PROGRESSIVE ESTATE TAXATION[*]

Emmanuel Farhi   Iván Werning

We present a model with altruistic parents and heterogeneous productivity. We derive two key properties for optimal estate taxation. First, the estate tax should be progressive, so that parents leaving a higher bequest face a lower net return on bequests. Second, marginal estate taxes should be negative, so that all parents face a marginal subsidy on bequests. Both properties can be implemented with a simple nonlinear tax on bequests, levied separately from the income tax. These results apply to other intergenerational transfers, such as educational investments, and are robust to endogenous fertility choices. Both estate or inheritance taxes can implement the optimal allocation, but we show that the inheritance tax has some advantages. Finally, when we impose an ad hoc constraint requiring marginal estate taxes to be nonnegative, the optimum features a zero tax up to an exception level, and a progressive tax thereafter.

I. Introduction

One of the biggest risks in life is the family one is born into. We partly inherit the luck, good or bad, of our parents through the wealth they accumulate. Behind the veil of ignorance, future generations value insurance against this risk. At the same time, parents are partly motivated by the impact their efforts can have on their children's well-being through bequests. This paper studies optimal estate taxation in an economy that captures the tradeoff between insurance for newborns and incentives for parents.

We begin with a simple economy with two generations. Parents live during the first period. In the second period each is replaced by a single child. Parents are altruistic towards their child, they work, consume and bequeath; children simply consume. Following Mirrlees [1971], parents first observe a random productivity draw and then exert work effort. Both productivity and work effort are private information; only output, the product of the two, is publicly observable.

Our first objective is to study the constrained efficient allocations and derive their implications for marginal tax rates. For this economy, if one takes the expected utility for parents as the social welfare objective, then Atkinson and Stiglitz's [1976] celebrated uniform-taxation result applies. It implies that the parent's intertemporal consumption choice should not be distorted. Thus, when no direct weight is placed on the welfare of children, labor income should be taxed non-linearly, but bequests should remain untaxed.

In terms of the allocation, this tax system induces the consumption of parent and child to vary one-for-one. In this sense, the luck of the parent's productivity is perfectly inherited by the child. There is no mean reversion across generations. In effect, from the perspective of the child's generation, their consumption is manipulated to provide their parent with incentives. They are offered no insurance against the risk of their parent's productivity. The resulting consumption inequality lowers their expected welfare, but this of no direct concern to the planner.

While this describes one efficient arrangement, the picture is incomplete. In this economy, parent and child are distinct individuals, albeit linked by altruism. In positive analyses it is common to subsume both in a single fictitious 'dynastic agent'. However, a complete normative analysis must distinguish the welfare of both parents and children [Phelan 2006; Farhi and Werning 2007]. Figure I depicts our economy's Pareto frontier, plotting the ex-ante expected utility for the child on the horizontal axis, and that of the parent on the vertical axis. The arrangement discussed in the previous paragraph corresponds to the peak, marked as point A, which is interior point due to parental altruism.

This paper explores other efficient allocations, represented by points on the downward

sloping section of the Pareto frontier. To the right of point A, a role for estate taxation emerges with two critical properties.

The first property concerns the shape of marginal taxes: we show that estate taxation should be progressive. That is, more fortunate parents with larger bequests, should face a higher marginal estate tax. Since more fortunate parents get a lower after-tax return on bequests than the less fortunate, this induces bequests to become more similar. Our stark conclusion regarding the progressivity of estate taxation contrasts with the well-known lack of sharp results regarding the shape of the optimal income tax schedule [Mirrlees 1971; Seade 1982; Tuomala 1990; Ebert 1992; Diamond 1998; Saez 2001].[1]

In terms of the allocation, as we move to the right of point A, the consumption inequality for children falls, which increases their expected welfare. The child's consumption still varies with the parent's consumption, but the relationship is now less than one-for-one. Consumption mean reverts across generations. In this sense, luck in only imperfectly inherited. Children are partly insured against the risk of their parent's productivity.

The second property concerns the level of marginal taxes. We find that estate taxation should be negative, imposing a marginal subsidy that declines with the size of bequests. A subsidy encourages bequests, which improves the consumption of newborns. This highlights that, in order to improve the average welfare of newborns, it is efficient to combine a reduction in inequality with an increase in average consumption. In a way, the first generation buys inequality, to improve incentives, from the second generation in exchange for higher average bequests.

The second main objective of the paper is to derive an explicit tax system that implements these efficient allocations. We prove that a simple system, that confronts parents with separate nonlinear schedules for income and estate taxes, works. The optimal estate tax schedule is decreasing and convex, reflecting our results for the sign and progressivity of the marginal tax. Thus, our results are not simply about implicit taxes or wedges, but also about marginal taxes in an explicit tax system. Of course, tax implementations are rarely unique and our model is no exception. For example, one other possible implementation combines labor income taxation with a regressive consumption tax on parents. We discuss this alternative later and argue why in our view, our estate tax implementation seems more natural.

We illustrate the flexibility of our basic model by extending it in a number of directions.

---

1. Mirrlees's [1971] seminal paper established that for bounded distributions of skills the optimal marginal income tax rates are regressive at the top [see also Seade 1982; Tuomala 1990; Ebert 1992]. More recently, Diamond [1998] has shown that the opposite—progressivity at the top—is possible if the skill distribution is unbounded [see also Saez 2001]. In contrast, our results on the progressivity of the estate tax do not depend on any assumptions regarding the distribution of skills.

We start by considering more general welfare criteria. Although our results rely on a Utilitarian welfare function for the children's generation, the welfare criterion for the parents' generation is irrelevant. We also explore a Rawlsian criterion for the children's generation. To implement the optimal allocation in this case, the estate tax can be replaced by a no-debt constraint preventing parents from leaving negative bequests. This type of constraint is common throughout the world. Interestingly, we show that a no-debt constraint induces implicit tax rates that are negative and progressive, and that it corresponds to a limiting case of our earlier estate tax results. In other words, our estate tax results can be viewed as generalizing the principle of non inheritable debt.

We also consider a simple extension with human capital investments. In the model, it is never optimal to distort the choice between bequests and this alternative source of intergenerational transfers. Thus, our estate tax results carry over, implying that human capital should be subsidized, with a higher marginal subsidy on lower investments. This is broadly consistent with actual educational policies.

Finally, we compare estate and inheritance taxes by considering heterogeneous fertility. We show that the optimal estate tax must condition on the number of children, while the optimal inheritances tax does not. In this sense, inheritance taxes are simpler. These results apply even when fertility is endogenous, as in Becker and Barro [1988].

Our results highlight two properties of optimal marginal tax rates on estates: they should be progressive and negative. In order to determine whether there is a separate role for each of these features, we impose an ad hoc constraint that rules out negative marginal taxes on estates. We show that the progressivity result survives. In particular, the optimal marginal tax on estates is zero up to an exemption level, and positive and increasing above this level. Interestingly, the exemption level depends on the degree to which the rate of return to capital responds to the capital stock. In the limiting case where the rate of return to capital is fixed, the exemption level tends to infinity and estate taxes converge to zero.

We close by studying an infinite-horizon version of our model. This framework provides a motivation for weighing the welfare of future generations. Indeed, allocations that maximize the expected utility for the very first generation are disastrous for the average welfare of distant generations. Atkeson and Lucas [1992] proves an immiseration result of this kind, showing that inequality rises steadily over time, with everyone's consumption converging to zero. In contrast, as shown by Farhi and Werning [2007], with a positive weight on future generations, a steady state exists where inequality is bounded and constant.

Tax implementations are necessarily more involved in our infinite horizon setting, but our main results extend. We provide an implementation where taxes on estates are linear, but the rates depend on current and past labor income. When future generations are not valued,

the expected tax rate is zero as in Kocherlakota [2005]. However, when future generations are valued, the expected tax rate is strictly increasing in the parent's consumption and it is negative. This progressivity induces mean reversion across generations, and plays a key role in moderating the evolution of inequality over generations.

Although our approach is normative, it is interesting to compare our prescriptions with actual policies. There are both similarities and differences. On the one hand, progressivity of marginal tax rates is, broadly speaking, a feature of actual estate tax policy in developed economies. For example, in the United States bequests are exempt up to a certain level, then taxed linearly at a positive rate. Our paper provides the first theoretical justification, to the best of our knowledge, for this common feature of policy. On the other hand, the explicit marginal tax on estates is typically positive or zero, not negative. One interpretation is that our normative model stresses a connection between progressive and negative marginal tax rates which may be overlooked in current thinking on estate tax policy. However, the comparison with actual policies is more nuanced. First, a large fraction of bequests may lie below the exemption level and face a zero marginal tax rate. Second, as explained above, restrictions on debt inheritability constitute an implicit marginal subsidy on bequests. Finally, educational policies constitute an explicit subsidy to intergenerational transfers.

It is worth stressing that, while we find that marginal estate taxes should be progressive, we do not attempt to derive the overall progressivity of the tax system, nor the extent of redistribution within the first generation. In particular, we do not characterize the shape of labor income taxes. In principle, the redistributive effect of a more progressive estate tax could be counterbalanced by adjusting the income tax schedule.[2]

Related Literature. Cremer and Pestieau [2001] also study optimal estate taxation in a two-period economy, but their results are quite different from ours. In particular, they find that marginal tax rates may be regressive and positive over some regions. These results are driven by their implicit assumption that parental consumption and work are complements, departing from the Atkinson-Stiglitz benchmark of separability, which is our starting point.[3] Kaplow [1995] and Kaplow [2000] discuss estate and gift taxation in an optimal taxation framework with altruistic donors or parents. These papers make the point that gifts or estates should be subsidized, but assume away unobserved heterogeneity and are therefore silent on the issue of progressivity.

Our work also relates to a number of recent papers that have explored the implications

---

2. Indeed, our proofs use such a readjustment to describe a set of feasible perturbation that leaves work incentives unchanged.

3. In the main body of the paper, Cremer and Pestieau [2001] studies a model without work effort, with an exogenous wealth shock that is privately observed by parents. However, in their appendix they develop a more standard Mirrlees model with the assumption that parental consumption and work are complements.

of including future generations in the welfare criterion. Phelan [2006] considered a planning problem that weighted all generations equally, which is equivalent to not discounting the future at all. Farhi and Werning [2007] considered intermediate cases, where future generations receive a geometrically declining weight. This is equivalent to a social discount factor that is less than one and higher than the private one. Sleet and Yeltekin [2006] have studied how such a higher social discount factor may arise from a utilitarian planner without commitment. However, none of these papers consider implications for estate taxation.

II. Parent and Child: A Two Period Economy

In our two-period economy a continuum of parents live during period $t = 0$. Each parent produces a single descendant, or child, that lives in period $t = 1$. Parents work and consume, while children simply consume. Each parent is altruistic towards her child.

At the beginning of period $t = 0$, parents first learn their productivity θ_0, and then produce n_0 efficiency units of labor. This requires n_0/θ_0 units of work effort. The utility of a parent with productivity θ_0 is given by

$$(1) \qquad v_0(\theta_0) = u(c_0(\theta_0)) - h\left(\frac{n_0(\theta_0)}{\theta_0}\right) + \beta v_1(\theta_0),$$

with $\beta < 1$. The child's utility is simply

$$(2) \qquad v_1(\theta_0) = u(c_1(\theta_0)).$$

The utility function $u(c)$ is increasing, concave, differentiable and satisfies Inada's conditions $u'(0) = \infty$ and $u'(\infty) = 0$; the disutility function $h(n)$ is increasing, convex and differentiable. In addition, we denote by $\bar{n}$, the possibly infinite maximum number of hours worked. Combining equations (1) and (2) gives $v_0(\theta_0) = u(c_0) + \beta u(c_1) - h(n_0/\theta_0)$.

