Estate Taxation with Altruism Heterogeneity*

Emmanuel Farhi
Harvard University

Iván Werning
MIT

We develop a theory of optimal estate taxation in a model where bequest inequality is driven by differences in parental altruism. We show that a wide range of results are possible, from positive taxes to subsidies, depending on redistributive objectives implicit in the cardinal specification of utility and social welfare functions. We propose a normalization that is helpful in classifying these different possibilities. We isolate cases where the optimal policy bans negative bequests and taxes positive bequests, features present in most advanced countries.

1 Introduction

Many people's ideas about estate taxes take the perspective of children, and build on the intuition that inheritances are pure luck—after all, children do nothing to deserve their parents—to conclude that bequests should be redistributed away to help level the playing field.

However, taking the perspective of parents, one can make a powerful argument against estate taxation on the grounds of fairness. This case is eloquently articulated in the form of a parable by Mankiw (2006):

> Consider the story of twin brothers – Spendthrift Sam and Frugal Frank. Each starts a dot-com after college and sells the business a few years later, accumulating a $10 million nest egg. Sam then lives the high life, enjoying expensive vacations and throwing lavish parties. Frank, meanwhile, lives more modestly. He keeps his fortune invested in the economy, where it finances capital accumulation, new technologies, and economic growth. He wants to leave most of his money to his children, grandchildren, nephews, and nieces.

*We thank Ali Shourideh and Stefanie Stantcheva for useful comments and suggestions. Lucas Manuelli provided valuable research assistance.

> Now ask yourself: Which millionaire should pay higher taxes?... What principle of social justice says that Frank should be penalized for his frugality? None that I know of.

In this paper, we offer a theory of estate taxation that reconciles these two philosophies. We analyze a model where parents with different degrees of altruism consume and leave bequests to their offspring. Altruism is private information, giving rise to a tradeoff between equality of opportunity for newborns and incentives for altruistic parents. We consider a wide class of social welfare functions and characterize both optimal nonlinear and linear estate tax systems.

In Farhi and Werning (2010) we formulated a similar optimal tax problem by taking a canonical Mirrleesian tax model—where skill differences are the only source of heterogeneity—and adding a bequest decision. In the model of that paper, more productive parents earn more, consume more and bequeath more.

Instead, in this paper we depart from the canonical optimal tax model, abstracting from parental earnings inequality to focus instead on differences in the degree of altruism.[1] Our main goal is to isolate what this different source for bequest inequality implies for estate taxation.

We find that optimal estate taxes depend crucially on redistributive objectives. Different welfare criteria lead to results ranging from taxes to subsidies. We identify a few useful benchmarks. First, optimal estate taxes are zero when no direct weight is placed on children and when parents welfare is summarized by a Utilitarian criterion using a normalization of utility (Proposition 1). This formalizes Mankiw's intuition. Second, when the Utilitarian criterion is augmented with a positive weight on children's welfare, subsidies on estates emerge (Proposition 2). Finally, a clear cut case for positive taxes on estates is possible when one adopts a more extreme preference for equality of opportunity of children. With a Rawlsian maximin criterion optimal policy taxes positive bequests and bans negative ones (Proposition 4). These two properties are consistent with most actual tax codes, providing one possible justification for their use. We provide both results for nonlinear taxes (Propositions 1–4) and results for linear taxes (Propositions 5–9).

[1] Piketty and Saez (2012) present a model with both altruism and productivity differences and study optimal linear taxes on savings/bequests. Our results suggest that altruism heterogeneity coupled with their choice of welfare functions may be key to understanding the simulations with positive and negative marginal tax rates that they report.

2 The Model

There are two generations, parents born at $t = 0$ and children born at $t = 1$; each living for one period. Parents are altruistic and each has exactly one offspring. There is a storage technology between periods with constant return R. Parents are heterogenous. A parent of type θ has strictly quasi-concave preferences represented by the utility function

$$U^p(c_0, c_1; \theta),$$

where U^p is increasing, strictly concave and twice differentiable in $(c_0, c_1; \theta)$.[2] The type θ is distributed in the population according to a continuous density $f(\theta)$ on the interval $[\underline{\theta}, \overline{\theta}]$. We make the following standard single-crossing condition assumption.

Assumption 1. *The parent's utility function U^p satisfies*

$$\frac{\partial}{\partial \theta}\left(\frac{U^p_{c_1}(c_0, c_1; \theta)}{U^p_{c_0}(c_0, c_1; \theta)}\right) > 0.$$

Higher types are more altruistic; lower types more selfish. Single crossing is an assumption about ordinal preferences, not cardinal utility. It will be useful to make a normalization regarding cardinal utility. Define the indirect utility function

$$V^p(I, R; \theta) \equiv \max_{c_0, c_1} U^p(c_0, c_1; \theta) \qquad \text{s.t.} \quad c_0 + \frac{1}{R}c_1 = I.$$

Assumption 2. *The parent's utility function U^p is such that marginal utility is constant without redistribution*

$$V^p_I(I, R; \theta) = V^p_I(I, R; \theta') \qquad \textit{for all } \theta, \theta', \textit{ and } I.$$

Assumption 2 amounts to a renormalization of cardinal utility, that does not change ordinal preferences (see the appendix for details). Nevertheless, it will prove useful to categorize different cases and results.

We maintain Assumptions 1 and 2 throughout the paper. For a few results we need the following additional assumption.

Assumption 3. *The parent's utility function U^p satisfies*

$$U^p_{c_0,\theta}(c_0, c_1; \theta) \leq 0 \qquad \textit{and} \qquad U^p_{c_1,\theta}(c_0, c_1; \theta) \geq 0.$$

[2]With a few additional assumptions, any strictly quasi concave function can be monotonically transformed into a strictly concave utility function, see e.g. Connell and Rasmusen (2012).

Assumption 3 implies the single crossing condition in Assumption 1. A simple example satisfying all three assumptions is $U^p(c_0, c_1; \theta) = (1-\theta)\log(c_0) + \theta \log(c_1)$.

We will employ a weighted Utilitarian criterion

$$\int \left(\lambda_\theta U^p(c_0(\theta), c_1(\theta); \theta) + \alpha_\theta U^c(c_1(\theta))\right) f(\theta) d\theta,$$

where λ_θ is the weight on a parent of type θ, α_θ is the weight on a child with parent of type θ and U^c is increasing, concave and differentiable. There are two interpretations of these weights. First, by varying the weights across types and generations one traces out the Pareto frontier. Under this interpretation we adopt the ordinal preferences of parents and children and simply place flexible weights on different members of society; cardinal utility is irrelevant. A second interpretation, especially for λ_θ, is possible if we imagine evaluating expected utility behind the veil of uncertainty, before θ is realized. In this case, we interpret cardinal utility for parents to be $\lambda_\theta U^p(c_0, c_1; \theta)$. Observed consumption-savings behavior only identifies ordinal, not cardinal, utility.[3] Thus, flexible weights $\lambda_\theta, \alpha_\theta$ are required to consider a wide range of different tastes for redistribution or specifications of cardinal utility.

With α_θ constant, the curvature of U^c captures a preference for equality of children's consumption. We also want to consider a welfare function with extreme egalitarian preferences for children. To this end, we combine a weighted utilitarian criterion for parents' welfare, $\int \lambda_\theta U^p(c_0(\theta), c_1(\theta); \theta) f(\theta) d\theta$, with a Rawlsian maximin criterion for children's welfare,

$$\min_\theta U^c(c_1(\theta)).$$

This delivers the same implications as the weighted-Utilitarian criterion for some appropriate endogenous weights α_θ.

We assume each parent's θ type is private information. This makes the first best unavailable and creates a tradeoff between redistribution and incentives. We follow both a Mirrleesian approach, with no exogenous restrictions on policy instruments beyond those implied by private information, and a Ramsey approach with restricted taxes.

3 Nonlinear Taxation

We begin with the Mirrleesian approach, without arbitrary restrictions on tax instruments, by studying the mechanism design problem that incorporates the incentive con-

[3]See Lockwood and Weinzierl (2012) for an application of this principle to the taxation of labor.

straints. Similar to Mirrlees (1971), the optimum can be implemented with a nonlinear tax of bequests. Parents are subject to the budget constraints

$$c_0 + B + T(B) = I_0 \tag{1}$$

$$c_1 = I_1 + RB \tag{2}$$

where T is a nonlinear tax on bequests. At points where T is differentiable, the marginal tax rate on bequests equals the implicit marginal tax rate on estates $T'\left(\frac{c_1(\theta)-I_1}{R}\right) = \tau(\theta)$, defined by $(1+\tau(\theta))U^p_{c_0}(c_0(\theta), c_1(\theta); \theta) \equiv RU^p_{c_1}(c_0(\theta), c_1(\theta); \theta)$. Next we characterize the optimal allocation and the associated implicit marginal tax rate.

