Bequest Taxation and Political Constraints

1 Bequest Taxation

1.1 Motivation

• Key: ex post versus ex ante perspective
  
• ex post (children): inheritances are pure luck (children do nothing to deserve their parents), high bequest taxes to level the playing field
  
• ex ante (parents): two individuals with identical lifetime earnings, but one of them consumes all of it, the other saves and prefers to leave most to his children and grandchildren. Who should pay higher lifetime taxes? Perhaps reason for low bequest taxes.

1.2 Farhi/Werning (2010)

• start with 2 generations: parents and children
  
• 2 periods \( t = 0, 1 \): parents work and consume in \( t = 0 \), leave bequests, children consume only in \( t = 1 \)
  
• ex ante heterogeneity \( \theta \) for parents
  
• parents’ preferences include altruism towards their children:
    
    \[
    v_0 = u(c_0(\theta)) - h(y(\theta)/\theta) + \beta v_1(c_1(\theta))
    \]
    
    I.e., importantly, no heterogeneity in altruism (only in productivity)
  
• children’s preferences: \( v_1(c_1(\theta)) = u(c_1(\theta)) \)
  
• linear technology, so aggregate resource constraint
    
    \[
    \int c_0(\theta)dF(\theta) + \frac{1}{R} \int c_1(\theta)dF(\theta) \leq \int y(\theta)dF(\theta) + e_0 + \frac{1}{R} e_1 \tag{1}
    \]
• parents’ incentive constraints

\[ u(c_0(\theta)) - h(y(\theta)/\theta) + \beta u(c_1(\theta)) \geq u(c_0(\theta')) - h(y(\theta')/\theta) + \beta u(c_1(\theta')) \forall \theta, \theta' \quad (2) \]

• utilitarian welfare within generations:

\[ V_t = \int v_t(\theta) dF(\theta), \quad t = 0, 1 \]

• Pareto problem

\[
\begin{align*}
\max_{c_0(\theta), c_1(\theta), y(\theta)} & \quad V_0 \\
\text{s.t.} & \quad V_1 \geq V_t, \quad (1) \text{ and } (2)
\end{align*}
\]

I.e. we can vary the perspective between parents and children

• dual

\[
\begin{align*}
\min_{c_0(\theta), c_1(\theta), y(\theta)} & \quad \int \left( c_0(\theta) + \frac{1}{R} c_1(\theta) - y(\theta) \right) dF(\theta) \\
\text{s.t.} & \quad \int (u(c_0(\theta)) + \beta u(c_1(\theta)) - h(y(\theta)/\theta)) dF(\theta) \geq V_0, \quad (3) \\
& \quad \int u(c_1(\theta)) dF(\theta) \geq V_1, \quad (4)
\end{align*}
\]

and (2)

• multiplier \( \eta \geq 0 \) on the constraint (4) to form Lagrangian

\[
\begin{align*}
\min_{c_0(\theta), c_1(\theta), y(\theta)} & \quad \int \left( c_0(\theta) + \frac{1}{R} c_1(\theta) - y(\theta) \right) dF(\theta) - \eta \int u(c_1(\theta)) dF(\theta) \\
\text{s.t.} & \quad (2) \text{ and } (3)
\end{align*}
\]

• in both constraints, \( c_0(\theta) \) and \( c_1(\theta) \) enter through the total consumption utility \( U(\theta) \equiv u(c_0(\theta)) + \beta u(c_1(\theta)) \). Hence, any solution must solve the subproblem of minimizing (5) s.t. \( u(c_0(\theta)) + \beta u(c_1(\theta)) = U(\theta) \forall \theta \) (inverse Euler variation).