In addition to production there is and endowment e_0 of goods in period 0 and an endowment e_1 of goods in period 1. Moreover, goods can be transferred between periods $t = 0$ and $t = 1$ with a linear savings technology with rate of return $R > 0$. An allocation is *resource feasible* if

$$K_1 + \int_0^\infty c_0(\theta_0)\, dF(\theta_0) \leq e_0 + \int_0^\infty n_0(\theta_0)\, dF(\theta_0),$$

$$\int_0^\infty c_1(\theta_0)\, dF(\theta_0) \leq e_1 + RK_1$$

where K_1 is capital. Combining these two inequalities yields the present-value resource

constraint

$$(3) \qquad \int_0^\infty c_0(\theta_0) \, dF(\theta_0) + \frac{1}{R} \int_0^\infty c_1(\theta_0) \, dF(\theta_0) \leq e_0 + \frac{1}{R} e_1 + \int_0^\infty n_0(\theta_0) \, dF(\theta_0).$$

We assume that productivity is privately observed by the parent. By the revelation principle, we can restrict attention to direct mechanisms, where agents report their productivity and receive an allocation as a function of this report. An allocation is *incentive compatible* if truthfully revelation is optimal:

$$(4) \quad u(c_0(\theta_0)) + \beta u(c_1(\theta_0)) - h\left(\frac{n_0(\theta_0)}{\theta_0}\right) \geq u(c_0(\theta_0')) + \beta u(c_1(\theta_0')) - h\left(\frac{n_0(\theta_0')}{\theta_0}\right) \quad \forall \theta_0, \theta_0'.$$

An allocation is *feasible* if it satisfies the resource constraint (3) and the incentive constraints (4).

Next, we define two utilitarian welfare measures

$$V_0 \equiv \int_0^\infty v_0(\theta_0) \, dF(\theta_0) \qquad \text{and} \qquad V_1 \equiv \int_0^\infty v_1(\theta_0) \, dF(\theta_0).$$

Note that

$$V_0 = \int_0^\infty (u(c_0(\theta_0) - h(y(\theta_0)/\theta_0)) \, dF(\theta_0) + \beta V_1,$$

so that the utilitarian welfare of the second generation, V_1, enters that of the first generation, V_0, through the altruism of parents. In addition to this indirect channel, we will allow the welfare of the second generation, V_1, to enter our planning problem directly.

These welfare measures lead us to consider the following planning problem:

$$\max V_0$$

subject to the resource constraint (3), the incentive-compatibility constraints (4) and

$$(5) \qquad\qquad\qquad V_1 \geq \underline{V}_1.$$

The planning problem is indexed by $\underline{V}_1$. For low enough values of $\underline{V}_1$ the constraint (5) is not binding and the planning problem then maximizes parental welfare V_0 subject to feasibility. Let V_1^* be the corresponding level of welfare obtained by the second generation in the planning problem when constraint (5) is not imposed. This corresponds to the peak on the Pareto frontier illustrated in Figure I. Constraint (5) is not binding for all $\underline{V}_1 \leq V_1^*$. The second generation obtains a finite level of welfare V_1^* because they are valued indirectly,

through the altruism of the first generation. For values of $\underline{V}_1 > V_1^*$, constraint (5) binds and the solution corresponds to the downward sloping section in the figure.

III. The Main Result: Progressive Estate Taxation

In this section we derive two main results for the two-period economy laid out in the previous section. For any allocation with interior consumption, define the implicit estate tax $\tau(\theta_0)$ by

(6)
$$(1 + \tau(\theta_0))u'(c_0(\theta_0)) \equiv \beta R u'(c_1(\theta_0)).$$

This identity defines a distortion so that the intergenerational Euler equation holds. Similarly, we can define the implicit inheritance tax $\hat{\tau}(\theta_0)$ by

(7)
$$u'(c_0(\theta_0)) \equiv \beta R (1 - \hat{\tau}(\theta_0))u'(c_1(\theta_0)).$$

Each of these wedges can be expressed as a function of each other: $\hat{\tau}(\theta_0) = \frac{\tau(\theta_0)}{1+\tau(\theta_0)}$ or $\tau(\theta_0) = \frac{\hat{\tau}(\theta_0)}{1-\hat{\tau}(\theta_0)}$. In the exposition, we choose to focus mostly on the implicit estate tax. We first derive properties for this implicit tax. We then construct an explicit tax system that implements efficient allocations.

III.A. Implicit Tax Rates

To derive an intertemporal-optimality condition, let ν be the multiplier on constraint (5), μ be the multiplier on the resource constraint (3) and form the corresponding Lagrangian:

$$L \equiv \int_0^\infty [v_0(\theta_0) + \nu v_1(\theta_0)] \, dF(\theta_0) - \mu \int_0^\infty [c_0(\theta_0) + c_1(\theta_0)/R - n_0(\theta_0)] \, dF(\theta_0)$$

so that the planning problem is equivalent to maximizing L subject to incentive constraints (4). Suppose an allocation is optimal and has strictly positive consumption. Consider the following perturbation, at a particular point θ_0. Let $c_0^\varepsilon(\theta_0) = c_0(\theta_0) + \varepsilon$ and define $c_1^\varepsilon(\theta_0)$ as the solution to $u(c_0^\varepsilon(\theta_0)) + \beta u(c_1^\varepsilon(\theta_0)) = u(c_0(\theta_0)) + \beta u(c_1(\theta_0))$. This construction ensures that the incentive constraints are unaffected by ε. A first order necessary condition is that the derivative of L with respect to ε be equal to zero. This yields

$$\frac{\beta R}{u'(c_0(\theta_0))} = \frac{1}{u'(c_1(\theta_0))} - R\frac{\nu}{\mu},$$

which shows that $c_0(\theta_0)$ and $c_1(\theta_0)$ are increasing functions of each other. Incentive compatibility implies that utility from consumption, $u(c_0(\theta_0)) + \beta u(c_1(\theta_0))$, is nondecreasing in

productivity θ_0. It follows that both consumption of parent and child $c_0(\theta_0)$ and $c_1(\theta_0)$ are nondecreasing in θ_0.

This equation can be rearranged in the following two useful ways:

$$(8) \qquad \left(1 - R\frac{\nu}{\mu}u'(c_1(\theta_0))\right)u'(c_0(\theta_0)) = \beta R u'(c_1(\theta_0))$$

and

$$(9) \qquad u'(c_0(\theta_0)) = \beta R\left(1 - \frac{1}{\beta}\frac{\nu}{\mu}u'(c_0(\theta_0))\right)u'(c_1(\theta_0)).$$

Our first result regarding taxes, derived from equation (8) with $\nu = 0$, simply echoes the celebrated Atkinson-Stiglitz uniform-commodity taxation result for our economy.

PROPOSITION 1. The optimal allocation with $\underline{V}_1 \leq V_1^*$ has a zero implicit estate tax, $\tau(\theta_0) = 0$ for all θ_0.

Atkinson and Stiglitz [1976] showed that if preferences over a group of consumption goods is separable from work effort, then the tax rates on these goods can be set to zero. In our context, this result applies to the consumption (c_0, c_1) and implies a zero implicit estate tax.

The Euler equation $u'(c_0) = \beta R u'(c_1)$ implies that dynastic consumption is smoothed. As a result, the optimum features perfect inheritability of welfare across generations. For example, if the utility function is CRRA $u(c) = c^{1-\sigma}/(1-\sigma)$ then $c_1(\theta_0) = (\beta R)^{\frac{1}{\sigma}}c_0(\theta_0)$, or equivalently

$$\log c_1(\theta_0) - \log c_0(\theta_0) = \frac{1}{\sigma}\log(\beta R).$$

Thus, the consumption of parent and child vary, across dynasties with different θ_0, one-for-one in logarithmic terms. Making the child's consumption depends on the parent's productivity θ_0 provides the parent with added incentives. The child's welfare is not valued directly in the planning problem. As a result, they are used to provide incentives. From their point of view, no insurance for the risk of their parent's productivity is provided.

In contrast, when $\underline{V}_1 > V_1^*$, so that $\nu > 0$, then equation (8) implies that the ratio of marginal utilities is not equalized across agents and the marginal estate tax must be nonzero. Indeed, since consumption increases with θ_0 estate taxation must be progressive: the implicit marginal estate tax rate $\tau(\theta_0)$ increases with the productivity θ_0 of the parent.

PROPOSITION 2. Suppose $\underline{V}_1 > V_1^*$ and that the optimal allocation has strictly positive consumption. Then the implicit estate tax is strictly negative and increasing in the

parent's productivity θ_0:

$$(10) \qquad \tau(\theta_0) = -R\frac{\nu}{\mu}u'(c_1(\theta_0)).$$

The proposition provides an expression for the implicit estate tax that relates it to the child's consumption. The progressivity of the estate tax is implied by the fact that $c_1(\theta_0)$ is increasing in θ_0. From equation (9) one can also derive the following formula for the implicit marginal inheritance tax $\hat{\tau}(\theta_0)$:

$$(11) \qquad \hat{\tau}(\theta_0) = \frac{\tau(\theta_0)}{1+\tau(\theta_0)} = -\frac{1}{\beta}\frac{\nu}{\mu}u'(c_0(\theta_0)).$$

This alternative expression is sometimes useful.

Returning to the CRRA example, equation (9) now implies

$$(12) \qquad \log c_1(\theta_0) - \log c_0(\theta_0) = \frac{1}{\sigma}\log\left(1+\frac{1}{\beta}\frac{\nu}{\mu}c_0(\theta_0)^{-\sigma}\right) + \frac{1}{\sigma}\log(\beta R).$$

As long as $\frac{\nu}{\mu} > 0$, the right hand side of equation (12) is strictly decreasing in $c_0(\theta_0)$. Thus, the child's consumption still varies with that of the parent's but less than one-for-one in logarithmic terms. In this way, the intergenerational transmission of welfare is imperfect, with consumption mean reverting across generations. Mean reversion serves to reduce inequality in the second generation's consumption. When the expected welfare of the second generation is considered in the planning problem, insurance is provided to reduce inequality.

The progressivity of the implicit estate tax reflects this mean reversion. Fortunate parents, with higher productivities, must face a lower net-of-tax return on bequests so that their dynastic consumption slopes downward. Likewise, poorer parents, with lower productivities, require higher net-of-tax returns on bequests so that their dynastic consumption slopes upward.