3.1 A Weighted Utilitarian Objective

The dual planning problem we study is

$$\min_{c_0, c_1, v} \int \left(c_0(\theta) + \frac{1}{R} c_1(\theta)\right) f(\theta) d\theta, \tag{3}$$

subject to $c_1(\theta)$ monotone increasing and

$$v(\theta) = U^p(c_0(\theta), c_1(\theta); \theta), \tag{4}$$

$$\dot{v}(\theta) = U^p_\theta(c_0(\theta), c_1(\theta); \theta), \tag{5}$$

$$\int \left(\lambda_\theta U^p(c_0(\theta), c_1(\theta); \theta) + \alpha_\theta U^c(c_1(\theta))\right) f(\theta) d\theta \geq V. \tag{6}$$

This problem minimizes the resources required to achieve a certain level of welfare subject to incentive compatibility. The second constraint is the envelope condition which, together with the monotonicity condition, is necessary and sufficient for incentive compatibility (see e.g. Milgrom and Segal, 2002).

Our first results focus on cases with no weight on children's welfare.

Proposition 1. *Suppose that Assumptions 1 and 2 hold, and that there is no weight on children $\alpha_\theta = 0$. Then (a) if λ_θ is constant the optimum coincides with the first best and estate taxes are zero, $\tau(\theta) = 0$; (b) if λ_θ is decreasing and in addition Assumption 3 holds, then marginal estate taxes are positive $\tau(\theta) \geq 0$; and (c) if λ_θ is increasing and in addition Assumption 3 holds, then marginal estate taxes are negative, $\tau(\theta) \leq 0$.*

When the weight on parents, λ_θ, is constant, the first-best allocation is incentive compatible and, hence, optimal. This sets up an important benchmark where bequests are not taxed. It formalizes the parable by Mankiw (2006) cited in the Introduction.

In contrast, when weights λ_θ are decreasing, favoring selfish parents, this creates a force for positive taxation of estates. The reverse is true when weights λ_θ are increasing, favoring altruistic parents, leading to a subsidy on estates. These results emphasizes that ordinal preferences cannot settle the sign of estate taxes, which depends crucially on the weights λ_θ. The specification of cardinal utility or social welfare functions is crucial.

We now analyze the case where we allow for arbitrary weights on children. In the appendix we show that at points where the monotonicity constraint is not binding the implicit marginal tax rate on estates equals

$$\tau(\theta) = -\nu\alpha_\theta R U^c_{c_1}(\theta) - \nu\frac{\mu(\theta)}{f(\theta)} R U^c_{c_1}(\theta)\left(\frac{U^p_{\theta,c_1}(\theta)}{U^p_{c_1}(\theta)} - \frac{U^p_{\theta,c_0}(\theta)}{U^p_{c_0}(\theta)}\right), \tag{7}$$

where $\nu > 0$ is the multiplier on the promise keeping constraint (6) and $\mu(\theta)$ is the co-state variable associated with (5), satisfying $\mu(\underline{\theta}) = \mu(\bar{\theta}) = 0$. This formula is equivalent to the one in Farhi and Werning (2010), except for the term involving $\mu(\theta)$.

Proposition 2. *Suppose that Assumptions 1 and 2 hold. Suppose no bunching at the extremes. Then (a) marginal tax rates are negative at the extremes $\tau(\underline{\theta}) < 0$, $\tau(\bar{\theta}) < 0$; (b) if α_θ is constant or decreasing then $\tau(\underline{\theta}) < \tau(\bar{\theta})$.*

These result indicates that, unless we place zero weight on children, a force for subsidies is always present. It also highlights a force for progressive taxation, in the sense of a rising marginal tax rate. Both features are in line with the main results in Farhi and Werning (2010). Indeed, there are parental weights that lead to exactly the same formula as in this canonical tax model. These parental weights are precisely those such that the first best is incentive compatible so that $\mu(\theta) = 0$.

Proposition 3. *For constant weights on children $\alpha_\theta = \alpha \geq 0$, there exists parental weights λ_θ such that $\tau(\theta) = -\nu\alpha_\theta R U^c_{c_1}(\theta)$ for all θ.*

Numerical Illustration. Figure 1 collects a few illustrative examples, using logarithmic utility $U^p(c_0, c_1; \theta) = (1-\theta)\log(c_0) + \theta\log(c_1)$, $U^c(c_1) = \log(c_1)$ and a uniform distribution for θ over $[0.1, 0.9]$.

The first panel in Figure 1 has constant positive weights on both parents and children. Proposition 2 leads us to expect negative tax rates near the extremes. In this example, tax rates remain negative throughout and are increasing in most of the range. This outcome is essentially as in Farhi and Werning (2010).

The second panel puts no weight on children, but assumes a decreasing weight on parents λ_θ. Tax rates are positive throughout, as expected from Proposition 1 part (b). The third panel combines this decreasing λ_θ with a constant and positive weight α_θ; tax rates are negative near the extremes, but positive over an interior interval.

3.2 A Rawlsian Criterion for Children

We now evaluate the welfare of children using a Rawlsian criterion. This amounts to studying the same planning problem in (3)–(6) with the additional constraint

$$U^c(c_1(\theta)) \geq \underline{u}.$$

Define θ^* to be the highest value of θ for which this constraint holds with equality. For high enough $\underline{u}$ we have $\theta^* > \underline{\theta}$. All types $\theta \in [\underline{\theta}, \theta^*]$ are bunched, so the implicit marginal tax $\tau(\theta)$ is increasing in θ by single crossing. Thus, $\tau(\theta) \leq \tau(\theta^*)$ for all $\theta \leq \theta^*$. Indeed, it is possible that $\tau(\theta) < 0$ for some $\theta < \theta^*$ even if $\tau(\theta) \geq 0$ for $\theta \geq \theta^*$. We now show that, indeed, tax rates are positive above θ^*.

Proposition 4. *Suppose Assumptions 1, 2 and 3 hold. Suppose further that λ_θ is constant and that c_1 is a normal good. Then marginal taxes are positive $\tau(\theta) \geq 0$ for $\theta \geq \theta^*$ and strictly positive over a positive measure of θ.*

Even though the weight on parents is constant, the optimum involves positive taxation wherever the Rawlsian constraint is slack. Intuitively, children with bequests above the minimum do not contribute towards the maximin criterion, so they are taxed to redistribute towards the poorest children, as well as their selfish parents, who may otherwise be hurt by the imposition to improve their children's welfare. The implicit marginal tax rate at the bottom may or may not be negative, but it is positive for $\theta \geq \theta^*$. Given Proposition 1 part (a), positive taxes can be entirely attributed to placing a positive weight on children.[4] The second panel in Figure 2 illustrates this result. In this example, the implicit tax in the bunching region indeed becomes negative for low enough θ.

The optimal allocation has bunching below θ^*, so the tax schedule T must feature a kink, with marginal tax rates jumping upward. Indeed, it may require a marginal subsidy, coming from the left. A simple alternative implementation can avoid this by imposing the same budget constraints (1)–(2) but adding the constraint that $B \geq \underline{B}$.[5] By a suitable

[4] Formula (7) can still be applied with endogenous positive weights on children α_θ that are decreasing in θ and are zero for all $\theta > \theta^*$; the costate $\mu(\theta)$ negative and zero at the extremes.

[5] This implementation is natural because it highlights that the optimal allocation will typically feature

choice of lump-sum transfers, determining I_0 and I_1, we can normalize $\underline{B} = 0$. The tax code then only imposes positive marginal tax rates, but negative implicit taxes may be generated by the borrowing constraint, $B \geq 0$. Strictly positive taxes and the outlawing of negative bequests are common features of policy across developed countries.

4 Linear Taxes and Limits to Borrowing

We now restrict estate taxes to be linear. The planner taxes bequests at a constant rate τ, balancing its budget with a lump-sum tax (or transfer). We also consider the imposition of constraints on borrowing that limit parents from passing on debt to their children. To keep things simple, we start by discussing the logarithmic utility case. We then provide tax formulas for general preferences.