• FOCs

\[
\begin{align*}
[c_0(\theta)] & \quad 1 = \lambda(\theta) u'(c_0(\theta)) \\
[c_1(\theta)] & \quad \frac{1}{R} - \eta u'(c_1(\theta)) = \lambda(\theta) \beta u'(c_1(\theta))
\end{align*}
\]
• combining
\[ u'(c_0(\theta)) = \beta R \left( 1 + \frac{\eta}{\beta} u'(c_0(\theta)) \right) u'(c_1(\theta)) \]  
(6)

• hence, the marginal estate tax is
\[ \tau(\theta) = -\frac{\eta}{\beta} u'(c_0(\theta)) < 0 \]
i.e. in fact the optimum always involves a subsidy on bequests

• moreover, the (negative) estate tax is **progressive** since \( c_0(\theta) \) is increasing in \( \theta \) and hence \( \tau(\theta) \) is increasing in \( \theta \) (i.e. the bequest subsidy is decreasing in \( \theta \))

• benchmark: suppose we put no separate welfare weight on children, but only value them through their parents’ preferences and altruism. I.e. we drop constraint (4) and hence \( \eta = 0 \). Then the optimum involves no estate tax: \( \tau(\theta) = 0 \) for all \( \theta \). This is just an application of Atkinson/Stiglitz (1976).

• E.g. with CRRA preferences, \( u'(c) = c^{-\sigma} \) and so (6) implies
\[ c_1(\theta) = (\beta R)^{1/\sigma} c_0(\theta), \]
i.e. let children’s consumption vary proportionally to their parents’ consumption

• but there is no reason to vary children’s consumption in period 1 except through the fact that it helps to provide incentives to their parents. Hence, effectively we are exploiting the children here just in order to make their parents work more.

• when we value children separately, then \( \eta > 0 \) and the progressive estate tax will lead to some mean reversion in the children’s consumption, so there will be less inequality in period 1

• in this case, we can also think of \( c_1 \) as consumption chosen by the parent, which exerts a positive externality on the child (the effect on social welfare is not only through \( \beta u(c_1) \), which is internalized by the parent, but also through \( \eta u(c_1) \) from the child’s utility). The Pigouvian tax (here, subsidy) \( \tau(\theta) \) on bequests makes the parents internalize this positive externality.

• The externality is larger for low-\( \theta \) individuals, hence the decreasing marginal subsidies (increasing, progressive marginal tax). Indeed, estate taxes belong to the most
progressive taxes in most countries (typically large exemption amounts, but significant marginal tax rates for large bequests).

- How to think of the bequest subsidy in practice? E.g. education subsidies. Farhi/Werning (2010) also show that it can be implemented with a *debt constraint* on the parents if the children’s welfare function is Rawlsian. I.e. a constraint that parents cannot leave debt to their children, as observed in most countries. If binding, the debt constraint will make the parents borrowing constrained, inducing a wedge in their Euler equation such that

\[ u'(c_0(\theta)) > \beta Ru'(c_1(\theta)) \]

and hence consistent with (6).

### 1.3 Piketty/Saez (2013)

- key point: there is heterogeneity not only in labor skill but also in altruism
- very different approach: linear labor and bequest taxes only, sufficient statistics formula, steady state welfare in infinite horizon
- general altruistic (“warm glow”) preferences \( V^{ti}(c, b, l) \) with \( b = Rb_{t+1i}(1 - \tau_{Bl+1}) \) and \( t \) indexes the generation, \( i \) the individual (each generation lives for one period only)
- note: allows for general preference heterogeneity in terms of bequests (altruism)
- given \( \tau_L, \tau_B \), individuals solve

\[
\max_{l_{ti}, c_{ti}, b_{t+1i}} \quad V^{ti}(c_{ti}, Rb_{t+1i}(1 - \tau_{Bl+1}), l_{ti})
\]

s.t.

\[ c_{ti} + b_{t+1i} = Rb_{ti}(1 - \tau_{Bl}) + w_{ti}l_{ti}(1 - \tau_{Lt}) + E_t, \]

where \( E_t \) is total tax revenue

- FOC

\[ V^{ti}_{c} = R(1 - \tau_B)V^{ti}_{b} \]

(7)