Another intuition is based on interpreting our economy with altruism as an economy with an externality. In the presence of externalities, corrective Pigouvian taxes are optimal. One difference is that, typically, externalities are modeled as being a function of the average consumption of a good, such as the pollution produced from gasoline consumption. As a result, the corrective Pigouvian tax is linear. In contrast, in our model the externality enters nonlinearly, resulting in an optimal tax that is also nonlinear. To see this, think of c_1 as a good that the parent enjoys and chooses, but that happens to have a positive externality on the child. Since the externality is positive, a Pigouvian subsidy is called for. However, according to the utilitarian welfare metric, the externality is not a function of aggregate

9

consumption $\int c_1(\theta_0)\, dF(\theta_0)$. Instead, it equals $\int u(c_1(\theta_0))\, dF(\theta_0)$. Since the utility function $u(c_1)$ is concave, the externality is strongest for children with lower consumption. Indeed, the subsidy is directly proportional to $u'(c_1)$. This explains the progressivity of the implicit tax $\tau(\theta_0)$.

Private information is not crucial for our results. In our model, private information creates inequality in the utility parents obtain from consumption goods, $u(c_0) + \beta u(c_1)$. Our results would also obtain if such inequality were simply assumed, or possibly derived for other reasons.

III.B. Explicit Tax Implementations

An allocation is said to be *implemented by a non-linear labor income tax* $T^y(n_0)$ *and estate tax* $T^b(b)$ if, for all θ_0, $(c_0(\theta_0), c_1(\theta_0), n_0(\theta_0))$ solves

$$\max_{c_0, c_1, n_0} \left(u(c_0) + \beta u(c_1) - h\left(\frac{n_0}{\theta_0}\right) \right)$$

subject to

$$c_0 + b \le e_0 + n_0 - T^b(b) - T^y(n_0),$$
$$c_1 \le e_1 + Rb.$$

The first-order condition for this problem, assuming $T^b(b)$ is differentiable, gives

$$(1 + T^{b\prime}(b))u'(c_0) = \beta R u'(c_1).$$

To find a candidate tax schedule, we match this first-order condition with equation (8) and use the budget constraint to substitute out $c_1 = e_1 + Rb$, to obtain

$$(13) \qquad T^{b\prime}(b) = -R\frac{\nu}{\mu}u'(e_1 + Rb).$$

For any arbitrary value $T^b(0)$, this gives $T^b(b) = T^b(0) + \frac{\nu}{\mu}u(e_1) - \frac{\nu}{\mu}u(e_1 + Rb))$.[4] Indeed, this candidate does implement the optimal allocation. The proof, contained in the appendix, exploits the fact that marginal tax rates are progressive. As a result, the parent's problem is convex in the bequest choice, ensuring that the first-order condition, which we used above to define marginal taxes, is sufficient for the parent's optimal bequest choice.

---

4. Note that our implementation features one degree of freedom in the level of income and estate taxes. Marginal tax rates are entirely pinned down by the optimal allocation, which is unique. However, only the sum of the income and estate tax schedules, $T^y(0) + T^b(0)$, is determined. Thus, the model does not uniquely pin down the sign of average estate taxes, nor the revenue generated by the estate tax.

PROPOSITION 3. Suppose $\underline{V}_1 > V_1^*$ and that the optimal allocation has strictly positive consumption. Then the optimal allocation is implementable with a non-linear income tax and an estate tax. The estate tax T^b is strictly decreasing and convex.

Under this implementation, parents face negative marginal tax rates simply because T^b is decreasing. In equilibrium, parents with higher productivity face higher tax rates because they choose to leave larger bequests and T^b is convex. Note that $b + T^b(b)$ is not monotone and has minimum where $T^{b\prime}(\underline{b}) = -1$; parents never leave bequests below $\underline{b}$.

For a given a utility function $u(c)$ and return R, the optimal estate tax schedule T^b belong to a space of functions indexed by a single parameter $\frac{\nu}{\mu}$. It is interesting to note that this space of functions is independent of the distribution of productivity $F(\theta_0)$ and the disutility of work effort $h(\frac{n}{\theta_0})$.[5] Of course these primitives may affect the relevant value of $\frac{\nu}{\mu}$, which pins down the tax function from this space.

A higher value of $\frac{\nu}{\mu}$ gives higher marginal taxes $T^{b\prime}(b)$. Interestingly, it has no impact on the ratio $T^{b\prime\prime}(b)/T^{b\prime}(b)$, a measures of local relative progressivity. For any value $\frac{\nu}{\mu}$, explicit marginal taxes $T^{b\prime}(b)$ have full range since $T^{b\prime}(\underline{b}) = -1$ and $\lim_{b\to\infty} T^{b\prime}(b) = 0$ (assuming Inada conditions for $u(c)$). However, in equilibrium the range of implicit marginal taxes $\tau(\theta_0) = T^{b\prime}(\frac{c_1(\theta_0)-e_1}{R})$ is typically more confined, since parents may stay away from entire sections of the $T^b(b)$ schedule. In particular, as long as the allocation has consumption bounded away from zero for parents, so that $\inf c_0(\theta_0) > 0$ then equilibrium implicit taxes $\tau(\theta_0)$ are bounded away from -1.

Inheritance Taxes. It is also possible to implement the allocation using inheritance taxes paid by the child. Here, the difference between an estate and inheritance tax is minor, but a starker contrast emerges with the extensions considered Section IV.D. An allocation is *implementable by non-linear income and inheritance taxes, $\hat{T}^y(n_0)$ and $\hat{T}^b(Rb)$,* if $(c_0(\theta_0), c_1(\theta_0), n_0(\theta_0))$ maximizes the utility for a parent with productivity θ_0 subject to

$$c_0 + b \le e_0 + n_0 - \hat{T}^y(n_0),$$
$$c_1 \le e_1 + Rb - \hat{T}^b(Rb).$$

Proceeding similarly, the first-order condition is now

$$u'(c_0) = \beta R(1 - T'(Rb))u'(c_1),$$

---

5. In contrast, the tax on labor is very sensitive to the specification of these elements [Saez 2001].

so that matching this expression with equation (8) leads to a differential equation

$$(14) \qquad \frac{\hat{T}^{b\prime}(Rb)}{1 - \hat{T}^{b\prime}(Rb)} = -R\frac{\nu}{\mu}u'(e_1 + Rb - \hat{T}^b(Rb)).$$

with any arbitrary value $T^b(0)$. As it turns out, one can show that with this inheritance tax, the budget set of affordable (c_0, c_1, n_0) is identical to that affordable with the proposed estate tax implementation. Thus, parents choose the same allocation in both implementations and Proposition 3 implies that this allocation is the optimum.

Interestingly, both T^b and $\hat{T}^b$ are defined by matching their derivatives to $\tau(\theta_0)$ in equation (10), but avoiding reference to the allocation as a function of productivity θ_0. Thus, unlike $\tau(\theta_0)$, these tax schedules do not directly depend on the optimal allocation $c_0(\theta_0)$ or $c_1(\theta_0)$, except indirectly through the ratio of the multipliers $\frac{\nu}{\mu}$.

Other Implementations. We have stated our results in terms of the implicit marginal tax rates, as well as a particular tax implementation. As is commonly the case, tax implementations are not unique. Two other implementations are worth briefly mentioning. First, the optimal allocation can also be implemented by a non-linear income tax and a progressive consumption tax, $T^{c_1}(c_1)$, in the second period. Second, the optimal allocation can also be implemented by combining a non-linear income tax with a regressive consumption tax, $T^{c_0}(c_0)$, in the first period. In this two period version of the model all these implementations seem equally plausible. However, with more periods, implementations relying on consumption taxes require marginal consumption tax rates to grow without bound. Although, formally, this is feasible, it seems unappealing for considerations outside the scope of the model, such as tax evasion.[6] In any case, all possible implementations share that the intertemporal choice of consumption will be distorted, so that the implicit marginal tax rate on estates is progressive and given by $\tau(\theta_0)$.

IV. Extensions

In this section, we consider some extensions that can address a number of relevant issues. They also help in a comparison of optimal policies, within the model, to actual real-world policies.

IV.A. General Welfare Functions

Our first extension is the most straightforward. In our basic setup, we adopted Utilitarian social welfare measures for both generations. We now generalize the welfare measures

---

6. Moreover, in a multi-period extension where each agent lives for more than one period a consumption tax on annual consumption would not work, because the progressive intertemporal distortions should only be introduced across generations, not across a lifetime.

considered.

Define two welfare measures W_0 and W_1 for parents and children, respectively, by

$$W_0 = \int_0^\infty \hat{W}_0(v_0(\theta_0), \theta_0) \, dF(\theta_0)$$

$$W_1 = \int_0^\infty \hat{W}_1(v_1(\theta_0)) \, dF(\theta_0)$$

where $v_0(\theta_0) = u(c_0(\theta_0)) + \beta u(c_1(\theta_0)) - h(n_0(\theta_0)/\theta_0)$ and $v_1(\theta_0) = u(c_1(\theta_0))$. Assume that $\hat{W}_1$ is increasing, concave and differentiable, and that $\hat{W}_0(\cdot, \theta_0)$ is increasing and differentiable for all θ_0. The Utilitarian case considered before corresponds to the identity functions, $\hat{W}_0(v) = v$ and $\hat{W}_1(v) = v$.

Since the welfare function $\hat{W}_0$ may depend on θ_0, it allows $\hat{W}_0 = \pi(\theta_0)v_0(\theta_0)$ with arbitrary Pareto weights $\pi(\theta_0)$. Thus, we only require Pareto efficiency in evaluating the welfare of the first generation. In contrast, our results do depend on the welfare criterion for the second generation. Importantly, the generalized Utilitarian criterion W_1 captures a preference for equality.

The planning problem maximizes W_0 subject to (3), (4) and $W_1 \geq \underline{W}_1$, for some $\underline{W}_1$. Using the same perturbation argument developed for the Utilitarian case, we find that if the optimal allocation features strictly positive consumption, the implicit estate tax is given by

$$\tau(\theta_0) = -R\frac{\nu}{\mu}\hat{W}_1'(u(c_1(\theta_0))u'(c_1(\theta_0)).$$

Since $\hat{W}_1$ is increasing and concave, it follows that $\tau(\theta_0)$ is negative and increasing in θ_0. The estate tax is progressive and negative. Interestingly, the marginal tax does not depend directly on the parent's welfare function $\hat{W}_0$, except indirectly through the ratio of multipliers $\frac{\nu}{\mu}$. In contrast, as the formula above reveals, the welfare function for children $\hat{W}_1$ has a direct impact on the shape of estate taxes. In particular, for given $\frac{\nu}{\mu}$, more concave welfare functions imply more progressive tax schedules.[7]

As in Proposition 3 the optimal allocation, as long as it features strictly positive consumption, is implementable with a non-linear income tax T^y and either an estate tax T^b or inheritance tax $\hat{T}^b$. The estate and inheritance tax schedules are decreasing and convex.

IV.B. Non-Inheritable Debt

In most countries, children are not liable for their parent's debt. Interestingly, our main results for estate taxation can be seen as a generalization of this widely accepted constraint

---

7. Of course, the welfare function for parents $\hat{W}_0$ still plays an important role in determining the income tax schedule and, hence, the overall progressivity of the tax system.

that parents cannot borrow against their children. Firstly, a no-debt constraint creates *implicit* marginal taxes that are progressive and negative. This is because parents with lower productivity find the constraint more binding. Secondly, when the welfare criterion for children is Rawlsian, instead of Utilitarian, a no-debt constraint implements the optimal allocation.

