Bequest Taxation

1 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:

\[
v_0 = u(c_0(\theta)) - h(y(\theta)/\theta) + \beta v_1(c_1(\theta))
\]

• 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

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

• utilitarian welfare within generations:

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

• Pareto problem

\[
\max_{c_0(\theta), c_1(\theta), y(\theta)} V_0 \ \text{s.t.} \ V_0 \geq V_1, \ (1) \text{ and } (2)
\]

• dual

\[
\min_{c_0(\theta), c_1(\theta), y(\theta)} \int \left(c_0(\theta) + \frac{1}{R} c_1(\theta) - y(\theta)\right)dF(\theta)
\]
\[
\int \left( u(c_0(\theta)) + \beta u(c_1(\theta)) - h(y(\theta)/\theta) \right) dF(\theta) \geq V_0, \tag{3}
\]
\[
\int u(c_1(\theta)) dF(\theta) \geq V_1, \tag{4}
\]

and (2)

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

\[
\min_{c_0(\theta), c_1(\theta), y(\theta)} \int \left( c_0(\theta) + \frac{1}{R} c_1(\theta) - y(\theta) \right) dF(\theta) - \eta \int u(c_1(\theta)) dF(\theta) \tag{5}
\]
s.t. (2) and (3)

- 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

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

\[
[c_1(\theta)]
\]
\[
\frac{1}{R} - \eta u'(c_1(\theta)) = \lambda(\theta) \beta u'(c_1(\theta))
\]

- combining

\[
u'(c_0(\theta)) = \beta R \left( 1 + \frac{\eta}{\beta} u'(c_0(\theta)) \right) u'(c_1(\theta)) \tag{6}\]

- 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

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

and hence consistent with (6).

2 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 preferences $V^{ti}(c, b, l)$ with $b = Rb_{t+1}(1 - \tau_{B_{t+1}})$ and $t$ indexes the generation, $i$ the individual

• note: allows for general preference heterogeneity in terms of bequests (altruism)

• given $\tau_L, \tau_B$, individuals solve

$$\max_{l_{ti}, c_{ti}, b_{t+1i}} V^{ti}(c_{ti}, Rb_{t+1}(1 - \tau_{B_{t+1}}), l_{ti})$$

s.t.

$$c_{ti} + b_{t+1i} = Rb_{ti}(1 - \tau_{B_{ti}}) + w_{ti}l_{ti}(1 - \tau_{L_{ti}}) + E_t,$$

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

$$\max_{\tau_L, \tau_B} \int \psi_{ti} V^{ti}(Rb_{ti}(1 - \tau_B) + w_{ti}l_{ti}(1 - \tau_L) + E - b_{t+1i}, Rb_{t+1}(1 - \tau_B), l_{ti})di$$

• keeping $E$ fixed, $\tau_L$ and $\tau_B$ adjust to satisfy the government budget constraint

$$E = \tau_B Rb_t + \tau_L y_t$$

with $y_{ti} = w_{ti}l_{ti}$

• elasticities

$$e_B = \frac{db_t}{d(1 - \tau_B)} \frac{1 - \tau_B}{b_t} \bigg|_E (i.e. \ here \ \tau_L \ \text{adjusts}) (8)$$

$$e_L = \frac{dy_t}{d(1 - \tau_L)} \frac{1 - \tau_L}{y_t} \bigg|_E (i.e. \ here \ \tau_B \ \text{adjusts}) (9)$$

• social marginal welfare weight

$$g_{ti} = \psi_{ti} \frac{V^{ti}_c}{\int \psi_{tj} V^{ti}_c dj} (10)$$
• a budget balanced change \( d\tau_B \) requires \( d\tau_L \) such that

\[
Rb_t d\tau_B + \tau_B Rdb_t = -d\tau_L y_t - \tau_L dy_t
\]

• substituting (8) and (9) yields

\[
Rb_t d\tau_B \left( 1 - \frac{\tau_B}{1 - \tau_B} e_B \right) = -d\tau_L y_t \left( 1 - \frac{\tau_L}{1 - \tau_L} e_L \right)
\]

