Department of Economics
Johannes Kepler University of Linz

Taxing bequests and consumption in the steady state

by

Johann K. BRUNNER and Susanne PECH

Working Paper No. 1717
October 2017

Johannes Kepler University of Linz
Department of Economics
Altenberger Strasse 69
A-4040 Linz - Auhof, Austria
www.econ.jku.at

johann.brunner@jku.at, susanne.pech@jku.at

Taxing bequests and consumption in the steady state

Johann K. Brunner* and Susanne Pech†

University of Linz‡

October 13, 2017

Revised Version

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. In accordance with empirical evidence, we assume that in any generation more able individuals are likely to have a more able parent, which implies that in the steady state they also tend to receive larger inheritances than less able individuals. We 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.

Keywords: Optimal taxation, estate tax, consumption tax, wealth transmission.

* *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 who differ in their labor productivity and who leave 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]

In contrast, in the present paper we analyze the welfare consequences of taxes in the steady-

[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 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.

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 her received inheritances for own consumption and for leaving bequests to her single child of the next generation, and this process goes on in the following periods. Under the assumptions of a joy-of-giving motive for bequeathing and of affine-linear Engel curves for consumption and bequests in each period, we are able to characterize the inheritances of each individual in the steady state. With a given budget share devoted to bequests in each ability group, the transition probabilities determine how the distribution of inheritances develops over time. Eventually, the process 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.

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, determined by maximization of expected social welfare of a generation. For this analysis, it is crucial how the distribution of labor incomes is related to the distribution of inheritances, which in turn depends on the individuals' probabilities of having ancestors of a certain ability. We make the assumption that in each generation the probability distribution of a more able individual over her parent's abilities first-order stochastically dominates the probability distribution of a less able individual. We are able to show that this property indeed implies that in the steady state the distribution of inheritances received by a more able individual first-order stochastically dominates the distribution of inheritances received by a less able individual. To put it differently, the property that in any generation a more able individual is likely to have a more able parent than an individual of lower ability entails that in the steady state the more able individual also tends to receive larger inheritances.

That this stochastic-dominance assumption 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.[5]

First, we analyze the optimal linear income tax in the extended model with endogenous inequality of inherited wealth. We find that the redistributive effect of this tax is more pronounced, compared to the purely static Mirrlees-model, because it affects not only labor income as such, but also indirectly the steady-state inheritances financed by labor income of the ancestors. On the other hand, the distortion of labor supply causes an indirect negative effect on the amount of what is left to the children, that is, on inheritances. Altogether, the desirability of a positive marginal tax rate is guaranteed if the labor supply elasticity is not too large.[6]

Next, we address the main question, namely whether the introduction of indirect taxes 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 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 find a specific role of indirect taxes in the steady state of the economy. To determine the sign of their welfare effect requires a deep-going analysis of how total inheritances in the steady state are affected by the imposition

[5] 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 father.

Moreover, there is a growing empirical literature that studies multigenerational persistence of socioeconomic status (that is, over three or more generations). It provides some evidence that inequality is more persistent than estimates from parent-child correlations would suggest (for an overview see Solon 2015 and Braun and Stuhler 2017).

[6] 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).

of indirect taxes in each single period when bequest decisions are made. It turns out that an additional property must hold in order to derive a clear-cut result. This new condition, which is closely related but not identical to the first-order stochastic dominance mentioned above, requires that individuals who receive higher inheritances from some specific previous period, also tend to receive higher inheritances from their immediate parents. We argue by the use of extensive simulations that this condition can typically be taken as being satisfied in our framework.

Given these assumptions, the bequest tax has a redistributive effect, on top of what is achieved by the optimal linear income tax. However, it also distorts the bequest decision and induces, thus, lower accumulated wealth in the steady state. The overall welfare 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 the consumption tax, introduced in addition to an optimally chosen tax on labor income, that under the same assumptions it has a redistributive potential in our model, 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 between them 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 welfare effect of a consumption tax is 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 increases social welfare, because it combines the redistributive effect of both separate taxes and causes no distortion.

Our study establishes a role for indirect taxation as an instrument for increasing equality of opportunity. In a related work, Piketty and Saez (2013) also consider the steady state of a dynamic model. They assume that utility functions as well as abilities follow two stochastic processes and study both, the case of joy-of-giving preferences and the case of altruistic preferences for bequests. Their central result is a formula for the optimal linear bequest tax in terms of estimable parameters such as the elasticity of bequests with respect to the tax. For the derivation, they take the amount of government transfers as given, and the formula states to which extent, depending on the parameter constellation, these transfers should be financed by a tax on bequests as opposed to a tax on labor income. Calibration of the model with data from France and the US, leads to the finding that a tax rate on bequests up to 50 percent may be optimal.

Farhi and Werning (2010) study a model with dynastic preferences and nonlinear taxes on income and bequests in a two-period economy, which later is extended to an infinite-horizon setting. Their main discussion refers to the case that the social planner values the future generation separately, in addition to the value given to it through the altruistic preferences of the present generation, whose utility is also part of the social welfare function (this is sometimes called double-counting). As already mentioned, they find that the optimal tax rate on bequests is negative (to correct for the positive externality associated with leaving bequests) and progressive (the marginal tax rate decreases with income in order to reduce inequality over generations). If for some reason marginal tax rates are restricted to be nonnegative, a rate of zero up to a certain threshold income is optimal.

Kopczuk (2001) considers the steady-state of a dynamic economy where inequality in inheritances arises as a result of differences in abilities, similar to our model. However, he makes the simplifying assumption that the ability levels within a dynasty are perfectly correlated (complete immobility of abilities over generations) and derives general formulae which characterize optimal nonlinear taxes on labour income and bequests and whose components reflect the role of these taxes for the reduction of inequality and for the correction of externalities.

In contrast to these studies, an important feature of our contribution is that we provide a precise model of intergenerational transfers, by formulating the parent-child relation of abilities as a Markov process. This model allows us to consider the joint distribution of income and inherited wealth in the steady-state, and to analyze in detail the long-run relationship between direct and indirect taxes. To our knowledge, this has not been done before. In particular, we are able to characterize under which conditions taxes on income and bequests indeed have a redistributive effect in the long run. Moreover, we also study the role of the consumption tax.

As to the issue of double counting, we should stress that our discussion of the steady-state effects means that double counting is present in our analysis and, thus, the above-mentioned positive external effect is fully taken into account.[7] That is, if some tax change influences bequests left by a generation, we measure the direct welfare effect on the donor (who has a joy-of-giving motivation for net bequests), but also the indirect effect because simultaneously we include the change of the received inheritances of this (steady-state) generation as well. As steady-state inheritances result as the sum of all prior bequests, we thus incorporate how a tax modifies the positive external effect caused by all prior bequests.

In the following Section 2 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. Section 5 describes the maximization problem of the government and introduces the stochastic-dominance property mentioned above. In Section 6 the optimal linear tax on labor income is characterized and in Section 7 the results on the desirability of additional taxes on bequests and consumption, respectively, are derived. Section 8 provides a 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, $i = 1, ..., n$. Preferences are described by the common 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 convex 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.

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

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

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 Lagrange variable associated with (2) read as

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

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

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

We assume that individuals need some fixed minimum consumption $c_a \geq 0$ and that the income-consumption curve is linear, originating in c_a. As a consequence, expenditures $(1+\tau_b)b_i$ and $(1+\tau_c)(c_i - c_a)$ can be written as shares γ and $1-\gamma$, with $\gamma \in (0,1)$, of the available budget, after deducting the expenditures $(1+\tau_c)c_a$ for minimum consumption. Hence, with $\gamma_b \equiv \gamma/(1+\tau_b)$ and $\gamma_c \equiv (1-\gamma)/(1+\tau_c)$ the demand functions for c_i and b_i can be written as

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

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

Here x_i denotes net income of the household, $x_i \equiv \alpha + (1-\sigma)w_i l_i$. 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, 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 of b_i and $c_i - c_a$, while η determines the constant elasticity of substitution $1/(1-\eta)$.[8]

Indirect utility of an individual i is given by evaluating $\varphi(c_i, b_i) - g(l_i)$ using (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

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

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 potential ability levels $w_1 < w_2 < .... < w_n$ 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

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

[8] Only in case of a Stone-Geary utility function $(c_i - c_a)^{1-\delta} b_i^{\delta}$ (η converges to zero), $\gamma = \delta$ is a constant.

Obviously, then also $f = fP^s$ for any $s = 1, 2, ...$, because $fP^s = fPP^{s-1} = fP^{s-1} = ... = fP$.

Next we study the transmission of bequests from generation to generation. In the following we introduce a time index $-s$ to characterize a generation that lived s periods before some particular generation indexed by 0. Moreover, to indicate the ability level of a member of generation $-s$ we use the index $r_{-s} = 1, ..., n$. Hence $e_{r_{-s}}$ and $b_{r_{-s}}$ denote bequests received and left, respectively, by an individual with ability $w_{r_{-s}}$ in period $-s$. Similarly, $x_{r_{-s}}$ denotes net income of individual of type r_{-s} in period $-s$. Now consider the generation in period 0. Each individual $i = 1, ..., n$ of this generation receives potential inheritances $b_{r_{-1}} = \gamma_b(e_{r_{-1}} + x_{r_{-1}} - (1 + \tau_c)c_a)$, where $r_{-1} = 1, ..., n$ can be any type in the previous period -1. Note that with homothetic preferences, bequests left by an individual r_{-1} in generation -1 can be separated into those out of received inheritances, $\gamma_b e_{r_{-1}}$, and those out of own labor income, $\gamma_b(x_{r_{-1}} - (1 + \tau_c)c_a)$.

In a next step, we consider inheritances $e_{r_{-1}}$ of individual r_{-1} in period -1, which arise as some $b_{r_{-2}} = \gamma_b(e_{r_{-2}} + x_{r_{-2}} - (1 + \tau_c)c_a)$ and can, thus, again be split into bequests out of received inheritances in period -2 and those out of own labor income in period -2. Plugging in $b_{r_{-2}}$ for $e_{r_{-1}}$ in the formula for bequests left by individual r_{-1} in period -1, one gets for inheritances e_i $(= b_{r_{-1}})$ of some individual i in period 0

$$e_i = \gamma_b^2 e_{r_{-2}} + \gamma_b^2(x_{r_{-2}} - (1 + \tau_c)c_a) + \gamma_b(x_{r_{-1}} - (1 + \tau_c)c_a). \tag{11}$$

Continuing this separation further into the past until some period $-t$, we find the general formula

$$\begin{aligned} e_i &= \gamma_b^t e_{r_{-t}} + \gamma_b^t(x_{r_{-t}} - (1 + \tau_c)c_a) + \gamma_b^{t-1}(x_{r_{-t+1}} - (1 + \tau_c)c_a) \\ &\quad + ... + \gamma_b^2(x_{r_{-2}} - (1 + \tau_c)c_a) + \gamma_b(x_{r_{-1}} - (1 + \tau_c)c_a). \end{aligned} \tag{12}$$

Here $r_{-t}, r_{-t+1}, ..., r_{-2}, r_{-1}$ denote some indices from the set $\{1, ..., n\}$. Neglecting for the moment the first term $\gamma_b^t e_{r_{-t}}$, we find that inheritances of an individual i in period 0 result from available labor income in the previous generations up to $-t$. It is important to note that what an individual of generation 0 inherits from some particular generation $-s$, $1 \leq s \leq t$, depends only on the type $r_{-s} \in \{1, ..., n\}$ of the individual who originated the bequests, but not on the types in the in-between generations $-s+1, -s+2, ..., -1$: these generations just pass forward a share γ_b of their inheritance. In addition, each of these in-between generations again originates bequests out of own available labor income. As a consequence, still leaving aside the term $\gamma_b^t e_{r_{-t}}$ in (12), all possible inheritances in period 0, from the periods $-1, -2, ..., -t$ can be written as the sum

$$\sum_{s=1}^{t} \gamma_b^s (x_{r_{-s}} - (1+\tau_c)c_a), \quad r_{-s} \in \{1, ..., n\} \text{ for all } s = 1, 2, ..., t. \tag{13}$$

Note that these possible inheritances are the same for all ability groups i (all potential inheritors) in period 0. However, the probabilities of the various realizations of the inheritances differ according to the type i of the receiving inheritor in period 0. They are determined by the probabilities that i has a specific type r_{-s} as her great-great-...grandparent in period $-s$, $s = 1, ..., t$. For a closer characterization, let μ_{ij} be the probability that an individual of type i has a parent of type j. 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

$$\mu_{ij} = \text{Prob}(\text{parent} = j \mid \text{descendant} = i) = \frac{f_j p_{ji}}{f_i}. \tag{14}$$

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

$$M = \begin{pmatrix} p_{11} & \frac{f_2 p_{21}}{f_1} & \cdots\cdots & \frac{f_n p_{n1}}{f_1} \\ \frac{f_1 p_{12}}{f_2} & p_{22} & \cdots\cdots & \frac{f_n p_{n2}}{f_2} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \frac{f_1 p_{1n}}{f_n} & \frac{f_2 p_{2n}}{f_n} & \cdots\cdots & p_{nn} \end{pmatrix}. \tag{15}$$

By definition, M is also a stochastic matrix, that is, all elements are nonnegative and $\sum_{j=1}^{n} \mu_{ij} = 1$, for all $i = 1, ..., n$. Going back one step further, $\mu_{ir_{-1}} \mu_{r_{-1}r_{-2}}$ is the probability for individual i in period 0 of having a parent of type $r_{-1} \in \{1, ..., n\}$ in period -1 together with a grandparent of type $r_{-2} \in \{1, ..., n\}$ in period -2. Generally, we define $\mu_i(r_{-1}, r_{-2}, ..., r_{-s})$ as the probability of individual i of having skill types $r_{-1}, r_{-2}, ..., r_{-s}$ as their ancestors up to generation $-s$ and get

$$\mu_i(r_{-1}, r_{-2}, ..., r_{-s}) = \mu_{ir_{-1}} \cdot \mu_{r_{-1}r_{-2}} \cdot ... \cdot \mu_{r_{-s+1}r_{-s}}. \tag{16}$$

Moreover, $\sum_{r_{-1}=1}^{n} \mu_{ir_{-1}} \mu_{r_{-1}r_{-2}}$ is the probability of individual i having a grandparent r_{-2} in period -2, allowing for all possible types as the parent. One observes immediately that this probability is just the element (i, r_{-2}) of the Matrix $M \cdot M = M^2$. Clearly, M^2 is again a stochastic matrix, that is, $\sum_{r_{-1}=1}^{n} \sum_{r_{-2}=1}^{n} \mu_{ir_{-1}} \mu_{r_{-1}r_{-2}} = 1$, for each i. The same holds for M^s with any number of periods s, thus $\sum_{r_{-1}=1}^{n} \sum_{r_{-2}=1}^{n} ... \sum_{r_{-s}=1}^{n} \mu_i(r_{-1}, r_{-2}, ..., r_{-s}) = 1$. The elements (i, r_{-s}) of the matrix M^s are denoted by $\mu_{s;ir_{-s}}$, that is,

$$\mu_{s;ir_{-s}} \equiv \sum_{r_{-1}=1}^{n} \sum_{r_{-2}=1}^{n} .... \sum_{r_{-s+1}=1}^{n} \mu_{ir_{-1}} \mu_{r_{-1}r_{-2}} \cdot ... \cdot \mu_{r_{-s+1}r_{-s}}, \quad i = 1, ..., n, \; r_{-s} = 1, ..., n. \tag{17}$$

Each $\mu_{s;ir_{-s}}$ describes the probability that an individual of type i in period 0 has an ancestor of type r_{-s} in period $-s$. Then $\sum_{r_{-1}=1}^{n} ... \sum_{r_{-s}=1}^{n} \mu_i(r_{-1}, ..., r_{-s}) = \sum_{r_{-s}=1}^{n} \mu_{s;ir_{-s}} = 1$.