4.1 A Weighted Utilitarian Objective

We first consider the case with the weighted Utilitarian criterion and no borrowing limits. The planning problem, stated in the online appendix, is relatively straightforward and maximizes our welfare criterion subject to the resource constraint. The first-order conditions deliver a useful tax formula.

Proposition 5. *Assume logarithmic utility $U^p(c_0, c_1; \theta) = (1-\theta)\log(c_0) + \theta \log(c_1)$, $U^c(c_1) = \log(c_1)$. The optimal linear estate tax is given by*

$$\frac{\tau}{1+\tau} = -\frac{\nu}{I}\frac{\frac{Cov(\theta, \lambda_\theta + \alpha_\theta)}{\int \theta(1-\theta)f(\theta)d\theta} + \frac{\int \alpha_\theta(1-\theta)f(\theta)d\theta}{\int \theta(1-\theta)f(\theta)d\theta}}{1 + \frac{Var(\theta)}{\int \theta(1-\theta)f(\theta)d\theta}}.$$

The numerator is the sum of a Ramsey covariance term and a Pigouvian average term. The term in the denominator is a Ramsey adjustment.[6] Roughly speaking, the Ramsey terms reflect the costs and benefits of redistribution across dynasties, while the Pigouvian term reflects the value of redistribution from parents to children. The Pigouvian term has a corrective nature because when social welfare places direct weight on children, parents necessarily undervalue bequests. When $\alpha_\theta = 0$, the Pigouvian term vanishes, leaving

parents below θ^* bunched to satisfy the Rawlsian constraint $U^c(c_1(\theta)) \geq \underline{u}$. These same allocation could be obtained with an appropriate kink in the T function, typically requiring a sufficiently high subsidy rate to the left of the bunching point.

[6] This adjustment term encapsulates the impact on tax revenues of the income effect associated with a marginal tax change.

only the Ramsey terms and the formula specializes to a version of the many-person Ramsey tax problem of Diamond (1975).

The Ramsey covariance term in the numerator may be positive or negative and neatly highlights the importance of the weights λ_θ and α_θ. The Pigouvian average term in the numerator is always negative or zero, providing a force for a subsidy as long as children have positive weight. The Ramsey adjustment term in the denominator only scales taxes proportionately but does not affect their sign.

If both weights are constant then the covariance term is zero and the second term takes over. If we further assume that $\alpha_\theta = 0$ then the optimal tax is zero, $\tau = 0$, in line with Proposition 1 part (a); if, on the contrary, we place a positive and constant weight on children the optimal tax is a subsidy: $\tau < 0$. This linear tax result is consistent with the nonlinear results on negative marginal tax rates at the extremes in Proposition 2.

When α_θ and λ_θ are not constant the covariance term is not zero and a decreasing weights provide a force for a tax. Whether or not the optimal tax is positive or negative depends on the net effect of the two terms in the numerator.

This formula can be generalized away from logarithmic utility. We define the after tax interest rate $\tilde{R} = \frac{R}{1+\tau}$, the uncompensated demand functions $c_0(I, \tilde{R}, \theta)$ and $c_1(I, \tilde{R}, \theta)$, the compensated elasticity $\varepsilon_{c_1,\tilde{R}}(I, \tilde{R}, \theta)$ of c_1 to the after tax interest rate $\tilde{R}$, the indirect utility function $V^p(I, \tilde{R}, \theta)$, and $W(I, \tilde{R}, \theta) = \lambda_\theta V^p(I, \tilde{R}, \theta) + \alpha_\theta U^c(c_1(I, \tilde{R}, \theta)))$.

Proposition 6. *For general preferences, the optimal linear estate tax is given by*

$$\frac{\tau}{1+\tau} = -\nu \frac{\frac{Cov(c_1(\theta), W_I(\theta))}{\int \varepsilon_{c_1,\tilde{R}}(\theta) c_1(\theta) f(\theta) d\theta} + \tilde{R} \frac{\int \alpha_\theta U^c_{c_1}(\theta) \varepsilon_{c_1,\tilde{R}}(\theta) c_1(\theta) f(\theta) d\theta}{\int \varepsilon_{c_1,\tilde{R}}(\theta) c_1(\theta) f(\theta) d\theta}}{1 + \frac{1}{\tilde{R}} \frac{Cov(c_1(\theta), c_{1,I}(\theta))}{\int \varepsilon_{c_1,\tilde{R}}(\theta) c_1(\theta) f(\theta) d\theta}}.$$

The formula takes the form of a ratio as in the logarithmic utility case, with similar terms in the numerator and denominator. The formula highlights the role of the interest rate elasticity of bequests. Basically, the Ramsey terms are hit by the inverse of the elasticity of bequests while the Pigouvian term is not. More precisely, the Pigouvian term is a weighted average, and the interest rate elasticity of bequests only affects the corresponding weights. In this sense, the inverse-elasticity rule applies to the Ramsey terms as in Diamond (1975), but not to the Pigouvian term. Indeed, the average Pigouvian term is best thought as representing a Pigouvian motive for taxation. And to a large extent, Pigouvian taxes do not depend on elasticities.

4.2 A Rawlsian Criterion for Children

We now evaluate children's welfare according to a Rawlsian maximin criterion, exactly as in Section 3.2. In addition to a linear tax on bequests we provide the planner with one additional instrument: a minimum bequest requirement $\underline{B}$. As in Section 3.2, appropriate intergenerational transfers allow us to normalize and set $\underline{B} = 0$, so we can interpret this as a constraint that outlaws parents passing on debt to their children. We assume that the Rawlsian constraint is binding, which is the case for high enough $\underline{u}$.

Proposition 7. *Assume logarithmic utility $U^p(c_0, c_1; \theta) = (1-\theta)\log(c_0) + \theta \log(c_1)$ and $U^c(c_1) = \log(c_1)$. Suppose λ_θ is constant and that children's welfare is evaluated by a Rawlsian maximin criterion. Then the optimum is such that the tax rate is strictly positive $\tau > 0$ and a borrowing constraint is strictly binding for some agents.*

The optimum features a tax coupled with a borrowing limit.[7] This result is a linear counterpart to the nonlinear conclusions in Proposition 4. The economic logic is similar: the revenue from a positive tax is used to improve the welfare of the poorest children, as well as the welfare of parents that are hurt by the imposition of the borrowing constraint.

Proposition 7 requires logarithmic utility. We now provide a more general related local result. Although it does not fully settle the sign, this result does suggest that positive estate taxes may be optimal for a wide class of preferences.

Proposition 8. *Suppose λ_θ is constant and that children's welfare is evaluated by a Rawlsian maximin criterion. Suppose that Assumptions 1 and 2 hold. In addition, assume that c_1 is a normal good. There exists a positive tax $\tau > 0$ that improves on the no-intervention equilibrium with $\tau = 0$.*

We also provide an optimal tax formula for general preferences. We need to adapt the definitions of the demand functions, the indirect utility function and the interest rate elasticity of bequests to incorporate a borrowing constraint (see the online appendix).

Proposition 9. *For general preferences, the optimal tax rate is given by*

$$\frac{\tau}{1+\tau} = -\nu \frac{\frac{Cov\left(c_1(\theta), \lambda_\theta V_I^p(\theta)\right)}{\int \varepsilon_{c_1,\tilde{R}}(\theta) c_1(\theta) f(\theta) d\theta}}{1 + \frac{1}{\tilde{R}} \frac{Cov(c_1(\theta), c_{1,I}(\theta))}{\int \varepsilon_{c_1,\tilde{R}}(\theta) c_1(\theta) f(\theta) d\theta}}.$$

This optimal tax formula features only Ramsey terms and no Pigouvian term: the Pigouvian motive for taxation is addressed entirely by the borrowing constraint.

[7] If λ_θ is decreasing in θ, there is an additional force for a tax.

5 Conclusions

We have singled out one case where optimal policy takes a simple form: a ban on negative bequests and a positive tax on positive ones. These properties are features of tax codes in most developed economies. However, this result applies to a particular, albeit defensible, combination of welfare criteria (maximin for children) and cardinal normalizations. The conclusions are sensitive to the form of redistributive tastes, embedded in assumptions on the cardinality of utility and social welfare functions, as well as the source of the inequality in bequests, such as altruism heterogeneity versus parental earnings heterogeneity.

References

Connell, Christopher and Eric B. Rasmusen, "Concavifying the QuasiConcave," July 2012.

