i (x - c_a^v)) - \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \varepsilon (1 - \gamma)(A_i + \gamma B_i)(x - c_a^v), \qquad (30)$$

where $\varepsilon \geq 0$ denotes the elasticity of substitution between b_i and $c_i - c_a$, and B_i is row i of the matrix $B \equiv (I - \gamma M)^{-1} AM$. The first part of (30), which is similar to the first line of (29), is positive, as is shown in the Appendix. Therefore, expression (30) is positive for $\varepsilon = 0$. For an illustration assume that there is complete immobility in our economy (all bequests remain within in the same ability group). That is, $P = M = I$, then $(I - \gamma M)^{-1} = 1/(1 - \gamma)I$ and, thus, $A_i = \gamma/(1 - \gamma)$ and $B_i(x - c_a^v) = E[e_i]/(1 - \gamma)$, and for $\varepsilon = 0$ the expression (30) reduces to $\lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (\gamma^2/(1 - \gamma))(\bar{e} - E[e_i])$ which is positive for the same reason as given for the first term in (29). Intuitively, a zero substitution elasticity means that the share spent for bequests does not change with the tax τ_b; the latter only reduces the available budget but does not distort the accumulation process.

Hence a bequest tax increases social welfare if the share of bequests does not react too much, that is if ε is small. For a larger value of the elasticity of substitution a positive welfare effect of the bequest tax arises if the inequality aversion is large and expected inheritances of the lowest income groups are low. The reason is that for large ρ only the low-ability types count in the the social welfare function, and in (30) the second term goes to zero, for any ε, if $A_i(x - c_a^v) = E[e_i]$ goes to zero for small i (thus also $B_i(x - c_a^v)$ goes to zero; net income just suffices to cover necessary consumption). As an illustration consider the limit case when the social welfare function is maximin (ρ goes to infinity), then the sums in the formula (29) of Proposition (2) reduce to the expressions for $i = 1$. Moreover assume that the lowest income group does not receive inheritances from any other group,

18

that is, $A_1 = (\gamma/(1-\gamma), 0, 0, 0)$ and $B_1 = (\gamma/(1-\gamma)^2, 0, 0, 0)$. With x_1 close to c_a the second, negative term in (30) is close to zero and the overall effect is positive.

It is interesting to note that in case of complete mobility over generations, when all $E[e_i]$ are equal to $\bar{e}$ (then also all $B_i(x - c_a^v)$ are identical), the welfare effect of the tax on bequests is zero or negative, depending on whether ε is zero or positive.[15] There is obviously no reason for a redistribution of inheritances, thus no positive welfare effect occurs, while the distortion of the bequest decision remains.

Next we turn to the welfare effect of a tax on consumption:

Proposition 3 *Let $\tau_b = 0$. The welfare effect of an introduction of a tax τ_c on consumption, given that the linear income tax is chosen optimally, is*

$$\left.\frac{\partial S}{\partial \tau_c}\right|_{\tau_c=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(1-\gamma)(\bar{e} - E[e_i])$$
$$+ \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}\left[\frac{\partial A}{\partial \tau_c}\right]_i (x - c_a^v) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(1-\gamma)A_i(\bar{e}^v + x - c_a^v), \quad (31)$$

The first term is positive, given that M fulfills property (C); the sign of the second term depends on the elasticity of substitution ε: it is negative for $\varepsilon < 1$, zero for $\varepsilon = 1$ and positive for $\varepsilon > 1$; the third term is positive.
The overall effect on social welfare is positive for any $\varepsilon \geq 0$, if M fulfills property (C).

Proof. See Appendix. ∎

There is obviously a close analogy between the Propositions (2) and (3). As in the case of the bequest tax, the welfare effect of introducing a tax τ_c on consumption can be split into two separate effects: the first line of (31) shows that the consumption tax allows additional redistribution of that part of consumption which is financed out of (given) expected inheritances. This effect is positive for the same reason as discussed above, provided that M fulfills property (C). The second line describes how the introduction of τ_c affects wealth accumulation and, thus, inheritances in the steady state. Again, the first

---

[15]Complete mobility means that all elements of f and of the transition matrix P are equal to $1/n$. It is straightforward to show that then all elements of A are identical, equal to $\gamma/((1-\gamma)n)$; hence the vector $A(x - c_a^v)$ has identical elements.

term refers to the change in the share of income devoted to bequests, while the second term refers to increased consumption which is made possible because revenues from the consumption tax are returned to the households and induce higher wealth. We show in the Appendix that the second line of (31) can be expressed as

$$\lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (1-\gamma)(A_i \bar{e}^v - \gamma B_i(x - c_a^v)) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \varepsilon (1-\gamma)(A_i + \gamma B_i)(x - c_a^v). \quad (32)$$

The first term of (32) is the same as the first term in (30) with $(1-\gamma)$ instead of γ and the identical argument tells us that it is positive given that M fulfills property (C). The second term of (32) is obviously nonnegative for any $\varepsilon \geq 0$. Intuitively, the overall conclusion is that on the one hand the redistributive effect of the consumption tax is positive and similar to that of the bequest tax, but on the other hand the tax affects the individual decision in the opposite way by penalizing consumption and favoring bequests. The resulting increase in steady-state inheritances outweighs the distorting effect, which is the ultimate reason for the unambiguously positive welfare effect. This logic does not apply for the bequest tax.

Let us finally note that in case of complete mobility (identical inheritances) the just-mentioned mechanism is still responsible for a positive welfare effect of the consumption tax, though there is clearly no redistributive effect.

To complete the analysis we also report the effect of a common proportional tax on all expenditures, on consumption as well as bequests. Let $\tau = \tau_b = \tau_c$ denote such a tax rate.

