Intergenerational Disagreement and Optimal Taxation of Parental Transfers*

Nicola Pavoni† and Hakki Yazici‡

July 13, 2015

Abstract

We study optimal taxation of bequests and inter vivos transfers in a model where altruistic parents and their offspring disagree on intertemporal trade-offs. We show that laissez-faire equilibrium is Pareto inefficient, and whenever offspring are impatient from their parents’ perspective, optimal policy involves a positive tax on parental transfers. Cautioned by the technical complications present in this class of models, our normative prescriptions do not rely on the assumption of differentiability of the agents’ policy functions.

JEL classification: E21, H21, D03, D91.

Keywords: Parental transfer taxation, intergenerational disagreement, altruism.

---

*We thank Michelle Tertilt, four anonymous referees, Per Krusell, Marco Ottaviani, Vincenzo Quadrini, seminar participants at Bogazici University, Goethe University in Frankfurt, Istanbul Technical University, IIES, the SED meetings in Ghent, University of Alicante, University of Bologna, University of Oxford, and the 8th CSEF-IGIER Symposium on Economics and Institutions for their comments and suggestions. Pavoni gratefully acknowledges financial support from the European Research Council, Starting Grant #210908. Yazici gratefully acknowledges financial support from the European Community Framework Programme through the Marie Curie International Reintegration Grant #268457.

†Department of Economics, Bocconi University, via Roentgen 1, 20136 Milan; IFS, IGIER, and CEPR.

‡Sabanci University, FASS, Orta Mahalle, Universite Caddesi No:27 Tuzla, 34956 Istanbul.
1 Introduction

Transfer taxation has long been a highly controversial issue. Except for a brief period of time, the U.S. government has maintained a positive bequest tax ever since it was first introduced in 1916. The tax was eliminated in 2010 and reintroduced in 2011. The nature of the policy debate, however, seems to go in only one direction: in all developed countries, we indeed observe either zero or positive taxation of parental transfers.\footnote{Estate tax rates in the United States have been varying since the time they were introduced in 1916. For detailed information on the evolution of estate taxes in the United States, see \textit{Jacobson, Raub, and Johnson} (2007). Recently, in 2013, after the expiration of the Bush tax cuts, the estate taxes increased from 35\% to 55\%. In the United Kingdom, the inheritance tax rate has held steady at 40\%. Both countries have an exemption level below which wealth transfers go untaxed. In the United States, this level has decreased from 5 million US dollars to 1 million US dollars in 2013. In the United Kingdom, this amount is 325,000 pounds. For detailed information on the UK inheritance tax system, visit \url{https://www.gov.uk/inheritance-tax}. For an overview on parental transfer taxation in OECD countries, see \textit{Cremer and Pestieau} (2010) and OECD (2008).}

This regularity on observed tax systems around the world contrasts with the lack of a clean theoretical justification for a positive tax on bequests from an efficiency point of view.\footnote{It is, of course, possible to justify taxation of parental transfers based on equality grounds. See, for instance, \textit{Piketty and Saez} (2013).} In virtually all traditional models of bequests, efficiency calls for either zero or a negative tax on parental transfers.\footnote{We discuss these models and their implications for bequest taxation in detail in the related literature section.} Perhaps the most widely analyzed model of bequest taxation is the altruistic model, where the motive for bequests comes from the assumption that parents care about their children’s welfare. A maintained assumption in the altruistic model is that different generations agree about intertemporal trade-offs. The implication of the standard altruistic model for intergenerational wealth transfers - such as bequests and inter vivos - is simple. An altruistic parent knows that the offspring will save according to his optimal plan, which is also optimal from the parent’s point of view. As a result, parents do not have any paternalistic concerns regarding the offspring’s savings choice and the whole dynasty acts as if it is a single individual (e.g., \textit{Bernheim} (1989)). If the society cares only about the parent’s welfare directly, this would imply that parental transfers are socially optimal and should remain undistorted, given that there are no other reasons for taxation such as redistribution or financing government expenditures. Following the same arguments, whenever society attaches direct welfare weight to future generations, parental transfers should actually be subsidized according to the altruistic model with intergenerational agreement.\footnote{See \textit{Kaplow} (1995) and \textit{Farhi and Werning} (2010) for a discussion of optimality of bequest subsidies.}
However, disagreements are common to most, if not all, parent-offspring relationships. Clarke, Preston, Raksin, and Bengtson (1999) studies a wide range of disagreement patterns between older parents and adult children, and develops a typology of disagreement issues. They consider a random sample of 496 parents (average age 62) and 641 children (average age 39) and ask about possible sources of disagreement. More than 70% of the respondents report disagreements (about the same percentage among children and parents). The largest category of responses about conflict (38% among parents and 30% among children) is labeled as Habits and Lifestyle Choices and it includes sexual activity/orientation, living arrangements, quality of life, and allocation of resources and or education. A conclusion the authors reach about this category of conflicts is that intertemporal allocation of resources is a common source of intergenerational disagreement. The following quote summarizes this point:

“There are also conflicts that seem to express a world view common to many in the older generation. This same father writes: ‘He [39] wants all things like his generation of baby boomers, right now - new cars, new houses, vacations - all of it on one income and that a blue collar job income.’ This is echoed elsewhere in another father’s (60) comment over his daughter’s (37) lack of ‘concern for saving something for a rainy day.’ Another father (71) reports that his son’s (37) ‘using credit cards to the limit’ is an area of disagreement.”

In this paper, we analyze optimal taxation of parental transfers in a world where parents and offspring disagree on the intertemporal allocation of resources. We show that there is a genuine efficiency reason for government intervention in the market in the presence of intergenerational disagreement. In particular, we find that, whenever offspring are impatient from parents’ perspective, optimal intervention involves a positive tax on parental transfers.

We study transfer taxation using an intergenerational model in which bequests are motivated by altruism. The key principle for efficient bequest taxation can be fully grasped in the baseline model where agents live for two periods. In the first period, people make consumption-saving decisions, and in the second period, they choose how much to consume and bequeath to their offspring. Then, they are replaced by their offspring who go through the same life cycle. We model disagreement in a way that departs from the standard model minimally: a parent and an offspring agree on everything but how the under social preferences that weigh future generations directly.

5The exact question was: ‘No matter how well two people get along, there are times they disagree or get annoyed about something. In the last few years, what are some things on which you have differed, disagreed, or been disappointed about (even if not openly discussed) with your child (or parent)?’
offspring should allocate his resources between his young and old age.

In this environment, we first characterize laissez-faire market equilibrium. We focus on Markov equilibria. Using this characterization, we prove that equilibrium allocation is Pareto inefficient. The inefficiency stems from the fact that, as long as there is disagreement between parents and offspring, the latter do not fully internalize the consequences of their saving decisions on the former.

Having uncovered an efficiency reason to intervene the market outcome, we next characterize optimal policy. We begin our analysis of optimal policy by focusing on a particular Pareto efficient allocation that we call the ‘Ramsey’ allocation. In this allocation, all the welfare weight is on the initial parent and future generations are valued by society only indirectly via the initial parent’s welfare. The Ramsey allocation is probably the most widely adopted benchmark in the literature. This choice of the benchmark efficient allocation is further motivated by the fact that in the standard model with no disagreements a Ramsey government would find it optimal not to distort equilibrium transfer decisions of parents at all. In this sense, any need for an intervention in the case with intergenerational disagreements will be coming from nowhere else but the existence of disagreements.

We find that, if children are less patient than what their parents want them to be, then parents bequeath too much in laissez-faire equilibrium relative to the Ramsey allocation, and it is optimal for the government to correct their behavior through bequest taxation. The brief intuition for this result is as follows. Due to disagreement, in the laissez-faire equilibrium, the offspring save less than the Ramsey level for their old age, which induces parents to transfer more than the Ramsey level in order to compensate for the offspring’s undersaving behavior. In order to make children save the amount dictated by the Ramsey allocation, the government uses linear subsidies on savings and uses lump-sum taxes to finance these subsidies. However, from the perspective of the parents, who take the saving subsidy as given, their offspring are still undersaving under the new - subsidized - interest rate. In other words, since parents take the lump-sum tax as given, they do not internalize the fact that the subsidy is there to discipline the saving behavior of the next generation and does not actually change the gross return to their offspring’s savings. As a result, the parents still transfer too much to their offspring, and hence, should be taxed. Even though we find the case where children are impatient from parents’ perspective more natural, we also analyze the case in which they act more patiently than what their parents would like,

---

6We cannot rule out the existence of multiple equilibria, but, importantly, all our results are valid for all Markov equilibria.
and find that bequests should be subsidized in that case.

Next, we generalize our results to the whole Pareto frontier by analyzing optimal bequest taxation in the case where society cares directly about future generations. We show that Pareto optimal bequest tax has a nice separable form between an efficiency component and an intergenerational redistribution component. The efficiency component is the same as the Ramsey bequest tax: it is only present when there is intergenerational disagreement and calls for a tax whenever offspring are impatient from parents’ perspective. The second component, also present in models without intergenerational disagreements, represents a subsidy to parental transfers due to intergenerational redistribution coming from a direct weight on future generations. The overall sign of the bequest tax depends on whether the efficiency wedge or the intergenerational redistribution wedge dominates. It is important though to note that - as long as offspring are impatient from parents’ perspective - the efficiency component is positive for any Pareto efficient allocation, and hence, calls for a tax on parental transfers.

Our results extend to the case where agents live for an arbitrary number of periods and parents co-exist with their offspring. The optimal tax on inter vivos transfers obeys the same principles as the bequest tax, and thus, has the same sign. We also study the role of horizontal (cross-sectional) income heterogeneity by embedding our model into an optimal labor income taxation framework à la Mirrlees (1971). We find that while the optimality of a positive bequest tax is still dictated by the presence of intergenerational disagreements, horizontal redistribution motive shapes the curvature of the tax schedule. In particular, when the income process is mean reverting across generations and society puts direct welfare weight on offspring, the optimal bequest tax rate is typically increasing in the amount of bequests. The progressivity of bequest taxes stems from a mechanism similar to that in Farhi and Werning (2010) except that they find that optimality calls for progressive subsidies on bequests and not taxes. This is due to the fact that they do not have intergenerational disagreement which is the key driving force of bequest taxes in our environment.