Diamond, P. A., "A many-person Ramsey tax rule," *Journal of Public Economics*, November 1975, *4* (4), 335–342.

Farhi, Emmanuel and Ivan Werning, "Progressive Estate Taxation," *The Quarterly Journal of Economics*, May 2010, *125* (2), 635–673.

Hellwig, Martin, "A Maximum Principle for Control Problems with Monotonicity Constraints," Working Paper, Max Planck Insittute for Research on Collective Goods August 2009.

Lockwood, Benjamin and Matthew Weinzierl, "De Gustibus non est Taxandum: Theory and Evidence on Preference Heterogeneity and Redistribution," Working Paper 17784, National Bureau of Economic Research January 2012.

Mankiw, Gregory N., "The Estate Tax Debate," Greg Mankiw's Blog June 2006.

Milgrom, Paul and Ilya Segal, "Envelope Theorems for Arbitrary Choice Sets," *Econometrica*, March 2002, *70* (2), 583–601.

Mirrlees, James A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38* (2), 175–208.

Piketty, Thomas and Emmanuel Saez, "A Theory of Optimal Capital Taxation," NBER Working Papers 17989, National Bureau of Economic Research, Inc April 2012.

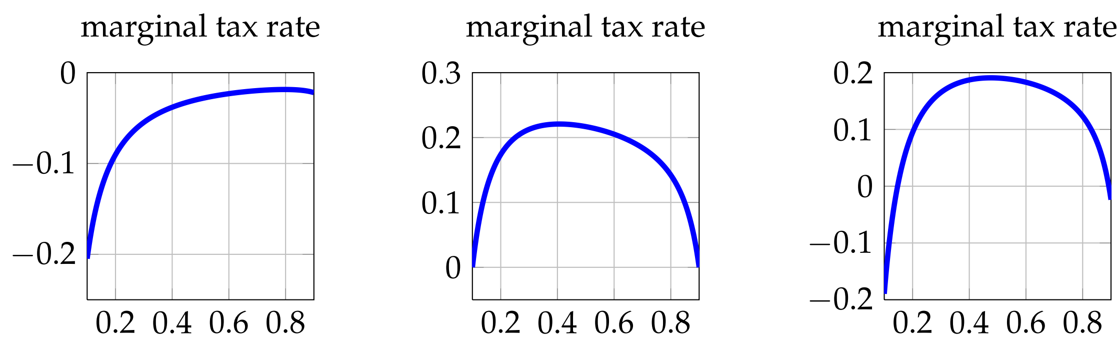

Figure 1: Optimal implicit marginal estate tax rates $\tau(\theta)$ as a function of θ for the weighted Utilitarian case. The Pareto weights λ_θ and α_θ are as follows: $\lambda_\theta = 1$ and $\alpha_\theta = 0.02$ (first panel), $\lambda_\theta = \frac{e^{-\theta}}{E[e^{-\theta}]}$ and $\alpha_\theta = 0$ (second panel) and $\lambda_\theta = \frac{e^{-\theta}}{E[e^{-\theta}]}$ and $\alpha_\theta = 0.02$ (third panel).

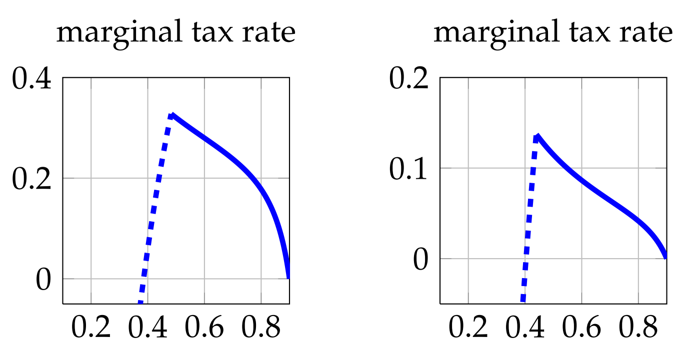

Figure 2: Optimal implicit marginal estate tax rates $\tau(\theta)$ as a function of θ with a Rawlsian criterion for generation 1. The Pareto weights λ_θ are as follows: $\lambda_\theta = \frac{e^{-\theta}}{E[e^{-\theta}]}$ (first panel), and $\lambda_\theta = 1$ (second panel). The dashed portion coincides with values θ for which the borrowing constraint is binding ($\theta \leq \theta^*$). For these values, the implicit marginal tax rate $\tau(\theta)$ is lower than the explicit marginal tax rate $\tau(\theta^*)$ that agents face in the implementation with a nonlinear tax and a borrowing limit, reflecting the binding borrowing constraint.

A Appendix

A.1 Proof of No Ordinal Restrictions in Assumption 2

Note that Assumption 2 is *not* required to hold at all levels of the gross interest rate, only the technological rate of return R, i.e. the return on storage without taxation.

Start with concave preferences $U(c_0, c_1; \theta)$ that satisfy Assumption 1. We construct a renormalization of preferences by applying the transformation $f^g(U(c_0, c_1; \theta); \theta)$. This renormalization ensures that Assumption 2 holds, and preserves concavity.

The functions $f^g(V; \theta)$ are constructed as follows. Fix θ^*. For any function g, for all I and θ, define $f^g(V; \theta)$ by

$$f^g_V(V(I, R; \theta); \theta) V_I(I, R; \theta) = g'(V(I, R; \theta^*)) V_I(I, R; \theta^*),$$

or

$$f^g_V(V(I, R; \theta); \theta) = g'(V(I, R; \theta^*)) \frac{V_I(I, R; \theta^*)}{V_I(I, R; \theta)}.$$

We can pick g so that $f^g_V(V; \theta)$ is decreasing in V for each θ so that concavity is preserved.

A.2 Proof of Proposition 1

Case (a) is immediate. We now prove case (b); case (c) is symmetric. Define $C_0(c_1, v; \theta)$ as $v = U^p(C_0(c_1, v; \theta), c_1; \theta)$. The planning problem is

$$\min \int \left(C_0(c_1(\theta), v(\theta); \theta) + \frac{1}{R} c_1(\theta) \right) f(\theta) d\theta$$

subject to c_1 non-decreasing and

$$\dot{v}(\theta) = U^p_\theta(C_0(c_1(\theta), v(\theta); \theta), c_1(\theta); \theta),$$

$$\int \lambda_\theta v(\theta) f(\theta) d\theta \geq V.$$

We study the relaxed planning problem

$$\min \int \left(C_0(c_1(\theta), v(\theta); \theta) + \frac{1}{R} c_1(\theta) \right) f(\theta) d\theta$$

subject to c_1 non-decreasing and

$$\dot{v}(\theta) = U^p_\theta(C_0(c_1(\theta), v(\theta); \theta), c_1(\theta); \theta) + r(\theta),$$

$$r(\theta) \geq 0,$$

$$\int \lambda_\theta v(\theta) f(\theta) d\theta \geq V.$$

The original problem imposes $r(\theta) = 0$ for all θ, but here we allow for $r(\theta) \geq 0$. Our first goal is to show that an interior solution to the relaxed problem features $r(\theta) = 0$ and, thus, coincide with the original planning problem.

Adapting Theorem 3.1 in Hellwig (2009) we form the Hamiltonian

$$H = \lambda_\theta v f(\theta) - \gamma \left(C_0(c_1, v; \theta) + \frac{1}{R} c_1 \right) f(\theta) + \mu(U^p_\theta(C_0(c_1, v; \theta), c_1, \theta) + r) + \chi q$$

where γ should be thought of the inverse of the multiplier ν on the promise keeping constraint. We obtain the following necessary conditions for an interior optimum:

$$\dot{\chi}(\theta) = \gamma \frac{1}{R} f(\theta) + \gamma f(\theta) C_{0,c_1}(\theta) - \mu(\theta) \left(U^p_{\theta,c_0}(\theta) C_{0,c_1}(\theta) + U^p_{\theta,c_1}(\theta) \right),$$
$$\dot{\mu}(\theta) = -\lambda_\theta f(\theta) + \gamma f(\theta) C_{0,v}(\theta) - \mu(\theta) U^p_{\theta,c_0}(\theta) C_{0,v}(\theta),$$

$$\mu(\underline{\theta}) = \mu(\bar{\theta}) = \chi(\underline{\theta}) = \chi(\bar{\theta}) = 0,$$

$$\mu(\theta) \leq 0 \quad \mu(\theta) r(\theta) = 0,$$

$$\chi(\theta) \leq 0,$$

$$\int \chi(\theta) dc_1(\theta) = 0.$$

The relaxed problem solves the original problem. To show that a solution to the relaxed problem features $r(\theta) = 0$ and, thus, coincide with the original planning problem, we argue by contradiction. Thus, assume that $r(\theta) > 0$ on a positive measure of points θ.