When the welfare criterion for children is Rawlsian, the planning problem maximizes W_0 subject to the resource constraint (3), the incentive-compatibility constraints (4) and

$$(15) \qquad\qquad u_1(\theta_0) \geq \underline{u}_1 \qquad \text{for all } \theta_0$$

where $\underline{u}_1$ parameterizes a minimum level of utility for children. Let $\underline{c}_1$ be the corresponding consumption level: $\underline{c}_1 = u^{-1}(\underline{u}_1)$.

For a high enough value of $\underline{u}_1$, the solution to this problem features a threshold $\underline{\theta}_0$ such that the constraint (15) is binding for all $\theta_0 < \underline{\theta}_0$ and slack for $\theta_0 > \underline{\theta}_0$. Moreover, for $\theta_0 \geq \underline{\theta}_0$ the implicit estate tax is zero. For $\theta_0 < \underline{\theta}_0$ we have $c_1(\theta_0) = \underline{c}_1$ so that

$$(16) \qquad\qquad \tau(\theta_0) = \beta R \frac{u'(\underline{c}_1)}{u'(c_0(\theta_0))} - 1.$$

Since $c_0(\theta_0)$ is nondecreasing in θ_0, it follows that $\tau(\theta_0)$ is nondecreasing and nonpositive.

In our implementation, an agent of type θ_0 faces the borrowing constraints

$$c_0 + b \leq e_0 + n_0 - \hat{T}^y(n_0),$$
$$c_1 \leq e_1 + Rb - \hat{T}_1^y$$
$$b \geq 0.$$

Under this implementation children pay a lump-sum tax $\hat{T}_1^y \equiv e_1 - \underline{c}_1$, so that when $b = 0$ they can consume $\underline{c}_1$.

PROPOSITION 4. Suppose that the welfare function for the children's generation is Rawlsian. Then the optimal allocation can be implemented with an income tax for parents, T^y, a lump-sum tax for the child, $\hat{T}_1^y$, and a no-debt constraint, $b \geq 0$.

Parents that find the debt constraint strictly binding, $b = 0$, the intergenerational Euler equation holds with strict inequality $u'(c_0(\theta_0)) > \beta R u'(\underline{c}_1)$. Thus, they face an implicit estate subsidy $\tau(\theta_0) < 0$. These parents would like to borrow against their children, but the implementation precludes it. The lower the productivity θ_0, the lower $c_0(\theta_0)$ and the stronger is this borrowing motive. As a result, the shadow subsidy is strictly increasing in

θ_0 over the range of parents that are at the debt limit.

This implementation highlights a feature of policy that is often overlooked: In most countries, children are not liable for their parent's debts and this alone contributes to progressive and negative implicit estate taxes, as in our model.

The Rawlsian case, and its no-debt constraint solution, can be obtained as a limit case of the previous analysis. To see this, consider a sequence of concave and continuously differentiable welfare functions $\{\hat{W}_{1,k}\}$ that becomes infinitely concave around $\underline{u}_1$ in the sense that

$$\lim_{k \to \infty} \lim_{u_1 \downarrow \underline{u}_1} \hat{W}'_{1,k}(u_1) = 0 \qquad \text{and} \qquad \lim_{k \to \infty} \lim_{u_1 \uparrow \underline{u}_1} \hat{W}'_{1,k}(u_1) = \infty.$$

In the limit, the solution with this sequence of welfare functions converges to that of the Rawlsian case. Similarly, along this sequence, the estate tax schedule $T^{b,k}(b)$ is convex and decreasing, as implied by our results. However, it converges to a schedule with an infinite tax on bequests below some threshold, effectively imposing a no-debt constraint, and a zero marginal tax rate above this same threshold. In this sense, our results extend the logic of a no-debt constraint to smoother welfare functions. An estate tax that is progressive and negative is a smoother version of a no-debt constraint.

We have taken the welfare function for children $\hat{W}_1$ as Rawlsian, but do not make any assumptions on the welfare function for parents $\hat{W}_0(\cdot, \theta_0)$. However, this does not formally consider the case where the welfare criterion for parents is also Rawlsian $W_0 = min_{\theta_0}(v_0(\theta_0))$. This case is more easily handled by considering the dual problem of minimizing the net present value of resources subject to the constraint that the utility $v_0(\theta_0)$ of every parent be above some bound $\underline{v}_0$. A very similar analysis then shows that Proposition 4 also applies when the welfare criterion for parents is Rawlsian. The details are contained in the Appendix.

IV.C. Educational Subsidies

In our model, bequests were the only transfer between one generation and the next. However, in reality, educational investments are an important form of giving by parents. We now explore the applicability of our results to these transfers by incorporating the simplest form of human capital.

Let x denote investment and $H(x)$ denote human capital, where H is a differentiable, increasing and concave function with Inada conditions $H'(0) = \infty$ and $H'(\infty) = 0$. Each unit of human capital produces a unit of the consumption good, so that the resource constraint becomes

$$\int_0^\infty \left(c_0(\theta_0) + \frac{c_1(\theta_0)}{R} \right) dF(\theta_0) \leq e_0 + \frac{e_1}{R} + \int_0^\infty \left(n_0(\theta_0) + \frac{H(x(\theta_0))}{R} - x(\theta_0) \right) dF(\theta_0)$$

15

Preferences are

$$v_0(\theta_0) = u(c_0(\theta_0)) - h\left(\frac{n_0(\theta_0)}{\theta_0}\right) + \beta v_1(\theta_0),$$

$$v_1(\theta_0) = U(c_1(\theta_0), H(x(\theta_0))),$$

where U is differentiable, increasing and concave in both arguments, and satisfies standard Inada conditions. This structure of preferences preserves the weak separability assumption required for the Atkinson-Stiglitz benchmark result. The assumption that H enters the utility function u_1 is a convenient way of ensuring that parents do not all make the exact same choice for H.[8] Indeed, we will assume that H is a normal good, so that richer parents invest more.

Following the same perturbation arguments as in Section III one finds that the formula for the implicit estate tax is unaffected and given by

$$\tau(\theta_0) = -R\frac{\nu}{\mu}U_{c_1}(c_1(\theta_0), H(\theta_0))$$

as long as the optimal allocation features strictly positive consumption.

Turning to the implicit tax on human capital, consider the following perturbation. Fix some θ_0. Increase investment $x^\varepsilon(\theta_0) = x(\theta_0) + \varepsilon$, leave parental consumption unchanged $c_0^\varepsilon(\theta_0) = c_0$, and set $c_1^\varepsilon(\theta_0)$ so that utility is unchanged:

$$U(c_1^\varepsilon(\theta_0), H(x^\varepsilon(\theta_0))) = U(c_1(\theta_0), H(x(\theta_0))).$$

This perturbation leaves utility for both the parent and the child unchanged, but impacts the resource constraint. This leads to the following first-order condition:

(17) $$R = H'(x(\theta_0))\left(1 + \frac{U_H(c_1(\theta_0), H(x(\theta_0)))}{U_{c_1}(c_1(\theta_0), H(x(\theta_0)))}\right).$$

This equation equalizes the rate of return on saving, R, to that on human capital, which features the purely monetary component, $H'(x)$, as well as a term due to the appreciation for human capital in utility. Equation (17) is also the first-order condition of

(18) $$V_1(e) = \max_{c_1, x} U(c_1, H(x)) \quad \text{s.t.} \quad c_1 - e_1 + Rx - H(x) \le e.$$

The quantity $c_1 - H(x)$ is the financial bequest received by the child, and x is human capital

---

8. Similar results would hold if instead H entered the parent's utility function directly.

investment. Equation (17) implies that it is optimal not to distort the choice between these two forms of transfers from parent to child. In what follows, we assume that financial bequest and human capital investment are both normal goods. That is, the optimal $c_1 - e_1 - H(x)$ and x in the maximization (18) are increasing in e.

We consider an implementation with three separate non-linear tax schedules: a nonlinear income tax schedule T^y, a non-linear estate tax T^b and non-linear human capital tax T^x. The parent maximizes

$$u(c_0) - h(n_0/\theta_0) + \beta U(c_1, H(x))$$

subject to

$$c_0 + b + x \leq e_0 + n_0 - T^y(n_0) - T^b(b) - T^x(x),$$
$$c_1 \leq e_1 + Rb + H(x).$$

As shown above, the choice between bequests and human capital should be undistorted. This requires equalizing marginal tax rates estate and human-capital, which suggests looking for candidates tax schedules with

$$T^{x\prime}(x(\theta)) = T^{b\prime}\left(\frac{c_1(\theta) - e_1 - H(x(\theta))}{R}\right) = \tau(\theta).$$

The next proposition establishes that this construction indeed works.[9]

PROPOSITION 5. Assume that financial bequests and human capital investment are normal goods in (18), and that the optimal allocation features strictly positive consumption. There exists three separate non-linear tax schedules T^y, T^b and T^x that implement the optimal allocation. In addition T^b and T^x are decreasing and convex. Moreover the marginal tax rates on bequests and human capital investment are equalized.

Many countries have policies towards education and other forms of human capital acquisition that are broadly consistent with these prescriptions. Governments help finance human capital investments. Typically, basic education is provided for free, while higher levels of education and other forms of training may be only partially subsidized. Furthermore, higher levels of education have an important opportunity cost component, which is not typically subsidized. In sum, these policies subsidize human capital investments, but provide a smaller marginal subsidy to those making larger investments.

---

9. If human capital does not enter utility, equation 17 reduces to $H'(x(\theta_0)) = R$. Thus, the optimum requires all parents to make the exact same human capital investment. Such an allocation cannot be implemented with three separate tax schedules. An implementation that does work in this case is to tax total wealth jointly, so that the child pays taxes as a function of total wealth $Rb + H(x)$.

Alternative arguments for subsidizing education have relied on a "good citizen" externality. As we explained in Section III, our economy with altruism can be interpreted as an economy with an externality. However, the "externality" in this case runs through the average welfare of the second generation, rather than through civic attitudes. Thus, we emphasize completely distinct issues. Interestingly, educational subsidies are often defended by appealing to "equality of opportunity". Our model captures a desire for equality by the central role played by the Utilitarian welfare of the next generation.

The generality of the results in Propostion 5 should not be overstated. Our simple model relies on strong assumptions. For example, it ignores labor supply choices by children, as well as uncertainty, heterogeneity and private information with respect to human capital returns. No doubt, in its current form, Proposition 5 is unlikely to survive all such extensions. Nevertheless, our model provides a simple benchmark that delivers sharp results, illustrating a mechanism that is likely to be at work in richer environments. [10]

IV.D. Estate versus Inheritance Taxes

In this section, we allow for fertility differences across households and show that our results are robust. Moreover, this richer setup allows for interesting comparisons of estate and inheritance taxes.

Exogenous fertility differences. We first assume that fertility is exogenous. Let m denote the number of children in a household. with joint distribution for fertility and productivity $\hat{F}(\theta_0, m)$. We assume that m is observable.

Preferences are as in Becker and Barro [1988]. A parent with productivity and fertility given by (θ_0, m) has utility $u(c_0) - h(n_0/\theta_0) + \sum_{j=1}^{m} \beta_m u(c_{1,j})$ where β_m is an altruism factor that may depend on m. Optimal allocations will be symmetric across children within a family, so that $c_{1,j} = c_1$ for all j.