Proposition 4 *The welfare effect of an introduction of a tax τ on expenditures for consumption and bequests, given that the linear income tax is chosen optimally, is*

$$
\begin{aligned}
\left. \frac{\partial S}{\partial \tau} \right|_{\tau=0} &= \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (\bar{e} - E[e_i]) \\
&\quad + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left[\frac{\partial A}{\partial \tau} \right]_i (x - c_a^v) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} A_i(\bar{e}^v + x - c_a^v). \quad (33)
\end{aligned}
$$

The first term is positive, given that M fulfills condition (C). The second term is negative

and the third term is positive.

The overall effect on social welfare is positive, given that M fulfills condition (C).

Proof. See Appendix. ∎

The interpretation of the terms is analogous to the corresponding terms in the previous propositions. Most importantly, the expenditure tax when combined with an optimally chosen income tax essentially allows redistribution of inherited wealth of the groups, as the first term in (33) shows.[16] Again, this redistribution has the desired effect only if inheritances and abilities are positively related. The second line of (33) is found as the sum of (30) and (32). The respective second terms cancel out and we know from earlier considerations that the first terms are positive. Thus, the whole welfare effect is unambiguously positive, it is clearly just the sum of the effects of the two single taxes presented in the above Propositions. Therefore it is larger than the welfare effect of the consumption tax alone if the bequest tax is welfare increasing. In case of complete mobility (identical inheritances) the welfare effect of the expenditure tax is zero.

The above formulas describing the welfare effects of the introduction of indirect taxes showed us the mechanisms at work. No important further insights can be derived from the conditions characterizing optimal tax rates τ_b and τ_c. Formally, one arrives at these conditions by setting the first derivatives of the Lagrangian function to zero. Intuitively, with tax rates τ_b and τ_c being larger than zero, their further increase causes increasingly negative effects on the respective tax basis, and in the optimum these are large enough to balance the positive redistributive effects.

6   Discussion and concluding remarks

What do we learn from these findings for the design of an optimal tax system? First of all, we have seen that the Aktinson-Stiglitz result on the redundancy of other taxes in addition to an optimal income tax does not hold any more if the process of wealth accumulation via saving and bequests is taken into account. It was already found in earlier studies (Saez

---

[16]This fact was discussed in more detail by 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.

2002, Brunner and Pech 2012ab) that a case for indirect taxes or capital taxes arises in a model where individuals differ in a second characteristic, not only in labor productivities. However, in the present model such taxes were shown to be useful although the only source of heterogeneity is again labor productivity. The reason are differences in inherited wealth which arise endogenously as a consequence of differing labor incomes. A basic assumption for our results is that the wealth transfer is characterized by the property that on average more able individuals have more able parents, thus they inherit more than less able individuals.

In our model the additional instruments are taxation of consumption and bequests, and we have established their social-welfare consequences, which involves weighing the redistributive potential against the utility loss from a distortion of individual decisions. It turned out that from this perspective a tax on consumption is the most powerful instrument. It performs redistribution, and its distortive effect favors leaving bequests. Note that the increase in bequests is amplified as we consider the long-run consequences: the increase in steady-state inheritances can be interpreted as the sum of all gains if the favorable effect occurs each year again. This amplification of inheritances (and consumption) outweighs the distortion. For this reason the consumption tax is a desirable supplement to an optimal income tax even if there is complete equality in inheritances.

In contrast, the bequest tax can be a useful instrument because of its redistributive potential. Its distortive effect penalizes bequests and the just described accumulation effect amplifies the decrease of steady-state inheritances. Therefore the reaction of bequests has to be small in order that the bequest tax increases social welfare. More generally, the bequest tax is desirable the more society wants to promote equality and the less the amount of inheritances going to low income groups.

It should be stressed again that for both taxes the ultimate cause for redistribution are differences in received steady-state inheritances of the groups. As can clearly be seen from the formulas of Propositions 2 and 3, what these taxes effectively redistribute is only that fraction of bequests (or consumption, respectively), which is financed out of (received) inheritances. It follows that an expenditure tax, imposed on both consumption and bequests at a uniform rate, effectively redistributes total steady-state inheritances.

Note also that the parameter γ determines the redistributive potential of the bequest tax vis-a-vis the consumption tax.[17]

Our results indicate that in an optimal tax system the income tax should be supplemented by a consumption tax and - particularly when society emphasizes redistribution - by a bequest tax. This fits well to the structure of actual tax systems in OECD countries (OECD 2012, p. 104) which indeed rely heavily on consumption taxation (such as the VAT in the European Union), in addition to the income tax (and to income-dependent social security contributions). Taxation of property or the transfer of property plays a minor role, though a bequest or inheritance tax exists in many countries (see OECD 2010, p. 32f). To our knowledge, prior contributions to optimal taxation theory did not provide an explanation why consumption taxation - in addition to income taxation - is so dominant.

In the last decades there has been an ongoing discussion on the merits of replacing income taxation by consumption taxation. This was advocated for instance by Kaldor (1955) and Meade (1978), with the main argument being the distortion of the savings decision caused by a comprehensive income tax whose base includes capital income. Our analysis has a different focus, namely the consequences of unequal inheritances, and it comes to a different conclusion as it provides a reason for taxing consumption in addition to labor income. Moreover, in our model the bequest tax can be interpreted as a kind of capital taxation, and it was shown to be a potential instrument for redistribution. One may argue that this role would be even more pronounced if the proportional tax were combined with a tax allowance or if progressive rates were applied. Such a schedule could be introduced because typically the overall amount of the estate is reported to the tax authority. In contrast, consumption is usually taxed at each single purchase of a good or service which is incompatible with a progressive schedule; the latter would require households to report total consumption as the difference between household income and savings (as suggested by Meade 1978).