In addition to its policy implications, this paper also makes a methodological contribution by deriving normative implications of models with intergenerational disagreements in the absence of differentiability assumptions. In the public finance literature, it is customary to compare equilibrium and efficient allocations by defining and measuring wedges which represent discrepancies between individual and social marginal returns to individual decisions. Our results, on the other hand, are presented over three levels of analysis. In addi-
tion to analyzing optimal wedges - which requires the assumption of differentiability - and optimal taxes - which requires further restrictive assumptions implying concavity of the agents’ problems - we provide a non-differentiable analysis of the discrepancies between the equilibrium allocation and the efficient allocation. Deriving normative prescriptions at this level of generality is particularly important for models with disagreements for at least two reasons. First, it is well known that, in general, these models may not have equilibria with differentiable policy functions. Second, even when a differentiable equilibrium exists, models with multiple selves often admit multiple (Markov) equilibria. It is important in such cases to understand whether a policy implication emerges from a general principle or, instead, it is linked to a specific equilibrium (or equilibrium property such as differentiability or linearity of the policy).

In our implementation result, we show that, when we focus our attention to linear Markov equilibria, the parental transfer wedge translates into a result on transfer taxation: efficient allocations can be implemented using linear wealth transfer taxes as long as the government has access to (linear) life-cycle saving taxation to offset offspring’s tendency to undersave. We also show that the linear Markov equilibria assumption is innocuous by proving that, under the constant elasticity of intertemporal substitution utility function (CEIS), such equilibria exist. It is important to note that the linear structure of taxes are not crucial for the optimality of bequest taxes. As long as the saving subsidy applied to offspring leaves their saving policy function strictly monotone in the amount of transfers they receive, parents have a motive to bequeath more than efficient, which implies it is optimal to tax bequests.

Related Literature. This paper is related to three strands of literature. First, it is related to the literature on optimal taxation of bequests and inter vivos transfers. Our contribution here is to provide a novel, pure efficiency argument for taxing parental transfers. In addition to the altruistic model that is already discussed, a widely used model of bequests is the warm-glow (or “joy of giving”) model. In this model too the optimal bequest tax is zero or netative, depending on whether society cares about the offspring directly (e.g., Kopczuk 2010). Another framework considered in the literature is the model with exchange motives for bequests. In this class of game-theoretical models, the normative

---

7Our model with intergenerational disagreements has an analytical structure that is quite similar to models with present-bias problems a la Laibson (1997), and it is well-known that in these models policy functions that describe people’s life-cycle saving behavior may not be differentiable. See Harris and Laibson (2002).

8See Cremer and Pestieau (2010), Kaplow (2001), and Kopczuk (2010) for excellent surveys on the literature on optimal transfer taxation.
predictions crucially depend on the details of the game played between the parents and the offspring (e.g., Laitner (1997)). Finally, we have the accidental bequests model, where taxing (accidental) bequests is non-distortionary. According to this model, bequest taxes are simply a good way to finance positive government expenditures when lump-sum taxes are not available. This model does not imply an optimal positive tax, at least not in the way we define optimality in this paper. Specifically, there is no equilibrium inefficiency to be corrected by taxes on bequests or gifts.

Our paper is also related to the literature on intergenerational disagreements. A seminal paper in this literature is Phelps and Pollak (1968) which analyzes national saving rate in an environment in which each generation lives for a single period and is imperfectly altruistic: the rate at which each generation discounts next generation’s consumption relative to their own consumption is higher than the rate at which they discount consumption across any two subsequent future generations. Under the assumptions that people have CEIS utility functions, returns to capital are linear and there is no depreciation, and focusing only on the equilibrium in which all generations save the same constant fraction of their income at all periods, the paper shows that equilibrium entails lower national saving compared to the Ramsey allocation. Doepke and Zilibotti (2014) analyzes an environment in which parents and children have preference disagreements, and parents can affect offspring’s choices in two ways: by influencing their preferences via education and by imposing direct restrictions on their choice sets. They use this model to explain the variation in parenting styles across industrialized countries and over time. Finally, Doepke and Tertilt (2009) analyzes a model where husbands and wives are imperfectly altruistic towards each other and disagree on the degree of their altruism towards their offspring. The authors use this model to explain the dramatic improvements in the legal rights of married women that occurred before the introduction of female suffrage. We contribute to this growing literature by analyzing policy implications of intergenerational disagreements.

One possible interpretation of our positive model is that offspring agree with their parents regarding how much they should save but they face self-control problems that prevent them from saving the right amount. Under this interpretation, the optimal tax problem is

---

9There is an obvious theoretical assumption - not yet carefully tested empirically - that would justify positive taxation of all sorts of wealth (not only of parental wealth transfers). It is the assumption that wealth concentration generates negative externalities. See Kopczuk (2010) for a discussion of negative wealth externalities.

10Following Strotz (1955), Laibson (1997) applies this framework to individual consumption saving problem over the life cycle under self-control problems and also finds undersaving behavior. See O’Donoghue and Rabin (1999) for more general application of this model to individual decision making.
a paternalistic one in the sense that taxes are used to correct the ‘wrong’ saving behavior of the offspring. This is the focus of a number of recent papers that have explored the implications of self-control problems for optimal taxation. O’Donoghue and Rabin (2003) analyze a model of paternalistic taxation for unhealthy goods. More closely related is Krusell, Kuruscu, and Smith (2010) which analyzes properties of linear taxes on life-cycle savings that implement the Ramsey allocation. Pavoni and Yazici (2015) also focus on optimal Ramsey taxation of life-cycle savings within a quasi-hyperbolic discounting model and allow for self-control problems that change with age. The current paper, on the other hand, does not assume that offspring’s true welfare coincide with that of the parents. The offspring is truly more impatient than what the parent wants him to be. In this environment, we show that, independent of which member society cares about, it is optimal to tax parental transfers. The optimality of bequest taxation does not stem from correcting people’s mistakes (because there are none), but rather from a positive transfer wedge that arises due to intergenerational disagreements. Since the analytical structures of the disagreement and self-control models are very similar, we also provide a methodological contribution to the self-control literature with our non-differential analysis of optimal policy.

The rest paper is organized as follows. Section 2 introduces the main model, and section 3 characterizes the equilibrium bequest behavior of parents in the absence of government intervention. In Section 4, we compare equilibrium and Ramsey bequest behavior and provide a tax implementation of the Ramsey allocation. In Section 5, we provide a number of important generalizations of our result including the Pareto efficient taxation of parental transfers. Section 6 concludes.

2 Model

The economy is populated by a continuum of a unit measure dynasties who live for a countable infinity of periods, \( t = 0, 1, \ldots \), where each agent within a dynasty is active for two periods. In the first period of their lives, agents are young adults and make consumption saving decisions. In the second period, they become parents, decide how much to consume and bequeath, and die. The next period their offspring becomes a young adult and goes through the same life cycle. This is a model of non-overlapping generations. People have

\[11\] Shall we thanks the editor for this interpretation?

\[12\] In Section 5.3, we allow for a longer life cycle and show that our main results regarding bequest taxation are robust to such extension. There, we also model periods in which parents and their offspring are alive
one unit of time endowment that they supply inelastically to the market every period. Parents bequeath because they are altruistic.

The economy begins with an initial parent in period 0. Every subsequent even period is a parenthood period whereas every odd period is a young adulthood period. Consider a parent in some calendar year \( t \). Her preference over dynastic allocation is given by

\[
V_t = u(c_t) + \gamma [u(c_{t+1}) + \delta V_{t+2}],
\]

where \( \delta, \gamma \in (0, 1) \), \( V_t \) represents the dynastic welfare of the parent in period \( t \), \( c_t \) is parental consumption, and \( c_{t+1} \) is the first period consumption of the offspring. The instantaneous utility function, \( u \), has the usual properties: strictly increasing, strictly concave, and twice differentiable, with \( \lim_{c \to 0} u'(c) = +\infty \). The parameter \( \delta \) represents the (long-term) discount factor that the parent thinks the offspring and all the future descendants should save according to between their young adulthood and old adulthood periods. The parameter \( \gamma \) is the altruism factor.

We model disagreement between parents and offspring in a way that departs from the standard dynastic framework in a minimalistic way: namely, the offspring agrees with the parent on everything except for how to allocate his wealth between young and old adulthood. Specifically, the offspring’s preference is given by:

\[
u(c_{t+1}) + \beta \delta V_{t+2}.
\]

In this formulation, as long as \( \beta \neq 1 \), the discount rate that the offspring uses between periods \( t + 1 \) and \( t + 2 \), \( \beta \delta \), is different from what is appropriate from his parent’s perspective, \( \delta \). Parents are sophisticated in the sense that they fully anticipate this discrepancy between own and offspring’s preferences.\(^{13}\)

We do not make an assumption about whether \( \beta \) is smaller or larger than one from the outset. However, both the Clarke, Preston, Raksin, and Bengtson (1999) study and anecdotal evidence seems to suggest that it is more natural to consider the case where \( \beta < 1 \). This is the case in which the offspring are impatient from the parents’ perspective. It is important to note that Pareto inefficiency of laissez-faire equilibrium allocation does together and analyze inter vivos transfer behavior and taxation. We show that the results regarding bequest taxation extend to inter vivos transfers.

\(^{13}\)In Appendix D, we analyze what happens if parents naively believe that offspring agree with them regarding intergenerational resource allocation and show that our main results carry over to this case as well.
not depend on whether the offspring are more or less patient than what the parents would like them to be. We report how the sign of optimal transfer taxes depend on whether $\beta$ is greater or less than unity throughout the paper.

There is an alternative interpretation of the benchmark model. Under this interpretation, even though the offspring’s true preference coincides with that of his parent’s, i.e., he evaluates his welfare according to $u(c_{t+1}) + \delta V_{t+2}$, the offspring faces self-control problems and saves according to $u(c_{t+1}) + \beta \delta V_{t+2}$. This interpretation is in line with the literature on self-control problems in the spirit of [Laibson (1997)]. In the current paper, on the other hand, we assume that the offspring and the parent truly disagree in the sense that $u(c_{t+1}) + \beta \delta V_{t+2}$ represents not only the behavioral preference of the offspring but also his true preference.