Let $p_{s;ij}$ be defined in a completely analogous way to $\mu_{s;ij}$. Then $\mu_{2;ij} = \sum_{k=1}^{n} \mu_{ik} \mu_{kj} = \sum_{k=1}^{n} f_k p_{ki} f_j p_{jk} / (f_i f_k) = (\sum_{k=1}^{n} p_{jk} p_{ki}) f_j / f_i = p_{2;ji} f_j / f_i$, and repeated application of this argument shows that generally

$$\mu_{s;ij} = \frac{f_j p_{s;ji}}{f_i}, \qquad \text{for any } s = 1, 2, ... \tag{18}$$

As a consequence, $\lim_{s \to \infty} M^s = W$. This follows from $\lim_{s \to \infty} P^s = W$, that is, for $s \to \infty$, $p_{s;ji} \to f_i$ which implies $\mu_{s;ij} \to f_j$, $j = 1, ..., n$, identical for all $i = 1, ..., n$.

We note for later use that with respect to the constant distribution f of the ability types, M has the same property as P:

$$f = f M^s, \qquad \text{for any } s = 1, 2, ... \tag{19}$$

To proof this, consider that $fM = (f_1(p_{11} + p_{12} + ... + p_{1n}), ..., f_n(p_{11} + p_{12} + ... + p_{1n})) = f$. Moreover, $fM^2 = fMM = f$, and the same logic applies for any s.

4 The steady-state

We already assumed without mentioning that the tax rates τ_c and τ_b on consumption and bequests, respectively, are constant over time. Extending this assumption to the income tax parameters α and σ implies that labor supply and net income of each type are also constant over time. Remember that 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 are identical in each period.

We define a steady state of this economy as a situation where the realizations of inheritances for the individuals of a generation remain constant over time. Thus, also all possible bequests left by the individuals do not change.

In view of the constancy of each type's net income over time, as a consequence of the fixed tax parameters, it follows immediately from (12) that all possible inheritances e_i of an individual i of a particular steady-state generation 0 are given by all possible infinite sums

$$\sum_{s=1}^{\infty} \gamma_b^s (x_{r_{-s}} - (1+\tau_c)c_a), \quad x_{r_{-s}} \in \{x_1, ..., x_n\} \text{ for all } s = 1, ... \tag{20}$$

Obviously, all these infinite sums indeed have a finite value as $\gamma_b < 1$ and all $x_{r_{-s}} \leq x_n$ which is a finite number.

5 The Government Problem

Formula (20) characterizes steady-state inheritances in terms of net incomes $x_{r_{-s}}$ (resulting from labor supply $l_{r_{-s}}$), $r_{-s} = 1, ..., n$, in all prior generations $-s$, and of 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 the following we study how the government should choose the tax parameters in order to maximize social welfare in the steady-state. In particular, we want to clarify the question of whether taxes on bequests or consumption (or on both) are appropriate supplements to an optimal linear income tax. In our model the individuals differ in labor productivities and in received inheritances, but the latter arise endogenously as a result of the exogenous differences in labor productivities. Therefore, it is a priori unclear whether taxes on bequests or consumption represent desirable additional instruments or are redundant as in the Atkinson-Stiglitz case (with productivity as the only dimension of heterogeneity).

The fact that individuals now differ with respect to an infinite number of potential inheritances, in addition to their skill types, makes the analysis of the long-run consequences of the taxes a rather complex task. In order to keep it tractable, we take into account only a finite, though arbitrarily large, number of previous generations, from which a steady-state generation actually receives inheritances. In other words, in each period those bequests which are initially left by a generation more than t periods before are discarded. This simplification is certainly justified by the fact that in (20) the value of inheritances declines with the distance s of earlier generations and approaches zero for s going to infinity. Moreover, as all involved functions are continuous, the sign of the welfare effect of a tax can never switch to the opposite when the number of previous generations taken into account goes to infinity.

Denote in the following by the index 0 an arbitrary period of our economy in the steady state. Let $t \in \mathbb{N}$ be arbitrarily large. Consider (20) for a finite number t of previous generations, then each realization of inheritances in period 0 is written as some

$$e(r_{-1}, ..., r_{-t}) \equiv \sum_{s=1}^{t} \gamma_b^s (x_{r_{-s}} - (1+\tau_c)c_a) \text{ with } r_{-s} \in \{1, 2, ..., n\} \text{ for all } s = 1, 2, ..., t. \quad (21)$$

Remember that the potential realizations $e(\cdot)$ are the same for all skill-types i, but the associated probabilities $\mu_i(r_{-1}, ..., r_{-t})$ differ according to the type, with $\mu_i(\cdot)$ given by (16). Let $E_i[e] = \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(r_{-1}, ..., r_{-t}) e(r_{-1}, ..., r_{-t})$ denote expected inheritances of a type-i individual of some steady-state generation 0, thus, $f_i E_i[e]$ represents actual aggregate inheritances of group i. Let further $V_i(r_{-1}, ..., r_{-t})$ denote indirect utility of an individual of ability type i, as given by (9), if this individual receives an inheritance $e_i = e(r_{-1}, ..., r_{-t})$.[9] We formulate the government's problem of maximizing steady-state social welfare in the following way: let $\rho > 0$ be the parameter of inequality aversion; the government maximizes the sum (weighted by the population shares f_i) of the groups' expected social utility, each written as $E_i[V_i^{1-\rho}/(1-\rho)] = \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(r_{-1}, ..., r_{-t}) V_i(r_{-1}, ..., r_{-t})^{1-\rho}/(1-\rho)$, given stochastic inheritances $e(r_{-1}, ..., r_{-t})$. It has to observe the budget constraint that the revenues from the income tax σ and the taxes on bequests (τ_b) and consumption (τ_c) cover the uniform subsidy

[9] Clearly, V_i also depends on the tax parameters $\alpha, \sigma, \tau_c, \tau_b$, but for simplicity of notation we do not indicate this explicitly.

α. Thus, the government's problem reads as

$$\max \sum_{i=1}^{n} f_i E_i[V_i^{1-\rho}/(1-\rho)], \tag{22}$$

$$\text{s.t. } \sigma \sum_{i=1}^{n} f_i w_i l_i - \alpha + \tau_b \sum_{i=1}^{n} f_i \gamma_b (E_i[e] + x_i - (1+\tau_c)c_a) +$$
$$\tau_c \sum_{i=1}^{n} f_i \{\gamma_c (E_i[e] + x_i - (1+\tau_c)c_a) + c_a\} \geq 0, \tag{23}$$

where the public budget constraint is formulated by use of the demand functions (8) and (7).

A crucial issue for the analysis of optimal taxes within this model is how the distribution of inheritances, over which the expectation is computed, is related to the distribution of abilities in each steady-state generation. Clearly, this in turn depends on the probabilities of wealth transmission across ability types from one generation to the next. From a social-welfare point of view, redistribution of income from high to low skill types only appears acceptable if low-able individuals do not typically receive larger estates than high-able individuals. As mentioned in the introduction, there is ample empirical evidence that this condition is fulfilled in reality, that is, a positive correlation holds between the size of received inheritances and own abilities.

A natural way to account for such a positive correlation in our model is to assume that each subsequent row of the matrix M first-order stochastically dominates the previous row, that is, in each generation an individual of type $i+1$ is likely to have a more able parent than an individual of type i.

Definition 1 ***(Row-FOSD)*** *An n-dimensional stochastic matrix M with elements μ_{ij} fulfills first-order stochastic dominance of subsequent rows if for each $i = 1, ..., n-1$, the row vector $(\mu_{i+1,1}, ..., \mu_{i+1,n})$ first-order stochastically dominates the row vector $(\mu_{i1}, ..., \mu_{in})$, that is, $\sum_{k=1}^{j} \mu_{ik} \geq \sum_{k=1}^{j} \mu_{i+1,k}$, for all $j = 1, ..., n$.*

As an important result, we find that this property of ability transition from one period to the next indeed implies for the steady state that overall received inheritances of a more-able individual tend to be larger than those of a less-able individual.

Proposition 1 *If M fulfills Row-FOSD then in the steady state the distribution of received inheritances of a type-$(i+1)$ individual first-order stochastically dominates the distribution of inheritances of a type-i individual.*

Proof. See Appendix B. ■

An immediate consequence of this fundamental characterization is that expected inheritances $E_i[e]$ and utility levels $E_i[V_i]$ increase with the abilities $i = 1, ..., n$. On the other hand, the welfare weight of an individual i, that is, her expected social marginal utility of income $\partial E_i[V_i^{1-\rho}/(1-\rho)]/\partial x_i = \lambda E_i[V_i^{-\rho}] = \lambda \sum_{r_{-1}=1}^{n} \cdots \sum_{r_{-t}=1}^{n} \mu_i(r_{-1}, ..., r_{-t}) V_i(r_{-1}, ..., r_{-t})^{-\rho}$ is decreasing with abilities (remember that $\lambda = \partial V_i/\partial x_i$ is the individual marginal utility of income, independent of an individual's wage w_i). These properties which are essential for the analysis of optimal taxes in the next sections, are summarized in the following.

Corollary 1 *If M fulfills Row-FOSD then for any $\sigma < 1$ the following inequalities apply: $E_i[e] < E_{i+1}[e]$, $E_i[V_i] < E_{i+1}[V_{i+1}]$ and $E_i[V_i^{-\rho}] > E_{i+1}[V_{i+1}^{-\rho}]$, $i = 1, ..., n-1$.*

Proof. See Appendix B. ■

The assumption of Row-FOSD is made throughout the upcoming analysis. As an important property we note

Lemma 1 *If M fulfills Row-FOSD, then M^s fulfills Row-FOSD as well.*

Proof. See Appendix B. ■

6 The linear income tax

As a first step towards an analysis of whether indirect taxes have a role in the steady state of this model, we need to study the welfare effect of the (linear) income tax. That is, we ask if and under which conditions a marginal increase of σ from zero, with a simultaneous increase of α to satisfy the resource constraint, is desirable according to the social welfare function, setting $\tau_b = \tau_c = 0$.

Let $\overline{wl} \equiv \sum_{i=1}^{n} f_i w_i l_i$ denote the average gross income and let, for any given $s = 1, ..., t$ and $r_{-s} = 1, ..., n$, $E_i[V_i^{-\rho} \mid r_{-s}] = \sum_{r_{-1}=1}^{n} \cdots \sum_{r_{-s+1}=1}^{n} \sum_{r_{-s-1}=1}^{n} \cdots \sum_{r_{-t}=1}^{n} \mu_i(r_{-1}, ..., r_{-t}) \cdot V_i(r_{-1}, ..., r_{-t})^{-\rho}/\mu_{s;ir_{-s}}$ be the expected social welfare weight of individual i, conditional on having an ancestor of type r_{-s} in period $-s$.

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

$$\begin{aligned}
&\lambda \sum_{i=1}^{n} f_i E_i[V_i^{-\rho}](\overline{wl} - w_i l_i) + \\
&\lambda \sum_{i=1}^{n} f_i \sum_{s=1}^{t} \sum_{r_{-s}=1}^{n} \mu_{s;ir_{-s}} E_i[V_i^{-\rho} \mid r_{-s}] \gamma^s (\overline{wl} - w_{r_{-s}} l_{r_{-s}}) + \\
&\lambda \sum_{i=1}^{n} f_i \sum_{s=1}^{t} \sum_{r_{-s}=1}^{n} \mu_{s;ir_{-s}} E_i[V_i^{-\rho} \mid r_{-s}] \gamma^s w_{r_{-s}} \frac{\partial l_{r_{-s}}}{\partial \sigma}.
\end{aligned} \tag{24}$$

The first two terms are positive and describe the immediate and the long-run welfare effect, respectively, of income redistribution. The third term is negative and represents the long-run distorting effect. The overall welfare effect is positive if the labor-supply reaction is not too strong.

Proof. See Appendix C. ■

The first line of (24) captures the immediate welfare effect, for given inheritances, thus ignoring the implications of income taxation in prior periods. It corresponds to the standard welfare effect of a linear income tax in a purely static Mirrlees-model without inheritances. A marginal increase of the tax rate σ (from zero) hits individuals according to their gross income $w_i l_i$, while an identical amount $\alpha = \overline{wl}$ (overall revenue) is returned to everyone, thus $\overline{wl} - w_i l_i$ indicates the net income change for a specific individual with wage w_i. This net income change is clearly positive for low-able and negative for high-able individuals; the associated welfare effect depends on the social evaluation of this redistribution, which is expressed by the social welfare weights of the groups, that is, their expected social marginal utility of income $\lambda E_i[V_i^{-\rho}]$. It was stated in Corollary 1 that these weights decrease with the ability level i if M has the property of Row-FOSD. Given increasing $w_i l_i$, this implies, together with $\sum_{i=1}^{n} f_i(\overline{wl} - w_i l_i) = 0$, that the first line of (24) is unambiguously positive. (Note that the immediate distorting effect of the tax is of second order, thus zero at $\sigma = 0$.) Clearly, the magnitude of this positive welfare effect increases with the spread of individual incomes and with the degree of inequality aversion ρ.