Modeling bequests as being motivated by joy of giving allows us to consider social

---

[17]A limitation of our model is that γ is the same for all households. One may assume that γ increases with income which is more likely to conform to reality. Presumably, such an assumption would strengthen the redistributive potential of the bequest tax and weaken redistribution by the consumption tax, especially in the case of inheritances being concentrated on high income.

welfare of each generation separately and to ask how it is affected by redistributive taxes. 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 descendants. It is difficult to see how in such a perfect-foresight framework the idea of wealth transmission across ability groups and the role of taxes can be adequately modeled.[18] Furthermore, empirical studies do not suggest that the altruistic model is more in accordance with actual behavior (see Kopczuk 2013b, among others).

A well-known issue when evaluating public policy in terms of intergenerational social welfare is double counting. In the present context, this concerns the question of whether the social evaluation of some tax should include the effect on both, welfare of future generations (the external effect) as well welfare of the bequeathing individual, or only on one of them. One may assume that when deciding about bequests the bequeathing individual already takes the welfare of future generations into account; then its separate consideration can be regarded as double-counting.[19] The latter implies, as a rule, that more significance is given to the negative effects from the distortion of the bequest decision. Let us note that our discussion of steady-state effects means that we calculate with "full" double counting. That is, if some tax change influences bequests left by a generation, we measure the direct welfare effect and we include all future welfare consequences by taking into account that received inheritances of this (steady-state) generation immediately change as well. Remember from Section 4 that steady-state inheritances result as the infinite sum of prior bequests, with decreasing weights. In this context note also that our analysis obviously does not say anything on the welfare trade-off between present and future generations, arising for individuals on the transition path from one steady state to another.

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

---

[18]Piketty and Saez (forthcoming) also develop their main formula in a model with a joy-of-giving motive motive for leaving bequests. Their analysis of a model with dynastic preferences draws on their assumption of stochastic utility functions which we do not employ.

[19]In the literature there is no unanimity whether double counting produces the correct representation of social welfare or not (see Cremer 2003, among others).

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.

Appendix

Proof of Lemma 3

(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 (10) on f we have $f_1(1 - p_{12}) + f_2 p_{21} = f_1$, or $f_2/f_1 = p_{12}/p_{21}$, which implies (see (12)) $\mu_{12} = f_2 p_{21}/f_1 = p_{12}$ 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 the main diagonal are equal to a constant, that is, $p_{ij} = d_i$ for $i \neq j$. Then the steady-state condition (10) reads as $f_j(1 - (n-1)d_j) + \sum_{i \neq j} f_i d_i = f_j$, or $f_j = (\sum_{i=1}^n f_i d_i)/(nd_j)$ for any $j = 1, ..., n$. This implies $f_j/f_i = d_i/d_j$ and further $\mu_{ij} = d_j d_i/d_j = p_{ij}, i, j = 1, ..., n, i \neq j$. Clearly, then also $\mu_{ii} = p_{ii}$ for all i.

Proof of Proposition 1

We determine the first-order condition for α, with $\mathcal{L}$ as the Lagrangian and ν as the Lagrange multiplier to (26):

$$\frac{\partial \mathcal{L}}{\partial \alpha} : \quad \sum_{i=1}^n f_i V_i^{-\rho} [\frac{\partial \varphi(.)}{\partial c_i} \gamma_c (\frac{\partial E[e_i]}{\partial \alpha} + 1 + (1 - \sigma)w_i \frac{\partial l_i}{\partial \alpha}) + \frac{\partial \varphi(.)}{\partial b_i} \gamma_b (\frac{\partial E[e_i]}{\partial \alpha} + 1$$
$$+ (1 - \sigma)w_i \frac{\partial l_i}{\partial \alpha}) - g' \frac{\partial l_i}{\partial \alpha}] + \nu (\sigma \sum_{i=1}^n f_i w_i \frac{\partial l_i}{\partial \alpha} - 1) = 0. \tag{34}$$

Using the first-order conditions (4) - (6) of the individual optimization problem, together with the definitions of γ_c and γ_b, and $\partial l_i/\partial \alpha = 0$, (34) can be transformed to

$$\frac{\partial \mathcal{L}}{\partial \alpha} : \quad \lambda \sum_{i=1}^n f_i V_i^{-\rho} (\frac{\partial E[e_i]}{\partial \alpha} + 1) - \nu = 0. \tag{35}$$

We proceed in a similar way, to derive the first-order condition for σ as

$$\frac{\partial \mathcal{L}}{\partial \sigma}: \quad \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left(\frac{\partial E[e_i]}{\partial \sigma} - w_i l_i \right) + \nu \sum_{i=1}^{n} f_i \left(w_i l_i + \sigma w_i \frac{\partial l_i}{\partial \sigma} \right) = 0. \tag{36}$$

To determine the welfare effect of introducing a linear income tax, we substitute for ν from (35) in $\partial \mathcal{L}/\partial \sigma$ and evaluate at $\sigma = 0$. After some transformations we obtain

$$\left. \frac{\partial \mathcal{L}}{\partial \sigma} \right|_{\sigma=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (\overline{wl} - w_i l_i) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left(\frac{\partial E[e_i]}{\partial \alpha} \overline{wl} + \frac{\partial E[e_i]}{\partial \sigma} \right). \tag{37}$$

In view of $\sum_{i=1}^{n} f_i (\overline{wl} - w_i l_i) = \overline{wl} - \overline{wl} = 0$, and $\overline{wl} - w_1 l_1 > \overline{wl} - w_2 l_2 > \cdots > \overline{wl} - w_n l_n$, the first term on the RHS of (37) is positive if M has the property (C), because then $V_1^{-\rho} > V_2^{-\rho} > \cdots > V_n^{-\rho}$ by Lemma 5. From the steady-state equation (27) for $E[e_i]$ we find

$$\frac{\partial E[e_i]}{\partial \alpha} = A_i \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \tag{38}$$

$$\frac{\partial E[e_i]}{\partial \sigma} = A_i \begin{pmatrix} -w_1 l_1 + (1-\sigma) w_1 \frac{\partial l_1}{\partial \sigma} \\ \vdots \\ -w_n l_n + (1-\sigma) w_n \frac{\partial l_n}{\partial \sigma} \end{pmatrix}. \tag{39}$$