We claim that if we can find a set E of positive measure such that for all $\theta \in E$ we have

$$r(\theta) > 0,$$
$$\mu(\theta) = 0,$$
$$\dot{\mu}(\theta) = 0,$$

and $c_1(\cdot)$ strictly increasing at θ then we are done. First we show that we can find such a set under the assumption that $r(\theta) > 0$ on a set of positive measure. Let E be a set of positive measure on which $r(\theta) > 0$. Then by the necessary conditions we know that $\mu(\theta) = 0$ on E. Since μ, v, χ are all absolutely continuous they are differentiable almost everywhere. Thus we can assume without loss of generality that μ, χ and v are all differentiable on E. Since $\mu(\theta) \leq 0$ we know that $\dot{\mu}(\theta) = 0$ on E. If we have that $c_1(\theta)$ is strictly increasing on a positive measure subset of E then we are done. Thus suppose not, so that c_1 is not strictly increasing at almost every of E. Without loss of generality we can suppose that c_1 is not strictly increasing at every point of E. In other words given $\theta \in E$ we can choose ϵ small enough so that $c_1(\theta - \epsilon) = c_1(\theta + \epsilon)$. Now by our preliminary fact we know that E contains an accumulation point, call it θ_0. By the preceding argument there exists an $\epsilon > 0$ so that c_1 is constant on the interval $(\theta_0 - \epsilon, \theta_0 + \epsilon)$. Let $\bar{c}_1 = c_1(\theta_0)$. Now since $U_{c_0} > 0$ and v is differentiable at θ_0 the implicit function theorem says that $c_0(\theta)$ defined by

$$v(\theta) = U^p(c_0(\theta), \bar{c}_1; \theta)$$

is differentiable at θ_0. Then

$$\dot{v}(\theta_0) = U^p_{c_1}(c_0(\theta_0), \bar{c}_1; \theta)\dot{c}_0(\theta_0) + U^p_\theta(c_0(\theta_0), \bar{c}_1; \theta).$$

We also know that

$$\dot{v}(\theta_0) = U^p_\theta(c_0(\theta_0), \bar{c}_1; \theta) + r(\theta_0) > U^p_\theta(c_0(\theta_0), \bar{c}_1; \theta).$$

Since $r(\theta_0) > 0$. Since $U^p_{c_0} > 0$ this implies that $\dot{c}_0(\theta_0) > 0$. Since θ_0 is an accumulation point of E there exists a sequence $\theta_n \in E$ with $\theta_n \to \theta_0$. We can further suppose that we have either $\theta_n \nearrow \theta_0$ or $\theta_n \searrow \theta_0$. We take the case $\theta_n \nearrow \theta_0$. Using the fact that for all $\theta \in E$ we have $\mu(\theta) = \dot{\mu}(\theta) = 0$ we have that

$$\gamma = \lambda(\theta) U^p_{c_0}(c_0(\theta), c_1(\theta); \theta).$$

Using the fact that $\dot{c}_0(\theta_0) > 0$ we know that for all large enough n we have $c_0(\theta_n) < c_0(\theta_0)$ and $c_1(\theta_n) = c_1(\theta)$. But then using the concavity of the utility function, the fact that $U^p_{c_0,\theta} < 0$ and λ_θ is decreasing we see that

$$\begin{aligned} \gamma &= \lambda_{\theta_0} U^p_{c_0}(c_0(\theta_0), c_1(\theta_0); \theta_0), \\ &\leq \lambda_{\theta_0} U^p_{c_0}(c_0(\theta_0), c_1(\theta_0); \theta_n), \end{aligned}$$

$$
\begin{aligned}
&< \lambda_{\theta_0} U^p_{c_0}(c_0(\theta_n), c_1(\theta_n); \theta_n), \\
&\leq \lambda_{\theta_n} U^p_{c_0}(c_0(\theta_n), c_1(\theta_n); \theta_n),
\end{aligned}
$$

which is a contradiction. Thus on the original set E we must have had a positive measure set of points A such that c_1 was strictly increasing at these points. Consider the new set $E' = E \cap A$ which also has positive measure. Then for all $\theta \in E'$ we have that

$$
\begin{aligned}
r(\theta) &> 0, \\
\mu(\theta) &= 0, \\
\dot{\mu}(\theta) &= 0,
\end{aligned}
$$

the function $c_1(\theta)$ is strictly increasing and μ, v, χ are all differentiable. Since c_1 is increasing at θ we know that we must have $\chi(\theta) = 0$. But since χ is differentiable at θ and $\chi(\theta) = 0$ and $\chi \leq 0$ it must be that $\dot{\chi}(\theta) = 0$.

From here onwards we restrict to points $\theta \in E'$. Using the fact that $\mu(\theta) = \dot{\mu}(\theta) = \dot{\chi}(\theta) = 0$ tells us that

$$
-\lambda_\theta f(\theta) + \gamma C_{0,v}(\theta) f(\theta) = 0.
$$

and

$$
\gamma \frac{1}{R} f(\theta) + \gamma f(\theta) C_{0,c_1}(\theta) = 0.
$$

Rearranging and using the definition of $C_0(c_1, v; \theta)$, these two equations are equivalent to

$$
\gamma = \lambda_\theta U^p_{c_0}(c_0(\theta), c_1(\theta); \theta), \tag{8}
$$

$$
R = \frac{U^p_{c_0}(c_0(\theta), c_1(\theta); \theta)}{U^p_{c_1}(c_0(\theta), c_1(\theta); \theta)}. \tag{9}
$$

The second equation tells us that we can write

$$
v(\theta) = V^p(I(\theta), R, \theta)
$$

for some $I(\theta)$ since $c_0(\theta)$ and $c_1(\theta)$ are chosen as they would be in the parent's optimal problem. Now we claim that $I(\theta)$ is decreasing on E'. This follows from the fact that

$$
\frac{\gamma}{\lambda(\theta)} = U^p_{c_0}(c_0(\theta), c_1(\theta); \theta) = V^p_I(I(\theta), R; \theta)
$$

and the fact that V is concave and λ is decreasing. Since E' has positive measure it contains a limit point θ. Suppose that there exists a sequence $\theta_n \in E$ with $\theta_n \searrow \theta$. The case

$\theta_n \nearrow \theta$ is symmetric. Then since v is differentiable at θ we know that

$$\dot{v}(\theta) = \lim_{n\to\infty} \frac{v(\theta_n) - v(\theta)}{\theta_n - \theta}$$

since V is differentiable in I and θ we see that

$$\begin{aligned} v(\theta_n) - v(\theta) &= V^p(I(\theta_n), R; \theta_n) - V^p(I(\theta), R; \theta) \\ &= V_I^p(I(\theta), R; \theta)(I(\theta_n) - I(\theta)) + V_\theta^p(I(\theta), R; \theta)(\theta_n - \theta) + \varepsilon(\theta_n - \theta) \end{aligned}$$

where the $\varepsilon(\theta_n - \theta)$ is a second order error term so that $\frac{\varepsilon(\theta_n - \theta)}{\theta_n - \theta} \to 0$. Thus

$$\begin{aligned} \dot{v}(\theta) &= \lim_{n\to\infty} \frac{V_I^p(I(\theta), R; \theta)(I(\theta_n) - I(\theta)) + V_\theta^p(I(\theta), R; \theta)(\theta_n - \theta)}{\theta_n - \theta} \\ &= V_\theta^p(I(\theta), R; \theta) + V_I^p(I(\theta), R; \theta) \cdot \lim_{n\to\infty} \frac{(I(\theta_n) - I(\theta))}{\theta_n - \theta} \\ &\leq V_\theta^p(I(\theta), R; \theta) \end{aligned}$$

since $I(\theta_n) - I(\theta) \leq 0$. Thus we see that $\dot{v}(\theta) \leq V_\theta^p(I(\theta), R; \theta)$. But since

$$\dot{v}(\theta) = V_\theta^p(I(\theta), R; \theta) + r(\theta) = U_{c_0}^p(c_0(\theta), c_1(\theta); \theta) + r(\theta)$$

and $r(\theta) > 0$ we have a contradiction. Thus it must have been that $r = 0$ almost surely so that the solution to the relaxed problem coincides with the solution to the original problem.