Production takes place at the aggregate level according to the function $F(k_t, l_t)$, where $k_t$ and $l_t$ are aggregate levels of capital stock and labor in period $t$, and $F$ is a neoclassical concave production function with the usual properties: $F_1, F_2 > 0$ and $F_{11}, F_{22} \leq 0$. Since each agent supplies one unit of labor inelastically, for all $t$, we have

$$l_t = 1.$$ 

Letting $\theta$ be the depreciation rate, this allows us to write the production function as

$$f(k) = F(k, 1) + (1 - \theta)k.$$ 

Letting $f(k_0)$ be the endowment of the initial parent in period zero, the feasibility for any $t \geq 0$ is

$$c_t + k_{t+1} = f(k_t).$$ 

As evident from the feasibility condition above, we assume there is one representative dynasty, which implies that in any calendar year there is only one age group alive. We could, instead, allow for members of different dynasties to be at different points in their life cycles. Moreover, we could also allow for income heterogeneity by assuming, for instance, that people have different skill levels and that effective labor is given by labor times the skill level, similar to [Mirrlees (1971)]. In the main body of the paper, we abstract from such ‘horizontal’ distribution issues in order to isolate our mechanism. We show in Section 5.2 that the mechanism behind our results is robust to income heterogeneity.
3 Laissez Faire

In this section, we characterize the equilibrium parental transfer behavior. Let $b_{t+1}$ and $b_{t+2}$ denote the bequest made by the parent in period $t$ and the offspring’s saving level in $t+1$, respectively. Let $R_t, w_t$ be the interest rate and the wage rate in period $t$. Let $Q := \{R_t, w_t\}_{t=0}^{\infty}$ be the sequence of prices that decision makers take as given. Finally, let $Q_t := \{R_s, w_s\}_{s=t}^{\infty}$ be continuation of prices from period $t$ onwards.

Define $V(a_t, Q_t)$ as the value of the problem of an agent who is a parent in calendar year $t$ with $a_t := R_t b_t + w_t$ units of wealth and who faces the price sequence $Q_t$. The parent’s problem is given by

$$V(a_t, Q_t) = \max_{b_{t+1} \geq -B(Q_{t+1})} u(c_t) + \gamma [u(c_{t+1}(b_{t+1}, Q_{t+1})) + \delta V(a_{t+2}(b_{t+1}, Q_{t+1}), Q_{t+2})],$$

subject to the budget constraints and the definition of wealth

$$c_t = a_t - b_{t+1},$$
$$c_{t+1}(b_{t+1}, Q_{t+1}) = R_{t+1} b_{t+1} + w_{t+1} - b_{t+2}(b_{t+1}, Q_{t+1}),$$
$$a_{t+2}(b_{t+1}, Q_{t+1}) := R_{t+2} b_{t+2}(b_{t+1}, Q_{t+1}) + w_{t+2},$$

together with the condition defining the policy of the offspring:

$$b_{t+2}(b_{t+1}, Q_{t+1}) = \arg \max_{b_{t+2} \geq -B(Q_{t+2})} u(R_{t+1} b_{t+1} + w_{t+1} - \tilde{b}_{t+2}) + \beta \delta V(R_{t+2} \tilde{b}_{t+2} + w_{t+2}, Q_{t+2}).$$

$B(Q_t)$ is the ‘natural’ (and never binding) borrowing limit defined by requiring consumption to be non-negative at all periods:

$$B(Q_t) := \sum_{s=t}^{\infty} \frac{w_s}{\Pi_{p=t}^{s} R_s}. $$

In equilibrium, prices are given by

$$R_t = f'(k_t),$$
$$w_t = f(k_t) - f'(k_t)k_t,$$

\[14\] To save notation, we indicate the policy as a function. In case there are multiple solutions to the offspring’s problem, $b_{t+2}(\cdot, Q_{t+1})$ should be intended as a selection from the policy correspondence.
and aggregate capital and saving levels satisfy the market clearing condition

\[ k_t = b_t. \]

The parent chooses his bequest level \( b_{t+1} \) taking into account the choice rule of his offspring, \( b_{t+2}(\cdot, Q_{t+1}) \), which describes how offspring’s saving choice changes as a function of parental bequests under a given price sequence. The parent is sophisticated in the sense that he correctly guesses his children’s choice, and that is why he takes (2) into account as a constraint in his problem. Define \( b_{t+1}(b_t, Q_t) \) as the policy function describing parental optimal bequeathing behavior as a function of his period \( t - 1 \) savings and the price sequence.

A Markov equilibrium consists of a sequence of capital levels \( \{k_t\}^\infty_{t=0} \), a sequence of prices \( Q \), value functions \( V(\cdot, Q_t) \) and policy functions \( \{b_{t+1}(\cdot, Q_t), b_{t+2}(\cdot, Q_{t+1})\}^\infty_{t=0,2,4,...} \) such that:

(i) the prices satisfy (3); (ii) the value function and the policies are consistent with the parent’s problem; (iii) markets clear: \( b_t = k_t \) for all \( t \).

Proposition 1 below characterizes equilibrium parental transfer behavior. Proving Proposition 1 would be relatively easier if we could assume the differentiability of policy function, \( b_{t+2}(\cdot, Q_{t+1}) \), in bequests received. However, the dynastic intertemporal resource allocation problem with disagreements across dynasty members implies time-inconsistency in the overall dynastic problem and it is well-known from the self-control literature that in such environments we cannot guarantee even the continuity of the policy functions even when we focus our attention to Markov equilibria.\(^{15}\) To ensure that Proposition 1 describes a general feature of economies with intergenerational disagreements and is not an artifact of differentiability assumptions, we prove it without making any differentiability or continuity assumptions about the value or policy functions.

**Proposition 1.** Suppose \( \beta < 1 \). Then, in equilibrium,

\[ u'(c_t) \geq R_{t+1} \gamma u'(c_{t+1}), \quad (4) \]

with strict inequality whenever the offspring’s optimal saving policy, \( b_{t+2}(\cdot, Q_{t+1}) \), is strictly

\(^{15}\)See Morris and Postlewaite (1997) and Harris and Laibson (2002) for examples of economies with quasi-hyperbolic discounters where policy functions are discontinuous. Krusell and Smith (2003) shows that even when we focus attention to Markov equilibria, it is not possible to rule out the existence of discontinuous equilibria. Notice that the discontinuity in these models is due to a ‘disagreement’ across an agent’s multiple selves whereas in our model the disagreement is across different generations.
monotone in the amount of the bequests received, \( b_{t+1} \).

If \( \beta > 1 \), then
\[
    u'(c_t) \leq R_{t+1} \gamma u'(c_{t+1}),
\]
with strict inequality whenever the offspring’s saving policy is strictly monotone in \( b_{t+1} \).

If \( \beta = 1 \), then
\[
    u'(c_t) = R_{t+1} \gamma u'(c_{t+1}).
\]

Proof. Relegated to Appendix A.1.

We provide an intuition for Proposition 1 only for the \( \beta < 1 \) case, the intuition for the \( \beta > 1 \) case is symmetric. To get a better grasp of what the proposition says, first focus on the case in which the child and parent agree on inter-temporal trade offs, meaning \( \beta = 1 \). In that case, the parent chooses the level of transfers to equate the marginal cost of his for-gone consumption (left-hand side of (6)) to the marginal benefit of his child’s increased consumption in period \( t+1 \) (right-hand side of (6)). However, when \( \beta < 1 \), then, as seen from (4), the parent keeps increasing transfers even after the marginal cost is equated to the marginal benefit from increased child consumption in period \( t+1 \). The parent does this because bequeathing has an additional marginal benefit from the parent’s perspective when the offspring is impatient. Intuitively, when \( \beta < 1 \), the offspring is undersaving from the parent’s perspective. The parental welfare goes up if the parent can make the offspring increase his savings, which is possible by increasing bequests as long as the offspring’s optimal saving policy is strictly increasing in the amount of bequests received. This is why increasing bequests carries an additional benefit. We will provide a sharper marginal character-ization of equilibrium bequest behavior in Section 3.2 where we assume differentiability of the value and policy functions. This sharper characterization will also enable us to sharpen the intuition explained above. Before that, we prove that laissez-faire equilibrium is inefficient in Section 3.1 for the general case without assuming differentiability.

3.1 The Inefficiency of Laissez-Faire Equilibrium

We show that the laissez-faire equilibrium is Pareto inefficient whenever \( \beta \neq 1 \). Importantly, the proof uses a perturbation argument and does not use any differentiability or continuity assumptions regarding the equilibrium value or policy functions. This is important because we know from Krusell and Smith (2003) that in models similar to ours there can be many Markov equilibria some of which might have discontinuous policy functions.
By not making any assumptions on continuity or differentiability, our proof ensures that Pareto inefficiency of equilibrium is a general phenomenon that does not only apply to differentiable equilibria.

It is important to understand that the sense in which equilibria are inefficient are very different between our disagreement economy and the related self-control economy. According to the self-control interpretation, offspring want to save according to $\delta$ discounting but lack the self-control and end up saving according to $\beta\delta$ discounting. In this context, increasing the offspring’s savings by a small amount improves the welfare of both the offspring and the parent, implying that original equilibrium cannot be efficient. The inefficiency stems from the fact that some agents cannot carry out actions that are optimal from everybody’s perspective. In our model with disagreement, the $\beta\delta$ discounting represents offspring’s true preference, and thus, a perturbation that simply increases offspring’s saving rate cannot Pareto improve over equilibrium because it makes the offspring worse off. The inefficiency of laissez-faire equilibria in the disagreement economy comes purely from the existence of a consumption externality that arises from the presence of disagreement.

Proposition 2 states and proves that as long as $\beta \neq 1$ the laissez-faire equilibrium allocation is Pareto inefficient. The proof provides a resource feasible perturbation of the equilibrium allocation that improves welfare of some agents strictly without hurting others.