Thus at $\sigma = 0$ the second expression in (37) reads as

$$\lambda \sum_{i=1}^{n} f_i V_i^{-\rho} A_i \begin{pmatrix} \overline{wl} - w_1 l_1 \\ \vdots \\ \overline{wl} - w_n l_n \end{pmatrix} + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} A_i \begin{pmatrix} w_1 \frac{\partial l_1}{\partial \sigma} \\ \vdots \\ w_n \frac{\partial l_n}{\partial \sigma} \end{pmatrix}. \tag{40}$$

If M has the property (C) then $A(w_1 l_1, ..., w_n l_n)^T$ has increasing components and the first term of (40) is positive, for the same reason as discussed above for the first term in (37). The second term in (40) is negative because $\partial l_i/\partial \sigma$ is negative.

Proof of Proposition 2

Using the Envelope Theorem, we get for the optimal value function $S(\tau_b, \tau_c)$ of the maximization problem (25) and (26)

$$
\begin{aligned}
\frac{\partial S}{\partial \tau_b} &= \sum_{i=1}^{n} f_i V_i^{-\rho} [\frac{\partial \varphi(\cdot)}{\partial c_i} \{\gamma_c(\frac{\partial E[e_i]}{\partial \tau_b} + (1-\sigma)w_i \frac{\partial l_i}{\partial \tau_b}) + \frac{\partial \gamma_c}{\partial \tau_b}(E[e_i] + x_i - (1+\tau_c)c_a)\} \\
&\quad + \frac{\partial \varphi(\cdot)}{\partial b_i}\{\gamma_b(\frac{\partial E[e_i]}{\partial \tau_b} + (1-\sigma)w_i \frac{\partial l_i}{\partial \tau_b}) + \frac{\partial \gamma_b}{\partial \tau_b}(E[e_i] + x_i) - (1+\tau_c)c_a\} - g'\frac{\partial l_i}{\partial \tau_b}] \\
&\quad + \nu \sum_{i=1}^{n} f_i [\sigma w_i \frac{\partial l_i}{\partial \tau_b} + (\gamma_b + \tau_b \frac{\partial \gamma_b}{\partial \tau_b})(E[e_i] + x_i - (1+\tau_c)c_a) + (\tau_b \gamma_b \\
&\quad + \tau_c \gamma_c)(\frac{\partial E[e_i]}{\partial \tau_b} + (1-\sigma)w_i \frac{\partial l_i}{\partial \tau_b})].
\end{aligned}
\tag{41}
$$

We use $\gamma_c = (1-\gamma)/(1+\tau_c)$, $\gamma_b = \gamma/(1+\tau_b)$, their respective derivatives $\partial \gamma_c/\partial \tau_b = -(\partial \gamma/\partial \tau_b)/(1+\tau_c)$ and

$$
\frac{\partial \gamma_b}{\partial \tau_b} = \frac{\partial \gamma/\partial \tau_b}{1+\tau_b} - \frac{\gamma}{(1+\tau_b)^2} = \frac{\partial \gamma/\partial \tau_b - \gamma_b}{1+\tau_b}
\tag{42}
$$

as well as the f.o.c.'s (4) - (6) of the individual optimization problem and evaluate (41) at $\tau_c = \tau_b = 0$. By this, (41) reduces to

$$
\begin{aligned}
\frac{\partial S}{\partial \tau_b}\Big|_{\tau_b=0} &= \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}[\frac{\partial E[e_i]}{\partial \tau_b} - \gamma(E[e_i] + \alpha + (1-\sigma)w_i l_i - c_a)] \\
&\quad + \nu \sum_{i=1}^{n} f_i(\sigma w_i \frac{\partial l_i}{\partial \tau_b} + \gamma(E[e_i] + \alpha + (1-\sigma)w_i l_i - c_a)),
\end{aligned}
\tag{43}
$$

where we have inserted $x_i = \alpha + (1-\sigma)w_i l_i$. By use of the condition (36) for optimal σ, multiplied by $\gamma(1-\sigma)$, and observing that $\partial l_i/\partial \tau_b = \gamma(1-\sigma)\partial l_i/\partial \sigma$[20], we can write (43)

---

[20]$\partial l_i/\partial \tau_b = (\partial l_i/\lambda)(\partial \lambda/\partial \tau_b)$, and from Section 2 we know that $\lambda = \varphi((1-\gamma) + c_a, \gamma/(1+\tau_b))$. Hence at $\tau_b = 0$ we have $\partial \lambda/\partial \tau_b = -(\partial \varphi/\partial c_i)(\partial \gamma/\partial \tau_b) + (\partial \varphi/\partial b_i)(\partial \gamma/\partial \tau_b - \gamma) = -\gamma \lambda$ (by use of (4) and (5)). Implicit differentiation of (6) with respect to σ and τ_b, respectively, gives the above relation.

as

$$\left.\frac{\partial S}{\partial \tau_b}\right|_{\tau_b=0} = -\lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(\gamma(E[e_i] + \alpha - c_a)) + \nu\gamma(\overline{e} + \alpha - c_a)$$
$$+\lambda \sum_{i=1}^{n} f_i V_i^{-\rho}\frac{\partial E[e_i]}{\partial \tau_b} - \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}\gamma(1 - \sigma)\frac{\partial E[e_i]}{\partial \sigma}, \qquad (44)$$

which, after substitution for ν by use of the condition (35) for optimal α, is further transformed to

$$\left.\frac{\partial S}{\partial \tau_b}\right|_{\tau_b=0} = \lambda\gamma \sum_{i=1}^{n} f_i V_i^{-\rho}(\overline{e} - E[e_i]) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}[\frac{\partial E[e_i]}{\partial \tau_b}$$
$$-\gamma(1 - \sigma)\frac{\partial E[e_i]}{\partial \sigma} + \gamma(\overline{e} + \alpha - c_a)\frac{\partial E[e_i]}{\partial \alpha}]. \qquad (45)$$