The relaxed problem features positive taxes. In the proof, we will make repeated use of the fact that over an interval where there is bunching, the tax rate τ is increasing. This is a direct consequence of the single crossing condition in Assumption 1.

Since μ, v, χ are all absolutely continuous they are differentiable almost everywhere. Thus there is a full measure set Ω of θ such that μ, χ and v are all differentiable. Consider $\theta \in \Omega$.

Suppose that c_1 is strictly increasing at θ. Then $\chi(\theta) = \dot{\chi}(\theta) = 0$, and we get using equation 11 together with the fact that $\mu(\theta) \leq 0$ and single crossing that $\tau(\theta) \geq 0$.

Now suppose that c_1 is not strictly increasing at θ. Consider the greatest interval around θ over which c_1 is constant. Then the tax rate τ is increasing over this interval so that it is comprised between its limit values at the bounds θ_l and θ_h of the interval.

If $\theta_l > \underline{\theta}$, then the function c_1 must be strictly increasing at θ_l so that $\tau(\theta_l) \geq 0$. If c_1

is continuous at θ_l, then so are c_0 and τ. We can conclude that $\tau(\theta) \geq 0$. Suppose now that c_1 is not continuous at θ_l. If $\lim_{\tilde{\theta}\to\theta_l^+} \tau(\tilde{\theta}) \geq 0$, then we have $\tau(\theta) \geq 0$. Otherwise $\lim_{\tilde{\theta}\to\theta_l^+} \tau(\tilde{\theta}) < 0$, from which we derive a contradiction by constructing a new allocation that satisfies the constraints of the relaxed planning problem but achieves lower cost. We first construct a new allocation $(\hat{c}_0, \hat{c}_1)$ which coincides with the old one except at points $\tilde{\theta} \in (\theta_l, \theta_h)$ such that $\tau(\tilde{\theta}) < 0$, in which case we pick $(\hat{c}_0(\tilde{\theta}), \hat{c}_1(\tilde{\theta}))$ so that $U^p(\hat{c}_0(\tilde{\theta}), \hat{c}_1(\tilde{\theta}); \tilde{\theta}) = U^p(c_0(\tilde{\theta}), c_1(\tilde{\theta}); \tilde{\theta})$ and $R = \frac{U^p_{c_0}(\hat{c}_0(\tilde{\theta}),\hat{c}_1(\tilde{\theta});\tilde{\theta})}{U^p_{c_1}(\hat{c}_0(\tilde{\theta}),\hat{c}_1(\tilde{\theta});\tilde{\theta})}$ so that $\hat{\tau}(\tilde{\theta}) = 0$. Then define an ironed version of this allocation by setting $\hat{\hat{c}}_1(\tilde{\theta}) = \hat{c}_1(\phi(\tilde{\theta}))$ and $\hat{\hat{c}}_0(\tilde{\theta}) = \hat{c}_0(\phi(\tilde{\theta}))$ where $\phi(\tilde{\theta}) = \arg\max_{\tilde{\theta}' < \tilde{\theta}} \hat{c}_1(\tilde{\theta}')$. Then this allocation satisfies the constraints of the relaxed planning problem but has a lower cost.

Suppose now that $\theta_l = \underline{\theta}$. We have $\chi(\underline{\theta}) = 0$. Then for every $\epsilon > 0$, we can find $\tilde{\theta}_\epsilon$ in $[\underline{\theta}, \underline{\theta} + \epsilon) \cap \Omega$, such that $\dot{\chi}(\tilde{\theta}) \leq 0$. Because $\lim_{\epsilon\to 0} \chi(\tilde{\theta}_\epsilon) = 0$, we conclude using equation (10) that $\lim_{\epsilon\to 0} \tau(\tilde{\theta}_\epsilon) \geq 0$. Since τ is increasing on $(\underline{\theta}, \theta_h)$, this allows us to conclude that $\tau(\theta) \geq \lim_{\epsilon\to 0} \tau(\tilde{\theta}_\epsilon) \geq 0$.

A.3 Derivation of Optimal Tax Formula and Proposition 2

Define $C_0(c_1, v; \theta)$ as $v = U^p(C_0(c_1, v; \theta), c_1; \theta)$. We have $C_{0,c_1} = -\frac{U^p_{c_1}}{U^p_{c_0}}$, and $C_{0,v} = \frac{1}{U^p_{c_0}}$. We adapt Theorem 3.1 in Hellwig (2009). We form the Hamiltonian

$$H = (\lambda_\theta v + \alpha_\theta U^c(c_1)) f(\theta) - \gamma \left(C_0(c_1, v; \theta) + \frac{1}{R} c_1 \right) f(\theta) + \mu U^p_\theta(C_0(c_1, v; \theta), c_1, \theta) + \chi q$$

We have the following necessary conditions:

$$\dot{\chi}(\theta) = -\alpha_\theta U^c_{c_1}(\theta) f(\theta) + \gamma \frac{1}{R} f(\theta) + \gamma f(\theta) C_{0,c_1}(\theta) - \mu(\theta) \left(U^p_{\theta,c_0}(\theta) C_{0,c_1}(\theta) + U^p_{\theta,c_1}(\theta) \right),$$
$$\dot{\mu}(\theta) = -\lambda_\theta f(\theta) + \gamma f(\theta) C_{0,v}(\theta) - \mu(\theta) U^p_{\theta,c_0}(\theta) C_{0,v}(\theta),$$

$$\mu(\underline{\theta}) = \mu(\bar{\theta}) = \chi(\underline{\theta}) = \chi(\bar{\theta}) = 0,$$

$$\chi(\theta) \leq 0,$$

and the complementary slackness condition

$$\int \chi(\theta) dc_1(\theta) = 0.$$

Using the definition for

$$\tau(\theta) = R\frac{U^p_{c_1}(\theta)}{U^p_{c_0}(\theta)} - 1$$

we can rewrite the first equation as

$$\gamma\tau(\theta) = -R\frac{\dot{\chi}(\theta)}{f(\theta)} - \alpha_\theta R U^c_{c_1}(\theta) - \frac{\mu(\theta)}{f(\theta)} R U^c_{c_1}(\theta)\left(\frac{U^p_{\theta,c_1}(\theta)}{U^p_{c_1}(\theta)} - \frac{U^p_{\theta,c_0}(\theta)}{U^p_{c_0}(\theta)}\right). \tag{10}$$

The result in Proposition 2 follows immediately from the fact that $\mu(\theta)$ is zero at the extremes. As long as $c_1(\theta)$ is strictly increasing, then we have $\chi(\theta) = 0$ and $\dot{\chi}(\theta) = 0$ so that using $\gamma = \frac{1}{\nu}$, we have

$$\tau(\theta) = -\nu\alpha_\theta R U^c_{c_1}(\theta) - \nu\frac{\mu(\theta)}{f(\theta)} R U^c_{c_1}(\theta)\left(\frac{U^p_{\theta,c_1}(\theta)}{U^p_{c_1}(\theta)} - \frac{U^p_{\theta,c_0}(\theta)}{U^p_{c_0}(\theta)}\right). \tag{11}$$

A.4 Proof of Proposition 3

We fix the weights α_θ, and solve the following system

$$\gamma\frac{1}{U^p_{c_0}(\theta)} = -\alpha_\theta\frac{U^c_{c_1}(\theta)}{U^p_{c_1}(\theta)} + \gamma\frac{1}{R}\frac{1}{U^p_{c_1}(\theta)}, \tag{12}$$

$$U^p(\theta) = v(\theta), \tag{13}$$

$$U^p_\theta(\theta) = \dot{v}(\theta). \tag{14}$$

Given $v(\theta)$, equations 12 and 13 pin down $c_0(\theta)$ and $c_1(\theta)$. Equation 14 can then be seen as a differential equation in $v(\theta)$.

If the solution of this system is such that $c_1(\theta)$ is increasing in θ, then the corresponding allocation is incentive compatible, and solves the planning problem for parental weights λ_θ given by

$$\gamma\frac{1}{U^p_{c_0}(\theta)} = \lambda_\theta.$$

For $\alpha_\theta = \alpha$ constant, we know that the allocation is incentive compatible. Indeed the allocation can be constructed by confronting agents with a nonlinear tax on bequests given by

$$T'\left(\frac{c_1 - I_1}{R}\right) = -\nu\alpha R U^c_{c_1}(c_1).$$

The bequest tax T is convex, and hence the resulting budget set is concave. The corresponding allocation is incentive compatible by construction.