**Proposition 2.** The laissez-faire equilibrium allocation is Pareto inefficient.

**Proof.** Relegated to Appendix A.2.

The intuition for the inefficiency of equilibrium is as follows. As it is evident from equation (7) below, the parent’s welfare, $V_t$, depends on how much the offspring saves. In this sense, there is a consumption externality. Obviously, this is true in the standard altruistic model without disagreements as well (see Bernheim (1989)). In the standard case, however, since offspring fully agrees with the parent, in a sense the offspring internalizes the consequences of his saving on the parent, so the externality does not have a consequence. This can be seen by setting $\beta = 1$ in equation (7) below:

$$V_t = u(c_t) + \gamma \left[ u(c_{t+1}) + \beta\delta V_{t+2} + (1 - \beta)\delta V_{t+2} \right]$$  \hspace{1cm} (7)

When $\beta \neq 1$, the offspring does not fully internalize the consequences of his saving on
parental welfare. The level of savings he chooses affects the term $(1 - \beta)\delta V_{t+2}$ in (7), but this term is external to the offspring. This externality is the reason why equilibrium is inefficient. In the case in which $\beta < 1$, the term that creates the externality asks for a higher saving rate. In this case, the planner can improve both agents’ welfare by forcing the offspring to increase his savings by a small amount. This creates a second order loss for the offspring since he was already at his optimal allocation. It creates a first order gain for the parent since the equilibrium level of offspring saving is strictly suboptimal from parent’s perspective. Then, the planner simply transfers a suitable amount from the parent to the offspring to compensate the second-order decline in offspring’s welfare, achieving the desired Pareto improvement.

3.2 Equilibrium Parental Behavior under Differentiability

In this section, we provide a marginal condition that characterizes equilibrium bequest behavior assuming differentiability of the policy functions which describe offspring’s savings. Recall that $b_{t+2}(\cdot, Q_{t+1})$ represents the offspring’s equilibrium choice under price sequence $Q_{t+1}$ as a function of the bequests he receives from his parent, $b_{t+1}$. Now consider a parent’s problem. The parent chooses $b_{t+1}$ subject to the flow budget constraints and the function $b_{t+2}(\cdot, Q_{t+1})$, defined by (2), which describes offspring saving decision. The parent’s first-order optimality condition with respect to the bequest decision, $b_{t+1}$, is

$$u'(c_t) = \gamma \left( u'(c_{t+1}) \left[ R_{t+1} - \frac{\partial b_{t+2}(b_{t+1}, Q_{t+1})}{\partial b_{t+1}} \right] + \delta V_1(a_{t+2}, Q_{t+2}) R_{t+2} \frac{\partial b_{t+2}(b_{t+1}, Q_{t+1})}{\partial b_{t+1}} \right),$$

where $V_1$ refers to the derivative of the value function with respect to its first argument and the derivatives are all evaluated at the equilibrium allocation.

The offspring’s first-order optimality condition for $b_{t+2}$ is given by

$$u'(c_{t+1}) = \beta \delta V_1(a_{t+2}, Q_{t+2}) R_{t+2}. \tag{9}$$

Using (9) in the parental optimality condition (8), we get the following proposition which describes equilibrium parental bequeathing behavior under differentiability.

**Proposition 3.** Suppose $b_{t+2}(\cdot, Q_{t+1})$ is differentiable in $b_{t+1}$. The equilibrium bequest behavior
is characterized by
\[
u'(c_t) = \gamma \left( R_{t+1} u'(c_{t+1}) + \frac{\partial b_{t+2}(b_{t+1}, Q_{t+1})}{\partial b_{t+1}} u'(c_{t+1}) \left[-1 + \frac{1}{\beta}\right] \right).
\]

Equation (10) is the usual savings optimality condition, with an additional term on the right-hand side. The left-hand side is the marginal cost of increasing bequests, which equals the utility loss from forgone parental consumption. The first term on the right-hand side is the usual marginal benefit of increasing saving – the utility gain from increased consumption in the period during which returns to savings are received. There is a second term on the right-hand side, however. One can see that this term does not show up in the solution to the usual savings problems where $\beta = 1$, meaning the saver and the person receiving savings agree on what the receiver will do with the savings (an implication of the Envelope condition). This additional term summarizes how increasing parental transfers affects parental welfare by affecting the consumption levels of the offspring. It is a multiplication of two terms: the first term,
\[
\frac{\partial b_{t+2}(b_{t+1}, Q_{t+1})}{\partial b_{t+1}} > 0,
\]
tells how the offspring’s saving is affected by an increase in bequests. In general, this derivative is weakly positive since increasing transfers increases the period $t + 1$ wealth of the offspring, which weakly increases his savings. As we show in Lemma [12] in Appendix [A.1], under the assumption of differentiability of $b_{t+2}(\cdot, Q_{t+1})$, this derivative is strictly positive.\(^{16}\)

The second term,
\[
= u'(c_{t+1}) \left[-1 + \frac{1}{\beta}\right],
\]
represents the utility value to the parent of increasing $b_{t+2}$ marginally and is positive (resp. negative) whenever $\beta < 1$ (resp. $\beta > 1$).

Intuitively, when $\beta < 1$, the parent knows that from his perspective the offspring is undersaving. So, parental welfare goes up if the parent can make the offspring increase his savings, which is possible by increasing bequests since $\frac{\partial b_{t+2}(b_{t+1}, Q_{t+1})}{\partial b_{t+1}} > 0$. As a result, the additional term in (10) is positive: there is an additional marginal benefit of increasing

\(^{16}\)To be precise, Lemma [12] proves that as long as the value function is differentiable, the policy is strictly monotone. The differentiability of the value function is implied by differentiability of the policy function.
transfers for the parent. It is this extra benefit of bequeathing that makes the parent behave according to (4). Observe that under differentiability the strict version of equation (4) holds.

4 Ramsey

In Proposition 2, we show that laissez-faire equilibrium is unambiguously Pareto inefficient. The optimal policy implications of this result might obviously depend on the particular Pareto efficient allocation that the policy targets. In this section, we start our optimal bequest tax policy analysis by targeting a widely adopted benchmark Pareto efficient allocation, namely the ‘Ramsey’ allocation. The Ramsey allocation is the efficient allocation that puts all the weight on the initial generation parent. Thus, it is given by the solution to a fictitious social planner’s consumption-saving problem where the planner discounts exponentially at rate $\delta$ between young and old adulthood and discounts future generations by altruism factor $\gamma$.

The Ramsey allocation has at least three desirable properties. First, it corresponds to the equilibrium allocation in absence of conflict about intertemporal trade offs. Thus, in the absence of disagreements, optimal policy is simply not to distort bequests at all. In that sense, in the case of Ramsey allocation, the optimality of bequest tax under disagreements is coming purely from the existence of disagreements. Second, it is unique and simple to characterize, as it is the dynastic solution to the standard Ramsey-Cass-Koopmans optimal growth problem. Finally, the saving behavior prescribed in the Ramsey allocation - saving according to discount factor $\delta$ - is optimal from the perspective of all the agents in the dynasty except for the young adult making the saving decision.  

We first characterize the Ramsey allocation. Then, we show that the Ramsey level of bequests does not satisfy the parent’s optimality condition in laissez-faire equilibrium: in particular, parents bequeath too much relative to the Ramsey level when offspring are impatient from parents’ perspective. This implies that implementing the Ramsey allocation in the market requires government intervention. We show that, under differentiability assumptions on offspring’s saving policy function, this result translates into an optimal positive bequest wedge. Finally, we provide an implementation of the Ramsey allocation in the market through linear taxes on savings and bequests for a special class of equilibria.

\footnote{Evaluating welfare from the perspective of the initial self is also the route taken in much of the related literature on self-control problems. See DellaVigna and Malmendier (2004), Gruber and Koszegi (2004) and O’Donoghue and Rabin (2006), for example.}
When $\beta < 1$, the optimal tax on bequests is positive.

In Section 5.1 we generalize our results to all the allocations on the Pareto frontier. There, we show that optimal bequest wedge has a nice separable form between an efficiency component and an intergenerational redistribution component. The efficiency component is identical to the Ramsey bequest wedge that we derive in Section 4.2. It stems from intergenerational disagreement: whenever $\beta = 1$, there is no efficiency wedge. The second component, also present in models without any intergenerational disagreements, represents a subsidy to parental transfers that is due to intergenerational redistribution coming from a direct weight on the offspring.\(^\text{18}\)

### 4.1 The Ramsey Allocation

The Ramsey allocation is unique, and given by the solution to a fictitious social planner’s consumption saving problem where the planner has altruism and discount factors $\gamma$ and $\delta$, respectively. The following Euler equations characterize the Ramsey levels of bequests and savings, which we denote with an asterisk:\(^\text{19}\) for all $t$ even,

\begin{align*}
  u'(c^*_t) &= \gamma f'(k^*_t + 1)u'(c^*_{t+1}), \quad (11) \\
  u'(c^*_{t+1}) &= \delta f'(k^*_{t+2})u'(c^*_{t+2}). \quad (12)
\end{align*}

### 4.2 The Ramsey Wedge

Now we turn to the implications of Proposition 1 for equilibrium wealth transfer behavior relative to the Ramsey allocation. Comparison of Ramsey condition for bequests, \((11)\), with the equilibrium condition, \((4)\), together with the equilibrium pricing condition $R_{t+1} = f'(k_{t+1})$ implies that the equilibrium level of parental transfers does not satisfy the Ramsey condition for bequests in general. In the Ramsey allocation, the offspring’s choice of savings is optimal from parents’ perspective, which implies that the only marginal benefit of bequests comes from increased offspring consumption in the next period. On the other hand, in equilibrium, whenever offspring’s policy is strictly monotone, parents find an additional benefit to bequeathing: compensating for offspring’s period $t+1$ undersaving behavior.

\(^{18}\)Along this line, also see Farhi and Werning (2010), Kaplow (1995), and Kaplow (2001).