Next, we use (27) to compute $\partial E[e_i]/\partial \tau_b$, and substitute this, together with (39) and (38), into (45), which yields

$$\left.\frac{\partial S}{\partial \tau_b}\right|_{\tau_b=0} = \lambda\gamma \sum_{i=1}^{n} f_i V_i^{-\rho}(\overline{e} - E[e_i]) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}\left\{\left[\frac{\partial A}{\partial \tau_b}\right]_i (x - c_a^v) + A_i \begin{pmatrix} (1 - \sigma)w_1\frac{\partial l_1}{\partial \tau_b} \\ \vdots \\ (1 - \sigma)w_n\frac{\partial l_n}{\partial \tau_b} \end{pmatrix} \right.$$
$$\left. + A_i\gamma \begin{pmatrix} -(1 - \sigma)^2 w_1\frac{\partial l_1}{\partial \sigma} + (1 - \sigma)w_1 l_1 + \alpha - c_a + \overline{e} \\ \vdots \\ -(1 - \sigma)^2 w_n\frac{\partial l_n}{\partial \sigma} + (1 - \sigma)w_n l_n + \alpha - c_a + \overline{e} \end{pmatrix} \right\}. \qquad (46)$$

By use of $\partial l_i/\partial \tau_b = \gamma(1 - \sigma)\partial l_i/\partial \sigma$ (see above) and $x_i = \alpha + (1 - \sigma)w_i l_i$, (46) reduces to

$$\left.\frac{\partial S}{\partial \tau_b}\right|_{\tau_b=0} = \lambda\gamma \sum_{i=1}^{n} f_i V_i^{-\rho}(\overline{e} - E[e_i]) +$$
$$+\lambda \sum_{i=1}^{n} f_i V_i^{-\rho}\left[\frac{\partial A}{\partial \tau_b}\right]_i (x - c_a^v) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}\gamma A_i(x - c_a^v + \overline{e}^v), \qquad (47)$$

which is the formula (29) in Proposition 2. The first term is positive if M fulfills condition (C), because then $V_1^{-\rho} > V_2^{-\rho} > \cdots > V_n^{-\rho}$ (Lemma 5), and we have $\overline{e} - E[e_1] > \overline{e} - E[e_2] > \cdots > \overline{e} - E[e_n]$, in addition to $\sum_{i=1}^{n} f_i(\overline{e} - E[e_i]) = \overline{e} - \overline{e} = 0$. The last term

$\lambda \sum_{i=1}^{n} \gamma A_i (x - c_a^v + \bar{e}^v)$ is positive. It remains to 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 τ_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 τ_b. We know that $(I - \gamma_b M)^{-1} = \sum_{t=0}^{\infty} (\gamma_b M)^t$, thus

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

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

$$\frac{\partial A}{\partial \tau_b} = \frac{\partial \gamma_b}{\partial \tau_b} \frac{1}{\gamma_b} [A + \gamma_b (I - \gamma M)^{-1} A M], \tag{50}$$

where $\partial \gamma_b / \partial \tau_b$ is given by (42). In order to analyze $\partial \gamma_b / \partial \tau_b$, we introduce the elasticity of substitution between bequests and consumption (above minimum consumption c_a), which is defined as

$$\varepsilon = -\frac{\partial [b_i / (c_i - c_a)]}{\partial [p_b / p_c]} \frac{p_b / p_c}{b_i / (c_i - c_a)} \geq 0 \tag{51}$$

with $p_b = 1 + \tau_b$ and $p_c = 1 + \tau_c$. By use of (7) and (8) $b_i / (c_i - c_a)$ reduces to $(\gamma p_c) / ((1 - \gamma) p_b)$. Moreover, at $\tau_c = 0$, $p_b / p_c = p_b$. Hence, ε is given by

$$\varepsilon = -\frac{\partial [\gamma / ((1 - \gamma) p_b)]}{\partial p_b} \frac{(1 - \gamma) p_b^2}{\gamma} = \frac{-(\partial \gamma / \partial p_b) p_b + \gamma (1 - \gamma)}{\gamma (1 - \gamma)}. \tag{52}$$

Solving (52) explicitly for $\partial \gamma / \partial p_b$ and observing that $\partial p_b / \partial \tau_b = 1$, we arrive at

$$\left. \frac{\partial \gamma}{\partial \tau_b} \right|_{\tau_b = 0} = (1 - \varepsilon) \gamma (1 - \gamma) \tag{53}$$

and, further, (see (42)) at $\partial\gamma_b/\partial\tau_b|_{\tau_b=0} = -\gamma(\gamma + \varepsilon(1 - \gamma))$. By use of this term and $B \equiv (I - \gamma M)^{-1}AM$ in (50) the second term of (47) can be rewritten as

$$\lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \left[\frac{\partial A}{\partial \tau_b}\right]_i (x - c_a^v) = -(\gamma + \varepsilon(1 - \gamma))\lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(A_i + \gamma B_i)(x - c_a^v), \quad (54)$$

which is negative (because all components of the matrix $A+\gamma B$ are nonnegative). Inserting (54) in (47), some transformations give us

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

$$- \lambda f_i V_i^{-\rho}\varepsilon(1 - \gamma) \sum_{i=1}^{n}(A_i + \gamma B_i)(x - c_a^v). \quad (55)$$

Note that $A_i\bar{e}^v = \gamma\bar{e}/(1 - \gamma)$ because the components of A_i add up to $\gamma/(1 - \gamma)$. Note further that the average of the $B_i(x-c_a^v)$ is $\sum_{i=1}^{n} f_i B_i(x-c_a^v) = \bar{e}/(1-\gamma),$[21] and if M fulfills property (C) then also B fulfills property (C). As a consequence, the elements $B_i(x - c_a^v)$ are increasing with i and the difference $A_i\bar{e}^v - \gamma B_i(x - c_a^v)$ is positive for low i, with larger weights $V_i^{-\rho}$, and negative for large i, with lower weights. From these considerations it follows that the second term of (55) is positive, while, obviously, the third term is negative (zero), if $\varepsilon > 0$ ($\varepsilon = 0$). Altogether, by continuity (55) is positive for sufficiently small values of ε.