The second line of (24) describes the long-run effect of income redistribution. As with the static part discussed above, the immediate income change for group r_{-s} in period $-s$, of income redistribution through the linear income tax, is given by the difference $\overline{wl} - w_{r_{-s}} l_{r_{-s}}$. However, the social weight of this redistribution is now more complex: it consists of the consequences of this income change in period $-s$ (that is, of the change $\gamma^s(\overline{wl} - w_{r_{-s}} l_{r_{-s}})$ of received inheritances in the generation 0), for welfare of all ability groups i in the steady-state generation 0. Clearly, the relevant welfare evaluation of each group i is $\mu_{s;ir_{-s}} E_i[V_i^{-\rho} \mid r_{-s}]$, i.e., its probability of having type r_{-s} as an ancestor multiplied by the associated conditional expected social welfare. Summing up over all $i = 1, ..., n$ and $r_{-s} = 1, ..., n$ gives us the welfare consequences of income redistribution in a single previous period $-s$, and the sum over s then describes the total long-run welfare effect.

It turns out that redistribution of income in previous periods in effect means redistribution of inheritances in the steady-state generation 0. This can formally be seen by writing the second line of (24) in an equivalent way as (see (52) in Appendix C)

$$\lambda \sum_{i=1}^{n} f_i E_i[V_i^{-\rho}(\overline{e} - e(\cdot))], \tag{25}$$

where $\overline{e} \equiv \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \cdots \sum_{r_{-t}=1}^{n} \mu_i(r_{-1}, .., r_{-t}) e(r_{-1}, .., r_{-t}) = \sum_{i=1}^{n} f_i E_i[e]$ denotes average inheritances received by generation 0. For each group i, the expectation term in (25) can be interpreted as describing the social-welfare evaluation if received inheritances in the steady-state are redistributed via a linear tax. This relates the analysis of the income tax to that of the bequest tax, as will be discussed more extensively in the next section.

As we show formally in Appendix C, the long-run welfare effect of income redistribution represented by the second line of (24) is positive. The essential idea is to split the effect of income redistribution in previous periods into one of inheritance redistribution across ability groups in generation 0, and an effect within groups. Intuitively, Row-FOSD of M implies that a more able individual receives larger expected inheritances, thus income redistribution in previous periods produces a desirable equalization of inheritances across ability groups in the steady state. In addition, also within each group, for individuals with the same skills, a desired equalization from those who receive large to those who receive small inheritances takes place. Altogether, the long-run redistributive effect of the introduction of an income tax is positive.

Finally, turning to the third line of (24), this effect is clearly negative because $\partial l_{r_{-s}}/\partial\sigma < 0$ for all $s = 1, ..., t$. Its structure is similar to that of the second line; it represents the long-run welfare loss caused by the disincentive effect of the income tax on labour supply in each period $-s$, which transforms into lower steady-state inheritances. Note that this effect drops out if redistribution within a generation alone is considered (as in the static Mirrlees model), because for each individual the 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 at $\sigma = 0$); nevertheless this disincentive effect occurs in the steady-state of our dynamic model, because the tax affects the size of bequests left by previous generations.

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. In other words, we assume that labour supply is not too elastic, such that for an initial increase of σ the disincentive effect (line three of (24)) is dominated by the positive effects. Obviously, with larger σ, 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 $g'(l_i) > 0$ for $l_i > 0$.

7 Optimal indirect taxes

For a precise analysis of whether an indirect tax can contribute to a desirable redistribution of wealth in the steady state, on top of what is achieved by an optimal income tax, we need to introduce a new property, in addition to Row-FOSD. Let $s \in \mathbb{N}$, $s \leq t-1$. Remember that $\mu_{s;jk}$ denotes the element (j,k) of M^s (thus, $\mu_{1;jk} = \mu_{jk}$), then clearly $\mu_{s+1;ik} = \sum_{j=1}^{n} \mu_{ij}\mu_{s;jk}$ is the element (i,k) of $MM^s = M^{s+1}$.

Definition 2 *An n-dimensional stochastic matrix M with elements μ_{ij} fulfills the condition of Path Dominance (abbreviated by PD) of order t, if for any $s = 1,...,t-1$, $i = 1,...,n$ and $k = 1,...,n-1$, the probability vector with elements $\mu_{ij}\mu_{s;j,k+1}/\mu_{s+1;i,k+1}$, $j = 1,...,n$, first-order stochastically dominates the probability vector with elements $\mu_{ij}\mu_{s;jk}/\mu_{s+1;ik}$, $j = 1,...,n$.*

By definition, $\sum_{j=1}^{n} \mu_{ij}\mu_{s;j,k+1}/\mu_{s+1;i,k+1} = 1 = \sum_{j=1}^{n} \mu_{ij}\mu_{s;jk}/(\sum_{j=1}^{n} \mu_{ij}\mu_{s;jk})$, for any i and k. As discussed above, the elements $\mu_{s+1;ik}$ of M^{s+1} describe the probability that an individual of skill type i has a type-k great-, great-,... grandparent, $s+1$ periods before. Thus, the $\mu_{ij}\mu_{s;jk}/\mu_{s+1;ik}$, $j = 1,...,n$ describe the probability that an individual of skill type i and with a type-k great-, great-,... grandparent, $s+1$ periods before, has an immediate parent of type-j, in period -1. Intuitively then, condition PD requires that if we consider two individuals of the same type i in the steady-state generation 0 with different great-, great-,... grandparents of type k and type $k+1$, respectively, $s+1$ periods before, the one with a type-$k+1$ great-, great-, ... grandparent tends to have a more-able *immediate* parent than the other. For a two-type economy we find:

Lemma 2 *If $n=2$, then PD is implied by Row-FOSD of M.*

Proof. We have to check that $\mu_{i1}\mu_{s;11}/(\mu_{i1}\mu_{s;11}+\mu_{i2}\mu_{s;21}) \geq \mu_{i1}\mu_{s;12}/(\mu_{i1}\mu_{s;12}+\mu_{i2}\mu_{s;22})$ which reduces to $\mu_{s;11}\mu_{s;22} \geq \mu_{s;12}\mu_{s;21}$. This inequality holds because Row-FOSD of M^s (see Lemma 1) for $n=2$ implies $\mu_{s;11} \geq \mu_{s;21}$ as well as $\mu_{s;12} \leq \mu_{s;22}$. ■

Though PD is related to Row-FOSD, it represents a separate condition. Indeed, for $n>3$ one can find matrices for which PD is not implied by Row-FOSD. On the other hand, Path Dominance is automatically fulfilled for large s (as is Row-FOSD), which follows immediately from the fact that $\lim_{s\to\infty} M^s = W$ (mentioned in Section 3), together with the property that all rows of W are identical and equal to the steady-state distribution f. The relation between Row-FOSD and Path Dominance of M is discussed in more detail below and in Appendix D.

7.1 The bequest tax

We begin with the analysis of the welfare effect of a proportional bequest tax and ask whether the latter can play a role as a complement to an optimally chosen labor income tax. Let for any $s=1,...,t-1$, $e_{-s}(r_{-s-1},...,r_{-t}) \equiv \sum_{\widehat{s}=s+1}^{t} \gamma_b^{\widehat{s}-s}(x_{r_{-\widehat{s}}}-(1+\tau_c)c_a)$ denote what generation $-s$ receives as inheritances from some series $r_{-s-1},...,r_{-t}$ of ancestors up to period $-t$, with average $\overline{e}_{-s} \equiv \sum_{r_{-s}=1}^{n} f_{r_{-s}} \sum_{r_{-s-1}=1}^{n} \cdots \sum_{r_{-t}=1}^{n} \mu_{r_{-s}}(r_{-s-1},..,r_{-t})e_{-s}(r_{-s-1},..,r_{-t})$. With ε as the elasticity of substitution between bequests and consumption, Ω as defined in (74) in Appendix C and $S(\tau_b,\tau_c)$ as the optimal value function of the maximization problem (22) - (23), we find for the bequest tax:

Proposition 3 *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*

$$\begin{aligned}\left.\frac{\partial S}{\partial \tau_b}\right|_{\tau_b=0} =&\lambda \sum_{i=1}^{n} f_i E_i[V_i^{-\rho}(\cdot)\gamma(\overline{e}-e(\cdot))] \\ &+\sum_{i=1}^{n} f_i \sum_{s=1}^{t-1} E_i[V_i^{-\rho}(\cdot)\gamma^s\gamma(\overline{e}_{-s}-e_{-s}(\cdot))]+\gamma\Omega- \\ &-\sum_{i=1}^{n} f_i E_i[V_i^{-\rho}(\cdot)\varepsilon(1-\gamma)(e(\cdot)+\sum_{s=1}^{t-1}\gamma^s e_{-s}(\cdot))]. \end{aligned} \tag{26}$$

The first line describes the immediate redistributive effect; it is positive. The second line describes the long-run redistributive effect; it is positive, given Path Dominance. The third line captures the long-run distortive effect and is negative (or zero).
The overall effect of τ_b on social welfare is positive (i) if Path Dominance holds and the elasticity of substitution between bequests and consumption is small, or (ii) if inequality aversion is large and inheritances at the lower end of the inheritance distribution are small.

Proof. See Appendix C. ■

As with the linear income tax, two separate effects can be distinguished as a consequence of the introduction of the bequest tax: the welfare effect when inheritances are held fixed, which is the term in the first line of the above formula, and the welfare effect due to changes in steady-state inheritances (as a result of changes in bequests left by previous generations), which is described by the second and the third lines of (26).

The immediate redistributive effect. This effect represents the main role of the bequest tax: it allows additional redistribution of $\gamma e(\cdot)$, i. e., of that particular part of bequests left by the steady-state generation 0, which is financed out of received inheritances. The important point is that in fact the bequest tax is an instrument for equalization of *received inheritances,* because these indeed represent a second characteristic of the individuals, in addition to abilities. This redistribution of received inheritances is performed indirectly, via the share γ which is used for bequests *left* to own heirs. Note that $\gamma(\overline{e}-e(\cdot))$ is clearly positive for individuals who inherit little while it is negative for those who inherit much. This effect also occurs in the static model with exogenous inheritances (Brunner and Pech 2012b). 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 this part no effect shows up in the first line of (26) because - as the Atkinson-Stiglitz result tells us - no further redistribution is possible in addition to that performed by the optimal income tax.

One observes that the immediate redistributive effect in line one of (26) is very similar to the long-run redistributive effect of the income tax, as already mentioned above (see (25)); it differs only by the factor γ. This relates to the common view that inheritance taxation may represent an undesirable double-taxation, because it affects the bequeathed part of income, and the later was already subjected to the income tax. Note, however, that this is clearly true for any indirect tax, thus also for a consumption tax (see below). More importantly, in the present context the essential point is that the bequest tax allows additional redistribution, above what is performed by the optimal income tax. The intuitive reason is that, compared to the latter, it causes a smaller distortion of the labor supply decision, as can immediately be seen from the relation $\partial l_i/\partial\tau_b = \gamma(1-\sigma)\partial l_i/\partial\sigma$ (see (59) in Appendix C), where $\gamma(1-\sigma) < 1$. On the other hand, in the longer run the bequest tax has a distorting effect on capital accumulation, as is discussed in the following.

Redistribution in previous periods. The first term in the line two of (26) comes from the fact that in each period $-s$ the bequest tax redistributes that particular part $\gamma e_{-s}(r_{-s-1}, ..., r_{-t})$ of bequests which is financed out of received inheritances from ancestors up to $-t$.[10] This is completely analogous to the mechanism discussed above for the redistribution of bequests financed out of inheritances in the steady state (described by the first line of (26)). Of this redistribution in $-s$, finally $\gamma^s\gamma(\overline{e}_{-s} - e_{-s}(r_{-s-1}, ..., r_{-t}))$ arrives at the generation 0. What matters for the social evaluation is how aggregate welfare of all groups $i = 1, ..., n$ in this steady-state generation 0 is affected by the total of this redistribution taking place in all previous periods $-s$, $-1 \geq -s > -t$, and this is described by the first term in the second line of (26).

To determine the sign of this term turns out to be a complex task. For a two-type economy, the Row-FOSD condition implies a clear positive effect, while for $n > 2$ the Path-Dominance condition on M is needed to guarantee that redistribution of bequests in any previous period $-s$ improves period-0 welfare. PD requires, essentially, that for an individual who receives

[10] Remember our assumption that only bequests from generations up to $-t$ arrive at generation 0. Therefore inheritances of generation $-s$ originating from ancestors prior to $-t$ do not occur in the formula.

higher inheritances from a specific generation $-s$, total inheritances from all previous periods together are higher as well. It must be stressed, however, that this condition is by no means necessary; the first term in the second line of (26) is quite likely to be positive, even if the first-order stochastic dominance, required in PD, is violated in some components. In fact, a number of simulation results which we present in Appendix D, corroborate this claim. These simulations show that for randomly chosen matrices M which fulfill Row-FOSD and whose rows are unimodal with the maximum being the diagonal element, 90 percent of them either fulfill PD or violate PD in less than 8 percent of all relevant comparisons. Moreover, the extent by which the required inequalities are not fulfilled when the comparisons fail, is small. As a consequence, these results indeed allow the conclusion that redistribution of bequests in previous periods has a positive impact on steady-state welfare.

As to the second term $\gamma\Omega$ in the second line of (26), we show in Appendix C that it is positive and goes to zero for t going to infinity, that is, the larger the number of generations whose bequests we actually take into account.