\(^{19}\)In Section 5, we denote any generic Pareto efficient allocation by asterisk in an effort to keep notational simplicity.
If we assume differentiability of policy functions, this argument translates into a positive bequest wedge for $\beta < 1$. To see this, first define the Ramsey bequest wedge as:

$$BW^*_t = 1 - \left( \frac{u'(c^*_t)}{\gamma u'(c^*_t)R^*_t} - \frac{\partial b_{t+2}(b^*_t, Q^*_t)}{\partial b_{t+1}} \left[ -1 + \frac{1}{\beta} \right] \right),$$

(13)

where $Q^*_{t+1}$ corresponds to the price sequence implied by the Ramsey allocation, that is $Q^*_{t+1} := \{ R^*_{t+s}, w^*_{t+s} \}_{s=1}^{\infty} = \{ f'(k^*_{t+s}), f(k^*_{t+s}) - f'(k^*_{t+s})k^*_{t+s} \}_{s=0}^{\infty}$, and the function $b_{t+2}(\cdot, Q^*_{t+1})$ is equilibrium policy of the offspring under that price sequence. Recall that parents bequeath according to (10) in equilibrium. The Ramsey bequest wedge in period $t$ measures the efficient marginal distortion that the planner needs to create in the return to bequests, $R^*_{t+1}$, in order to make the parent choose the Ramsey level of bequests in equilibrium. A positive (negative) $BW^*_t$ means that a parent facing the Ramsey interest rate, $R^*_{t+1} = f'(b^*_{t+1})$, would like to increase (decrease) his bequests marginally above (below) the Ramsey level if there is no government intervention.

**Corollary 4.** Suppose $b_{t+2}(\cdot, Q_{t+1})$ is differentiable in $b_{t+1}$ and $\beta < 1$. Then $BW^*_t > 0$.

**Proof.** Using equation (11) in the definition of $BW^*_t$, we get

$$BW^*_t = \frac{\partial b_{t+2}(b^*_t, Q^*_t)}{\partial b_{t+1}} \left[ -1 + \frac{1}{\beta} \right].$$

(14)

Lemma 12 in Appendix A.1 shows that when $b_{t+2}(\cdot, Q_{t+1})$ is differentiable in $b_{t+1}$, then $\frac{\partial b_{t+2}(b^*_t, Q^*_t)}{\partial b_{t+1}} > 0$. $BW^*_t > 0$ then follows from $0 < \beta < 1$. \(\square\)

Both the Ramsey planner and the parents in equilibrium face a marginal return equal to $R^*_{t+1}$ when bequeathing. For the parent, there is an additional return though: since in equilibrium the offspring is undersaving from parent’s perspective, a marginal increase in bequests increases parental welfare by increasing the offspring’s savings. In order to align parent’s marginal return with the Ramsey return, the planner needs to decrease equilibrium interest rate by (14), which is equal to that additional return normalized by Ramsey return. Notice that the bequest wedge is strictly positive since, under differentiability assumption, offspring’s saving is strictly increasing in the bequests received which implies that the additional return to bequeathing is strictly positive.
4.3 Implementation: Ramsey Taxation of Bequests

In this section, we want to implement the Ramsey allocation through a linear tax system on life-cycle savings and parental wealth transfers. Let \( \tau_{t+1} \) denote the linear tax rate on returns to period \( t \) savings, \( b_{t+1} \). If \( t \) is a period of parenthood, then \( \tau_{t+1} \) is a tax on bequests. Tax proceeds are rebated in a lump-sum manner in every period, so that the government balances its budget period by period. Letting \( T_t \) denote lump-sum taxes in period \( t \),

\[
T_t = R_t \tau_t b_t.
\]

Let \( Y := \{\tau_t, T_t\}_{t=0}^\infty \) be the sequence of taxes that the government chooses and commits to at the beginning of time and \( Y_t := \{\tau_s, T_s\}_{s=t}^\infty \). Let \( Y^* := \{\tau^*_t, T^*_t\}_{t=0}^\infty \) denote a tax system that implements the Ramsey allocation. We are interested in the Ramsey taxes on wealth transfers.

Letting \( \Psi := (Q, Y) \) be the joint sequence of prices and taxes, let \( \Psi_t := (Q_t, Y_t) \). Define \( V_t(a_t, \Psi_t) \) as the problem of a parent with wealth level \( a_t \) in calendar year \( t \) facing \( \Psi_t \), where the wealth level is \( a_t := R_t b_t (1 - \tau_t) + T_t + w_t \). The parent’s problem is given by

\[
V(a_t, \Psi_t) = \max_{b_{t+1} \geq -B(\Psi_{t+1})} u(c_t) + \gamma \left[ u(c_{t+1}(b_{t+1}, \Psi_{t+1})) + \delta V(a_{t+2}(b_{t+1}, \Psi_{t+1}, \Psi_{t+2})) \right],
\]

subject to the budget constraints

\[
\begin{align*}
c_t &= a_t - b_{t+1}, \\
c_{t+1}(b_{t+1}, \Psi_{t+1}) &= R_{t+1} b_{t+1} (1 - \tau_{t+1}) + T_{t+1} + w_{t+1} - b_{t+2}(b_{t+1}, \Psi_{t+1}), \\
a_{t+2}(b_{t+1}, \Psi_{t+1}) &= R_{t+2} b_{t+2} (1 - \tau_{t+2}) + T_{t+2} + w_{t+2},
\end{align*}
\]

and the offspring’s policy function is defined as

\[
\begin{align*}
b_{t+2}(b_{t+1}, \Psi_{t+1}) &= \arg \max_{b_{t+2} \geq -B(\Psi_{t+2})} u(\tilde{c}_{t+2}) + \beta \delta V(\tilde{a}_{t+2}, \Psi_{t+2}),
\end{align*}
\]

subject to

\[
\begin{align*}
\tilde{c}_{t+2} &= R_{t+1} b_{t+1} (1 - \tau_{t+1}) + T_{t+1} + w_{t+1} - \tilde{b}_{t+2}, \\
\tilde{a}_{t+2} &= R_{t+2} \tilde{b}_{t+2} (1 - \tau_{t+2}) + T_{t+2} + w_{t+2}.
\end{align*}
\]
The natural debt limit under $\Psi_t$ is given by

$$B(\Psi_t) := \sum_{s=t}^{\infty} \frac{w_s + T_s}{\prod_{p=t}^{s} R_p (1 - \tau_s)}.$$  

Note that the offspring’s optimal policy is also a function of taxes.

In general, an agent’s problem at any age is not convex since each parent faces a constraint describing the offspring’s policy, which may potentially break the convexity of the constraint set. Therefore, showing that the first-order optimality conditions of agents are satisfied by the Ramsey allocation under a tax system does not guarantee that the tax system implements the Ramsey allocation. As a result, Proposition 4 does not automatically imply that there is a linear tax system that implements the Ramsey allocation. Therefore, we restrict attention to Markov equilibria with policy functions that are linear in current wealth. The linearity of the policy functions guarantees that agents’ constraint sets are convex, thus implying that their problems are concave. Hence, we have the following implementation result.

**Proposition 5.** Suppose (Markov) equilibrium (with taxes) admits policies that are linear in current wealth. Then, there is a linear tax system that implements the Ramsey allocation. In this system, policies are strictly increasing, and optimal bequest taxes are strictly positive if and only if $\beta < 1$. They are given by

$$\tau^*_{t+1} = \left(-1 + \frac{1}{\beta}\right) M_{t+2}(\Psi^*_{t+1}) \frac{1}{R^*_{t+1}} > 0,$$

where

$$M_{t+2}(\Psi^*_{t+1}) = \frac{\partial b_{t+2}(b^*_{t+1}, \Psi^*_{t+1})}{\partial b^*_{t+1}} > 0$$

is the coefficient of offspring’s (linear) policy function under the Ramsey tax system and prices implied by the Ramsey allocation.

**Proof.** The linearity of the policy functions implies that each agent’s problem is concave, which implies that, once feasibility is guaranteed, the parent’s first-order optimality conditions are necessary and sufficient for the equilibrium. It is easy to derive the optimality condition for parental bequest choice under taxes, analogous to (10):

$$u'(c_t) = \gamma u'(c_{t+1}) \left( R_{t+1} (1 - \tau_{t+1}) + \frac{\partial b_{t+2}(b_{t+1}, \Psi_{t+1})}{\partial b^*_{t+1}} \left(-1 + \frac{1}{\beta}\right) \right).$$
Substituting in the Ramsey allocation and using \[(11)\] gives the expression for optimal bequest tax in the statement of Proposition 5 above. The fact that $M_{t+2}(\Psi_{t+1}^*)$ is strictly positive follows from the differentiability of offspring’s policy function.

From the linearity of the budget in $\tau_t$, it is immediate to see that if we define $a_{t+1} := (1 - \tau_{t+1})b_{t+1}$ the gross of taxes we have
\[
\frac{\partial b_{t+2}(b_{t+1}^*, \Psi_{t+1}^*)}{\partial a_{t+1}} = \frac{1}{1 - \tau_{t+1}} \frac{\partial b_{t+2}(b_{t+1}^*, \Psi_{t+1}^*)}{\partial b_{t+1}} = \frac{M_{t+2}(\Psi_{t+1}^*)}{1 - \tau_{t+1}}.
\]
An equivalent formula for bequest taxes is hence
\[
\frac{\tau_{t+1}^*}{1 - \tau_{t+1}^*} = \left(-1 + \frac{1}{\beta}\right) \frac{\partial b_{t+2}(b_{t+1}^*, \Psi_{t+1}^*)}{\partial a_{t+1}} \frac{1}{R_{t+1}^*}.
\]

Next, we show that when the utility function is of the CEIS form, there is always a Markov equilibrium with policy functions that are linear in current wealth.\[20\]

**Proposition 6.** Suppose period utility is of the CEIS form, meaning
\[
\begin{align*}
  u(c) &= \frac{c^{1-\rho}}{1 - \rho}, \text{ for } \rho \in (0,1) \text{ and } \rho > 1; \\
  &= \log(c), \text{ for } \rho = 1.
\end{align*}
\]
Then, if an equilibrium with taxes exists, there is an equilibrium in which consumption in each period is a linear function of the net present value of wealth as of that period.