- \( w_{ti} \) allows for general skill heterogeneity
• general Pareto weights \( \psi_{ti} \). Hence, in a steady state, the government solves
\[
\max_{\tau_L, \tau_B} \int \psi_{ti} V^{ti}(R b_{ti}(1 - \tau_B) + w_{ti} l_{ti}(1 - \tau_L) + E - b_{t+1i}, R b_{t+1i}(1 - \tau_B), l_{ti}) di
\]
• keeping \( E \) fixed, \( \tau_L \) and \( \tau_B \) adjust to satisfy the government budget constraint
\[
E = \tau_B R b_t + \tau_L y_t
\]
with \( y_{ti} = w_{ti} l_{ti} \) and \( b_t \) and \( y_t \) denote aggregates
• elasticities
\[
e_B = \left. \frac{d b_t}{d(1 - \tau_B)} \frac{1 - \tau_B}{b_t} \right|_E (\text{i.e. here } \tau_L \text{ adjusts}) \tag{8}
\]
\[
e_L = \left. \frac{d y_t}{d(1 - \tau_L)} \frac{1 - \tau_L}{y_t} \right|_E (\text{i.e. here } \tau_B \text{ adjusts}) \tag{9}
\]
These elasticities can be interpreted as long run policy elasticities that measure the change in the aggregate bequest flow and labor income with respect to budget neutral changes in \( (\tau_L, \tau_B) \) (i.e. they incorporate own- and cross-tax effects)
• social marginal welfare weight
\[
g_{ti} = \psi_{ti} \frac{V^{ti}_{ci}}{\int \psi_{ij} V^{ji}_{ci} dj} \tag{10}
\]
• a budget balanced change \( d\tau_B \) requires \( d\tau_L \) such that
\[
R b_t d\tau_B + \tau_B R d b_t = -d\tau_L y_t - \tau_L d y_t
\]
• substituting (8) and (9) yields
\[
R b_t d\tau_B \left( 1 - \frac{\tau_B}{1 - \tau_B} e_B \right) = -d\tau_L y_t \left( 1 - \frac{\tau_L}{1 - \tau_L} e_L \right) \tag{11}
\]
• the welfare effect of this change is (using the envelope theorem)
\[
dSWF = \int \left( \psi_{ti} V^{ti}_{ci}(R d b_{ti}(1 - \tau_B) - R b_{ti} d\tau_B - y_{ti} d\tau_L + \psi_{ti} V^{ti}_{bi}(-R b_{t+1i} d\tau_B)) \right) di
\]
since the indirect effects through \( l_{ti}, b_{t+1i} \) vanish (but \( b_{ti} \) is taken as given)
• at the optimum, this has to be zero, so using (7) and (10),

\[
0 = \int g_{tt} \left( Rdb_{ti}(1 - \tau_B) - Rb_{ti}d\tau_B - y_{ti}d\tau_L - b_{t+1i}d\tau_B \frac{1}{1 - \tau_B} \right) di
\]

\[
= \int g_{tt} \left( Rdb_{ti}(1 - \tau_B) - Rb_{ti}d\tau_B + Rb_{ti}d\tau_B \frac{1 - \frac{\tau_B}{1 - \tau_B} \hat{e}_B y_{ti}}{1 - \frac{\tau_B}{1 - \tau_B} \hat{e}_L y_t} - b_{t+1i}d\tau_B \frac{1}{1 - \tau_B} \right) di
\]

\[
= \int g_{tt} \left( -Rb_{ti}d\tau_B(1 + e_{Bti}) + Rb_{ti}d\tau_B \frac{1}{1 - \frac{\tau_B}{1 - \tau_B} \hat{e}_B y_{ti}} - b_{t+1i}d\tau_B \frac{1}{1 - \tau_B} \right) di
\]

where the second step uses (11) and the third (8)

• dividing through by \( Rb_{ti}d\tau_B \) yields

\[
0 = -\int g_{tt} \frac{b_{ti}}{b_{t}} (1 + e_{Bti}) di + \frac{1 - \hat{e}_B}{1 - \frac{\tau_B}{1 - \tau_B} \hat{e}_L} \int g_{tt} \frac{y_{ti}}{y_t} di - \int g_{tt} \frac{b_{t+1i}}{b_{t}} di \frac{1}{R(1 - \tau_B)}
\]

\[
= -\int g_{tt} \frac{b_{ti}}{b_{t}} di - \int g_{tt} \frac{b_{ti}}{b_{t}} e_{Bti} di + \frac{1 - \hat{e}_B}{1 - \frac{\tau_B}{1 - \tau_B} \hat{e}_L} \frac{\overline{b}^{left}}{R(1 - \tau_B)}
\]

\[
= -\overline{b}^{received}(1 + \hat{e}_B) + \frac{1 - \hat{e}_B}{1 - \frac{\tau_B}{1 - \tau_B} \hat{e}_L} \frac{\overline{b}^{left}}{R(1 - \tau_B)}
\]

where we defined

\[
\overline{b}^{left} \equiv \int g_{tt} \frac{b_{t+1i}}{b_{t}} di,
\]

\[
\overline{b}^{received} \equiv \int g_{tt} \frac{b_{ti}}{b_{t}} di,
\]

and

\[
\overline{y} = \int g_{tt} \frac{y_{ti}}{y_t} di.
\]