The distortion of the bequest decision and the overall effect. The third line of (26) describes the social evaluation of the substitution effect associated with the introduction of the bequest tax in previous periods. It is negative or zero because of $\varepsilon \geq 0$. Given that redistribution in previous periods has a positive effect (line two of (26)), the overall effect is clearly positive for $\varepsilon = 0$. Intuitively, a zero substitution elasticity means that the ratio $b_i/(c_i - c_a)$ does not change with the tax τ_b; the latter only redistributes the available budget but does not distort the accumulation process. In other words, it only causes an income effect but no substitution effect. By continuity, we can extend this conclusion and state that 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 ρ, the parameter of inequality aversion, is large and inheritances at the lower end of the distribution are small. The reason is that for large ρ only the individuals with lowest incomes and lowest inheritances count in the social welfare function. If their inheritances are close to zero (all their ancestors in the periods $-s = -1, ..., -t$ had net incomes just sufficient to cover necessary consumption c_a) then obviously the inheritances of their ancestors are also close to zero and, thus, the third line of (26) is close to zero for any ε. As an illustration consider the limit case when the social welfare function is maximin (ρ goes to infinity), then in the formula (26) of Proposition 3 only the expressions for $i = 1$ and $r_{-s} = 1$, for all $s = 1, ..., t$, occur, and the first two lines are obviously positive, because $\overline{e} > e(1, ..., 1)$ and $\overline{e}_{-s} > e_{-s}(1, ..., 1)$. Moreover assume that $x_1 - c_a$ is zero then the steady-state inheritance $e(1, ..., 1) = \sum_{s=1}^{t} \gamma^s (x_1 - c_a)$ in period 0 is zero, as is the inheritance $e_{-s}(1, ..., 1) = \sum_{\widehat{s}=s+1}^{t} \gamma^{\widehat{s}-s} (x_1 - c_a)$ in any previous period $-s$. This implies that the third line of (26) is zero as well, and the overall effect is positive.

7.2 The consumption and the expenditure tax

Next we turn to the welfare effect of a tax on consumption, and we follow the same procedure and use the same notation as in the above analysis of the bequest tax.

Proposition 4 *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*

$$\begin{aligned} \left.\frac{\partial S}{\partial \tau_c}\right|_{\tau_c = 0} =& \lambda \sum_{i=1}^{n} f_i E_i [V_i^{-\rho}(\cdot)(1-\gamma)(\overline{e} - e(\cdot))] \\ &+ \sum_{i=1}^{n} f_i \sum_{s=1}^{t-1} E_i [V_i^{-\rho}(\cdot) \gamma^s (1-\gamma)(\overline{e}_{-s} - e_{-s}(\cdot))] + (1-\gamma)\Omega + \\ &+ \sum_{i=1}^{n} f_i E_i [V_i^{-\rho}(\cdot) \varepsilon (1-\gamma)(e(\cdot) + \sum_{s=1}^{t-1} \gamma^s e_{-s}(\cdot))]. \end{aligned} \tag{27}$$

The first line describes the immediate redistributive effect; it is positive. The second line represents the long-run redistributive effect; it is positive given Path Dominance. The third line describes the long-run distortive effect and is positive (or zero).
Given Path Dominance, the overall effect of τ_c on social welfare is positive for any $\varepsilon \geq 0$.

Proof. See Appendix C. ∎

There is obviously a close similarity between the formulae in the Propositions 3 and 4. In the first and second lines of the latter, only γ is replaced by $(1-\gamma)$, while line three is identical but with the opposite sign. As in the case of the bequest tax, the welfare effect of introducing a tax τ_c on consumption can again be split into two separate effects: the first line of (27) shows that the consumption tax allows additional redistribution of that part of consumption, $(1-\gamma)e$, which is financed out of given inheritances. This effect is positive for the same reason as stated for τ_b; the consumption tax effectively is an instrument for equalization of received inheritances.

The second line describes the welfare implication for generation 0 of this redistribution taking place in previous periods, as a consequence of the introduction of the consumption tax τ_c. $(1-\gamma)\Omega$ is positive, and for the first term exactly the same considerations apply as mentioned above for the bequest tax: a positive sign is guaranteed by the Path-Dominance condition, but our numerical simulations show that it is very likely to hold for any matrix M fulling Row-FOSD, even if Path Dominance is not strictly fulfilled. Finally, the distorting effect (the substitution term in line three of (27)) now favors the accumulation process, hence its welfare consequence is positive.[11]

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

Corollary 2 *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*

$$\begin{aligned} \left.\frac{\partial S}{\partial \tau}\right|_{\tau=0} &= \lambda \sum_{i=1}^{n} f_i E_i[V_i^{-\rho}(\cdot)(\overline{e}-e(\cdot))] \\ &\quad + \sum_{i=1}^{n} f_i \sum_{s=1}^{t-1} E_i[V_i^{-\rho}(\cdot)\gamma^s(\overline{e}_{-s}-e_{-s}(\cdot))] + \Omega. \qquad (28) \end{aligned}$$

The immediate redistributive effect (described by the first line) is positive and the long-run redistributive effect (described by the second line) is positive, given Path Dominance.

Proof. Combine (26) and (27), as the welfare effect of a uniform tax rate $\tau = \tau_b = \tau_c$ on both, bequests and consumption, is simply the sum of the welfare effects of τ_b and τ_c. ■

The interpretation of the terms is analogous to the corresponding terms in the previous propositions. The welfare effect is clearly just the sum of the effects of the two single taxes presented in the above Propositions. Most importantly, this means - in accordance with the findings for the single taxes - that by being imposed on *all expenditures* this tax actually captures all *inherited wealth* and allows its redistribution (when combined with an optimally

[11] Let us note as an illustration that in case of complete mobility of abilities over generations (identical inheritances of all individuals) the just-mentioned mechanism inducing increased accumulation would imply a positive welfare effect of the consumption tax, though there is clearly no redistributive effect.

chosen income). The first line in (28) shows the welfare consequences of this redistribution in period 0, while the second line refers to the consequences from redistribution in previous periods $-s$, $s = 1, ..., t-1$. Again, the second line is guaranteed to be positive under the Path-Dominance assumption, but is very likely to be positive otherwise. The respective third lines in (26) and (27) cancel out. Obviously, the welfare effect of τ is larger than that of the consumption tax alone if the bequest tax is welfare increasing.

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.

8 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 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.[12] 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 more able individuals tend to 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

[12] **A positve tax rate on capital income was also shown to be optimal in a model with infinitely lived individuals, incomplete insurance marekts and borrowing constraints (Aiyagari 1995 as well as in the New Dynamic Public Finance framework with stochastic skills and intertemporal incentive constraints (Golosov et al. 2003, Diamond and Mirrlees 1978).**

potential against the utility loss from a distortion of individual decisions. It turned out that from this perspective a tax on consumption is a suitable 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 period again. This amplification of inheritances (and consumption) outweighs the distortion.

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 bequest tax increases social welfare, if this reaction of bequests is sufficiently small. **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.**

Note that the marginal propensity to save (for bequests) determines the redistributive potential of the bequest tax vis-a-vis the consumption tax. The larger this propensity, the larger the redistributive effect of the bequest tax. In our model, in order to derive a tractable formulation of steady-state inheritances, the marginal propensity to save is assumed to be the same for all households, while a propensity which increases with income is more likely to conform to reality. Presumably, such an assumption would lead to a more unequal distribution of inheritances and would, thus, 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 individuals. **Moreover, again for the sake of tractability we have abstracted away from income effects on the labor supply which implies that there is no effect of receiving inheritances on the labor supply of the recepients. However with the probably more plausible assumption of leisure being a normal good, a bequest tax increases labor supply of the next generation, and we speculate that the corresponding increase in the labor incomes should mitigate (counteract?) to some extent the negative effect of the bequest tax on wealth accumulation found in our model (see Kopczuk 2013a??).**

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 3 and 4, 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 while leaving wealth accumulation undistorted.

Our results indicate that in an optimal tax system the income tax should be supplemented by a consumption tax and - particularly when a society emphasizes redistribution - by a bequest tax. This fits well to the structure of actual tax systems in OECD countries (OECD 2016, p. 103) 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 of 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 indeed be implemented 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 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.[13] Furthermore, empirical studies do not suggest that the altruistic model is more in accordance with actual behavior (see Kopczuk 2013b, among others).

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.

Appendix A. Lemmas 3 - 5

Four useful results needed later on are proved in the following.

Lemma 3 *Let M fulfill Row-FOSD and let $\delta_1, ..., \delta_n \in \mathbb{R}$, $\delta_1 \geq \delta_2 \geq .... \geq \delta_n$. Then $\sum_{k=1}^n \mu_{jk}\delta_k \geq \sum_{k=1}^n \mu_{j+1,k}\delta_k$, $j = 1, ..., n-1$.*

Proof. Interpret $-\delta_1 \leq -\delta_2 \leq .... \leq -\delta_n$ as realizations of a discrete random variable for which $\mu_{j1}, ..., \mu_{jn}$ and $\mu_{j+1,1}, ..., \mu_{j+1,n}$ are two probability distributions. The above inequality in "$\leq$"-form follows immediately from the well-known result that the expected value of $-\delta_1, ..., -\delta_n$ is larger with the second distribution than with the first, if the second distribution first-order dominates the first. Multiplication by -1 gives the required inequality. ■

Lemma 4 *Let M fulfill Row-FOSD and let $A_k, B_k, k = 1, ..., n$ be nonnegative real numbers with the properties: $A_k > A_{k+1}$, $B_k > B_{k+1}$ for $k = 1, ..., n-1$, and $A_k > B_k$ for $k = 1, ..., n$. Then $\sum_{k=1}^n \mu_{jk}A_k > \sum_{k=1}^n \mu_{j+1,k}B_k$, $j = 1, ..., n-1$.*

Proof. Lemma 3 with $B_k = \delta_k$ gives us $\sum_{k=1}^n \mu_{jk}B_k \geq \sum_{k=1}^n \mu_{j+1,k}B_k$. Replacing the B_k by larger numbers A_k increases the LHS. ■

[13] Piketty and Saez (2013) 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.

Lemma 5 *Given Path Dominance of M, in addition to Row-FOSD of M and P, then*

$$G(r_{-s}) \equiv \frac{1}{f_{r_{-s}}} \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \dots \sum_{r_{-s+1}=1}^{n} \sum_{r_{-s-1}=1}^{n} \dots \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho}. \tag{29}$$

is decreasing r_{-s} for any $s = 2, ..., t$.

Proof. We introduce, for $s < t$ and fixed r_{-s}, as well as for a given series $r_{-1}, r_{-2}, ..., r_{-s+1}$, the expected social welfare weight, where the expectation is computed over over the remaining possible paths $r_{-s-1}, ..., r_{-t}$ and the positive constant λ is neglected:

$$E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+1}, r_{-s}] \equiv$$
$$\sum_{r_{-s-1}=1}^{n} \dots \sum_{r_{-t}=1}^{n} \mu_i^c(r_{-s-1}, ..., r_{-t} \mid r_{-1}, r_{-2}, ..., r_{-s+1}, r_{-s}) V_i(r_{-1}, ..., r_{-s+1}, r_{-s}, r_{-s-1}, ..., r_{-t})^{-\rho}.$$

Here the conditional probabilities are given as $\mu_i^c(r_{-s-1}, ..., r_{-t} \mid r_{-1}, r_{-2}, ..., r_{-s+1}, r_{-s}) = \mu_{ir_{-1}} ... \mu_{r_{-s+1}r_{-s}} \mu_{r_{-s}r_{-s-1}} ... \mu_{r_{-t+1}r_{-t}} / (\mu_{ir_{-1}} ... \mu_{r_{-s+1}r_{-s}}) = \mu_{r_{-s}r_{-s-1}} \cdot ... \cdot \mu_{r_{-t+1}r_{-t}}$. Then (29) can be written as

$$\begin{aligned} G(r_{-s}) \quad = \quad & \frac{1}{f_{r_{-s}}} \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \mu_{ir_{-1}} \sum_{r_{-2}=1}^{n} \mu_{r_{-1}r_{-2}} \dots \\ & \sum_{r_{-s+1}=1}^{n} \mu_{r_{-s+2}r_{-s+1}} \mu_{r_{-s+1}r_{-s}} E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+1}, r_{-s}]. \end{aligned} \tag{30}$$

Note that on the RHS of (30), r_{-s} is fixed, the sums are computed over the remaining indices $r_{-1}, ..., r_{-s+1}$. Obviously, for $s = t$, the expected value is just equal to $V_i(\cdot)^{-\rho}$. To prove that $G(r_{-s}) > G(r_{-s} + 1)$, we first show that this inequality holds for the conditional expectation, i.e., we show, for any $i = 1, ..., n$, and fixed $r_{-1}, r_{-2}, ..., r_{-s+1}$:

$$E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+1}, r_{-s}] > E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+1}, r_{-s} + 1]. \tag{31}$$

That is, if type i has an ancestor of type $r_{-s}+1$ in period $-s$, her welfare weight is lower (utility is higher) than if the ancestor is of type r_{-s}. To see that this holds, we write inheritances of type i in the two cases as

$$\begin{aligned} e|\, r_{-1},...,r_{-s+1},r_{-s} &= e(r_{-1},...,r_{-s+1}) + \gamma^s(x_{r_{-s}} - (1-\tau_c)c_a) + e_{-s}(r_{-s-1},...,r_{-t}) \\ e|\, r_{-1},...,r_{-s+1},r_{-s}+1 &= e(r_{-1},...,r_{-s+1}) + \gamma^s(x_{r_{-s}+1} - (1-\tau_c)c_a) + e_{-s}(r_{-s-1},...,r_{-t}) \end{aligned}$$

The first term on the RHS is identical, the second term is clearly larger for $r_{-s}+1$, and for the third term we know from Proposition 1 that the probability distribution $F_{r_{-s}+1}(e_{-s})$ first-order stochastically dominates the probability distribution $F_{r_{-s}}(e_{-s})$. Altogether, as $V_i^{-\rho} \mid r_{-1},r_{-2},...,r_{-s+1},r_{-s}$ is a decreasing function of $e \mid r_{-1},...,r_{-s+1},r_{-s}$, the inequality (31) follows. We also note that $E_i[V_i^{-\rho} \mid r_{-1},r_{-2},...,r_{-s+1},r_{-s}]$ is obviously decreasing in i as well as in any $r_{-\widehat{s}}$, $\widehat{s}=1,...,s-1$.