**Proof.** Relegated to Appendix [A.3].\]

It is interesting to note that even though the government corrects offspring’s saving behavior through saving taxes, parental transfers should still be taxed to achieve the Ramsey allocation. Here, one might ask: given that from the Ramsey perspective the offspring is saving the right amount (thanks to corrective taxes), why does the parent still bequeath more than the Ramsey level? This occurs because, from the parent’s perspective, the offspring is undersaving. The taxes that are levied on the offspring create a diversion between the parent’s and the planner’s perception of what is optimal for the offspring. The government sets the lump-sum taxes to balance its budget in period $t + 2$, meaning sets $T_{t+2} = R_{t+2} \tau_{t+2} b_{t+2}$. Thus, from the government’s perspective, the return to offspring’s savings in period $t + 1$ is actually $R_{t+2}$, and the taxes are there only to drive offspring

\[20\] For a special case of our model economy with partial equilibrium and constant prices, Phelps and Pollak (1968) and Laibson (1994) have shown the existence of linear equilibria under CEIS utility.
to Ramsey behavior. The parent, on the other hand, takes lump-sum taxes as given and, hence, sees the return as $R_{t+2}(1 - \tau_{t+2})$, and wants the child to save optimally according to this return. This is why the offspring, who save at the Ramsey level under optimal saving subsidy, are still undersaving from the parents’ perspective. As a result, parents still have a motive to transfer more than the Ramsey level. To discourage this, the government finds it optimal to tax bequests. Another way of interpreting this result is as follows. By bequeathing an extra unit to the offspring, parents increase offspring’s savings, which increases the cost of financing the subsidy for the government. Hence, bequests generate a negative fiscal externality. The tax on bequest is a Pigouvian tax needed to internalize this externality.\(^{21}\)

If the government (or the parents) could command the offspring to the Ramsey allocation, then there would be no need to distort parents’ bequest decision. A natural question that follows, then, is whether the bequest tax result is peculiar to the assumption that the government is restricted to use linear taxes to discipline offspring’s savings. The proof of Proposition 1 shows that as long as offspring react to an increase in bequests by increasing their savings, parents do have an extra return to bequeathing, and hence, will bequeath too much if there is no government intervention. Therefore, as long as the government uses a kind of tax policy that leaves offspring’s equilibrium savings strictly increasing in the bequests they receive, parents will bequeath too much if they are not taxed. We conclude that the optimality of the bequest tax is not peculiar to the implementation in which there are linear taxes on children’s savings; bequest taxes would remain optimal for any tax system that does not completely eliminate the monotonicity of offspring’s optimal saving policies in the amount of bequests they receive.

Finally, note that, implicit in the Markovianity assumption, we do not allow for conditional bequests. In particular, parents are not allowed to condition the payment of the bequest to a specific level of savings by the offspring. The lack of such - potentially welfare improving - arrangements might be justified by the fact that they are difficult to enforce in reality. Recall, indeed, that bequest payments - by definition - occur after parents die and saving decisions might be difficult to monitor, especially by third parties. In this sense, the “Rotten Kid Theorem” of \cite{Becker1974} does not hold in our environment by assumption.\(^{21}\)

\(^{21}\)Shall we thank the referee for this interpretation?
4.4 A Back of the Envelope Quantitative Analysis

Our main result states that, as long as offspring - the recipients of intergenerational transfers - are too impatient from their parents’ perspective, parents transfer too much, and hence, to restore efficiency, bequests should be taxed.

In this section we aim at giving a rough idea of the quantitative importance of our mechanism for bequest taxation. To do so, we compute optimal bequest taxes that implement the Ramsey allocation for a parameterized version of our economy. We focus on logarithmic utility which allows us to find closed-form solutions for optimal bequest taxes.

**Proposition 7.** Suppose \( u(c) = \log(c) \). Then, the optimal bequest tax in any period of parenthood \( t \) is given by the following formula:

\[
\tau_t^* = \frac{\delta(1 + \gamma)(1 - \beta)}{1 + \delta}.
\]

**Proof.** Relegated to Appendix [A.4] \( \square \)

Observe that, under the logarithmic utility function assumption, the tax formula does not depend on the shape of the production function, \( F \), or the depreciation rate, \( \theta \). Therefore, we do not need to specify values for these parameters. The only parameters that are needed to calculate optimal taxes are the discount factor \( \delta \), the altruism factor, \( \gamma \), and the parameter \( \beta \) which represents the degree of disagreement. We assume perfect altruism in the sense that parents discount offspring’s consumption as much as they discount their own consumption, i.e., \( \gamma = \delta \). In this case, we have \( \tau_t^* = \delta(1 - \beta) \). In quantitative work, it is common to calibrate the discount factor using annual real interest rate as a target. Under the assumption that the interest rate is constant at \( R \) and equilibrium allocation is stationary in the sense that consumption across young adulthood periods and parenthood periods are constant, one can show that \( \beta, \delta \) and \( R \) have to jointly satisfy

\[
\frac{\delta^2 \beta}{1 - \delta + \beta \delta} R^2 = 1 \tag{15}
\]

\(^{22}\)Using the usual envelope condition, offspring’s equilibrium optimality condition for saving \( [9] \) becomes \( u'(c_{t+1}) = R \delta u'(c_{t+2}) \). This condition, together with parental optimality condition \( [10] \) evaluated at \( \gamma = \delta \) and the steady-state requirement that \( c_t = c_{t+2} \), implies that

\[
\delta^2 \delta R^2 \left( 1 + \frac{\partial b_{t+2} (b_{t+1}, Q_{t+1})}{\partial b_{t+1}} \frac{-1 + 1/\beta}{R} \right) = 1.
\]
Table 1: Optimal Bequest Taxes

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibrated $\delta$</td>
<td>0.571</td>
<td>0.545</td>
<td>0.524</td>
<td>0.506</td>
<td>0.491</td>
<td>0.484</td>
</tr>
<tr>
<td>Bequest tax</td>
<td>28.55%</td>
<td>21.80%</td>
<td>15.72%</td>
<td>10.12%</td>
<td>4.91%</td>
<td>2.42%</td>
</tr>
</tbody>
</table>

We take the annual real interest rate to be 3%, and assume each period in our model corresponds to 25 years, implying $R = 1.03^{25}$. Table 1 below reports calibrated $\delta$ in the second row and the corresponding optimal bequest tax rates in the third row for different values of $\beta$. A glance at Table 1 seems to indicate that the optimal taxes generated by our mechanism has a potential of being quantitatively significant.

5 Extensions

In this section, we consider several extensions of the baseline framework. First and foremost, we show that the optimality of creating a positive bequest wedge shown in Section 4 for the case of Ramsey allocation holds for other allocations on the Pareto frontier. Second, we extend our model by introducing horizontal inequality and show that bequest taxation is still optimal. Moreover, we show that optimal bequest taxes are typically progressive. Finally, we show that our normative results are also robust to extending people’s life cycle to any finite periods and allowing parents and offspring to coexist in the same period. This also allows us to show that it is optimal to tax inter vivos transfers as well. The optimality of transfer taxation in the presence of illiquid bequests is also discussed.

5.1 Intergenerational Redistribution: Pareto Efficient Bequest Wedges

The Ramsey allocation is one of the many possible Pareto efficient allocations. Specifically, it is the efficient allocation that arises when the planner only cares about the initial parent directly. In this section, we study the whole Pareto frontier by allowing the planner to care about all generations directly, and characterize the general Pareto efficient bequest wedge.

We begin by characterizing the Pareto frontier of the economy. Let $U = \{U_n\}_{n=1}^{\infty}$ be any

It follows from equation (43) in the proof of Proposition 7 in Appendix A.4 that, under $\gamma = \delta, \frac{\partial b_{n+2}(h_{n+1}, Q_{n+1})}{\partial h_{n+1}} = \frac{\delta^2}{1 - \delta + \delta \beta}$. Together these imply condition (13).
given sequence of utilities where \( U_n \in \mathbb{R} \) represents the minimum level of utility that needs to be delivered to \( n^{th} \) generation agents (who are born in calendar time \( 2n - 1 \)) starting with the agent born in period 1. Let \( \mathcal{U} \) be the set of such utility sequences that are achievable by feasible allocations.\(^{23}\)

We compute a Pareto efficient allocation solving a planning problem where we maximize the utility of one agent, initial parent, subject to delivering other agents \( U \) in the set of all feasible utility sequences \( \mathcal{U} \), the solution to this planning problem traces the whole Pareto frontier of the economy.

We use a simple recursive formulation of the planning problem to characterize the Pareto frontier of the economy. If \( k \) is the level of capital at the beginning of the parenthood period in a given generation, \( W : K \times \mathcal{U} \to R \), is the value function that denotes the maximal utility for current parents that can be achieved subject to providing future generations a sequence \( U \) of utilities.\(^{24}\) The value function is given by the solution to the following problem:

\[
W(k, U) = \max_{y, k'} u(f(k) - y) + \gamma [u(f(y) - k') + \delta W(k', U')] \\
\text{s.t.} \quad 0 \leq y \leq f(k), \quad 0 \leq k' \leq f(y), \quad U' = C(U); \\
u(f(y) - k') + \beta \delta W(k', U') \geq U. \tag{16}
\]

In the above notation, \( U' = C(U) \) indicates the ‘continuation sequence’ of the sequence \( U \).\(^{25}\) The choice variables \( y \) and \( k' \) represent, respectively, the bequest decision of the parent and the saving decision of the offspring. Constraint (16) guarantees that each new generation gets a level of utility \( U \).

Since the problem is concave in the choice variables for each \((k, U)\), it can be shown that the value function \( V \) is concave in \( k \) for each \( U \). Since \( u \) is differentiable, an application of

\(^{23}\)A key feature restricting the set \( \mathcal{U} \) is the domain of \( u : \mathbb{R}_+ \to \mathbb{R} \). If \( u(0) > -\infty \), to be achievable, a sequence of utilities \( U \) must admit a sequence of capital levels \( \{k_t\}_{t=0}^\infty \) such that, for all \( t \), we have \( 0 \leq k_{t+1} \leq f(k_t) \) and for generation \( n \) who are born in period \( s = 2n - 1 \)

\[
U_n = u(f(k_s) - k_{s+1}) + \beta \delta \left\{ u(f(k_{s+1}) - k_{s+2}) + \sum_{t=s+1}^\infty \gamma^{t-s} \delta^{t-s-1} \left[ u(f(k_{2t-1}) - k_{2t}) + \delta u(f(k_{2t}) - k_{2t+1}) \right] \right\}.
\]

\(^{24}\)\( K \equiv [0, \bar{k}] \) is the domain of the level of capital stock. It is well known that, with a positive rate of depreciation and Inada on \( f \) at infinity, there is a \( \bar{k} \) such that \( f(k) < k \) \forall k \geq \bar{k}.