Next consider the limiting case $\rho \rightarrow \infty$, where the social welfare function (25) approaches the maximin form. Then the objective function of the maximization problem (25) - (27) is indirect utility V_1 of the group with the lowest ability; and all the sums in (55) reduce to $i = 1,$[22] i.e.,

$$\frac{\partial S}{\partial \tau_b}\bigg|_{\tau_b=0} = \lambda\gamma(\bar{e} - E[e_1]) + \lambda\gamma(A_1\bar{e}^v - \gamma B_1(x - c_a^v)) - \lambda\varepsilon(1 - \gamma)(A_1 + \gamma B_1)(x - c_a^v). \quad (56)$$

---

[21] By definition, $\bar{e}^v = fA(x - c_a^v) = f\sum_{t=1}^{\infty}(\gamma M)^t(x - c_a^v) = \sum_{t=1}^{\infty}\gamma^t f(x - c_a^v) = (\gamma/(1 - \gamma))f(x - c_a^v)$, where we have used the steady-state property $fM = f$ (Lemma 4). By similar considerations we get $fB(x - c_a^v) = f(\sum_{t=0}^{\infty}(\gamma M)^t)(\sum_{t=1}^{\infty}(\gamma M)^t M(x - c_a^v) = (\gamma/(1 - \gamma)^2)f(x - c_a^v)$.

[22] Note that the optimal linear income tax is derived by maximizing indirect utility V_1 of the group with the lowest ability subject to (26) and (27). As also in this case in the optimum the marginal tax rate σ on labor income is smaller than 1, net incomes increase with ability i, as does indirect utility V_i (see Lemma 5). Consequently, the welfare effect of introducing τ_b is derived by differentiating the Lagrangian with V_1 as the objective function.

The first two terms are positive by the same reasons as above. Let $E[e_1] = A_1(x - c_a^v)$ go to zero, which implies that $x_1 - c_a$ converges to zero, and let A_1 converge to $(\gamma/(1-\gamma), 0, 0, 0, ...)$. The latter in turn implies that B_1 converges to $(\gamma/(1-\gamma)^2, 0, 0, 0, ...)$. Therefore, if $E[e_1] \to 0$, the third term in (56) converges to zero for any $\varepsilon > 0$. By continuity, the overall welfare effect is also positive for small values of $E[e_1]$ and for large but finite values of ρ.

Proof of Proposition 3

We proceed as in the proof of Proposition 2 and determine the derivative of the optimal value function $S(\tau_b, \tau_c)$ with respect to τ_c

$$
\frac{\partial S}{\partial \tau_c} = \sum_{i=1}^{n} f_i V_i^{-\rho} [\frac{\partial \varphi(\cdot)}{\partial c_i} \{\gamma_c(\frac{\partial E[e_i]}{\partial \tau_c} + (1-\sigma)w_i\frac{\partial l_i}{\partial \tau_c} - c_a) + \frac{\partial \gamma_c}{\partial \tau_c}(E[e_i] + x_i - (1+\tau_c)c_a)\}
$$
$$
+ \frac{\partial \varphi(\cdot)}{\partial b_i}\{\gamma_b(\frac{\partial E[e_i]}{\partial \tau_c} + (1-\sigma)w_i\frac{\partial l_i}{\partial \tau_c} - c_a) + \frac{\partial \gamma_b}{\partial \tau_c}(E[e_i] + x_i) - (1+\tau_c)c_a)\}
$$
$$
- g\prime\frac{\partial l_i}{\partial \tau_c}] + \nu \sum_{i=1}^{n} f_i[\sigma w_i \frac{\partial l_i}{\partial \tau_c} + (\gamma_c + \tau_c\frac{\partial \gamma_c}{\partial \tau_c})(E[e_i] + x_i - (1+\tau_c)c_a) + c_a
$$
$$
+ (\tau_b \gamma_b + \tau_c \gamma_c)(\frac{\partial E[e_i]}{\partial \tau_c} + (1-\sigma)w_i\frac{\partial l_i}{\partial \tau_c} - c_a)]. \tag{57}
$$

By use of the individual f.o.c.'s (4) - (6), the definitions of x_i, γ_c and γ_b as well as $\partial \gamma_c/\partial \tau_c = -\partial \gamma/\partial \tau_c/(1+\tau_c) - (1-\gamma)/(1+\tau_c)^2 = -(\partial \gamma/\partial \tau_c + \gamma_c)/(1+\tau_c)$, $\partial \gamma_b/\partial \tau_c = (\partial \gamma/\partial \tau_c)/(1+\tau_b)$ and evaluation at $\tau_b = \tau_c = 0$ we obtain after some transformations

$$
\frac{\partial S}{\partial \tau_c}\Big|_{\tau_c=0} = (1-\gamma)(1-\sigma)[\lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(-w_i l_i) + \nu \sum_{i=1}^{n} f_i w_i l_i]
$$
$$
+ \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}[\frac{\partial E[e_i]}{\partial \tau_c} - (1-\gamma)(E[e_i] + \alpha - c_a) - c_a]
$$
$$
+ \nu \sum_{i=1}^{n} f_i[\sigma w_i \frac{\partial l_i}{\partial \tau_c} + (1-\gamma)(E[e_i] + \alpha - c_a) + c_a], \tag{58}
$$

Observing that $\partial l_i/\partial \tau_c = (1-\gamma)(1-\sigma)\partial l_i/\partial \sigma$[23] and substituting for the first term in the square bracket by use of condition (36) for optimal σ as well as for ν by use of condition

---

[23]This follows from analogous considerations as shown in footnote (20) for τ_b instead of τ_c. Note that at $\tau_b = \tau_c = 0$ we find $\partial \lambda/\partial \tau_c = (\partial \varphi/\partial c_i)(-\partial \gamma/\partial \tau_c - (1-\gamma)) + (\partial \varphi/\partial b_i)(\partial \gamma/\partial \tau_c)$, which is equal to $-(1-\gamma)\lambda$.