Next, we proceed in two steps (the following applies for $s>2$; if $s=2$, (34) and (35) in step 2 below are defined immediately):

Step 1: We now take the expected value also over the possible r_{-s+1}, that is, we define, for given $r_{-1},...,r_{-s+2}$ (and, still, for fixed r_{-s}),

$$E_i[V_i^{-\rho} \mid r_{-1},...,r_{-s+2},r_{-s}] \equiv \frac{\sum\limits_{r_{-s+1}=1}^{n} \mu_{r_{-s+2}r_{-s+1}}\mu_{r_{-s+1}r_{-s}}}{\mu_{2;r_{-s+2}r_{-s}}} E_i[V_i^{-\rho} \mid r_{-1},...,r_{-s+1},r_{-s}]$$

and in the same way, by replacing r_{-s} by $r_{-s}+1$, we define $E_i[V_i^{-\rho} \mid r_{-1},r_{-2},...,r_{-s+2},r_{-s}+1]$. From above we know that (31) holds and that both $E_i[V_i^{-\rho} \mid r_{-1},r_{-2},...,r_{-s+1},r_{-s}]$ and $E_i[V_i^{-\rho} \mid r_{-1},r_{-2},...,r_{-s+1},r_{-s}+1]$ are decreasing in i as well as in any $r_{-\widehat{s}}$, $\widehat{s}=1,...,s-1$, hence in particular, in r_{-s+1}. Therefore, in view of PD which implies the following first-order stochastic dominance

$$\frac{\sum\limits_{r_{-s+1}=1}^{j} \mu_{r_{-s+2}r_{-s+1}}\mu_{r_{-s+1}r_{-s}}}{\mu_{2;r_{-s+2}r_{-s}}} \geq \frac{\sum\limits_{r_{-s+1}=1}^{j} \mu_{r_{-s+2}r_{-s+1}}\mu_{r_{-s+1}r_{-s}+1}}{\mu_{2;r_{-s+2}r_{-s}+1}}, \qquad \text{for all } j=1,...,n,$$

we find by means of Lemma 4 (with $A_k = E_i[V_i^{-\rho} \mid r_{-1},...,r_{-s+2},k,r_{-s}]$, $B_k = E_i[V_i^{-\rho} \mid r_{-1},...r_{-s+2},k,r_{-s}+1]$) that

$$E_i[V_i^{-\rho} \mid r_{-1},r_{-2},...,r_{-s+2},r_{-s}] > E_i[V_i^{-\rho} \mid r_{-1},r_{-2},...,r_{-s+2},r_{-s}+1]. \tag{32}$$

Both, $E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+2}, r_{-s}]$ and $E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+2}, r_{-s}+1]$ are obviously decreasing in i as well as in any $r_{-\widehat{s}}$, $\widehat{s} = 1, ..., s-2$.

By the definition of $E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+2}, r_{-s}]$, we can write (30) in the form

$$\begin{aligned} G(r_{-s}) &= \frac{1}{f_{r_{-s}}} \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \mu_{ir_{-1}} \sum_{r_{-2}=1}^{n} \mu_{r_{-1}r_{-2}} \cdots \\ &\quad \sum_{r_{-s+2}=1}^{n} \mu_{r_{-s+3}r_{-s+2}} \mu_{2;r_{-s+2}r_{-s}} E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+2}, r_{-s}] \end{aligned}$$

and replacing r_{-s} by $r_{-s}+1$, $G(r_{-s}+1)$ reads analogously. If $s-2=1$, we proceed with step 2 and derive (34) and (35).

If $s-2>1$ we follow the same logic as above and take the expected value also over the possible r_{-s+2}. That is, we define, for given $r_{-1}, ..., r_{-s+3}$ (and r_{-s})

$$E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+3}, r_{-s}] \equiv \frac{\sum\limits_{r_{-s+2}=1}^{n} \mu_{r_{-s+3}r_{-s+2}} \mu_{2;r_{-s+2}r_{-s}}}{\mu_{3;r_{-s+3}r_{-s}}} E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+2}, r_{-s}]$$

as well as $E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+3}, r_{-s}+1]$. Again, we know that (32) holds and that these expected values are decreasing in i as well as in any $r_{-\widehat{s}}$, $\widehat{s} = 1, ..., s-2$, hence in particular, in r_{-s+2}, and PD together with Lemma 4 gives us

$$E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+3}, r_{-s}] > E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+3}, r_{-s}+1]. \tag{33}$$

As before, these expected values are decreasing in i as well as in any $r_{-\widehat{s}}$, $\widehat{s} = 1, ..., s-3$. (30) can be written in the form

$$\begin{aligned} G(r_{-s}) &= \frac{1}{f_{r_{-s}}} \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \mu_{ir_{-1}} \sum_{r_{-2}=1}^{n} \mu_{r_{-1}r_{-2}} \cdots \\ &\quad \sum_{r_{-s+3}=1}^{n} \mu_{r_{-s+4}r_{-s+3}} \mu_{3;r_{-s+3}r_{-s}} E_i[V_i^{-\rho} \mid r_{-1}, r_{-2}, ..., r_{-s+3}, r_{-s}] \end{aligned}$$

and the analogous form applies for $G(r_{-s}+1)$. If $s-3>1$ we go through the same transformations until after $s-1$ repetitions we arrive at

Step 2, where we define

$$E_i[V_i^{-\rho} \mid r_{-s}] \equiv \frac{\sum\limits_{r_{-1}=1}^{n} \mu_{ir_{-1}}\mu_{s-1;r_{-1}r_{-s}}}{\mu_{s;ir_{-s}}} E_i[V_i^{-\rho} \mid r_{-1}, r_{-s}] \tag{34}$$

and

$$G(r_{-s}) = \frac{1}{f_{r_{-s}}} \sum_{i=1}^{n} f_i \mu_{s;ir_{-s}} E_i[V_i^{-\rho} \mid r_{-s}], \tag{35}$$

where the $E_i[V_i^{-\rho} \mid r_{-s}]$ are decreasing in r_{-s} and i, in view of PD. Noting that $f_i\mu_{s;ir_{-s}}/f_{r_{-s}} = p_{s;r_{-s}i}$ (see (18)), $G(r_{-s}) = \sum_{i=1}^{n} p_{s;r_{-s}i} E_i[V_i^{-\rho} \mid r_{-s}]$. First-order stochastic dominance of row $r_{-s}+1$ over row r_{-s} of P and Lemma 4 with $A_i = E_i[V_i^{-\rho} \mid r_{-s}]$, $B_i = E_i[V_i^{-\rho} \mid r_{-s}+1]$ indeed imply that the $G(r_{-s})$ are decreasing in r_{-s}. ■

Lemma 6 *The average inheritances of the steady-state generation* 0 *and of any prior generation* $-s = -1, ..., -t+1$ *can be expressed as*

$$\overline{e} = \sum_{s=1}^{t} \gamma_b^s (\overline{x} - (1+\tau_c)c_a) \tag{36}$$

and

$$\overline{e}_{-s} = \sum_{\widehat{s}=s+1}^{t} \gamma_b^{\widehat{s}-s} (\overline{x} - (1+\tau_c)c_a), \tag{37}$$

respectively, with $\overline{x} \equiv \sum_{i=1}^{n} f_i x_i = \sum_{r_{-s}=1}^{n} f_{r_{-s}} x_{r_{-s}}$ *denoting the average net income of generation* 0 *and of any prior generation* $-s = -1, ..., -t$.

Proof. We use (21) to rewrite $\overline{e} \equiv \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(r_{-1}, .., r_{-t}) e(r_{-1}, .., r_{-t})$ as

$$\overline{e} = \sum_{s=1}^{t} \gamma_b^s (\sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(r_{-1}, ..., r_{-t}) x_{r_{-s}} - (1+\tau_c)c_a). \tag{38}$$

We consider $\sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(r_{-1}, ..., r_{-t}) x_{r_{-s}}$ for some $s \in \{1, ..., t\}$, and rearrange it by use of (16) to

$$\sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-s}=1}^{n} \mu_{ir_{-1}} \cdot ... \cdot \mu_{r_{-s+1}r_{-s}} x_{r_{-s}} \sum_{r_{-s-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_{r_{-s}r_{-s-1}} \cdot ... \cdot \mu_{r_{-t+1}r_{-t}}$$

which reduces to $\sum_{i=1}^{n} f_i \sum_{r_{-s}=1}^{n} \mu_{s;ir_{-s}} x_{r_{-s}}$ by use of (17) and $\sum_{r_{-s-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_{r_{-s}r_{-s-1}} \cdot ... \cdot \mu_{r_{-t+1}r_{-t}} = 1$. Further we know from Section 3 that $fM^s = f$ (see (19)), thus $\sum_{i=1}^{n} f_i \mu_{s;ir_{-s}} =$

$f_{r_{-s}}$, and we can write

$$\sum_{i=1}^{n} f_i \sum_{r_{-s}=1}^{n} \mu_{s;ir_{-s}} x_{r_{-s}} = \sum_{r_{-s}=1}^{n} \sum_{i=1}^{n} f_i \mu_{s;ir_{-s}} x_{r_{-s}} = \sum_{r_{-s}=1}^{n} f_{r_{-s}} x_{-s} = \overline{x}.$$

Hence, $\sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(r_{-1}, ..., r_{-t}) x_{r_{-s}} = \overline{x}$ for any $s = 1, ..., t$. Using this result in (38) gives us (36).

Analogously, we use $e_{-s}(r_{-s-1}, ..., r_{-t})$, defined as $\sum_{\widehat{s}=s+1}^{t} \gamma_b^{\widehat{s}-s}(x_{r_{-\widehat{s}}} - (1+\tau_c)c_a)$, $s = 1, ..., t-1$, in $\overline{e}_{-s} \equiv \sum_{r_{-s}=1}^{n} f_{r_{-s}} \sum_{r_{-s-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_{r_{-s}}(r_{-s-1}, ..., r_{-t}) e_{-s}(r_{-s-1}, ..., r_{-t})$, to obtain

$$\overline{e}_{-s} = \sum_{\widehat{s}=s+1}^{t} \gamma_b^{\widehat{s}-s} \Big(\sum_{r_{-s}=1}^{n} f_{r_{-s}} \sum_{r_{-s-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_{r_{-s}}(r_{-s-1}, .., r_{-t}) x_{r_{-\widehat{s}}} - (1+\tau_c)c_a \Big). \tag{39}$$

We consider $\sum_{r_{-s}=1}^{n} f_{r_{-s}} \sum_{r_{-s-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_{r_{-s}}(r_{-s-1}, .., r_{-t}) x_{r_{-\widehat{s}}}$ for some $\widehat{s} > s$ and rearrange it to, by similar considerations as above,

$$\sum_{r_{-s}=1}^{n} f_{r_{-s}} \sum_{r_{-\widehat{s}}=1}^{n} \mu_{\widehat{s}-s;r_{-s}r_{-\widehat{s}}} x_{r_{-\widehat{s}}} = \sum_{r_{-\widehat{s}}=1}^{n} \sum_{r_{-s}=1}^{n} f_{r_{-s}} \mu_{\widehat{s}-s;r_{-s}r_{-\widehat{s}}} x_{r_{-\widehat{s}}} = \sum_{r_{-\widehat{s}}=1}^{n} f_{r_{-\widehat{s}}} x_{r_{-\widehat{s}}} = \overline{x}.$$

By use of this result in (39) we obtain (37). ■

Appendix B. Proofs of Proposition 1, Corollary 1 and Lemma 1

Proof of Proposition 1. Let $\widehat{e}(\widehat{r}_{-1}, .., \widehat{r}_{-t})$ be some particular realization of inheritances in period 0, determined by the abilities of parents, grandparents, ... Define for any series $(r_{-1}, r_{-2}, ..., r_{-t+1})$ the number $R(r_{-1}, r_{-2}, ..., r_{-t+1}) \equiv \max\{r_{-t} \in \{1, ..., n\} \mid \sum_{s=1}^{t-1} \gamma_b^s (x_{r_{-s}} - (1+\tau_c)c_a) + \gamma_b^t (x_{r_{-t}} - (1+\tau_c)c_a) \leq \widehat{e}\}$. (If the set in this definition is empty, then let $R(r_{-1}, r_{-2}, ..., r_{-t+1}) \equiv 0$.) That is, for any series of types $(r_{-1}, r_{-2}, ..., r_{-t+1})$ in the prior generations up to $-t+1$, $R(r_{-1}, r_{-2}, ..., r_{-t+1})$ denotes the highest skill type in generation $-t$ such that total inheritances do not exceed $\widehat{e}$. Therefore, with $F_i(\widehat{e})$ denoting the probability distribution of inheritances received by individual i, we have

$$F_i(\widehat{e}) = \sum_{r_{-1}=1}^{n} \mu_{ir_{-1}} \sum_{r_{-2}=1}^{n} \mu_{r_{-1}r_{-2}} \sum_{r_{-3}=1}^{n} \mu_{r_{-2}r_{-3}} \cdot ... \cdot \sum_{r_{-t+1}=1}^{n} \mu_{r_{-t+2}r_{-t+1}} \sum_{r_{-t}=1}^{R(r_{-1}, r_{-2}, ..., r_{-t+1})} \mu_{r_{-t+1}r_{-t}} \tag{40}$$

Note that the last sum term in (40) is zero if $R(\cdot) = 0$. Now consider this last term for two consecutive indices (ability types) r_{-t+1} and $r_{-t+1}+1$, corresponding to two ancestors with net incomes $x_{r_{-t+1}} < x_{r_{-t+1}+1}$ in the last but one period $-t+1$. FOSD of row $r_{-t+1}+1$ of M over the row r_{-t+1} means $\sum_{r_{-t}=1}^{j} \mu_{r_{-t+1}r_{-t}} \geq \sum_{r_{-t}=1}^{j} \mu_{r_{-t+1}+1,r_{-t}}$ for all $j = 1,...,n$. Moreover, by definition, $R(r_{-1}, r_{-2}, ..., r_{-t+1}) \geq R(r_{-1}, r_{-2}, ..., r_{-t+1} + 1)$, because individual $r_{-t+1} + 1$ leaves a higher estate than individual r_{-t+1} and, thus, the set of ancestors r_{-t} such that total inheritances do not exceed $\widehat{e}$, is smaller for the former (or equal). As a consequence,

$$\sum_{r_{-t}=1}^{R(r_{-1},r_{-2},...,r_{-t+1})} \mu_{r_{-t+1}r_{-t}} \geq \sum_{r_{-t}=1}^{R(r_{-1},r_{-2},...,r_{-t+1}+1)} \mu_{r_{-t+1}+1,r_{-t}}. \tag{41}$$