\(^{25}\)In calendar time notation, if \( \{\bar{U}_t\}_{s=t}^\infty \) is a sequence of lifetime utilities for all agents from an even date \( t \) onwards, the sequence of offspring utilities is \( U = \{\bar{U}_{t+2s}\}_{s=0}^\infty \) and \( C(U) := \{\bar{U}_{t+2s}\}_{s=1}^\infty \) its ‘continuation’.
Benveniste and Sheinkman lemma implies that $V$ is differentiable in the first argument for all $U \in \mathcal{U}$\footnote{See Stokey, Lucas, Jr., and Prescott (1989), Theorem 4.10.}. Formally, let $g : K \times \mathcal{U} \to R_+$ be the policy function for $y$ and $h : K \times \mathcal{U} \to R_+$ represent the saving policy of the offspring. The envelope theorem implies that all $U$ associated to constraint (16), the first order conditions for optimality are:

$$W_1(k, U) = f'(k)u'(f(k) - g(k, U)),$$

whenever $0 < g(k, U) < f(k)$, where $W_1$ refers to the derivative of $W$ with respect to first argument.

Concave problems admit Kuhn-Tucker multipliers. If we denote by $\lambda$ the multiplier associated to constraint \cite{16}, the first order conditions for optimality are:

$$y \text{ (bequest)} : \quad u'(f(k) - y) = (\gamma + \lambda)f'(y)u'(f(y) - k')$$

\tag{17}

$$k' \text{ (savings)} : \quad u'(f(y) - k') = \delta\frac{\gamma + \lambda \beta}{\gamma + \lambda}f'(k')u'(f(k') - y').$$

\tag{18}

Once we solve for the policy functions $g$ and $h$, we can construct the Pareto efficient sequence of capital stocks in the usual way. First, let $U_1 = \{U_n\}_{n=1}^{\infty}$ be the sequence that represents a particular Pareto efficient allocation. Define $U_m = \{U_n\}_{n=m}^{\infty}$. Then, define capital allocation under this efficient allocation recursively as: $k_1^* = g(k_0, U_1)$, $k_2^* = h(k_1^*, U_1)$, $k_3^* = g(k_2^*, U_2)$, $k_4^* = h(k_3^*, U_2)$, and so on. We then use $\{k_{t+1}^*\}$ together with (17) to compute the multiplier, $\lambda_{t+1}^*$,

$$u'(f(k_i^*) - k_{t+1}^*) = (\gamma + \lambda_{t+1}^*)f'(k_{t+1}^*)u'(f(k_{t+1}^*) - k_{t+2}^*),$$

\tag{19}

where $\lambda_{t+1}^*$ represents the multiplier on constraint \cite{16} under $k = k_i^*$ and the particular utility sequence $U$ that indexes the Pareto efficient allocation.

We are now ready to compute the bequest wedge defined by (13) for a Pareto efficient allocation indexed by $U$:

$$BW_t^*(U) := 1 - \left(\frac{u'(c_t^*)}{\gamma u'(c_t^*)f'(k_{t+1}^*)} - \frac{\partial b_{t+2}(k_{t+1}^*, Q_{t+1}^*)}{\partial b_{t+1}} \left[\frac{-1 + \frac{1}{\beta}}{f'(k_{t+1}^*)}\right]\right),$$

where $Q_{t+1}^*$ is the price sequence generated from the Pareto efficient allocation as before and $\frac{\partial b_{t+2}(k_{t+1}^*, Q_{t+1}^*)}{\partial b_{t+1}}$ represents derivative of period $t + 1$ offspring’s policy with respect to
the bequests received at $k_{t+1}^*$ and $Q_{t+1}^*$. Using (19) in $BW_t^*(U)$ above gives

$$BW_t^*(U) = -\frac{\lambda_{t+1}}{\gamma} + \frac{\partial b_{t+2}(k_{t+1}^*, Q_{t+1}^*)}{\partial b_{t+1}} \left[\frac{-1 + \frac{1}{\beta}}{f'(k_{t+1}^*)}\right].$$

(20)

The bequest wedge characterized by (20) has two components. First, consider the second term on the right-hand side of (20). This term represents the efficiency component of the bequest wedge that arises due to intergenerational disagreements. A comparison of this component with the Ramsey bequest wedge (14) shows that they are essentially identical. The efficiency component is positive, meaning it calls for a tax on bequests, as long as the offspring is impatient from the parent’s perspective ($\beta < 1$). The intuition is the same as in the Ramsey case. In particular, observe that the sole reason for the efficiency component is the presence of disagreement: the wedge disappears when there is no disagreement ($\beta = 1$).

In the Ramsey case, the efficiency wedge is the only component of the wedge. In general, Pareto efficient bequest wedges have another component though. The first term on the right-hand side of (20) represents this component which comes from intergenerational redistribution. The intergenerational redistribution wedge is also present in the standard altruistic model of bequests without disagreements.

It is indeed easy to see from (20) that this component exists even when $\beta = 1$. The intergenerational redistribution wedge is negative, and thus, calls for a subsidy on bequests. It is strictly negative as long as the multiplier, $\lambda_{t+1}$, is strictly positive. This is true as long as the constraint (16) binds which is equivalent to the planner putting a direct welfare weight on the offspring born in period $t + 1$. Intuitively, the offspring’s welfare enters the planner’s objective through two channels: one indirectly through the parent’s welfare and the other directly. Thus, the planner cares about the offspring more than the parent does, and for this reason the parent bequeaths too little from planner’s perspective. It is then optimal to subsidize parents. The magnitude of the intergenerational redistribution wedge depends on the Pareto efficient allocation itself. In the formula in (20), this dependence is implicit as $\lambda_{t+1}$ depends on the sequence of utility promises, $U$, which indexes Pareto efficient allocations. Of course, despite the fact that the laissez-faire equilibrium allocation is never efficient, it might be the case that bequests must be optimally taxed at zero rate. This would happen if the intergenerational disagreement and intergenerational redistribution components of the tax formula

Table 2: Optimal Bequest Taxes

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
<th>0.8</th>
<th>0.9</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$</td>
<td>0.758</td>
<td>0.745</td>
<td>0.734</td>
<td>0.724</td>
<td>0.718</td>
<td>0.711</td>
</tr>
</tbody>
</table>

exactly offset each other.

**Ramsey allocation as Pareto Improvement: An example with Log utility.** The Ramsey allocation and the taxes that achieve this allocation might as well improve the utility to all agents in the economy compared to the laissez-faire equilibrium. In this section, we focus on an example where the instantaneous utility function is logarithmic and production in period $t$ is given by $f_t(k) = R_t k + w_t$. In Proposition 14 in Appendix A.4, we derive an analytic condition regarding when Ramsey allocation Pareto improves over laissez-faire equilibrium allocation that depends only on $\gamma, \delta$ and $\beta$. Using this condition, in the following proposition, we show that, if people are perfectly altruistic ($\gamma = \delta$), then for any level of disagreement between parents and offspring that is between zero and one, Ramsey allocation dominates the equilibrium allocation as the discount factor approaches one.\footnote{We focus to the case with $\beta < 1$ in order to have well defined lifetime utilities for all $\delta \leq 1$.}

**Proposition 8.** Set $\gamma = \delta$. Then, for any $\beta \in (0,1)$, the Ramsey allocation Pareto improves over laissez-faire allocation as $\delta \to 1$.

**Proof.** Relegated to the end of Appendix A.4

Intuition for the proposition???

Next, we are interested in less extreme cases. For each $\beta$ value given in the first row of Table 2, we compute the minimum value of $\delta = \gamma$ above which the Ramsey allocation Pareto dominates the equilibrium allocation for that $\beta$ using Proposition 14 in Appendix A.4. The results are reported in the second row in Table 2. Comment on the numbers?

5.2 **Horizontal Redistribution**

An important simplifying assumption we make so far in the paper is that all dynasties in the economy are identical. In this section, we relax this assumption and allow people to differ in their skill levels, which translates into income inequality in equilibrium.
A key issue that arises when the population is divided among different income groups is horizontal redistribution, that is, redistribution across different families. In this section, we incorporate Mirrlees (1971)’s model of horizontal redistribution into our intergenerational model and analyze how the normative predictions of the latter change under inequality and redistribution. First, we show that in two important cases all the normative results derived in the model without heterogeneity continue to hold exactly. These are: (i) whenever redistribution can be performed via non-distortionary taxation (which corresponds to the case where skills are publicly observable), and (ii) even if redistribution is limited because people’s skills are private information, whenever we are interested in the Ramsey allocation of Section 4, i.e., in the notation of Section 5.1, when \( \lambda = 0 \). Second, we show that, when the planner cares directly about the offspring’s welfare (i.e., when \( \lambda > 0 \)) and, in addition, skills are private information to the agents, horizontal redistribution introduces an element of progressivity into optimal bequest taxation. We show that under certain conditions (e.g., log utility) the optimal bequest tax is \textit{progressive}.

To reduce the notational burden, we present our results using a three period framework. In order to analyze the efficient allocation allowing for imperfect horizontal distribution, we follow the Mirrleesian tradition assuming that labor income is publicly observable while agent’s productivity is not. Whenever labor is supplied inelastically (e.g., because agents do not evaluate leisure as in our baseline model), horizontal redistribution is perfect and the planner is able to implement the same allocation as that with observable skills, which in turn corresponds to the allocation we studied so far in Section 4 and then extended in Section 5.1. What does this paragraph add?

\textbf{A Mirrleesian Model} Suppose the economy lasts three periods. At \( t = 0 \), skill level \( s \geq 0 \) is drawn from a set \( S \) according to a distribution \( H \) and parents choose labor \( l_0 \in [0, 1] \) with flow utility

\[
u(c_0) - v(l_0),
\]

where the function \( v \) is increasing and convex, with \( v'(0) = 0 \) and \( v'(1) = \infty \). To simplify the analysis, we assume that in periods 1 and 2 labor is supplied inelastically at \( \bar{l}_t = 1 \) as in the previous sections of the paper. Income in period \( t = 0 \) equals \( y_0 = w_0sl_0 \) which is publicly observable as is the real wage per unit of skill \( w_0 \). In periods \( t = 1, 2 \), all agents have a common labor income \( y_t \), that is the offspring do not inherit the skills of their parents. This is an extreme form of mean reversion in skills which will be somewhat important for the
progressivity result.