A.5 Proof of Proposition 4

The planning problem is (normalizing $\lambda_\theta = 1$)

$$\min \int \left(C_0(c_1(\theta), v(\theta); \theta) + \frac{1}{R} c_1(\theta) \right) f(\theta) d\theta$$

subject to c_1 non-decreasing and

$$c_1(\theta) \geq \underline{c},$$
$$\dot{v}(\theta) = U^p_\theta(C_0(c_1(\theta), v(\theta); \theta), c_1(\theta); \theta),$$
$$\int v(\theta) dF(\theta) \geq V.$$

As in the proof of Proposition 1, we study the relaxed planning problem

$$\min \int \left(C_0(c_1(\theta), v(\theta); \theta) + \frac{1}{R} c_1(\theta) \right) f(\theta) d\theta$$

subject to c_1 non-decreasing and

$$c_1(\theta) \geq \underline{c},$$
$$\dot{v}(\theta) = U^p_\theta(C_0(c_1(\theta), v(\theta); \theta), c_1(\theta); \theta) + r(\theta),$$
$$r(\theta) \geq 0,$$
$$\int v(\theta) dF(\theta) \geq V.$$

The original problem imposes $r(\theta) = 0$ for all θ, but here we allow for $r(\theta) \geq 0$. The necessary conditions can be derived by adapting Theorem 3.1 in Hellwig (2009). Indeed, his setup explicitly allows for a constraint such as $c_1(\theta) \geq \underline{c}$.

We claim that a solution to the relaxed problem must feature $r(\theta) = 0$ and, thus, coincide with the original planning problem. The proof of this claim is essentially identical to that laid out in the proof of Proposition 1. The presence of the new constraint $c_1(\theta) \geq \underline{c}$ does not change the key arguments involved. Indeed, the necessary conditions are identical except that $\chi(\underline{\theta})$ is not required to be zero.

Following the proof of Proposition 1 it follows immediately that $\tau(\theta) \geq 0$ for all $\theta \geq \theta^*$. Finally we can use Proposition 8 to conclude that we cannot have $\tau(\theta) = 0$ almost surely for $\theta \geq \theta^*$.

A.6 Proof of Proposition 5

The proposition is a direct application of Proposition 6 specialized to the logarithmic utility case.

A.7 Proof of Proposition 6

We define the after-tax interest rate $\tilde{R} = \frac{R}{1+\tau}$. The planning problem is[8]

$$\min \int \left(c_0(I, \tilde{R}, \theta) + \frac{1}{R} c_1(I, \tilde{R}, \theta) \right) f(\theta) d\theta$$

subject to

$$\int \left(\lambda_\theta V^p(I, \tilde{R}, \theta) + \alpha_\theta U^c(c_1(I, \tilde{R}, \theta)) \right) f(\theta) d\theta \geq V,$$

where $c_0(I, \tilde{R}, \theta)$ and $c_1(I, \tilde{R}, \theta)$ are the uncompensated demand functions and $V^p(I, \tilde{R}, \theta)$ is the indirect utility function. We denote by $W(I, \tilde{R}, \theta) = \lambda_\theta V^p(I, \tilde{R}, \theta) + \alpha_\theta U^c(c_1(I, \tilde{R}, \theta)))$. Finally we denote by $c_0^c(u, \tilde{R}; \theta)$ and $c_1^c(u, \tilde{R}; \theta)$ the compensated demand functions, and by $\varepsilon_{c_1, \tilde{R}}(I, \tilde{R}, \theta)$ the compensated elasticity of c_1 to the after tax interest rate $\tilde{R}$.

We now proceed to prove Proposition 6. We use

$$c_{0,I}(\theta) + \frac{1}{\tilde{R}} c_{1,I}(\theta) = 1,$$

$$c_{0,\tilde{R}}(\theta) + \frac{1}{\tilde{R}} c_{1,\tilde{R}}(\theta) - \frac{1}{\tilde{R}^2} c_1(\theta) = 0,$$

$$c_{1,\tilde{R}}(\theta) = c^c_{1,\tilde{R}}(\theta) + \frac{1}{\tilde{R}^2} c_1(\theta) c_{1,I}(\theta),$$

$$c_{0,\tilde{R}}(\theta) = c^c_{0,\tilde{R}}(\theta) + \frac{1}{\tilde{R}^2} c_1(\theta) c_{0,I}(\theta),$$

$$V^p_{\tilde{R}}(\theta) = \frac{1}{\tilde{R}^2} V^p_I(\theta) c_1(\theta).$$

We find

$$\int \left[\left(\frac{1}{R} - \frac{1}{\tilde{R}} \right) c_{1,\tilde{R}}(\theta) + \frac{1}{\tilde{R}^2} c_1(\theta) \left(1 - \nu V^p_I(\theta) \right) - \alpha_\theta \nu U^c_{c_1}(\theta) c_{1,\tilde{R}}(\theta) \right] f(\theta) d\theta = 0,$$

[8]If $\alpha_\theta = 0$ this case amounts to a many-person Ramsey tax problem, as in Diamond (1975).

$$\left(\frac{1}{\tilde{R}}-\frac{1}{R}\right)=\frac{\int\left[\frac{1}{\tilde{R}^2}c_1(\theta)\left(1-\nu V_I^p(\theta)\right)-\alpha_\theta\nu U_{c_1}^c(\theta)c_{1,\tilde{R}}(\theta)\right]f(\theta)d\theta}{\int c_{1,\tilde{R}}(\theta)f(\theta)d\theta}.$$

Similarly, we have

$$\int\left[1+\left(\frac{1}{R}-\frac{1}{\tilde{R}}\right)c_{1,I}(\theta)-\nu V_I^p(\theta)-\alpha_\theta\nu U_{c_1}^c(\theta)c_{1,I}(\theta)\right]f(\theta)d\theta=0,$$

$$\left(\frac{1}{\tilde{R}}-\frac{1}{R}\right)=\frac{\int\left[1-\nu V_I^p(\theta)-\alpha_\theta\nu U_{c_1}^c(\theta)c_{1,I}(\theta)\right]f(\theta)d\theta}{\int c_{1,I}(\theta)f(\theta)d\theta}.$$

After some manipulations, we get

$$\frac{1}{R}-\frac{1}{\tilde{R}}=\frac{\frac{\nu}{\tilde{R}^2}\text{Cov}\left(c_1(\theta),W_I(\theta)\right)+\int\alpha_\theta\nu U_{c_1}^c(\theta)c_{1,\tilde{R}}^c(\theta)f(\theta)d\theta}{\int c_{1,\tilde{R}}^c(\theta)f(\theta)d\theta+\frac{1}{\tilde{R}^2}\text{Cov}\left(c_1(\theta),c_{1,I}(\theta)\right)},$$

which can be transformed into

$$\frac{\tau}{1+\tau}=-\nu\tilde{R}\frac{\frac{1}{\tilde{R}^2}\text{Cov}\left(c_1(\theta),W_I(\theta)\right)+\int\alpha_\theta U_{c_1}^c(\theta)c_{1,\tilde{R}}^c(\theta)f(\theta)d\theta}{\int c_{1,\tilde{R}}^c(\theta)f(\theta)d\theta+\frac{1}{\tilde{R}^2}\text{Cov}\left(c_1(\theta),c_{1,I}(\theta)\right)},$$

and finally

$$\frac{\tau}{1+\tau}=-\nu\frac{\frac{\text{Cov}(c_1(\theta),W_I(\theta))}{\int\varepsilon_{c_1,\tilde{R}}(\theta)c_1(\theta)f(\theta)d\theta}+\tilde{R}\frac{\int\alpha_\theta U_{c_1}^c(\theta)\varepsilon_{c_1,\tilde{R}}(\theta)c_1(\theta)f(\theta)d\theta}{\int\varepsilon_{c_1,\tilde{R}}(\theta)c_1(\theta)f(\theta)d\theta}}{1+\frac{1}{\tilde{R}}\frac{\text{Cov}(c_1(\theta),c_{1,I}(\theta))}{\int\varepsilon_{c_1,\tilde{R}}(\theta)c_1(\theta)f(\theta)d\theta}}.$$