(41) tells us that the last sum term in (40) is nonincreasing with $r_{-t+1} = 1,...,n$. Thus, denoting this sum term (for any given series $r_{-1}, r_{-2}, ..., r_{-t+2}$) by $\delta_{r_{-t+1}}(r_{-1}, r_{-2}, ..., r_{-t+2}) \equiv \sum_{r_{-t}=1}^{R(r_{-1},r_{-2},...,r_{-t+1})} \mu_{r_{-t+1}r_{-t}}$, we have $\delta_{r_{-t+1}}(\cdot) \geq \delta_{r_{-t+1}+1}(\cdot)$ and we can apply Lemma 3 (see Appendix A) to derive for the last two sum terms in (40) that for any two consecutive indices (ability types) r_{-t+2} and $r_{-t+2}+1$ in period $-t+2$ (note that the $\mu_{r_{-t+2}+1,r_{-t+1}}$ first-order stochastically dominate the $\mu_{r_{-t+2}r_{-t+1}}$, where $r_{-t+1} = 1,...,n$):

$$\sum_{r_{-t+1}=1}^{n} \mu_{r_{-t+2}r_{-t+1}} \sum_{r_{-t}=1}^{R(r_{-1},r_{-2},...,r_{-t+1})} \mu_{r_{-t+1}r_{-t}} \geq \sum_{r_{-t+1}=1}^{n} \mu_{r_{-t+2}+1,r_{-t+1}} \sum_{r_{-t}=1}^{R(r_{-1},r_{-2},...,r_{-t+1})} \mu_{r_{-t+1}r_{-t}}.$$

Hence, the last two sum terms in (40) are nonincreasing with r_{-t+2}. Denoting again these two sum terms (for any given series $r_{-1}, r_{-2}, ..., r_{-t+3}$) by $\delta_{r_{-t+2}}(r_{-1}, r_{-2}, ..., r_{-t+3}) \equiv \sum_{r_{-t+1}=1}^{n} \mu_{r_{-t+2}r_{-t+1}} \sum_{r_{-t}=1}^{R(r_{-1},r_{-2},...,r_{-t+1})} \mu_{r_{-t+1}r_{-t}}$, we can again apply Lemma 3 to see that the last three terms in (40) are nonincreasing with r_{-t+3}. Repeated application of the same argument tells us finally that the RHS in (40) is nonincreasing in i, that is,

$$\sum_{r_{-1}=1}^{n} \mu_{ir_{-1}} \sum_{r_{-2}=1}^{n} \mu_{r_{-1}r_{-2}} \sum_{r_{-3}=1}^{n} \mu_{r_{-2}r_{-3}} \cdot ... \cdot \sum_{r_{-t+1}=1}^{n} \mu_{r_{-t+2}r_{-t+1}} \sum_{r_{-t}=1}^{R(r_{-1},r_{-2},...,r_{-t+1})} \mu_{r_{-t+1}r_{-t}} \geq$$
$$\sum_{r_{-1}=1}^{n} \mu_{i+1,r_{-1}} \sum_{r_{-2}=1}^{n} \mu_{r_{-1}r_{-2}} \sum_{r_{-3}=1}^{n} \mu_{r_{-2}r_{-3}} \cdot ... \cdot \sum_{r_{-t+1}=1}^{n} \mu_{r_{-t+2}r_{-t+1}} \sum_{r_{-t}=1}^{R(r_{-1},r_{-2},...,r_{-t+1})} \mu_{r_{-t+1}r_{-t}},$$

which means $F_i(\widehat{e}) \geq F_{i+1}(\widehat{e})$. ■

Proof of Corollary 1. Obviously, the number of potential (different) realizations of e is larger than 1, because with $\sigma < 1$, more-able individuals have higher net incomes and leave higher bequests. Then Proposition 1 immediately implies $E_i[e] < E_{i+1}[e]$, as a general consequence of first-order stochastic dominance of the distribution F_{i+1} over F_i.

From (9), $E_i[V_i]$ can be written as $\lambda E_i[e] + \lambda(x_i - c_a(1+\tau_c)) - g(l_i)$. We know that $E_i[e] < E_{i+1}[e]$. Moreover, $\lambda(x_i - c_a(1+\tau_c)) - g(l_i)$, which describes utility in case of zero inheritances, clearly increases with the ability type i, given a marginal tax rate $\sigma < 1$. This proves $E_i[V_i] < E_{i+1}[V_{i+1}]$.

Next, note that $(\lambda e(\cdot) + Z)^{-\rho}$ is a decreasing function of $e(\cdot)$ for any constant Z. Therefore, an elementary result for FOSD tells us that $E_i[(\lambda e(\cdot) + Z)^{-\rho}] > E_{i+1}[(\lambda e(\cdot) + Z)^{-\rho}]$. Finally, using again that $\lambda(x_i - c_a(1+\tau_c)) - g(l_i) < \lambda(x_{i+1} - c_a(1+\tau_c)) - g(l_{i+1})$, the inequality remains true if Z is replaced by these two terms, which gives $E_i[(\lambda(e(\cdot) + x_i - c_a(1+\tau_c)) - g(l_i))^{-\rho}] > E_{i+1}[(\lambda(e(\cdot) + x_{i+1} - c_a(1+\tau_c)) - g(l_{i+1}))^{-\rho}]$. ∎

Proof of Lemma 1. Consider any matrix $R = AB$, where both $n \times n$ matrices A and B, with elements a_{ij} and b_{ij} fulfill Row-FOSD. The required inequality $\sum_{j=1}^{l} r_{ij} \geq \sum_{j=1}^{l} r_{i+1,j}$, for any $i = 1, ..., n-1$ and $l = 1, ..., n$ is written as $\sum_{j=1}^{l} \sum_{k=1}^{n} a_{ik} b_{kj} \geq \sum_{j=1}^{l} \sum_{k=1}^{n} a_{i+1,k} b_{kj}$ which is equivalent to $\sum_{k=1}^{n} a_{ik} \sum_{j=1}^{l} b_{kj} \geq \sum_{k=1}^{n} a_{i+1,k} \sum_{j=1}^{l} b_{kj}$. Set $\delta_k \equiv \sum_{j=1}^{l} b_{kj}$, then $\delta_k \geq \delta_{k+1}$, because of Row-FOSD of B. Lemma 3 (see Appendix A) together with Row-FOSD of A imply the required inequality. ∎

Appendix C. Proofs of Propositions 2 - 4

Proof of Proposition 2. Let $\tau_b = \tau_c = 0$. The first derivatives with respect to α and σ of the Lagrangian $\mathcal{L}$ to the problem (22), (23) (with ν as the Lagrange multiplier to (23)) are:

$$\frac{\partial \mathcal{L}}{\partial \alpha} = \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(r_{-1}, ..., r_{-t}) V_i(r_{-1}, ..., r_{-t})^{-\rho} \frac{\partial V_i(\cdot)}{\partial \alpha} + \nu(\sigma \sum_{i=1}^{n} f_i w_i \frac{\partial l_i}{\partial \alpha} - 1), \quad (42)$$

$$\frac{\partial \mathcal{L}}{\partial \sigma} = \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho} \frac{\partial V_i(\cdot)}{\partial \sigma} + \nu \sum_{i=1}^{n} f_i (w_i l_i + \sigma w_i \frac{\partial l_i}{\partial \sigma}). \quad (43)$$

By application of the Envelope Theorem for (1) - (3) we have

$$\frac{\partial V_i(\cdot)}{\partial \alpha} = \lambda(\frac{\partial e(\cdot)}{\partial \alpha} + 1), \quad (44)$$

$$\frac{\partial V_i(\cdot)}{\partial \sigma} = \lambda(\frac{\partial e(\cdot)}{\partial \sigma} - w_i l_i). \tag{45}$$

Using (44), together with $\partial l_i/\partial \alpha = 0$ in (42) and setting it equal to zero, gives the first-order condition for α as

$$\lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho}(\frac{\partial e(\cdot)}{\partial \alpha} + 1) - \nu = 0. \tag{46}$$

Substituting (45) into (43) we obtain

$$\frac{\partial \mathcal{L}}{\partial \sigma} = \lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho}(\frac{\partial e(\cdot)}{\partial \sigma} - w_i l_i) + \nu \sum_{i=1}^{n} f_i(w_i l_i + \sigma w_i \frac{\partial l_i}{\partial \sigma}). \tag{47}$$

To determine the welfare effect of introducing a linear income tax (increase of σ with α adapted appropriately), we substitute for ν from the condition (46) for optimal α in $\partial \mathcal{L}/\partial \sigma$ and evaluate at $\sigma = 0$. After some transformations we obtain

$$\begin{aligned} \left.\frac{\partial \mathcal{L}}{\partial \sigma}\right|_{\sigma=0} &= \lambda \sum_{i=1}^{n} f_i E_i[V_i^{-\rho}](\overline{wl} - w_i l_i) + \\ &\quad \lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho}(\frac{\partial e(\cdot)}{\partial \alpha}\overline{wl} + \frac{\partial e(\cdot)}{\partial \sigma}). \end{aligned} \tag{48}$$

In addition, we use the steady-state equation (21) for $e(r_{-1}, ..., r_{-t})$ to determine

$$\frac{\partial e(r_{-1}, ..., r_{-t})}{\partial \alpha} = \sum_{s=1}^{t} \gamma_b^s, \tag{49}$$

$$\frac{\partial e(r_{-1}, ..., r_{-t})}{\partial \sigma} = \sum_{s=1}^{t} \gamma_b^s(-w_{r_{-s}} l_{r_{-s}} + (1-\sigma) w_{r_{-s}} \frac{\partial l_{r_{-s}}}{\partial \sigma}). \tag{50}$$

From (49) and (50) we get, at $\sigma = 0$ (and $\tau_b = \tau_c = 0$),

$$\frac{\partial e(\cdot)}{\partial \alpha}\overline{wl} + \frac{\partial e(\cdot)}{\partial \sigma} = \sum_{s=1}^{t} \gamma^s(\overline{wl} - w_{r_{-s}} l_{r_{-s}} + w_{r_{-s}} \frac{\partial l_{r_{-s}}}{\partial \sigma}).$$

Substituting this expression into (48), we obtain

$$\begin{aligned}\left.\frac{\partial \mathcal{L}}{\partial \sigma}\right|_{\sigma=0} &= \lambda \sum_{i=1}^{n} f_i E_i[V_i^{-\rho}](\overline{wl} - w_i l_i) + \\ &\lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \cdots \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho} \sum_{s=1}^{t} \gamma^s (\overline{wl} - w_{r_{-s}} l_{r_{-s}}) + \\ &\lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \cdots \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho} \sum_{s=1}^{t} \gamma^s w_{r_{-s}} \frac{\partial l_{r_{-s}}}{\partial \sigma}, \end{aligned} \tag{51}$$

which after changing the order of the sum terms and using the expression for $E_i[V_i^{-\rho} \mid r_{-s}]$ gives us the formula (24) in the text.

Next, we determine the signs of each of the three terms of (24). As to the first term, $\lambda \sum_{i=1}^{n} E_i[V_i^{-\rho}] f_i(\overline{wl} - w_i l_i)$, remember that $\sum_{i=1}^{n} f_i w_i l_i = \overline{wl}$, therefore $\sum_{i=1}^{n} f_i(\overline{wl} - w_i l_i) = 0$. In view of $\overline{wl} - w_1 l_1 > \overline{wl} - w_2 l_2 > \cdots > \overline{wl} - w_n l_n$, the first term in (24) is positive because $E_i[V_i^{-\rho}]$ is decreasing in the ability type i, as is shown in Corollary 1.

To show that the second line of (24) is positive too, note first that at $\tau_b = \tau_c = 0$ and $\sigma = 0$ (and thus also $\alpha = 0$), (21) reads as $e(\cdot) = \sum_{s=1}^{t} \gamma^s (w_{r_{-s}} l_{r_{-s}} - c_a)$ and (36) reads as $\overline{e} = \sum_{s=1}^{t} \gamma^s (\overline{wl} - c_a)$. Thus, the second term of (24) (used in the form of the second line of (51)) can be rewritten as

$$\lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \cdots \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho} (\overline{e} - e(\cdot)). \tag{52}$$

We split $\overline{e} - e(\cdot)$ into variations of expected inheritances of the groups and variations within the groups and get

$$\begin{aligned} &\lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \cdots \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho} (\overline{e} - E_i[e]) + \\ &\lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \cdots \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho} (E_i[e] - e(\cdot)) \end{aligned} \tag{53}$$

where we note that $\overline{e} = \sum_{i=1}^{n} f_i E_i[e]$. The first line of (53) can be written as $\lambda \sum_{i=1}^{n} E_i[V_i^{-\rho}] f_i(\overline{e} - E_i[e])$. From Lemma 2 it follows that $E_i[e] < E_{i+1}[e]$. Moreover, $\sum_{i=1}^{n} f_i(\overline{e} - E_i[e]) = 0$, hence decreasing weights $E_i[V_i^{-\rho}]$ (Corollary 1) imply that the first line in (53) is positive. As to the second line in (53) we have $\sum_{r_{-1}=1}^{n} \cdots \sum_{r_{-t}=1}^{n} \mu_i(\cdot)(E_i[e] - e(\cdot)) = 0$, for any $i = 1, ..., n$. In addition, $V_i^{-\rho}(r_{-1}, ..., r_{-t}) > V_i^{-\rho}(\widetilde{r}_{-1}, ..., \widetilde{r}_{-t})$ if $e(r_{-1}, ..., r_{-t}) < e(\widetilde{r}_{-1}, ..., \widetilde{r}_{-t})$. Thus the welfare weights are decreasing in $e(r_{-1},...,r_{-t})$, that is, in the second line of (53), positive terms have larger weights than negative terms for each $i = 1,...n$. Therefore, the second line is positive.