(11)

• the welfare effect of this change is (using the envelope theorem)

\[
dSWF = \int \left( \psi_{ti} V^{ti}_c (Rdb_t (1 - \tau_B) - Rb_t d\tau_B - y_{ti} d\tau_L) + \psi_{ti} V^{ti}_b (-Rb_{t+1} d\tau_B) \right) di
\]

since the indirect effects through \( l_{ti}, b_{t+1} \) vanish (but \( b_{ti} \) is taken as given)

• at the optimum, this has to be zero, so using (7) and (10),

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

\[
= \int g_{ti} \left( Rdb_t (1 - \tau_B) - Rb_t d\tau_B + Rb_t d\tau_B \frac{1 - \tau_B}{1 - \tau_B} e_B y_{ti} - b_{t+1} d\tau_B \frac{1}{1 - \tau_B} \right) di
\]

\[
= \int g_{ti} \left( -Rb_t d\tau_B (1 + e_{Bti}) + Rb_t d\tau_B \frac{1 - \tau_B}{1 - \tau_B} e_B y_{ti} - b_{t+1} d\tau_B \frac{1}{1 - \tau_B} \right) di
\]

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

• dividing through by \( Rb_t d\tau_B \) yields

\[
0 = - \int g_{ti} \frac{b_{ti}}{b_t} (1 + e_{Bti}) di + \frac{1 - e_B \frac{\tau_B}{1 - \tau_B}}{1 - e_L \frac{\tau_L}{1 - \tau_L}} \int g_{ti} \frac{y_{ti}}{y_t} di - \int g_{ti} \frac{b_{t+1} di}{b_t} \frac{1}{R(1 - \tau_B)}
\]

\[
= - \int g_{ti} \frac{b_{ti}}{b_t} di - \int g_{ti} \frac{b_{ti}}{b_t} e_{Bti} di + \frac{1 - e_B \frac{\tau_B}{1 - \tau_B}}{1 - e_L \frac{\tau_L}{1 - \tau_L}} \int g_{ti} \frac{y_{ti}}{y_t} di - \frac{b^{left}}{R(1 - \tau_B)}
\]

\[
= - \bar{b}^{received} (1 + \bar{e}_B) + \frac{1 - e_B \frac{\tau_B}{1 - \tau_B}}{1 - e_L \frac{\tau_L}{1 - \tau_L}} \frac{\bar{b}^{left}}{R(1 - \tau_B)}
\]

where we defined

\[
\bar{b}^{left} \equiv \int g_{ti} \frac{b_{ti}}{b_t} di,
\]

\[
\bar{b}^{received} \equiv \int g_{ti} \frac{b_{ti}}{b_t} di
\]
\[ \hat{e}_B = \int g_{ti} \frac{b_{ti} e_{Bti}}{b_t} di \int g_{ti} \frac{b_{ti}}{b_t} di. \]

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

- 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 \( g_{ti} = 0 \) for all positive bequest receivers and so \( \bar{b}^{\text{received}} = 0 \). Solving for \( \tau_B \) yields

\[
\tau_B = \frac{1 - \left(1 - \frac{e_L r}{1 - e_B}\right) \bar{b}^{\text{left}}}{1 + e_B},
\]  

(12)

where \( \bar{b}^{\text{left}} \) and \( \bar{y}_L \) are now simply the ratios of the average bequests left and the labor income of zero-receivers relative to the population averages.

- if zero bequests receivers leave much smaller bequests than average, then \( \bar{b}^{\text{left}} \to 0 \) and thus \( \tau_B = 1/(1 + e_B) \), 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 \( \tau_B = 1 - \bar{b}^{\text{left}}/\bar{y}_L \). 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 \( \bar{b}^{\text{left}}/\bar{y}_L = 0.1 \) and so \( \tau_B = 90\% \).

- On the other hand, sticking to the case with \( e_L = 0 \) and \( R = 1 \), if \( \bar{b}^{\text{left}} > \bar{y}_L \), 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.