The welfare measures are

$$V_0 = \int_0^\infty v_0(\theta_0, m) \, d\hat{F}(\theta_0, m) \qquad\qquad V_1 = \int_0^\infty v_1(\theta_0, m) \, m d\hat{F}(\theta_0, m)$$

with $v_0(\theta_0) = u(c_0(\theta_0, m)) + m\beta_m u(c_1(\theta_0, m)) - h(n_0(\theta_0)/\theta_0)$ and $v_1(\theta_0, m) = u(c_1(\theta_0, m))$.

We assume child rearing costs of $\kappa \geq 0$ per child. The present-value resource constraint

---

10. Some recent treatments of optimal taxation in environments with endogenous human capital incorporating some of these features include Kapicka [2006], Kapicka [2008] and Grochulski and Piskorski [2007]. These papers focus on taxation within a lifetime.

becomes

(19)
$$\int_0^\infty \left(c_0(\theta_0, m) + m\left(\kappa + \frac{1}{R}c_1(\theta_0, m) \right) \right) d\hat{F}(\theta_0, m) \leq e_0 + \frac{1}{R}e_1 + \int_0^\infty n_0(\theta_0, m)\, d\hat{F}(\theta_0, m).$$

The incentive-compatibility constraints are

$$(20) \quad u(c_0(\theta_0, m)) + m\beta_m u(c_1(\theta_0, m)) - h\left(\frac{n_0(\theta_0, m)}{\theta_0} \right) \geq$$
$$u(c_0(\theta_0', m)) + m\beta_m u(c_1(\theta_0', m)) - h\left(\frac{n_0(\theta_0', m)}{\theta_0} \right) \quad \forall m, \theta_0, \theta_0'.$$

Note that m is on both sides of the inequality, incorporating the fact that m is observable. The planning problem maximizes V_0 subject to (19), (20) and $V_1 \geq \underline{V}_1$, for some $\underline{V}_1$.

The same variational argument used before can be applied conditioning on (θ_0, m), giving the following expression for the implicit estate tax if the optimal allocation features strictly positive consumption:

$$(21) \qquad\qquad \tau(\theta_0, m) = -R\frac{\nu}{\mu}u'(c_1(\theta_0, m)).$$

In this context, it is not possible to implement the optimal allocation with a non-linear income tax $T^{y,m}$ and an estate tax T^b that is independent on the number of children m. To see this, suppose parents faced such a system. Parents choose an estate of total size b and equally divide it among children to provide them each with consumption $c_1 = e_1 + Rb/m$. But then families with different number of children m and the same estate b would face the same marginal tax rate contradicting equation (21), which says that the marginal tax rate should be a function of child consumption c_1, and, thus, lower (i.e. a greater subsidy) for the larger family.

It is possible to implement the optimal allocation if the estate tax schedule is allowed to depend on family size m, so that parents face a tax schedule $T^{b,m}$. However, since the implicit tax in equation (21) depends on θ_0 and m only trough $c_1(\theta_0, m)$, it is possible to do the same with an inheritance tax that is independent of family size m. In this implementation, a parent with m children faces the budget constraint $c_0 + \sum_{j=1}^m b_j + m\kappa \leq e_0 + n_0 - \hat{T}^{y,m}(n_0)$. Each child is then subject to the budget constraint $c_{1,j} \leq e_1 + Rb_j - T^b(Rb_j)$. The proof of the next proposition is omitted, but proceeds exactly as that of Proposition 3.[11]

---

11. We also explored an extension where parent care more about one child than the other; we omit the details but discuss the main features briefly. In this model, we assumed parents had two children indexed by $j \in \{L, H\}$ and let the altruism coefficient for child j be β^j, with $\beta^H \geq \beta^L$. The preference for one child over another may reflect the effects of birth order, gender, beauty, physical or intellectual resemblance.

PROPOSITION 6. Suppose that the optimal allocation features strictly positive consumption. Then there exists two separate non-linear tax schedules, a income tax $\hat{T}^{y,m}$ that depends on family size and an inheritance tax $\hat{T}^b$ independent of family size, that implement the optimal allocation. In addition $\hat{T}^b$ is decreasing and convex.

Endogenous fertility choice. Consider now endogenous fertility choice as in Becker and Barro [1988]. Normative models with endogenous fertility may raise some conceptual problems, such as taking a stand on the utility of an unborn child. We are able to sidestep this issue, by solving a subproblem over $(c_0(\theta_0),\ c_1(\theta_0),\ n_0(\theta_0))$ for any given $m(\theta_0)$.

Most of the economy's elements are the same as in the exogenous fertility case. Utility, social welfare measures and the resource constraint are exactly the same, except for taking into account that the joint distribution over (θ_0, m) has support on the locus $(\theta_0, m(\theta_0))$. The only difference is that the incentive-compatibility constraint becomes

$$(22)\quad u(c_0(\theta_0)) + m(\theta_0)\beta_{m(\theta_0)}u(c_1(\theta_0)) - h\left(\frac{n_0(\theta_0)}{\theta_0}\right) \geq$$
$$u(c_0(\theta_0')) + m(\theta_0')\beta_{m(\theta_0')}u(c_1(\theta_0')) - h\left(\frac{n_0(\theta_0')}{\theta_0}\right) \quad \forall \theta_0, \theta_0'.$$

to reflect the fact that fertility m is chosen, just like consumption c_0, c_1 and labor n. The planning problem maximizes V_0 subject to (20), (22) and $V_1 \geq \underline{V}_1$, for some $\underline{V}_1$.

For each θ_0, the same perturbation argument over consumption applies, with not change $n(\theta_0)$ nor $m(\theta_0)$. As a result, the optimal allocation features the same implicit taxes on estates.

The implementation also works just as in the exogenous fertility case. As long as the optimal allocation features strictly positive consumption, it can be implemented with an income tax that depends on m and an inheritance tax that is independent on the number of children m. This might seem surprising given that inheritance taxes or subsidies are bound to affect the tradeoff between the quantity and the quality of children. The key insight is that the impact of progressive inheritance taxes on fertility choice can be undone by the income tax $T^{y,m}$ which depends both on income n_0 and the number of children m.[12]

Our results are easily extended to this model. Once again, the model favors inheritance taxes over estate taxes because with an estate tax, the marginal tax on the two children would be equalized, even when their consumption is not. However, just as in equation (21) marginal tax rates should depend on the child's consumption.

12. The argument runs as follows. Think of the parent's problem in two stages. In the second stage, given a choice for (m, n), the parent maximizes over bequests $b_j = b$. In the first stage, the parent chooses over (m, n). Now, the results from the previous section apply directly to the second stage. In particular, the optimal choice of bequests will be independent of θ_0 and depend only on (m, n). Now, as for the first stage, the income tax schedule $T^{y,m}(n)$ can tax prohibitively any combination of (m, n) that is not prescribed by

V. Imposing Positive Marginal Estate Taxes

Our model delivers two prescriptions: marginal taxes should be progressive and negative. To study the role of the former without the latter we now impose an ad hoc restriction that rules out negative marginal tax rates.

This requires adding the inequality constraints

(23) $$u'(c_0(\theta_0)) \leq \beta R u'(c_1(\theta_0)) \quad \forall \theta_0$$

to the planning problem. Although the constraint set is no longer convex, the first-order conditions are still necessary for optimality and can be used to obtain formulas for the optimal implicit tax on estates.

As long as the optimal allocation features strictly positive consumption, equation (11) for the implicit marginal estate tax becomes

$$\frac{\tau(\theta_0)}{1+\tau(\theta_0)} = \max\left\{0, -\frac{1}{\beta}\frac{\nu}{\mu}u'(c_0(\theta_0))\right\}.$$

Of course, this implies that $\tau(\theta_0) = 0$ for all θ_0. Thus, when marginal tax rates are restricted to be nonnegative, the constraint binds and they are optimally set to zero. This illustrates a connection between our two results, the progressivity and negativity of marginal tax rates. The former is not optimal without the latter. However, we now show that such a tight connection depends crucially on the simplifying assumption of a linear savings technology.

Suppose instead that if K_1 goods are invested at $t = 0$ then $G(K_1)$ goods are available at $t = 1$, with G weakly concave and twice differentiable. Constraint (3) must be replaced by

$$\int_0^\infty c_1(\theta_0)\, dF(\theta_0) \leq e_1 + G\left(e_0 + \int_0^\infty n_0(\theta_0)\, dF(\theta_0) - \int_0^\infty c_0(\theta_0)\, dF(\theta_0)\right),$$

where $K_1 = e_0 + \int_0^\infty n_0(\theta_0)\, dF(\theta_0) - \int_0^\infty c_0(\theta_0)\, dF(\theta_0) \geq 0$. Additionally, instead of taking R as a parameter in the nonnegativity constraint (V), we must impose $R = G'(K_1)$.

Letting $\phi(\theta_0)dF(\theta_0)$ denote the multiplier on inequality (23), and assuming that the optimal allocation features strictly positive consumption, the first-order conditions now imply:

(24) $$\frac{\tau(\theta_0)}{1+\tau(\theta_0)} = \max\left\{0, -\frac{1}{\beta}\frac{\nu}{\mu}u'(c_0(\theta_0)) - \beta\frac{G''(K_1)}{\lambda}\int\frac{\phi(\theta_0)}{u'(c_0(\theta_0))}\,dF(\theta_0)\right\}$$

The new term in condition (24) reflects the fact that an increase in $\tau(\theta_0)$, which discourages

---

the optimal allocation. Combining both observations, the parent's faces a problem in the first stage that is essentially equivalent to the incentive constraints.

bequests for a parent with productivity θ_0, now has spillover effect on all parents. When one parent lowers bequests, this contributes to lowering aggregate capital K_1, which in turn raises the pretax return $R = G'(K_1)$ and relaxes the new constraints (23) for all θ_0. This effect is only present if $G''(K_1) < 0$.

The new term is positive and independent of θ_0. The next proposition follows immediately from these observations.

PROPOSITION 7. Consider the planning problem with the additional constraint that $\tau(\theta_0) \geq 0$ for all θ_0 and the more general savings technology $G(K_1)$ and suppose that the optimal allocation features strictly positive consumption. Then, there exist a threshold θ_0^* such that for $\tau(\theta_0) = 0$ for all $\theta_0 \leq \theta_0^*$, and $\tau(\theta_0)$ is strictly positive, increasing in θ_0 for all $\theta_0 > \theta_0^*$.[13]

To gain some intuition for this result it is useful to discuss briefly the extreme case of an economy with no savings technology. Indeed, this corresponds to the limit where $G(K)$ is infinitely concave. In an economy with no savings technology, we can still consider a market where parents can borrow and save with pretax return R. Given taxes, market clearing determines a pretax return R. Parents only care about the after tax return $R(1 - \tau(\theta_0))$. Thus, the overall level of marginal taxes is irrelevant in the following sense. If we change $(1 - \tau(\theta_0))$ proportionally across θ_0, then the new equilibrium pretax level of R changes by the inverse of this proportion, so that the after tax return is unchanged; thus, the allocation is completely unaffected. By this logic, only the difference in marginal taxes across agents affects the allocation. As a result, progressive taxation is still optimal, but the level of taxation is not pinned down.[14] In particular, imposing $\tau(\theta_0) \geq 0$ is not constraining: an equilibrium with high R achieves the same after tax returns $R(1 - \tau(\theta_0))$ with positive marginal tax rates.