Finally, the third term (24) is negative as $\partial l_{r_{-s}}/\partial\sigma < 0$. It follows immediately that the overall welfare effect is positive given that this third negative term is dominated by the positive terms in the first two lines. ■

Proof of Proposition 3. *Derivation of formula (26).* Using the Envelope Theorem, we get for the optimal value function $S(\tau_b, \tau_c)$ of the maximization problem (22) and (23)

$$\begin{aligned}\frac{\partial S}{\partial \tau_b} &= \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(r_{-1},...,r_{-t})V_i(r_{-1},...,r_{-t})^{-\rho}\frac{\partial V_i(\cdot)}{\partial \tau_b} \\ &+\nu \sum_{i=1}^{n} f_i\{\sigma w_i \frac{\partial l_i}{\partial \tau_b} + (\gamma_b + \tau_b\frac{\partial \gamma_b}{\partial \tau_b})(E_i[e] + x_i - (1+\tau_c)c_a) \\ &+(\tau_b\gamma_b + \tau_c\gamma_c)(E_i[\frac{\partial e(\cdot)}{\partial \tau_b}] + (1-\sigma)w_i\frac{\partial l_i}{\partial \tau_b})\}. \end{aligned} \tag{54}$$

Applying the Envelope Theorem for (1) - (3) and using demand function (8) (with e_i being replaced by $e(\cdot)$) gives us

$$\frac{\partial V_i(\cdot)}{\partial \tau_b} = \lambda(-b_i + \frac{\partial e(\cdot)}{\partial \tau_b}) = \lambda(-\gamma_b(e(\cdot) + x_i - (1+\tau_c)c_a) + \frac{\partial e(\cdot)}{\partial \tau_b}). \tag{55}$$

Substituting (55) and $x_i = \alpha + (1-\sigma)w_i l_i$ into (54), we get, at $\tau_c = \tau_b = 0$,

$$\begin{aligned}\left.\frac{\partial S}{\partial \tau_b}\right|_{\tau_b=0} =&\lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(\cdot)V_i(\cdot)^{-\rho}\{-\gamma(e(\cdot) + \alpha + (1-\sigma)w_i l_i - c_a) + \frac{\partial e(\cdot)}{\partial \tau_b}\} \\ &+ \nu \sum_{i=1}^{n} f_i\{\sigma w_i \frac{\partial l_i}{\partial \tau_b} + \gamma(E_i[e] + \alpha + (1-\sigma)w_i l_i - c_a)\}. \end{aligned} \tag{56}$$

Setting (47) equal to zero, gives us the first-order condition for optimal σ at $\tau_c = \tau_b = 0$:

$$\lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(\cdot)V_i(\cdot)^{-\rho}(\frac{\partial e(\cdot)}{\partial \sigma} - w_i l_i) + \nu \sum_{i=1}^{n} f_i(w_i l_i + \sigma w_i \frac{\partial l_i}{\partial \sigma}) = 0. \tag{57}$$

Implicit differentiation of (6) with respect to σ and τ_b, respectively, gives the relation $\partial l_i/\partial\tau_b = -(\partial l_i/\partial\sigma)(\partial\lambda/\partial\tau_b)(1-\sigma)/\lambda$ and from Section 2 we know that $\lambda = \widetilde{\varphi}(\gamma_c, \gamma_b)$. Thus, we have $\partial\lambda/\partial\tau_b = (\partial\widetilde{\varphi}/\partial(c_i - c_a))(\partial\gamma_c/\partial\tau_b) + (\partial\widetilde{\varphi}/\partial b_i)(\partial\gamma_b/\partial\tau_b)$. Writing the objective function (1) in terms of $\widetilde{\varphi}(c_i - c_a, b_i)$ and using the f.o.c.'s (4) and (5) we obtain $\partial\lambda/\partial\tau_b = \lambda(1+\tau_c)(\partial\gamma_c/\partial\tau_b) + \lambda(1+\tau_b)(\partial\gamma_b/\partial\tau_b)$, which can be transformed further to $\partial\lambda/\partial\tau_b = -\lambda\gamma_b$ by use of $\partial\gamma_c/\partial\tau_b = -(\partial\gamma/\partial\tau_b)/(1+\tau_c)$, and of

$$\frac{\partial\gamma_b}{\partial\tau_b} = \frac{\partial\gamma/\partial\tau_b}{1+\tau_b} - \frac{\gamma}{(1+\tau_b)^2} = \frac{\partial\gamma/\partial\tau_b - \gamma_b}{1+\tau_b}; \tag{58}$$

hence we have

$$\frac{\partial l_i}{\partial\tau_b} = \gamma_b(1-\sigma)\frac{\partial l_i}{\partial\sigma}. \tag{59}$$

Using this relation at $\tau_b = 0$ as well as (57) multiplied by $\gamma(1-\sigma)$, we can write (56) as (remember that $\overline{e} = \sum_{i=1}^n f_i E_i[e]$)

$$\begin{aligned}\left.\frac{\partial S}{\partial\tau_b}\right|_{\tau_b=0} =& -\lambda\sum_{i=1}^n f_i \sum_{r_{-1}=1}^n \cdots \sum_{r_{-t}=1}^n \mu_i(\cdot)V_i(\cdot)^{-\rho}\gamma(e(\cdot)+\alpha-c_a) + \nu\gamma(\overline{e}+\alpha-c_a)\\ &+\lambda\sum_{i=1}^n f_i \sum_{r_{-1}=1}^n \cdots \sum_{r_{-t}=1}^n \mu_i(\cdot)V_i(\cdot)^{-\rho}\{\frac{\partial e(\cdot)}{\partial\tau_b} - \gamma(1-\sigma)\frac{\partial e(\cdot)}{\partial\sigma}\},\end{aligned} \tag{60}$$

which, after substitution for ν by use of the condition (46) for optimal α, is further transformed to

$$\left.\frac{\partial S}{\partial\tau_b}\right|_{\tau_b=0} = \lambda\sum_{i=1}^n f_i \sum_{r_{-1}=1}^n \cdots \sum_{r_{-t}=1}^n \mu_i(\cdot)V_i(\cdot)^{-\rho}\{\gamma(\overline{e}-e(\cdot)) + Q\}, \tag{61}$$

with Q defined as

$$Q \equiv \frac{\partial e(\cdot)}{\partial\tau_b} - \gamma(1-\sigma)\frac{\partial e(\cdot)}{\partial\sigma} + \gamma(\overline{e}+\alpha-c_a)\frac{\partial e(\cdot)}{\partial\alpha}. \tag{62}$$

Next, we use (21) to compute $\partial e(\cdot)/\partial\tau_b$ at $\tau_c = 0$:

$$\frac{\partial e(\cdot)}{\partial\tau_b} = \sum_{s=1}^t (s\gamma_b^{s-1}\frac{\partial\gamma_b}{\partial\tau_b}(x_{r_{-s}} - c_a) + \gamma_b^s(1-\sigma)w_{r_{-s}}\frac{\partial l_{r_{-s}}}{\partial\tau_b}). \tag{63}$$

With (63) at $\tau_b = 0$, (49) and (50), as well as using (59) and $x_{r_{-s}} = \alpha + (1-\sigma)w_{r_{-s}}l_{r_{-s}}$, Q becomes

$$Q = \gamma\sum_{s=1}^t \gamma^s(\overline{e} + x_{r_{-s}} - c_a) + \sum_{s=1}^t s\gamma^{s-1}\frac{\partial\gamma_b}{\partial\tau_b}(x_{r_{-s}} - c_a), \tag{64}$$

where $\partial\gamma_b/\partial\tau_b$ is given by (58). Next, we introduce the elasticity of substitution between bequests and consumption (above minimum consumption c_a), defined as

$$\varepsilon = -\frac{\partial[b_i/(c_i - c_a)]}{\partial[p_b/p_c]} \frac{p_b/p_c}{b_i/(c_i - c_a)} \geq 0 \tag{65}$$

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

$$\varepsilon = -\frac{\partial[\gamma/((1-\gamma)p_b)]}{\partial p_b} \frac{(1-\gamma)p_b^2}{\gamma} = \frac{-(\partial\gamma/\partial p_b)p_b + \gamma(1-\gamma)}{\gamma(1-\gamma)}. \tag{66}$$

Solving (66) explicitly for $\partial\gamma/\partial p_b$ and observing that $\partial p_b/\partial\tau_b = 1$, we arrive at

$$\left.\frac{\partial\gamma}{\partial\tau_b}\right|_{\tau_b=0} = (1-\varepsilon)\gamma(1-\gamma) \tag{67}$$

which further gives us (see (58))

$$\left.\frac{\partial\gamma_b}{\partial\tau_b}\right|_{\tau_b=0} = -\gamma(\gamma + \varepsilon(1-\gamma)). \tag{68}$$

By use of (36) and (68), we write (64) as

$$Q = \gamma(\sum_{s=1}^{t} \gamma^s)^2(\overline{x} - c_a) - \gamma \sum_{s=1}^{t} (s-1)\gamma^s(x_{r_{-s}} - c_a) - \varepsilon(1-\gamma) \sum_{s=1}^{t} s\gamma^s(x_{r_{-s}} - c_a). \tag{69}$$

$(\sum_{s=1}^{t} \gamma^s)^2$ can be decomposed into $\sum_{s=1}^{t}(s-1)\gamma^s + \sum_{s=1}^{t}(t-s+1)\gamma^{t+s}$,[14] which we use to transform (69) to

$$Q = \gamma \sum_{s=1}^{t} (s-1)\gamma^s(\overline{x} - x_{r_{-s}}) + \gamma^{1+t} \sum_{s=1}^{t} (t-s+1)\gamma^s(\overline{x} - c_a) - \varepsilon(1-\gamma) \sum_{s=1}^{t} s\gamma^s(x_{r_{-s}} - c_a). \tag{70}$$

Next, for $s < t$ and for some series $r_{-s-1}, ..., r_{-t}$, we use the definition of $e_{-s}(r_{-s-1}, ..., r_{-t})$ at $\tau_b = \tau_c = 0$ to compute $\gamma^s e_{-s}(\cdot) = \sum_{\widehat{s}=s+1}^{t} \gamma^{\widehat{s}}(x_{r_{-\widehat{s}}} - c_a)$ and further

$$\begin{aligned} \sum_{s=1}^{t-1} \gamma^s e_{-s}(\cdot) &= \sum_{s=1}^{t-1} \sum_{\widehat{s}=s+1}^{t} \gamma^{\widehat{s}}(x_{r_{-\widehat{s}}} - c_a) \\ &= \sum_{s=1}^{t} (s-1)\gamma^s(x_{r_{-s}} - c_a). \end{aligned} \tag{71}$$

[14] $(\sum_{s=1}^{t} \gamma^s)^2 = (\gamma + \gamma^2 + ... + \gamma^{t-1} + \gamma^t)(\gamma + \gamma^2 + ... + \gamma^{t-1} + \gamma^t) = \gamma^2 + 2\gamma^3 + 3\gamma^4 ... + (t-1)\gamma^t + t\gamma^{t+1}$ $+(t-1)\gamma^{t+2} + (t-2)\gamma^{t+3} + ... + 3\gamma^{2t-2} + 2\gamma^{2t-1} + \gamma^{2t} = \sum_{s=1}^{t}(s-1)\gamma^s + \sum_{s=1}^{t}(t+1-s)\gamma^{t+s}$.

Analogously, by use of (37) we find that

$$\sum_{s=1}^{t-1} \gamma^s \overline{e}_{-s} = \sum_{s=1}^{t} (s-1)\gamma^s (\overline{x} - c_a). \tag{72}$$

(71) and (72), together with (21), allows us to write (70) as

$$Q = \gamma \sum_{s=1}^{t-1} \gamma^s (\overline{e}_{-s} - e_{-s}(\cdot)) + \gamma^{1+t} \sum_{s=1}^{t} (t-s+1)\gamma^s (\overline{x} - c_a) - \varepsilon(1-\gamma)(e(\cdot) + \sum_{s=1}^{t-1} \gamma^s e_{-s}(\cdot)). \tag{73}$$

Writing Q in (61) in the form (73) and defining

$$\Omega \equiv \lambda \sum_{i=1}^{n} f_i E_i[V_i^{-\rho}] \gamma^t \sum_{s=1}^{t} (t-s+1)\gamma^s (\overline{x} - c_a) \tag{74}$$

we obtain formula (26) in Proposition 3.

Determination of the signs of the terms in (26). Note that the first line of (26) is equal to (52), multiplied by γ, and is, thus, positive as shown in the Proof of Proposition 1. Line three of (26) is negative (or zero) because of $\varepsilon \geq 0$.

In the second line, Ω defined in (74) is immediately seen to be positive; it goes to zero with increasing t. It remains to show that the first term, the long-run redistributive effect of τ_b is positive in case that the condition of Path Dominance is satisfied. We go back to (61) and write Q in the version (70) which is equivalent to (73). This allows us to rewrite the first term in the second line of (26) as

$$\lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \dots \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho} \gamma \sum_{s=1}^{t} (s-1)\gamma^s (\overline{x} - x_{r_{-s}}). \tag{75}$$

We consider (75) separately for each $s > 1$. (Note that for $s = 1$ the corresponding term is zero.) By reordering we obtain

$$\lambda\gamma(s-1)\gamma^s \sum_{r_{-s}=1}^{n} (\overline{x} - x_{r_{-s}}) \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} \dots \sum_{r_{-s+1}=1}^{n} \sum_{r_{-s-1}=1}^{n} \dots \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho}, \quad s = 2, ..., t, \tag{76}$$

and further, by use of $G(r_{-s})$, defined in (29),

$$\lambda\gamma(s-1)\gamma^s \sum_{r_{-s}=1}^{n} (\overline{x} - x_{r_{-s}}) f_{r_{-s}} G(r_{-s}), \quad s = 2, ..., t. \tag{77}$$

By definition, $\sum_{r_{-s}=1}^{n}(\overline{x}-x_{r_{-s}})f_{r_{-s}}=0$, and $\overline{x}-x_{r_{-s}}$ decreases in r_{-s} for any $s=2,...,t$. Moreover, it is shown in Lemma 5 (see Appendix A) that the weights $G(r_{-s})$ of $\overline{x}-x_{r_{-s}}$ also decrease in r_{-s} for any $s=2,...,t$, if Path Dominance of M holds, in addition to Row-FOSD of M and P. As a consequence, (77) is positive for any $s=2,...,t$ and, thus, also (75). Note that, though we use Row-FOSD of P as a separate condition, it is in fact implied by our general assumption of Row-FOSD of M together with Path Dominance of M, as our extensive simulations with randomly chosen matrices show, see Appendix D.

The overall effect. (i) Given PD, the first two lines of (26) are known to be positive, therefore the overall welfare effect of τ_b is positive as well, for sufficiently small values of ε.