31

(35) for optimal α, (58) can be rewritten as

$$
\begin{aligned}
\frac{\partial S}{\partial \tau_c}\bigg|_{\tau_c=0} = & \ \lambda(1-\gamma)\sum_{i=1}^{n} f_i V_i^{-\rho}(\overline{e} - E[e_i]) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}[\frac{\partial E[e_i]}{\partial \tau_c} \\
& -(1-\gamma)(1-\sigma)\frac{\partial E[e_i]}{\partial \sigma} + \{(1-\gamma)(\overline{e}+\alpha-c_a)+c_a\}\frac{\partial E[e_i]}{\partial \alpha}]. \quad (59)
\end{aligned}
$$

From the steady-state equation (27) for $E[e_i]$ one gets

$$
\frac{\partial E[e_i]}{\partial \tau_c} = A_i \begin{pmatrix} (1-\sigma)w_1\frac{\partial l_1}{\partial \tau_c} - c_a \\ \vdots \\ (1-\sigma)w_n\frac{\partial l_n}{\partial \tau_c} - c_a \end{pmatrix} + \left[\frac{\partial A}{\partial \tau_c}\right]_i (x - c_a^v). \quad (60)
$$

By use of (60), (39), (38) and $\partial l_i/\partial \tau_c = (1-\gamma)(1-\sigma)\partial l_i/\partial \sigma$, (59) is transformed to the formula (31) in Proposition 3. We know already from the proof of Proposition 2 that the first term in (31) is positive, given condition (C) for M. Hence, the first term in (31) is positive, and the third term in (31) is positive as well. Moreover, from similar considerations as shown in the Proof of Proposition 2 for τ_b we obtain (remember that $\gamma_b = \gamma$ at $\tau_b = 0$)

$$
\frac{\partial A}{\partial \tau_c} = \frac{\partial \gamma}{\partial \tau_c}\frac{1}{\gamma}[A + \gamma(I - \gamma M)^{-1}AM]. \quad (61)
$$

Again we express $\partial \gamma/\partial \tau_c$ in terms of the elasticity ε of substitution, defined by (51). Affine-linear Engel curves imply that proportional price changes do not alter the demand ratio of two goods which in our case (where $\tau_b = \tau_c = 0$) means $\partial \gamma/\partial \tau_b + \partial \gamma/\partial \tau_c = 0$ or $\partial \gamma/\partial \tau_c = -\partial \gamma/\partial \tau_b$. Using this in (53) we get

$$
\frac{\partial \gamma}{\partial \tau_c}\bigg|_{\tau_c=0} = (\varepsilon - 1)\gamma(1 - \gamma), \quad (62)
$$

Obviously, $\partial \gamma/\partial \tau_c \lessgtr 0$, if $\varepsilon \lessgtr 1$, and the same applies to the second term in (31), which by use of (61) and (62) can be rewritten as

$$
\lambda \sum_{i=1}^{n} f_i V_i^{-\rho}\left[\frac{\partial A}{\partial \tau_c}\right]_i (x - c_a^v) = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(\varepsilon - 1)(1 - \gamma)(A_i + \gamma B_i)(x - c_a^v). \quad (63)
$$

32

Substituting this expression into (31), some rearrangement gives us

$$
\left.\frac{\partial S}{\partial \tau_c}\right|_{\tau_c=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} (1-\gamma)(\bar{e} - E[e_i]) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(1-\gamma)(A_i \bar{e}^v - \gamma B_i(x - c_a^v))
$$
$$
+ \lambda \sum_{i=1}^{n} f_i V_i^{-\rho} \varepsilon(1-\gamma)(A_i + \gamma B_i)(x - c_a^v). \tag{64}
$$

We know from the Proof of Proposition 2, that $\sum_{i=1}^{n} f_i V_i^{-\rho}(A_i \bar{e}^v - \gamma B_i(x - c_a^v))$ is positive given condition (C) for M, while the third term is nonnegative for any $\varepsilon \geqslant 0$. This completes the proof that $\partial S/\partial \tau_c|_{\tau_c=0} > 0$ for any $\varepsilon \geqslant 0$, given that M fulfills condition (C).

Proof of Proposition 4

The welfare effect of a common tax rate $\tau = \tau_b = \tau_c$ on both, bequests and consumption, is simply the sum of the welfare effect of both tax rates τ_b and τ_c at $\tau_b = \tau_c$. Hence, $\partial S/\partial \tau = \partial S/\partial \tau_b + \partial S/\partial \tau_c$, which by use of (29) and (31) can be written as the formula (33) in Proposition 4. There $[\partial A/\partial \tau]_i = [\partial A/\partial \tau_b]_i + [\partial A/\partial \tau_c]_i$ denotes row i of the sum of the partial derivatives of the matrix A with respect to τ_b and τ_c. By use of (50) and (61), this sum can be written as

$$
\frac{\partial A}{\partial \tau} = \frac{\partial A}{\partial \tau_b} + \frac{\partial A}{\partial \tau_c} = \left(\frac{\partial \gamma_b}{\partial \tau_b} + \frac{\partial \gamma_b}{\partial \tau_c}\right)\frac{1}{\gamma_b}[A + \gamma(I - \gamma M)^{-1}AM], \tag{65}
$$