When the savings technology is concave the situation is intermediate between the linear technology case and the case without a savings technology. Now, imposing $\tau(\theta_0) \geq 0$ is constraining, but this constraint does not binding for all θ_0. As explained above, a positive marginal tax on a subset of high θ_0 parents increases the return R. This then raises the after-tax return on bequests for low θ_0 parents, even without subsidies.

The implementation is the same as before, so we omit the details. A tax system with a nonlinear income tax and an estate tax (or an inheritance tax) works. The estate (or

---

13. In general, as long as $G'' < 0$, a non-trivial threshold, $F(\theta_0^*) < 1$, is possible. Indeed, no matter how close to zero G'' is, if the optimal allocation has $c_1(\theta_0) \to \infty$ as $\theta_0 \to \infty$ and if $u'(c) \to 0$ when $c \to \infty$, then we necessarily have $F(\theta_0^*) < 1$.

14. In an earlier version of this paper, the model had no savings technology throughout. Thus, the only result we reported was progressivity of estate taxation, not the sign of marginal estate taxes.

inheritance) tax schedule remains convex, but is now weakly increasing due to the new constraint on nonnegative marginal taxes. The schedule is flat below some bequest level, associated with the threshold θ^*, but strictly increasing and strictly convex above that level.

In the basic model of Section II, assuming a linear savings technology had no effect on any of our results. This is consistent with the often made observation that only first derivatives of technology appear in optimal tax formulae. Because of this, a linear technology is commonly adopted as a simplifying assumption in public finance.[15] It is therefore noteworthy, that, in the present context, with the nonnegativity constraint (23), the second derivative of technology $G''(K_1)$ does appear in the tax formula (24). Of course, this non-standard result is driven by the nonstandard constraint (23) that features the first derivative $R = G'(K_1)$.

VI. A Mirrleesian Economy with Infinite Horizon

In this section we extend the model to an infinite horizon. We state the results and sketch the proofs. A detailed analysis is available in an online Appendix.

An infinite horizon planning problem. An individual born into generation t has ex-ante welfare v_t with

$$(25) \qquad v_t = \mathbb{E}_{t-1}[u(c_t) - h(n_t/\theta_t) + \beta v_{t+1}] = \sum_{s=0}^{\infty} \beta^s \, \mathbb{E}_{t-1}\big[u(c_{t+s}) - h(n_{t+s}/\theta_{t+s})\big],$$

where θ_t indexes the agent's productivity type and $\beta < 1$ is the coefficient of altruism.[16] We assume that types θ_t are independently and identically distributed across dynasties and generations $t = 0, 1, \ldots$ With innate talents assumed non-inheritable, intergenerational transmission of welfare is not mechanical linked through the environment but may arise to provide incentives for altruistic parents. Productivity shocks are assumed to be privately observed by individuals and their descendants.

We identify dynasties by their initial utility entitlement v with distribution ψ in the population. An allocation is a sequence of capital stocks $\{K_t\}$ and a sequence of functions $\{c_t^v, n_t^v\}$ for each v that represent consumption and effective units of labor as a function of a history of reports $\hat{\theta}^t \equiv (\hat{\theta}_0, \hat{\theta}_1, \ldots, \hat{\theta}_t)$. For any given initial distribution of entitlements ψ, we say that an allocation $(\{c_t^v, n_t^v\}, \{K_t\})$ is *feasible* if: (i) $\{c_t^v, n_t^v\}$ incentive compatible –i.e.

---

15. For example, Mirrlees [1976] adopts a linear technology and defends this assumption: "This is not a serious restriction. The linear constraint can be thought of as a linear approximation to production possibilities in the neighborhood of the optimum [...]. So long as first-order necessary conditions are at issue, it does not matter [...]".

16. We assume that the utility function satisfies the Inada conditions $u'(0) = \infty$, $u'(\infty) = 0$, $h'(0) = 0$ and $h'(\bar{n}) = \infty$, where $\bar{n}$ is the (possibly infinite) upper bound on work effort.

if truth-telling is optimal – and delivers expected utility v and (ii) it satisfies the resource constraints

$$(26) \qquad\qquad C_t + K_{t+1} \leq F(K_t, N_t) \qquad t = 0, 1, \ldots$$

where $C_t \equiv \int_0^\infty \sum_{\theta^t} c_t^v(\theta^t) \Pr(\theta^t) \, d\psi(v)$ and $N_t \equiv \int \sum_{\theta^t} n_t^v(\theta^t) \Pr(\theta^t) \, d\psi(v)$ are aggregate consumption and labor, respectively.[17]

Given ψ and $\underline{V}$, efficient allocations minimize the required initial capital stock K_0 over feasible allocations $(\{c_t^v, n_t^v\}; \{K_t\})$ that verify a sequence of admissibility constraints requiring that average continuation utility across the population for every future generation be higher than $\underline{V}$. This is a Pareto problem between current and future generations. Let $V^* \equiv (u(0) - \mathbb{E}[h((\bar{n})/\theta)])/(1 - \beta)$ be the welfare associated with misery.

When $\underline{V} = V^*$, the admissibility constraints are slack and future generations are taken into account only through the altruism of the first generation. This is the case studied by Atkeson and Lucas [1995] and Kocherlakota [2005]. When $\underline{V} > V^*$, the admissibility constraints are binding at times.

Let $\beta^t \mu_t$ and $\beta^t \nu_t$ denote the multipliers on the resource constraint and the admissibility constraint at date t. At an interior solution, the first order necessary conditions for consumption and capital can be rearranged to give

$$(27) \qquad \frac{1}{u'(c^v(\theta^t))} = \frac{1}{\beta F_K(K_{t+1}, N_{t+1})} \mathbb{E}_t \left[\frac{1}{u'(c^v(\theta^{t+1}))} \right] - \frac{\nu_{t+1}}{\mu_t}.$$

When $\nu_{t+1} = 0$ this optimality condition is known as the Inverse Euler equation.[18] Consequently, we refer to equation (27) as the Modified Inverse Euler equation. It generalizes equation (8) to incorporate uncertainty regarding the descendants consumption.

Linear inheritance taxes. A simple implementation proceeds along the lines of Kocherlakota [2005] and features linear taxes on inherited wealth. Consider an efficient, interior, allocation $\{c_t^v(\theta^t), n_t^v(\theta^t)\}$. The tax implementation works as follows. In each period, conditional on the history of their dynasty's reports $\hat{\theta}^{t-1}$ and any inherited wealth, individuals report their current shock $\hat{\theta}_t$, produce, consume, pay taxes and bequeath wealth subject to the budget constraints

$$(28) \qquad c_t(\theta^t) + b_t(\theta^t) \leq W_t n_t^v(\hat{\theta}^t) - T_t^v(\hat{\theta}^t) + (1 - \hat{\tau}_t^v(\hat{\theta}^t)) R_{t-1,t} b_{t-1}(\theta^{t-1})$$

---

17. We assume the production function $F(K, N)$ is strictly increasing and continuously differentiable in both of its argument, exhibits constant returns to scale and satisfies the usual Inada conditions.

18. This condition is familiar in dynamic Mirrleesian models [Diamond and Mirrlees 1978; Rogerson 1985; Golosov, Kocherlakota, and Tsyvinski 2003; Albanesi and Sleet 2006].

where $W_t = F_N(K_t, N_t)$ is the wage, $R_{t-1,t} = F_K(K_t, N_t)$ is the interest rate, and initially $b_{-1} = K_0$. Individuals are subject to two forms of taxation: a labor income tax $T_t^v(\hat{\theta}^t)$, and a proportional tax on inherited wealth $R_{t-1,t}b_{t-1}$ at rate $\hat{\tau}_t^v(\hat{\theta}^t)$.[19]

The idea is to devise a tax policy that induces all agents to be truthful and to bequeath $b_t = K_t$. Following Kocherlakota [2005], set the linear tax on inherited wealth to

$$\hat{\tau}_t^v(\theta^t) = 1 - \frac{1}{\beta R_{t-1,t}} \frac{u'(c_{t-1}^v(\theta^{t-1}))}{u'(c_t^v(\theta^t))}.$$

Choose the labor income tax so that the budget constraint holds with equality

$$T_t^v(\theta^t) = W_t n_t^v(\theta^t) + (1 - \tau_t^v(\theta^t))R_{t-1,t}K_t - c_t^v(\theta^t) - K_{t+1}.$$

These choices work because for any reporting strategy , the agent's consumption Euler equation holds. Since the budget constraints hold with equality, this bequest choice is optimal regardless of the reporting strategy. The allocation is incentive compatible by hypothesis, so it follows that truth telling is optimal. Resource feasibility ensures that the markets clear.

The assignment of consumption and labor at in any period depends on the history of reports in a way that can be summarized by the continuation utility $v_t(\theta^{t-1})$. Therefore, the inheritance tax $\hat{\tau}^v(\theta^{t-1}, \theta_t)$ can be expressed as a function of $v_t(\theta^{t-1})$ and θ_t; abusing notation we denote this by $\hat{\tau}_t(v_t, \theta_t)$. Similarly write $c_{t-1}(v_t)$ for $c_{t-1}^v(\theta^{t-1})$. The average inheritance tax rate $\bar{\tau}_t(v_t)$ is then defined by $\bar{\tau}_t(v_t) \equiv \sum_\theta \hat{\tau}(v_t, \theta) \Pr(\theta)$. Using the modified inverse Euler equation (27) we obtain

(29)
$$\bar{\tau}_t(v_t) = -\frac{\nu_t}{\mu_{t-1}} u'(c_{t-1}(v_t)).$$

Formula (29) is the exact analog of equation (11). Note that in the Atkeson-Lucas benchmark where the welfare of future generations is only taken into account through the altruism of the first generation, the average inheritance tax is equal to zero exactly as in Kocherlakota [2005]. Both the negative sign and the progressivity of average inheritance taxes derive directly from the desire to insure future generations against the risk of being born to a poor family.

---

19. In this formulation, taxes are a function of the entire history of reports, and labor income n_t is mandated given this history. However, if the labor income histories $n^t: \Theta^t \to \mathbb{R}^t$ being implemented are invertible, then by the taxation principle we can rewrite T and τ as functions of this history of labor income and avoid having to mandate labor income. Under this arrangement, individuals do not make reports on their shocks, but instead simply choose a budget-feasible allocation of consumption and labor income, taking as given prices and the tax system. See Kocherlakota [2005].

Discussion: Long Run Inequality and Estate Taxation. We now turn to the implications of our results for the dynamics of inequality. For this purpose, it is useful to organize the discussion around the concept of steady states. We specialize to the logarithmic utility case, $u(c) = \log(c)$. This simplifies things because $1/u'(c) = c$, which is the expression that appears in the first-order optimality condition (27).

A steady state consists of a distribution of utility entitlements ψ^* and a welfare level $\underline{V}^*$ such that the solution to the planning problem features, in each period, a cross-sectional distribution of continuation utilities v_t that is also distributed according to ψ^*. We also require the cross-sectional distribution of consumption and work effort and consumption to replicate itself over time. As a result, all aggregates are constant at a steady state. In particular, $K_t = K^*$, $N_t = N^*$ and $R_t = R^*$, etc.