(ii) Next consider the limiting case $\rho\to\infty$, where the social welfare function (22) approaches the maximin form. Then the objective function of the maximization problem (22) - (23) is indirect utility $V_1(1,...,1)$ of the group with the lowest ability w_1 and the lowest inheritances $e(1,...,1)=\sum_{s=1}^{t}\gamma^s(x_1-c_a)$; and all the sums in (61), with Q written in the form (70) reduce to $i=1$, $r_{-s}=1$ for all $s=1,...,t$ which yields[15]

$$\left.\frac{\partial S}{\partial\tau_b}\right|_{\tau_b=0} = \lambda\gamma(\overline{e}-e(1,...,1))+\lambda\gamma\sum_{s=1}^{t}(s-1)\gamma^s(\overline{x}-x_1)+ \lambda\gamma^{1+t}\sum_{s=1}^{t}(t-s+1)\gamma^s(\overline{x}-c_a)-\lambda\varepsilon(1-\gamma)\sum_{s=1}^{t}s\gamma^s(x_1-c_a). \quad (78)$$

The first term is positive because $\overline{e}>e(1,...,1)$, the second term is positive, because $\overline{x}>x_1$, and, obviously, the third term is positive as well. Moreover, let $e(1,..,1)$ be arbitrarily small, which means that x_1-c_a is close to zero. Then the fourth term in (78) is also close to zero for any $\varepsilon>0$, and the overall welfare effect is positive. By continuity, the overall welfare effect is also positive for large but finite values of ρ and for small values of $e(\cdot)$ at the lower end of the inheritance distribution. ■

Proof of Proposition 4. We proceed as in the proof of Proposition 3 and determine the

[15]Note that the optimal linear income tax is derived by maximizing indirect utility $V_1(1,..,1)$ of the group with the lowest ability w_1 and the lowest inheritances $e(1,..,1)$ subject to (23). As also in this case the optimal marginal tax rate σ on labor income is smaller than 1, net incomes increase with ability i. Consequently, $V_1(1,..,1)$ continues to be the lowest utility in the economy and the welfare effect of introducing τ_b is derived by differentiating the Lagrangian with $V_1(1,..,1)$ as the objective function.

derivative of the optimal value function $S(\tau_b, \tau_c)$ with respect to τ_c

$$\begin{aligned}\frac{\partial S}{\partial \tau_c} \quad = \quad & \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(r_{-1},..,r_{-t}) V_i(r_{-1},..,r_{-t})^{-\rho} \frac{\partial V_i(\cdot)}{\partial \tau_c} \\ & +\nu \sum_{i=1}^{n} f_i \{\sigma w_i \frac{\partial l_i}{\partial \tau_c} + (\gamma_c + \tau_c \frac{\partial \gamma_c}{\partial \tau_c})(E_i[e] + x_i - (1+\tau_c)c_a) \\ & +c_a + (\tau_b \gamma_b + \tau_c \gamma_c)(E_i[\frac{\partial e(\cdot)}{\partial \tau_c}] + (1-\sigma) w_i \frac{\partial l_i}{\partial \tau_c} - c_a)],\end{aligned} \tag{79}$$

where we obtain $\partial V_i(\cdot)/\partial \tau_c$ by applying the Envelope Theorem as

$$\frac{\partial V_i(\cdot)}{\partial \tau_c} = \lambda(-c_i + \frac{\partial e(\cdot)}{\partial \tau_c}) = \lambda(-\gamma_c(e(\cdot) + x_i - (1+\tau_c)c_a) - c_a + \frac{\partial e(\cdot)}{\partial \tau_c}), \tag{80}$$

and we have inserted (7). By use of (80) and of the definitions of x_i, $\overline{wl}$ and $\overline{e}$, we write (79), evaluated at $\tau_b = \tau_c = 0$

$$\begin{aligned}\left.\frac{\partial S}{\partial \tau_c}\right|_{\tau_c=0} =& \lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho} \{-(1-\gamma)(e(\cdot) + \alpha + (1-\sigma) w_i l_i - c_a) - c_a \\ & + \frac{\partial e(\cdot)}{\partial \tau_c}\} + \nu \sum_{i=1}^{n} f_i \sigma w_i \frac{\partial l_i}{\partial \tau_c} + \nu\{(1-\gamma)(\overline{e} + \alpha + (1-\sigma)\overline{wl} - c_a) + c_a\}.\end{aligned} \tag{81}$$

Using (57), multiplied by $(1-\gamma)(1-\sigma)$, and observing that $\partial l_i/\partial \tau_c|_{\tau_c=0} = (1-\gamma)(1-\sigma)\partial l_i/\partial \sigma$[16], (81) becomes

$$\begin{aligned}\left.\frac{\partial S}{\partial \tau_c}\right|_{\tau_c=0} =& \lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho} \{\frac{\partial e(\cdot)}{\partial \tau_c} - (1-\gamma)(1-\sigma)\frac{\partial e(\cdot)}{\partial \sigma} \\ & - (1-\gamma)(e(\cdot) + \alpha - c_a) - c_a\} + \nu\{(1-\gamma)(\overline{e} + \alpha - c_a) + c_a\},\end{aligned} \tag{82}$$

and, further, by substituting for ν from (46)

$$\left.\frac{\partial S}{\partial \tau_c}\right|_{\tau_c=0} = \lambda \sum_{i=1}^{n} f_i \sum_{r_{-1}=1}^{n} ... \sum_{r_{-t}=1}^{n} \mu_i(\cdot) V_i(\cdot)^{-\rho} \{(1-\gamma)(\overline{e} - e(\cdot)) + D\}, \tag{83}$$

[16] We proceed analogously as in the derivation of (59) in the Proof of Proposition 3: Implicit differentiation of (6) with respect to τ_c and σ gives $\partial l_i/\partial \tau_c = -(\partial l_i/\partial \sigma)(1-\sigma)(\partial \lambda/\partial \tau_c)/\lambda$. By use of $\lambda = \widetilde{\varphi}(\gamma_c, \gamma_b)$ we have $\partial \lambda/\partial \tau_c = (\partial \widetilde{\varphi}/\partial(c_i - c_a))(\partial \gamma_c/\partial \tau_c) + (\partial \widetilde{\varphi}/\partial b_i)(\partial \gamma_b/\partial \tau_c) = -\lambda \gamma_c$ (use (4), (5) with $\widetilde{\varphi}(c_i - c_a, b_i)$ in the objective function (1), $\partial \gamma_c/\partial \tau_c = -\partial \gamma/\partial \tau_c/(1+\tau_c) - (1-\gamma)/(1+\tau_c)^2$ and $\partial \gamma_b/\partial \tau_c = (\partial \gamma/\partial \tau_c)/(1+\tau_b)$).

where D is defined as

$$D \equiv \frac{\partial e(\cdot)}{\partial \tau_c} - (1-\gamma)(1-\sigma)\frac{\partial e(\cdot)}{\partial \sigma} + ((1-\gamma)(\overline{e}+\alpha-c_a)+c_a)\frac{\partial e(\cdot)}{\partial \alpha} \tag{84}$$

From the equation (21) for $e(r_{-1},..,r_{-t})$ one gets at $\tau_b = 0$

$$\frac{\partial e(\cdot)}{\partial \tau_c} = \sum_{s=1}^{t}(s\gamma^{s-1}\frac{\partial \gamma}{\partial \tau_c}(x_{r_{-s}} - (1+\tau_c)c_a) + \gamma^s((1-\sigma)w_{r_{-s}}\frac{\partial l_{r_{-s}}}{\partial \tau_c} - c_a). \tag{85}$$

Inserting (85) at $\tau_c = 0$, (49), (50), together with $\partial l_i/\partial \tau_c|_{\tau_c=0} = (1-\gamma)(1-\sigma)\partial l_i/\partial \sigma$ and $x_{r_{-s}} = \alpha + (1-\sigma)w_{r_{-s}}l_{r_{-s}}$, (84) can be transformed to

$$D = (1-\gamma)\sum_{s=1}^{t}\gamma^s(\overline{e}+x_{r_{-s}}-c_a) + \sum_{s=1}^{t}s\gamma^{s-1}\frac{\partial \gamma}{\partial \tau_c}(x_{r_{-s}}-c_a). \tag{86}$$

As in the Proof of Proposition 3 we express $\partial\gamma/\partial\tau_c$ in terms of the elasticity ε of substitution, defined by (66). Affine-linear Engel curves imply that proportional price changes do not alter the demand ratio between bequests and consumption (above minimum consumption c_a), which in our case (where $\tau_b = \tau_c = 0$) means $\partial\gamma/\partial\tau_b + \partial\gamma/\partial\tau_c = 0$ or $\partial\gamma/\partial\tau_c = -\partial\gamma/\partial\tau_b = (\varepsilon-1)\gamma(1-\gamma)$, by use of (67). We substitute this term, together with (36) and $(\sum_{s=1}^{t}\gamma^s)^2 = \sum_{s=1}^{t}(s-1)\gamma^s + \sum_{s=1}^{t}(t+1-s)\gamma^{t+s}$ (see footnote 15) into (86), which yields

$$D = (1-\gamma)\sum_{s=1}^{t}(s-1)\gamma^s(\overline{x}-x_{r_{-s}}) + (1-\gamma)\gamma^t\sum_{s=1}^{t}(t+1-s)\gamma^s(\overline{x}-c_a) + \varepsilon(1-\gamma)\sum_{s=1}^{t}s\gamma^s(x_{r_{-s}}-c_a),$$

which we transform further by use of (71), (72) and (21) to

$$D = (1-\gamma)\sum_{s=1}^{t-1}\gamma^s(\overline{e}_{-s}-e_{-s}(\cdot)) + (1-\gamma)\gamma^t\sum_{s=1}^{t}(t-s+1)\gamma^s(\overline{x}-c_a) + \varepsilon(1-\gamma)(e(\cdot)+\sum_{s=1}^{t-1}\gamma^s e_{-s}(\cdot)). \tag{87}$$

Substituting (87) into (83) and using (74) gives us formula (27) in Proposition 4. The first term of (27) is equal to (52), multiplied by $(1-\gamma)$, and is, thus, positive as shown in the Proof of Proposition 2. The second line in (27) is the same as the second line of (26) with $(1-\gamma)$ instead of γ, and thus positive, given PD as shown in the Proof of Proposition 3. Finally, line three of (27) is nonnegative for any $\varepsilon \geqslant 0$. This completes the proof that $\partial S/\partial\tau_c|_{\tau_c=0} > 0$ given PD. ■

Appendix D. Simulations on the Path-Dominance condition

To our knowledge, there exists no precise analytical characterization of this condition and its relation to Row-FOSD of the stochastic matrix M. Instead, in order to get a deeper insight into this property, we ran a large number of simulations. We randomly created stochastic matrices M with unimodal rows (their maximal element being in the diagonal), and chose those matrices which fulfilled Row-FOSD. For these matrices we checked whether they fulfilled Path Dominance and, if not, to which degree they deviate from it. Remember that Path Dominance requires that first-order stochastic dominance holds in a number of pairwise comparisons of probability vectors, and we examined how many of these comparisons did not hold. The following table presents the results of these computations.

Table 1

	dimension of matrix								
	2	3	4	5	6	7	8	9	10
90^{th} percentile of negative PD checks for $s = 1$	0	0.056	0.083	0.070	0.067	0.065	0.058	0.051	0.047
Ratio of negative PD checks for $s = 1$	0	0.020	0.013	0.010	0.010	0.007	0.006	0.006	0.005
90^{th} percentile of negative PD checks for $s = 2$	0	0.056	0.021	0.020	0.006	0	0	0	0
90^{th} percentile of negative PD checks for $s = 3$	0	0	0	0	0	0	0	0	0
90^{th} percentile of negative Row-FOSD checks of P	0	0	0.111	0.063	0.080	0.056	0.061	0.047	0.037
Negative Row-FOSD checks of P given PD	0	0	0	0	0	0	0	0	0
Number of randomly chosen matrices		2000	2000	2000	2000	2000	1000	500	100

The first row of this table shows that ninety percent of all randomly chosen matrices, as described above, either fulfill the Path-Dominance check for $s = 1$ or fail in at most 8 percent of the checks required in the definition. The second row only refers to those cases where at least one of these checks fails, that is, where the difference between two respective terms is negative but should be positive. The numbers in this row tell us that in these cases the sum (in absolute value) of all these negative differences relative to the total sum of the differences of all required checks is 2 percent or less.

The next two rows present the same values for Path Dominance as the first row, when the required comparisons are performed for $s = 2$ and $s = 3$ instead of $s = 1$ (remember that in the definition of PD, the comparisons have to be fulfilled up to the order t). It turns out that the share of failed checks goes down quite rapidly, as expected from the fact that for $s \to \infty$ they are automatically fulfilled. Thus, the first four rows confirm that even if Path Dominance is not fulfilled in the strict sense, the extent of the violations is rather small. For matrices of dimension two these properties are all always satisfied, as was shown in Lemma 2.

The fifth row shows that for ninety percent of the randomly chosen matrices M the associated Markov transition matrix P either fulfills Row-FOSD or fails in at most 11 percent of all checks. More importantly, the simulation results presented in row six provide clear evidence that for all chosen matrices M which also fulfill Path Dominance, the associated matrix P fulfills Row-FOSD, as claimed in the Proof of Proposition 3.

These simulation results make us quite confident that the respective terms describing the long-run effect of redistribution via indirect taxes is indeed positive.

References

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

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

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

Braun, Sebastian T. and Jan Stuhler (2017), The Transmission of Inequality Across Multiple Generations: Testing Recent Theories with Evidence from Germany. The Economic Journal doi:10.1111/ecoj.12453.

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

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

Brunner, Johann K. and Susanne Pech (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, Helmuth, Pierre Pestieau and Jean-Charles Rochet (2001), Direct versus Indirect Taxation: The Design of the Tax Structure Revisited, *International Economic Review*, Vol. 42., 781-799.

Cremer, Helmuth, Pierre Pestieau and Jean-Charles 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, Emmanuel and Iván Werning, (2010), Progressive Estate Taxation, *Quarterly Journal of Economics 125,* 635-673.

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

Hellwig, Martin 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, Louis (2001), A Framework for Assessing Estate and Gift Taxation, in: William G. Gale, James R. Hines Jr. and Joel Slemrod (eds.), *Rethinking Estate and Gift Taxation*, Washington D. C.: Brookings Institution Press, 164-204.

Kopczuk, Wojciech (2001), Optimal Estate Taxation in the Steady State, Mimeo, 2001.

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, James E. (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 (2016), *Revenue Statistics 2016*, OECD Publishing.

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

Piketty, Thomas and Emmanuel Saez (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.

Solon, Gary (2015), What do we know so far about multigenerational mobility?, National Bureau of Economic Research Working Paper 21053.

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.