where $\partial \gamma_b/\partial \tau_b + \partial \gamma_b/\partial \tau_c = (\partial \gamma/\partial \tau_b - \gamma_b + \partial \gamma/\partial \tau_c)/(1+\tau) = -\gamma_b/(1+\tau)$, where we have used $\partial \gamma/\partial \tau_b + \partial \gamma/\partial \tau_c = 0$. Hence, at $\tau = 0$ (65) reduces to

$$
\frac{\partial A}{\partial \tau} = -[A + \gamma(I - \gamma M)^{-1}AM], \tag{66}
$$

from which it is immediate that the second term of (33) is negative. Moreover, by use of (66), (33) can be rewritten as

$$
\left.\frac{\partial S}{\partial \tau}\right|_{\tau=0} = \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(\bar{e} - E[e_i]) + \lambda \sum_{i=1}^{n} f_i V_i^{-\rho}(A_i \bar{e}^v - \gamma B_i(x - c_a^v)), \tag{67}
$$

33

from which it follows that $\partial S/\partial \tau|_{\tau=0} > 0$, if that M fulfills condition (C) (remember that we have shown in the Proof of Proposition 2 that both terms are positive, given M fulfills condition C).

References

Atkinson, A. B. and Stiglitz, J. E. (1976), The Design of Tax Structure: Direct versus Indirect Taxation, *Journal of Public Economics 6*, 55-75.

Black, Sandra D. and Devereux, Paul J. (2011), Recent Developments in intergenerational mobility, in Ashenfelter, Orley and Card, David (eds.), *Handbook of Labor Economics,* Vol. 4, Part B, Chapter 14, Amsterdam: North–Holland, 1591-1969.

Boadway, R., Marchand and M. Pestieau (2000), Redistribution with Unobservable Bequests: A Case for Taxing Capital Income, *Scandinavian Journal of Economics,* Vol. 102, 253-267.

Brunner, Johann K. (1989), *Theory of Equitable Taxation*, Springer Verlag.

Brunner, Johann K. and Pech, Susanne (2012a), Optimal Taxation of Bequests in a Model with Initial Wealth, *Scandinavian Journal of Economics 114*, No. 4, 1368-1392.

Brunner, Johann K. and Pech, Susanne (2012b), Optimal Taxation of Wealth Transfers When Bequests are Motivated by Joy of Giving, *B.E. Journal of Economic Analysis and Policy* (Topics), Vol. 12, Issue 1, Article 8.

Cremer, H., Pestieau, P. and J.-Ch. Rochet (2001), Direct versus Indirect Taxation: The Design of the Tax Structure Revisited, *International Economic Review*, Vol. 42., 781-799.

Cremer, H., Pestieau, P. and J.-Ch. Rochet (2003), Capital income taxation when inherited wealth is not observable, *Journal of Public Economics 87*, 2475-2490.

Deaton, Angus (1979), Optimally uniform commodity taxes, *Economics Letters 2*, 357-361.

Farhi, E. and Werning, I. (2010), Progressive Estate Taxation, *Quarterly Journal of Economics 125*, 635-673.

Gale, W. G. and Slemrod, J. (2001), Overview, in: W. G. Gale, J. R. Hines Jr. and J. Slemrod (eds.), *Rethinking Estate and Gift Taxation*, Washington D. C.: Brookings Institution Press, 1-64.

Hellwig, M. F. (1986), The optimal linear income tax revisited, *Journal of Public Economics 31*, 163-179.

Jäntti, Markus, Bernt Bratsberg, Knut Røed, Oddbjørn Raaum, Robin Naylor, Österbacka, Eva, Björklund Anders and Tor Eriksson (2006), American exceptionalism in a new Light: A comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States, *Discussion paper no. 1938*, (Institute for the Study of Labor (IZA), Bonn).

Kaldor, Nicholas (1955). *An Expenditure Tax.* London: George Allen & Unwin.

Kaplow, L. (2001), A Framework for Assessing Estate and Gift Taxation, in: W. G. Gale, J. R. Hines Jr. and J. Slemrod (eds.), *Rethinking Estate and Gift Taxation*, Washington D. C.: Brookings Institution Press, 164-204.

Kopczuk, Wojciech (2013a), Incentive Effects of Inheritances and Optimal Estate Taxation, *American Economic Review: Papers and Proceedings*, 103(3), 472-77.

Kopczuk, Wojciech (2013b), Taxation of Intergenerational Transfers and Wealth, in A. J. Auerbach, R. Chetty and A. Saez (eds.), *Handbook of Public Economics*, Vol. 5, 329-390.

Leigh, Andrew (2007), Intergenerational mobility in Australia, *B.E. Journal of Economic Analysis and Policy* (Contributions), Vol. 7, Issue 2, Article 6.

Meade, J. (1978), *The Structure and Reform of Direct Taxation*, Report of a Committee chaired by Professor J. E. Meade for the Institute for Fiscal Studies, London: George Allen & Unwin.

OECD (2010), *Tax Policy Reform and Economic Growth*, OECD Publishing.

OECD (2012), *Revenue Statistics 2012*, OECD Publishing.

Piketti, Thomas (2011), On the long-run evolution of inheritance: France 1820 - 2050, *The Quarterly Journal of Economics 126*, 1071–1131.

Piketty, Thomas and Saez, Emmanuel (2013), A Theory of Optimal Inheritance Taxation, *Econometrica 81*, 1851–1886.

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

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

Wolff, Edward N. (2002), Inheritances and Wealth Inequality, 1989-1998, *American Economic Review: Papers and Proceedings 92(2)*, 260-264.

Zimmerman, David J. (1992), Regression toward mediocrity in economic stature, *American Economic Review 82*, 409-429.