Consider first the case where $\underline{V} = -\infty$. Suppose that there exists an invariant distribution ψ, and let R be the associated interest rate. The admissibility constraints are slack and $\nu_t = 0$ giving the standard Inverse Euler equation

$$(30) \qquad c_t^v(\theta^t) = \frac{1}{\beta R^*}\mathbb{E}_t[c_{t+1}^v(\theta^{t+1})].$$

Integrating over v and θ^t, it follows that $C_{t+1} = \beta R^* C_t$, which is consistent with a steady state only if $\beta R^* = 1$. However, equation (30) then implies that consumption is a positive martingale. By the Martingale Convergence Theorem, consumption must converge almost surely to a finite constant. Indeed, one can argue that $c_t \to 0$ and $v_t \to -\infty$.[20] We conclude that no steady state exists in this case, echoes the immiseration result in Atkeson and Lucas [1992].

Now suppose that $\underline{V} > -\infty$. At a steady state, the admissibility constraints are binding and μ_t/ν_t is equal to a strictly positive constant. To be compatible with some constant average consumption $\bar{c}$, equation (27) requires $R^* < 1/\beta$ and can be rewritten as

$$\mathbb{E}_t[c_{t+1}^v] = \beta R^* c_t^v + (1 - \beta R^*)\bar{c},$$

Consumption is an autoregressive process, mean reverting towards average consumption $\bar{c}$ at rate $\beta R^* < 1$. Just as in the two period case, the intergenerational transmission of welfare is imperfect. Indeed, the impact of the initial entitlement of dynasties dies out over generations and $\lim_{j \to \infty} \mathbb{E}_t c_{t+j} \to \bar{c}$. Indeed, one can show that a steady state may exist with bounded inequality. Moreover, at the steady state there is a strong from of social mobility in that, regardless of their ancestor's welfare position v_t, the probabilistic conditional distribution at

---

20. This follows because consumption c_t is a monotone function of v_{t+1}. However, if v_{t+1} converges to a finite value then the incentive constraints must be slack. This can be shown to contradict optimality.

t for v_{t+j} of distant descendants converges to ψ^* as $j \to \infty$.

VII. Concluding Remarks

Our analysis delivers sharp results for the optimal estate tax. We explored a number of extensions and conjecture that the mechanism we isolate here will remain important in other settings.

We close by briefly mentioning two important issues omitted in the present paper. Firstly, in our model, the lifetime of parents and children do not overlap. If this simplifying assumption is dropped, inter-vivo transfers would have to be considered alongside bequests. Secondly, the focus in this paper was entirely normative. However, in an intergenerational context, questions of political economy and lack of commitment arise naturally. Farhi and Werning [2008] explores such a model and finds that taxation remains progressive but that the marginal tax may be positive.

Harvard University and Toulouse School of Economics
Massachusetts Institute of Technology

Appendix

A. Proof of Proposition 3

Equation (13) implies that T^b is decreasing and convex, and that

$$(31) \qquad T^{b\prime}(R^{-1}(c_1(\theta_0) - e_1)) = \tau(\theta_0).$$

Furthermore, since $c_0(\theta_0)$ and $c_1(\theta_0)$ are nondecreasing in θ_0 it follows from incentive compatibility that there are functions $\hat{c}_0(n)$ and $\hat{c}_1(n)$ such that $\hat{c}_0(n_0(\theta_0)) = c_0(\theta_0)$ and $\hat{c}_1(n_0(\theta_0)) = c_1(\theta_0)$.

Let $N = \{n : \exists \theta_0 \text{ s.t. } n = n_0(\theta_0)\}$ be the equilibrium set of labor choices, in efficiency units. Next, define the income tax function as

$$T^y(n) \equiv n + e_0 - \hat{c}_0(n) + R^{-1}(e_1 - \hat{c}_1(n)) - T^b(R^{-1}(\hat{c}_1(n) - e_1)),$$

if $n \in N$ and $T^y(n) = \infty$ if $n \notin N$.

We now show that the constructed tax functions, $T^y(n)$ and $T^b(b)$, implement the optimal allocation. Clearly, parents cannot choose $n \notin N$. For any given $n \in N$ the parent's subproblem over consumption choices is

$$V(n) \equiv \max_{c_0, c_1} \{u(c_0) + \beta u(c_1)\}$$

subject to $c_0 + R^{-1}(c_1 - e_1) + T^b(R^{-1}(c_1 - e_1)) \leq n - T^y(n)$. Using the fact that T^b is convex, it follows that the constraint set is convex. The objective function is concave. Thus, the first-order condition

$$(1 + T^{b\prime}(R^{-1}(c_1 - e_1)))u'(c_0) = \beta R u'(c_1)$$

is sufficient for an interior optimum. Combining equations (6) and (31) it follows that $\hat{c}_0(n), \hat{c}_1(n)$ are optimal. Hence $V(n) = u(\hat{c}_0(n)) + \beta u(\hat{c}_1(n))$.

Next, consider the parent's maximization over n given by

$$\max_{n \in N}\{V(n) - h(n/\theta_0)\}.$$

We need to show that $n_0(\theta_0)$ solves this problem, which implies that the allocation is implemented since consumption would be given by $\hat{c}_0(n_0(\theta_0)) = c_0(\theta_0)$ and $\hat{c}_1(n_0(\theta_0)) = c_1(\theta_0)$. Now, from the previous paragraph and our definitions it follows that

$$n_0(\theta_0) \in \arg\max_{n \in N}\{V(n) - h(n/\theta_0)\}$$

$$\Leftrightarrow n_0(\theta_0) \in \arg\max_{n \in N}\{u(\hat{c}_0(n)) + \beta u(\hat{c}_1(n)) - h(n/\theta_0)\}$$

$$\Leftrightarrow \theta_0 \in \arg\max_{\theta}\{u(c_0(\theta)) + \beta u(c_1(\theta)) - h(n_0(\theta)/\theta_0)\}$$

Thus, the first line follows from the last, which is guaranteed by the assumed incentive compatibility of the allocation, conditions (4). Hence, $n_0(\theta_0)$ is optimal and it follows that $(c_0(\theta_0), c_1(\theta_0), n_0(\theta_0))$ is implemented by the constructed tax functions.

B. Proof of Proposition 4

We can implement this allocation with an income with an income tax $T^y(n)$, an income tax $T_1^y = e_1 - \underline{c}_1$ and a no-debt constraint mandating that $b \geq 0$.

Exactly as in the proof of Proposition 3, we can define the functions $\hat{c}_0(n)$ and $\hat{c}_1(n)$ such that $\hat{c}_0(n_0(\theta_0)) = c_0(\theta_0)$ and $\hat{c}_1(n_0(\theta_0)) = c_1(\theta_0)$. Let $N = \{n : \exists \theta_0 \text{ s.t. } n = n_0(\theta_0)\}$ be the equilibrium set of labor choices, in efficiency units. Next, define the income tax function as

$$T^y(n) \equiv n + e_0 - \hat{c}_0(n) + R^{-1}(\underline{c}_1 - c_1(n)),$$

if $n \in N$ and $T^y(n) = \infty$ if $n \notin N$.

Clearly, parents cannot choose $n \notin N$. For any given $n \in N$ the parent's subproblem over consumption choices is

$$V(n) = \max_{c_0, c_1}\{u(c_0) + \beta u(c_1)\}$$

28

subject to

$$c_0 + (c_1 - \underline{c_1})/R \leq n - T^y(n)$$

$$c_1 \geq \underline{c_1}.$$

This is a concave problem with solution $(\hat{c}_0(n), \hat{c}_1(n))$, so that $V(n) = u(\hat{c}_0(n)) + \beta u(\hat{c}_1(n))$.

The parent's maximization problem over n is

$$W(\theta_0) = \max_{n \in N} \{V(n) - h(n/\theta_0)\}.$$

We need to prove that $n_0(\theta_0)$ solves this problem, i.e. that

$$n_0(\theta_0) \in \arg\max_{n \in N}\{V(n) - h(n/\theta_0)\}$$

$$\Leftrightarrow n_0(\theta_0) \in \arg\max_{n \in N}\{u(\hat{c}_0(n)) + \beta u(\hat{c}_1(n)) - h(n/\theta_0)\}$$

$$\Leftrightarrow \theta_0 \in \arg\max_{\theta}\{u(c_0(\theta)) + \beta u(c_1(\theta)) - h(n_0(\theta)/\theta_0)\}$$

Thus, the first line follows from the last, which is guaranteed by the assumed incentive compatibility of the allocation, conditions (4). Hence, $n_0(\theta_0)$ is optimal and it follows that $(c_0(\theta_0), c_1(\theta_0), n_0(\theta_0))$ is implemented by the constructed tax functions and the no-debt constraint.

C. Proof or Proposition 5

We can separate the planning problem into two steps: first, solve the optimal allocation in terms of the reduced allocation $\{c_0(\theta_0), e(\theta_0), n_0(\theta_0)\}$; second, solve $c_1(\theta_0)$ and $x(\theta_0)$ using the program (18). The reduced allocation $\{c_0(\theta_0), e(\theta_0), n_0(\theta_0)\}$ is the solution of the following planning program:

$$\max \int_0^\infty \left[u(c_0(\theta_0)) - h(n_0(\theta_0)/\theta_0) + \beta V_1(e(\theta_0))\right] dF(\theta_0)$$

subject to the resource constraint

$$\int_0^\infty \left[c_0(\theta_0) + \frac{e(\theta_0)}{R}\right] dF(\theta_0) \leq e_0 + \int_0^\infty n_0(\theta_0) dF(\theta_0),$$

the incentive compatibility constraints

$$u(c_0(\theta_0)) + \beta V_1(e(\theta_0)) - h\left(\frac{n_0(\theta_0)}{\theta_0}\right) \geq u(c_0(\theta_0')) + \beta V_1(e(\theta_0')) - h\left(\frac{n_0(\theta_0')}{\theta_0}\right) \quad \forall \theta_0, \theta_0'$$

29

and the promise keeping constraint

$$\int_0^\infty V_1(e(\theta_0))dF(\theta_0) \geq \underline{V_1}.$$

This problem is the exact analog of our original planning problem with $c_1(\theta_0)$ replaced by $e(\theta_0)$ and child utility $u(c_1(\theta_0))$ replaced by $V_1(e(\theta_0))$. Therefore we know that $c_0(\theta_0)$ and $e(\theta_0)$ are increasing in θ_0. We also know that $c_1(\theta_0) - e_1 - H(x(\theta_0))$ and $x(\theta_0)$ are increasing in θ_0.

We use the generalized inverse of $x(\theta)$, namely $x^{-1}(x) = \inf\{\theta_0, x(\theta_0) \leq x\}$, to define

$$T^{x\prime}(x) = \tau((x)^{-1}(x))$$

and set any value $T^x(x^*)$ for the intercept at some $x^* > 0$.

We use the generalized inverse of $(c_1 - e_1 - H(x))(\theta)$, namely $(c_1 - e_1 - H(x))^{-1}(z) = \inf\{\theta_0, (c_1 - e_1 - H(x))(\theta_0) \leq z\}$ to define

$$T^{b\prime}(b) = \tau\left((c_1 - e_1 - H(x))^{-1}(Rb)\right)$$

and set any value $T^b(0)$ for the intercept at $b = 0$.

Note that by the monotonicity of $\tau(\theta)$, $x(\theta)$ and $(c_1 - e_1 - H(x))(\theta)$, the functions T^b and T^x are convex.