A.8 Proof of Proposition 7

We apply Proposition 9 to the logarithimic utility case to prove Proposition 7. We have

$$c_1(I,\tilde{R},\theta)=\min(I\tilde{R}\theta,\exp(\underline{u})),$$

$$\lambda_\theta V^p(I,\tilde{R},\underline{u},\theta)=\begin{cases}\lambda_\theta\left[(1-\theta)\log\left(I-\frac{\exp(\underline{u})}{\tilde{R}}\right)+\theta\underline{u}\right] & \text{if}\quad I\tilde{R}\theta\leq\exp(\underline{u}),\\ \lambda_\theta\left[\log I+(1-\theta)\log(1-\theta)+\theta\log(\tilde{R}\theta)\right] & \text{if}\quad I\tilde{R}\theta\geq\exp(\underline{u}),\end{cases}$$

so that

$$\lambda_\theta V_I^p(I,\tilde{R},\underline{u},\theta)=\begin{cases}\frac{\lambda_\theta(1-\theta)}{I-\frac{\exp(\underline{u})}{\tilde{R}}} & \text{if}\quad I\tilde{R}\theta\leq\exp(\underline{u}),\\ \frac{\lambda_\theta}{I} & \text{if}\quad I\tilde{R}\theta\geq\exp(\underline{u}).\end{cases}$$

Note that for $I\tilde{R}\theta \leq \exp(\underline{u})$, we have

$$\frac{\lambda_\theta(1-\theta)}{I - \frac{\exp(\underline{u})}{\tilde{R}}} = \frac{\lambda_\theta(1-\theta)\tilde{R}}{I\tilde{R}(1-\theta) + I\tilde{R}\theta - \exp(\underline{u})} = \frac{\lambda_\theta}{I + \frac{I\tilde{R}\theta - \exp(\underline{u})}{\tilde{R}(1-\theta)}} > \frac{\lambda_\theta}{I},$$

and $\frac{I\tilde{R}\theta - \exp(\underline{u})}{\tilde{R}(1-\theta)}$ is negative and increasing in θ. Hence $\mathrm{Cov}(c_1(\theta), V_I^p(\theta))$ is negative even when λ_θ is constant. Also, $c_{1,I}(\theta) = \tilde{R}\theta$ for $\theta \geq \frac{\exp(\underline{u})}{I\tilde{R}}$ and 0 otherwise. Thus, both $c_1(\theta)$ and $c_{1,I}(\theta)$ are increasing in θ, implying that $\mathrm{Cov}\left(c_1(\theta), c_{1,I}(\theta)\right) \geq 0$.

A.9 Proof of Proposition 8

We first show that $V_I^p(I, R, \underline{u}; \theta)$ is higher for types such that the constraint $U^c(c_1(I, R, \underline{u}; \theta)) = \underline{u}$ binds. Let C_1^c be the inverse function of U^c. We have

$$V^p(I, R, \underline{u}; \theta) = \max U(I - \frac{1}{R}c_1, c_1)$$

subject to

$$c_1 \geq C_1^c(\underline{u}).$$

The FOC implies that

$$\frac{U_{c_1}}{U_{c_0}} \leq \frac{1}{R},$$

with an equality if $c_1(I, R, \underline{u}; \theta) > C_1^c(\underline{u})$, i.e. if $\theta > \theta^*$. Consider $\theta < \theta^*$ so that $c_1(I, R, \underline{u}; \theta) = C_1^c(\underline{u})$. We have

$$V_I^p(I, R, \underline{u}; \theta) = U_{c_0}^p(I - \frac{1}{R}C_1^c(\underline{u}), C_1^c(\underline{u})).$$

Hence

$$U_{c_1}^c \frac{d}{d\underline{u}} V_I^p = -\frac{1}{R}U_{c_0,c_0}^p + U_{c_0,c_1}^p,$$

where we have omitted the arguments for brevity.

Since $\frac{U_{c_1}^p}{U_{c_0}^p} \leq \frac{1}{R}$ and $U_{c_0,c_0}^p \leq 0$, we get

$$U_{c_1}^c \frac{d}{d\underline{u}} V_I^p \geq -\frac{U_{c_1}^p}{U_{c_0}^p} U_{c_0,c_0}^p + U_{c_0,c_1}^p \geq 0,$$

where the second inequality follows from the assumption that c_1 is a normal good. Hence $\frac{d}{d\underline{u}} V_I^p \geq 0$. Using the fact then when $\underline{u}$ is low enough, the constraint $U^c(c_1(I, R, \underline{u}; \theta)) = \underline{u}$ ceases to bind and then $V_I^p(I, R, \underline{u}; \theta)$ is independent of θ, we conclude that $V_I^p(I, R, \underline{u}; \theta)$

is higher for types such that the constraint $U^c(c_1(I, R, \underline{u}; \theta)) = \underline{u}$ binds.

We now use this observation to prove that a small positive tax is beneficial. We find it easier to work in the prime where we maximize welfare subject to a resource constraint rather than in the dual. Imagine that we change R to $R + dR$. To satisfy the resource constraint, we need to change I to $I + dI$, where $dI = -\frac{dR}{R^2} \int c_1(I, R, \underline{u}; \theta) f(\theta) d\theta$. The change in welfare is given by

$$dW = \int \lambda_\theta (V_R^p(\theta) dR + V_I^p(\theta) dI) f(\theta) d\theta,$$

which using $V_R^p = \frac{c_1}{R^2} V_I^p$, we can rewrite as

$$dW = \int \lambda_\theta (c_1(\theta) V_I^p(\theta) \frac{dR}{R^2} - V_I(\theta) \frac{dR}{R^2} \int c_1(\theta') f(\theta') d\theta') f(\theta) d\theta,$$

or

$$dW = \frac{dR}{R^2} \int \lambda_\theta (c_1(\theta) V_I^p(\theta) - V_I^p(\theta) \int c_1(\theta') f(\theta') d\theta') f(\theta) d\theta,$$

which when $\lambda_\theta = \bar{\lambda}$ is constant, can be rewritten as

$$dW = \frac{dR}{R^2} \bar{\lambda} \mathrm{Cov}(c_1(\theta), V_I^p(\theta)).$$

Our previous result shows that $\mathrm{Cov}(c_1(\theta), V_I^p(\theta)) < 0$, assuming normality of good c_1. We conclude that starting with no tax, a small positive tax increases welfare.

A.10 Proof of Proposition 9

The planning problem is

$$\min \int \left(c_0(I, \tilde{R}, \underline{u}, \theta) + \frac{1}{R} c_1(I, \tilde{R}, \underline{u}, \theta) \right) f(\theta) d\theta$$

subject to

$$\int \lambda_\theta V^p(I, \tilde{R}, \underline{u}, \theta) f(\theta) d\theta \geq V,$$

where we have defined $\underline{u} = U^c(R\underline{B})$ and $V^p(I, \tilde{R}, \underline{u}, \theta) = \max_{c_0, c_1} U^p(c_0, c_1; \theta)$ subject to $c_0 + \frac{c_1}{\tilde{R}} \leq I$ and $U^c(c_1(\theta)) \geq \underline{u}$, with demands $c_0(I, \tilde{R}, \underline{u}, \theta)$ and $c_1(I, \tilde{R}, \underline{u}, \theta)$. The dual problem $\min_{c_0, c_1} \left(c_0 + \frac{c_1}{\tilde{R}} \right)$ subject to $U = U^p(c_0, c_1; \theta)$ and $U^c(c_1(\theta)) \geq \underline{u}$ gives associated compensated demands $c_0^c(U, \tilde{R}, \underline{u}, \theta)$ and $c_1^c(U, \tilde{R}, \underline{u}, \theta)$, and by $\varepsilon_{c_1, \tilde{R}}(I, \tilde{R}, \theta)$ the compensated elasticity of c_1 to the after tax interest rate $\tilde{R}$.

.

The proof of Proposition 9 follows exactly the same steps as that of Proposition 6. Indeed, this follows from the fact that the Slutsky relations between uncompensated and compensated demands are still valid $c^c_{1,\tilde{R}} = -\frac{c_1}{\tilde{R}^2} c_{1,I} + c_{1,\tilde{R}}$.[9]

[9]Indeed using a variation of dI and $d\tilde{R}$ such that $V_I \frac{dI}{d\tilde{R}} + V_{\tilde{R}} = 0$, we get $c^c_{1,\tilde{R}} = -\frac{V_{\tilde{R}}}{V_I} c_{1,I} + c_{1,\tilde{R}}$, with $\frac{V_{\tilde{R}}}{V_I} = \frac{c_1}{\tilde{R}^2}$.