These are the ratios of the welfare weighted averages of bequests left, bequests received, and labor income relative to their unweighted average (and would therefore be below one if the variable is lower for those with high social marginal welfare weight). Moreover,

\[
\hat{e}_B = \int g_{tt} b_{ti} e_{Bti} di / \int g_{tt} b_{ti} di
\]

is the average bequest elasticity weighted by \( g_{tt} b_{ti} \).

• For simplicity, let us focus on the meritocratic case where we put all the welfare weight on zero bequest receivers, e.g. because we believe that any inequality for which individuals themselves are not responsible, such as from bequests, should be
equalized (in contrast to inequality from labor income). Then $g_t = 0$ for all positive bequest receivers and so $b_{\text{received}} = 0$. Solving for $\tau_B$ yields

$$
\tau_B = \frac{1 - \left(1 - \frac{e_L \tau_L}{1 - \tau_L}\right) b_{\text{left}}}{1 + e_B}, \tag{12}
$$

where $b_{\text{left}}$ and $\overline{y}$ are now simply the ratios of the average bequests left and the labor income of zero-receivers relative to the population averages.

- if zero bequest receivers leave much smaller bequests than average, then $b_{\text{left}} \rightarrow 0$ and thus $\tau_B = \frac{1}{1 + e_B}$, which is just the revenue maximizing rate.

- $\tau_B$ is increasing in $e_L$ as more elastic labor makes it desirable to shift taxation to bequests rather than labor. It is decreasing in $e_B$ by the same argument.

- in the inelastic case with $e_L = e_B = 0$ and $R = 1$, it simplifies to $\tau_B = 1 - \frac{b_{\text{left}}}{\overline{y}_L}$. i.e. it only depends on distributional parameters, namely the relative position of zero bequest receivers in the distribution of bequests left as opposed to labor income. E.g. if the zero bequest receivers expect to leave bequests that are only 1/10 of the average bequests but they expect to earn the same average labor income, then $b_{\text{left}} / \overline{y}_L = 0.1$ and so $\tau_B = 90\%$.

- On the other hand, sticking to the case with $e_L = 0$ and $R = 1$, if $b_{\text{left}} > \overline{y}_L$, we have $\tau_B < 0$, so negative bequest taxes as in Farhi/Werning (2010) are still possible. But it now all depends on the welfare weights as well as distributional statistics, and the latter can be informed by data. → calibrations in the paper

## 2 Political Economy of Inequality

### 2.1 Lack of Commitment

- have already discussed time inconsistency problem with capital taxation
- but could also emerge with labor taxation
- deviation from promised plan due to
  - ex post Pareto improvement (particularly strong incentive to deviate)
- gains from redistribution

- horizon
  - finite: bad outcomes
  - infinite: reputation mechanisms (trigger strategies), better outcomes

- concepts
  - equilibrium without commitment (more complicated)
  - at a given date, if we could commit from now on, would there be an incentive to deviate from the original plan? (easier)

- $U_{\text{no commitment}} \leq U_{\text{commitment}}$

2.2 Farhi/Sleet/Werning/Yeltekin (2012)

- back to 2 period model $t = 0, 1$

- ex ante heterogeneity $\theta$

- individuals consume and work in $t = 0$, only consume in $t = 1$

- preferences
  $$u(c_0(\theta)) - h(y(\theta)/\theta) + \beta u(c_1(\theta))$$

- linear technology, so aggregate resource constraints
  $$\int c_0(\theta)dF(\theta) + K_1 \leq \int y(\theta)dF(\theta) + RK_0$$
  $$\int c_1(\theta)dF(\theta) \leq RK_1$$

- combine to intertemporal resource constraint
  $$\int c_0(\theta)dF(\theta) + \frac{1}{R} \int c_1(\theta)dF(\theta) \leq \int y(\theta)dF(\theta) + RK_0$$ (13)