Recall that $c_0(\theta_0)$, $c_1(\theta_0)$, $x(\theta_0)$ and $n(\theta_0)$ are increasing functions of θ_0. Moreover, these functions are constant on the same intervals, if such an interval exists. As a result, exactly as in the proof of Proposition 3, we can define the functions $\hat{c}_0(n)$, $\hat{c}_1(n)$ and $\hat{x}(n)$ such that $\hat{c}_0(n_0(\theta_0)) = c_0(\theta_0)$, $\hat{c}_1(n_0(\theta_0)) = c_1(\theta_0)$ and $\hat{x}(n_0((\theta_0)) = x(\theta_0)$. Let $N = \{n : \exists\theta_0 \text{ s.t. } n = n_0(\theta_0)\}$ be the equilibrium set of labor choices, in efficiency units. Next, define the income tax function as

$$T^y(n) \equiv n + e_0 - \hat{x}(n) - \hat{c}_0(n) + R^{-1}(e_1 + H(\hat{x}(n)) - c_1(n)) - $$
$$T^b((\hat{c}_1(n) - e_1 - H(\hat{x}(n)))/R) - T^x(\hat{x}(n)),$$

if $n \in N$ and $T^y(n) = \infty$ if $n \notin N$.

We now show that the constructed tax functions implement the allocation. Clearly, parents cannot choose $n \notin N$. For any given $n \in N$ the parent's subproblem over consumption choices is

$$V(n) \equiv \max\{u(c_0) + \beta U(c_1, H(x))\}$$

30

subject to $c_0 + R^{-1}(c_1 - e_1 - H(x)) + x + T^b((c_1 - H(x))/R) + T^x(x) \leq n - T^y(n)$. This problem is convex, the objective is concave and the constraint set is convex, since T^b and T^x are convex. It follows that the first-order conditions

$$1 = \frac{\beta R}{1 + T^{b\prime}(c_1 - e_1 - H(x))} \frac{U_{c_1}(c_1, H(x))}{u'(c_0)}$$

$$1 = \frac{\beta}{1 + T^{x\prime}(x)} \frac{U_H(c_1, H(x))}{u'(c_0)}$$

are sufficient for an interior optimum. It follows from the construction of the tax functions T^b and T^x that these conditions for optimality are satisfied by $\hat{c}_0(n), \hat{c}_1(n), \hat{x}(n)$. Hence $V(n) = u(\hat{c}_0(n)) + \beta U(\hat{c}_1(n), H(\hat{x}(n)))$.

Next, consider the parent's maximization over n given by

$$\max_n \{V(n) - h(n/\theta_0)\}.$$

We need to show that $n_0(\theta_0)$ solves this problem, i.e. that

$$n_0(\theta_0) \in \arg\max_{n \in N} \{V(n) - h(n/\theta_0)\}$$

$$\Leftrightarrow n_0(\theta_0) \in \arg\max_{n \in N} \{u(\hat{c}_0(n) + \beta U(\hat{c}_1(n), H(\hat{x}(n))) - h(n/\theta_0)\}$$

$$\Leftrightarrow \theta_0 \in \arg\max_{\theta} \{u(c_0(\theta)) + \beta U(c_1(\theta), H(x(\theta_0))) - h(n_0(\theta)/\theta_0)\}$$

Thus, the first line follows from the last, which is guaranteed by the assumed incentive compatibility of the allocation. Hence, $n_0(\theta_0)$ is optimal and so it follows that the optimal allocation $(c_0(\theta_0), c_1(\theta_0), n_0(\theta_0), x(\theta_0))$ is implemented by the constructed tax functions.

D. Proof of Result with Rawlsian Welfare Function for Parents

The corresponding planning problem is to maximize $min_{\theta_0}(v_0(\theta_0))$ subject to the resource constraint (3), the Rawlsian constraint for children (15) and the incentive-compatibility constraints (4). The problem is that there is no representation of the objective function $W_0 = min_{\theta_0}(v_0(\theta_0))$ of the form $W_0 = \int_0^\infty \hat{W}_0(v_0(\theta_0), \theta_0) \, dF(\theta_0)$. This difficulty is easily overcome by noting that the planning problem is concave. There is a one to one correspondence between the solutions of this problem and those of the dual problem of minimizing

$$\int_0^\infty c_0(\theta_0) \, dF(\theta_0) + \frac{1}{R} \int_0^\infty c_1(\theta_0) \, dF(\theta_0)$$

subject to the Rawlsian constraint for parents

$$(32) \qquad\qquad v_0(\theta_0) \geq \underline{v}_0 \qquad \text{for all } \theta_0,$$

the Rawlsian constraint for children (15) and the incentive-compatibility constraints (4). The dual problem is more tractable because the objective function is differentiable. Exactly as above, it can be shown that there exists $\underline{\theta}_0 > 0$ such that equation (15) is binding for all $\theta_0 \leq \underline{\theta}_0$. Then for all $\theta_0 \geq \underline{\theta}_0$, the implicit estate tax is zero. When $\theta_0 < \underline{\theta}_0$, the implicit estate tax is given by (16). The rest of the analysis follows. The dual problem also allows us to tackle the case where $\underline{v}_0 = \underline{u}_1$, which would lead to a welfare function that is Rawlsian both across and within generations. There again, Proposition 4 applies.

E. Proof of Equation (24)

Let $\phi(\theta_0)dF(\theta_0)$ denote the multiplier on inequality (23). The first-order conditions can then be rearranged to obtain the implicit marginal estate tax rate:

$$\frac{\tau(\theta_0)}{1+\tau(\theta_0)} = -\frac{1}{\beta}\frac{\nu}{\mu}u'(c_0(\theta_0)) + \frac{\phi(\theta_0)}{\lambda}\frac{\beta R c''(u(c_0(\theta_0))) + \beta^{-1}c''(u(c_1(\theta_0)))}{c'(u(c_0(\theta_0)))}$$
$$-\frac{G''(K_1)}{\lambda}\int \beta c'(u(c_0(\theta_0)))\phi(\theta_0)\,dF(\theta_0),$$

where the function $c(u)$ is the inverse of the utility function $u(c)$, implying that $c'(u) > 0$ and $c''(u) > 0$, $R \equiv G'(K_1)$ and $K_1 = e_0 + \int_0^\infty n_0(\theta_0)\,dF(\theta_0) - \int_0^\infty c_0(\theta_0)\,dF(\theta_0)$. Together with $\tau(\theta_0) \geq 0$, $\phi(\theta_0) \geq 0$ and the complementary slackness condition $\nu(\theta_0)\tau(\theta_0) = 0$, this implies the formula in text.

References

Albanesi, Stefania and Christopher Sleet "Dynamic Optimal Taxation with Private Information." *Review of Economic Studies* LXXIII(1) (2006), 1–30.

Atkeson, Andrew and Robert E. Lucas, Jr. "On Efficient Distribution with Private Information." *Review of Economic Studies* LIX(3) (1992), 427–453.

———— "Efficiency and Equality in a Simple Model of Unemployment Insurance." *Journal of Economic Theory* LXVI (1995), 64–88.

Atkinson, A.B. and J.E. Stiglitz "The Design of Tax Structure: Direct vs. Indirect Taxation." *Journal of Public Economics* VI (1976), 55–75.

Becker, Gary S. and Robert J. Barro "A Reformulation of the Economic Theory of Fertility." *The Quarterly Journal of Economics* CIII(1) (1988), 1–25.

Cremer, H. and P. Pestieau "Non-linear taxation of bequests, equal sharing rules and the tradeoff between intra- and inter-family inequalities." *Journal of Public Economics* LXXIX (2001), 35–53.

Diamond, P. A. and J. A. Mirrlees "A model of social insurance with variable retirement." *Journal of Public Economics* X(3) (1978), 295–336.

Diamond, Peter A. "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates." *American Economic Review* LXXXVIII(1) (1998), 83–95.

Ebert, Udo "A reexamination of the optimal nonlinear income tax." *Journal of Public Economics* XLIX(1) (1992) 47–73.

Farhi, Emmanuel and Iván Werning "Inequality and Social Discounting." *Journal of Political Economy* CXV(3) (2007), 365–402.

——— "The Political Economy of Non-Linear Capital Taxation." Mimeo (2008).

Golosov, Mikhail, Narayana Kocherlakota, and Aleh Tsyvinski "Optimal Indirect and Capital Taxation." *Review of Economic Studies* LXX(3) (2003), 569–587.

Grochulski, Borys and Tomasz Piskorski "Risky Human Capital and Deferred Capital Income Taxation." Mimeo (2007).

Kapicka, Marek "Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping." *Review of Economic Dynamics* IX(4) (2006), 612–639.

——— "The Dynamics of Optimal Taxation when Human Capital is Endogenous." Mimeo (2008).

Kaplow, Louis "A Note on Subsidizing Gifts." *Journal of Public Economics* LVIII (1995), 469–477.

——— "A Framewok for Assessing Estate and Gift Taxation." NBER Working Paper 7775 (2000).

Kocherlakota, Narayana "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation." *Econometrica* LXXIII (2005), 1587–1621.

Mirrlees, J. A. "Optimal tax theory : A synthesis." *Journal of Public Economics* VI(4) (1976), 327–358.

Mirrlees, James "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* XXXVIII(2) (1971), 175–208.

Phelan, Christopher "Opportunity and Social Mobility." *Review of Economic Studies* LXXIII(2) (2006), 487–505.

Rogerson, William P. "Repeated Moral Hazard." *Econometrica* LIII(1) (1985), 69–76.

Saez, Emmanuel "Using Elasticities to Derive Optimal Income Tax Rates." *Review of Economic Studies* LXVIII(1) (2001), 205–29.

Seade, Jesus "On the Sign of the Optimum Marginal Income Tax." *Review of Economic Studies* XLIX(4) (1982), 637–43.

Sleet, Christopher and Sevin Yeltekin "Credibility and Endogenous Social Discounting." *Review of Economic Dynamics* IX(4) (2006) 410–437.

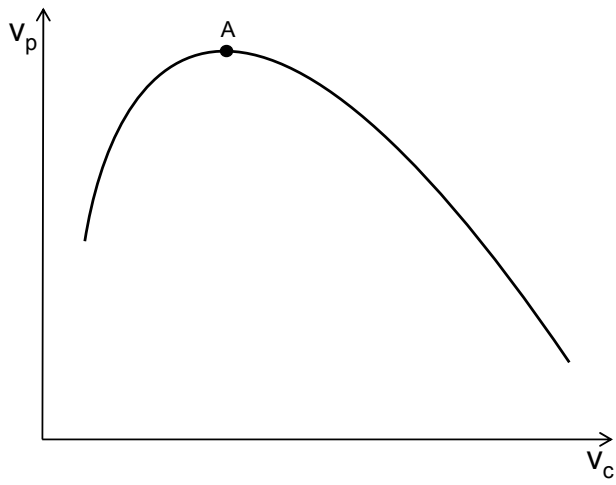Tuomala, Matti *Optimal Income Taxation and Redistribution.* Oxford University Press, Clarendon Press (1990).

FIGURE I: PARETO FRONTIER BETWEEN EX-ANTE UTILITY FOR PARENT, v_p, AND CHILD, v_c.