- incentive constraints
  $$u(c_0(\theta)) - h(y(\theta)/\theta) + \beta u(c_1(\theta)) \geq u(c_0(\theta')) - h(y(\theta')/\theta) + \beta u(c_1(\theta')) \forall \theta, \theta'$$ (14)
• consider utilitarian government with full commitment

• in period 0, propose

\[
\max_{c_0(\theta), c_1(\theta), y(\theta)} \int (u(c_0(\theta)) + \beta u(c_1(\theta)) - h(y(\theta)/\theta)) dF(\theta)
\]

s.t. (13) and (14)

• preferences over consumption in both periods \((c_0, c_1)\) are separable from work effort, so Atkinson/Stiglitz (1976) applies: no capital taxation

• now consider lack of commitment. In particular, suppose in period 1, the government can reform the tax system at some resource cost \(\kappa\)

• faces resource constraint

\[
\int c_1(\theta)dF(\theta) \leq RK_1 - \kappa \tag{15}
\]

• since it is utilitarian, it maximizes in period 1

\[
\max_{c_1(\theta)} \int u(c_1(\theta))dF(\theta)
\]

s.t. (15)

• idea: in \(t = 0\), promise to use \(c_1\) in addition to \(c_0\) to provide additional incentives for effort in \(t = 0\). But then once we are in \(t = 1\), effort is sunk so this is not optimal anymore.

• if a reform takes place, consumption will be equalized across all types and thus

\[
c_1(\theta) = RK_1 - \kappa \forall \theta
\]

• lack of commitment arises here due to gains from redistribution rather than an ex post Pareto improvement

• it is always in the interest of the government to propose a policy in period \(t = 0\) that will not be reformed in \(t = 1\). Otherwise, the government could have done better by offering an allocation that offers the same constant allocation for consumption in \(t = 1\), which would have saved the fixed cost \(\kappa\)
• reform can be avoided if
\[ \int u(c_1(\theta))dF(\theta) \geq u(RK_1 - \kappa) \]

• \( \kappa = \infty \rightarrow \) full commitment, \( \kappa = 0 \rightarrow \) no commitment, intermediate cases with limited commitment (\( \kappa \) will be endogenized with infinite horizon)

• planning problem in \( t = 0 \) with limited commitment becomes
\[
\max_{c_0(\theta),c_1(\theta),y(\theta)} \int (u(c_0(\theta)) + \beta u(c_1(\theta)) - h(y(\theta)/\theta)) dF(\theta)
\]
subject to (13) and (14) and the credibility constraint
\[ \int u(c_1(\theta))dF(\theta) \geq u \left( \int c_1(\theta)dF(\theta) - \kappa \right), \quad (16) \]
which rules out a reform in \( t = 1 \)

• dual
\[
\min_{c_0(\theta),c_1(\theta),y(\theta)} \int \left( c_0(\theta) + \frac{1}{R}c_1(\theta) - y(\theta) \right) dF(\theta)
\]
s.t.
\[ \int (u(c_0(\theta)) + \beta u(c_1(\theta)) - h(y(\theta)/\theta)) dF(\theta) \geq V, \quad (17) \]
(14) and (16)

• multiplier \( \eta \geq 0 \) on the credibility constraint (16) to form Lagrangian
\[
\min_{c_0(\theta),c_1(\theta),y(\theta)} \int \left( c_0(\theta) + \frac{1}{R}c_1(\theta) - \eta u(c_1(\theta)) - y(\theta) \right) dF(\theta) + \eta \left( \int c_1(\theta)dF(\theta) - \kappa \right)
\]
s.t. (14) and (17)

• in both constraints, \( c_0(\theta) \) and \( c_1(\theta) \) enter through the total consumption utility \( U(\theta) = u(c_0(\theta)) + \beta u(c_1(\theta)) \). Hence, any solution must solve the subproblem of minimizing (18) s.t. \( u(c_0(\theta)) + \beta u(c_1(\theta)) = U(\theta) \forall \theta \) (inverse Euler variation).

• FOCs
\[
[c_0(\theta)]
\]
\[ 1 = \lambda(\theta)u'(c_0(\theta)) \]
\[ [c_1(\theta)] \]

\[
\frac{1}{R} + \eta \left( u'(RK_1 - \kappa) - u'(c_1(\theta)) \right) = \lambda(\theta) \beta u'(c_1(\theta))
\]

- combining

\[
u'(c_0(\theta)) = \beta R \frac{1}{1 + R \eta \left( u'(RK_1 - \kappa) - u'(c_1(\theta)) \right)} u'(c_1(\theta))
\]

- hence, the marginal tax rate on capital satisfies

\[
1 - T'_k(Rk_1(\theta)) = \frac{1}{1 + R \eta \left( u'(RK_1 - \kappa) - u'(c_1(\theta)) \right)}
\]

- progressive capital taxation since \( T'_k(Rk_1(\theta)) \) is increasing in \( c_1(\theta) \)

- in fact, \( T'_k > 0 \) at the top since

\[
RK_1 - \kappa = \int c_1(\theta) dF(\theta) - \kappa < \max_{\theta} c_1(\theta)
\]

and \( T'_k < 0 \) at the bottom since

\[
\min_{\theta} c_1(\theta) < RK_1 - \kappa
\]

whenever the credibility constraint is binding (since otherwise \( c_1(\theta) > RK_1 - \kappa \) for all \( \theta \) and hence \( \int u(c_1(\theta)) dF(\theta) > u(RK_1 - \kappa) \))

- idea: \( u'(RK_1 - \kappa) - u'(c_1(\theta)) \) determines whether an additional unit of capital saved by type \( \theta \) tightens or loosens the credibility constraint. Increasing \( k_1(\theta) \) raises the LHS of (16) since it increases individual consumption, but it also increases the RHS, the value of reform. For high \( \theta \) agents, the increase in the LHS is smaller than the increase in the RHS due to their low marginal utility, and vice versa for low \( \theta \) agents. Thus, saving by high \( \theta \) agents tightens the credibility constraint, whereas it relaxes it for low \( \theta \) agents. The optimal progressive capital tax makes them internalize this effect of their savings behavior on the future commitment problem.

- paper generalizes to infinite horizon, can endogenize \( \kappa \) as repetitional cost

### 2.3 Scheuer/Wolitzky (2016)

- Allow for general reform threats to the capital tax schedule in second period
• Assume that the government must maintain a coalition that supports the status quo and is large enough to block reform: At least share $\alpha \in (0, 1)$ must prefer the status quo to the reform, otherwise reform will occur in $t = 1$.

• Reform cost $\kappa$ no longer needed

• New no-reform constraint (in the case of an equalizing reform)

$$\int 1 [u(c_1(\theta)) \geq u(RK_1)] dF(\theta) \geq \alpha,$$

which effectively requires that an individual at the $1 - \alpha$-quantile of the status quo after-tax wealth distribution $\{c_1(\theta)\}$ in period 1 must have at least the mean wealth $RK_1$.

• What is the optimal coalition for the government to build in order to make its policy sustainable? How does this shape tax policy?

• Show that this generally leads to U-shaped marginal capital taxes when the reform threat is progressive (i.e. more equalizing relative to the status quo). Intuition:

  – The rich are safe supporters of the status quo, the poor are safe supporters of the reform, the middle class is closest to indifferent and therefore the most responsive

  – Subsidize savings for the middle class to make them just indifferent between status quo and reform in $t = 2$. Do this to ensure sufficient popular support for the status quo.

  – Poor and rich are not responsive in terms of their political support, but taxing their savings has the advantage of reducing the capital stock in $t = 1$ that is available for redistribution under the reform.

  – Results in positive capital taxes for rich and poor and capital subsidies for the middle class

• Hence, the optimal coalition to make dynamic tax policy is the rich and the middle class against the poor.

• Examples consistent with U-shaped marginal capital taxes: Tax favored retirement savings, mortgage interest deduction, subsidies to savings for college education, phase-out of transfer programs to the poor that depend on wealth (which leads to high effective marginal tax rates at the bottom)
• Related to *Director’s Law* (Stigler, 1970): redistribution from rich to middle class rather than very poor.

• Historical examples: introduction of pensions in Bismarck’s Germany in late 19th century in response to threat of socialist movement, designed to ensure stability of the established order; introduction of Social Security in the US