Given an allocation, define \( z_t(s) := u(c_t(s)) \) as the utility an agent from family \( s \) in period \( t \) receives from consuming \( c_t(s) \). Since \( u(\cdot) \) is one-to-one, planner assigning a consumption allocation is the same as planner assigning utility. We find it convenient to write down the planning problem where planner chooses utilities. Letting \( \lambda \) be the direct welfare weight the planner puts on the offspring, the planner’s objective function is:

\[
\max_{\{z_t(\cdot)\}_{t=0}^{\infty}} \int_S \left\{ z_0(s) - v \left( \frac{y_0(s)}{w_0s} \right) + (\gamma + \lambda)z_1(s) + \delta(\gamma + \lambda\beta)z_2(s) \right\} dH(s).
\] (21)

For each \( t = 0, 1, 2 \), define

\[
C_t := \int_S c_t(s)dH(s) = \int_S g(z_t(s))dH(s), \quad L_0 := \int_S l_0(s)sdH(s), \quad K_t := \int_S k_t(s)dH(s)
\]

and \( L_1 = L_2 = 1 \), where the function \( g \) represents the inverse of the utility function \( u \) (i.e., \( g := u^{-1} \)) and the variables \( C_t, L_t \) and \( K_t \) represent the aggregate amounts of consumption, labor input in efficiency units, and capital in period \( t \). As before, factors of production are priced competitively, and in particular, \( w_t := F_L(K_t, L_t) \), where \( F_L \) refers to the partial derivative of \( F \) with respect to labor. Aggregate feasibility is then given by, for each \( t = 0, 1, 2 \),

\[
C_t + K_{t+1} \leq F(K_t, L_t) + (1 - \theta)K_t, \quad \text{with} \quad K_0 \quad \text{given.} \tag{22}
\]

We indicate by \( \mu_t \) the multiplier associated to the feasibility constraint in period \( t \).

We assume that skill level, \( s \), is privately observed by the parent. This implies that the planner also faces a set of incentive constraints in period zero, namely for all \( s, \hat{s} \in S \):

\[
z_0(s) - v \left( \frac{y_0(s)}{sw_0} \right) + \gamma [z_1(s) + \delta z_2(s)] \geq z_0(\hat{s}) - v \left( \frac{y_0(\hat{s})}{sw_0} \right) + \gamma [z_1(\hat{s}) + \delta z_2(\hat{s})]. \tag{23}
\]

Notice that since all information about skill types is revealed once and for all in period 0, there are only period 0 incentive constraints.

Given \( \lambda \), the planner’s problem then is to maximize (21) subject to (22) and (23). By varying \( \lambda \in [0, \infty) \), one can trace the whole set of Pareto efficient allocations. Therefore, \( \lambda \) indexes Pareto efficient allocations.

Now, we characterize Pareto efficient allocations. For any \( \lambda \geq 0 \), consider the following perturbation of a Pareto efficient allocation: for a given \( s \), set \( z_0^\varepsilon(s) = z_0^\ast(s) - \gamma \varepsilon, z_1^\varepsilon(s) = z_1^\ast(s) - \delta \varepsilon, z_2^\varepsilon(s) = z_2^\ast(s) - \gamma \delta \varepsilon \).
\( z_1^*(s) + \varepsilon, \) and \( z_2^*(s) = z_2^*(s). \) Notice that since

\[
z_0^*(s) + \gamma [z_1^*(s) + \delta z_2^*(s)] = z_0^*(s) + \gamma [z_1^*(s) + \delta z_2^*(s)],
\]

all such perturbations are by construction incentive compatible. This implies that for this set of perturbations, one can disregard the incentive constraint while solving the planner’s problem. The efficient allocation has to solve the first-order optimality condition for the planner’s problem over the choices of \( \varepsilon \) in a neighbourhood of zero, i.e., the following condition has to hold:

\[
\lambda + \frac{\mu_0 \gamma}{u'(c_0^*(s))} - \frac{\mu_0}{u'(c_1^*(s))} R_1^* = 0,
\]

where \( R_1^* := F_K(K_1^*, 1) + (1 - \theta) \) us the return to capital in period 1. The above condition can be rearranged as:

\[
\frac{\lambda}{\mu_0 \gamma} u'(c_0^*(s)) = \frac{u'(c_0^*(s))}{\gamma R_1^* u'(c_1^*(s))} - 1.
\]

(24)

How do you know this perturbation is resource feasible?

Recall that bequest wedge for a particular Pareto efficient allocation is the wedge that the planner needs to create in order to make the parent choose that Pareto efficient allocation in equilibrium. Following the definition of the bequest wedge introduced by (13), we define bequest wedge for a family of type \( s \) under Pareto index \( \lambda \) to be:

\[
BW^*(\lambda, s) = 1 - \left( \frac{u'(c_0^*(s))}{\gamma R_1^* u'(c_1^*(s))} - \frac{\partial b_2(b_1^*, Q_1^*)}{\partial b_1} \left[ -1 + \frac{1}{p} \right] \right).
\]

Plugging (24) in the bequest wedge definition above implies

\[
BW^*(\lambda, s) = -\lambda \frac{u'(c_0^*(s))}{\mu_0 \gamma} + \frac{\partial b_2(b_1^*, Q_1^*)}{\partial b_1} \left[ -1 + \frac{1}{p} \right] \frac{1}{R_1^*}.
\]

(25)

How do we know \( b_2() \) does not depend on \( s \)?

Similar to the bequest wedge characterized by (20) in Section 5.1, the bequest wedge characterized by (25) has two components. The first component is due to intergenerational redistribution and always implies a subsidy on bequests whereas the second component is the correction due to intergenerational disagreements and always implies a tax.

---

29Using the taxation principle, the wedge can also be written as a function of parent’s income \( y_0 \).
There are two important cases in which the bequest wedge in (25) simplifies exactly to the bequest wedge expressions we found earlier on. First, if skills are publicly observable, then since \( u'(c^*_0(s)) = \mu_0 \) for all \( s \) (full redistribution), we have:

\[
BW^*(\lambda, s) = -\frac{\lambda}{\gamma} + \frac{\partial b_2(b_1^*, Q_1^*)}{\partial b_1} \left[ -1 + \frac{1}{\beta} \right].
\]

This expression does not depend on \( s \), and is identical to (20), which is the expression for bequest wedge in Section 5.1 when there is no heterogeneity. It turns out when skills are publicly known, the bequest wedge is flat. Under CEIS assumption, an equilibrium with linear strategies exists, and following the proof of Proposition 5, one can show that there is a corresponding flat optimal tax or subsidy. Intuitively, since labor income taxes are non-distortionary, for the purposes of horizontal redistribution, the planner only uses labor income taxation. The bequest tax is only used to correct the inefficiency stemming from intergenerational disagreements and to facilitate intergenerational redistribution which do not depend on people’s skills levels.

Second, if the planner only cares about the parent directly (\( \lambda = 0 \)) - then, even in presence income heterogeneity and imperfect redistribution, the bequest wedge given by (25) reduces to expression (14) in Section 4:

\[
BW^*(0, s) = \frac{\partial b_2(b_1^*, Q_1^*)}{\partial b_1} \left[ -1 + \frac{1}{\beta} \right] R_1^* > 0.
\]

This means that if the planner does not put any direct weight on future generations, then the bequest wedge is positive and flat. Again under CEIS assumption, Proposition 5 implies that there is a corresponding flat optimal bequest tax that implements this particular Pareto efficient allocation.

Now, we analyze the general case where future generation is directly weighted in the planner’s objective and skills are private information. It is straightforward to show that incentive compatibility implies that \( c^*_0(s) \) is increasing in \( s \). This implies that the intergenerational redistribution component of the bequest wedge given in (25) calls for a subsidy decreasing with \( s \). The mechanism behind the progressivity of the intergenerational redis-

---

30 Although the exact coincidence of the two wedges does not survive to allowing for unobservable skill shocks and flexible labor supply after period 0, our taxation principle remains. In this more general environment, the new component of the tax will be purely dictated by the forces studied by the so called ‘new dynamic public finance literature’ (e.g., Kocherlakota (2005)). H: This is not clear.
tribution component is the same as in [Farhi and Werning (2010)](Farhi and Werning (2010)). Intuitively, as long as \( \lambda > 0 \), the planner’s preference for \( c_1 \) relative to \( c_0 \) is stronger than that of parents’. This is easy to see comparing planner’s objective function, (21), with incentive constraints, (23). This means that the planner will find it more effective to generate redistribution by reducing inequality in \( c_1 \) compared to reducing inequality in \( c_0 \). In turn, this implies that the optimal cross-sectional distribution of \( c_1 \) is more compressed that that of \( c_0 \). As a consequence, the efficient allocation displays increasing consumption (between period \( t = 0 \) and \( t = 1 \)) for agents with low \( s \) and decreasing consumption for agents with high \( s \). Since all parents face the same interest rate, \( R^*_1 \), but must be content with different intertemporal patterns of consumption, a natural implementation of the efficient allocation would prescribe higher bequest taxes to rich parents (as otherwise they will have incentive to deviate from the recommended consumption plan by increasing bequests) and lower bequest taxes to poor parents.

With logarithmic preferences, an increasing intergenerational redistribution wedge implies that the overall bequest wedge is also increasing in \( s \) (since under logarithmic preferences the disagreement component is constant in \( s \)). Thus, depending on which component dominates, the bequest wedge implies either a subsidy that is decreasing with \( s \) or a tax that is increasing with \( s \). In Appendix B (Proposition 15), we provide a simple tax system that implements Pareto efficient allocations. In the logarithmic utility case, this tax system features progressive bequest taxation.

### 5.3 Longer Life Cycle, Coexistence, and Inter Vivos Taxation

We assume that people live for two periods in our benchmark model. In this section, we extend our model by allowing each agent within a dynasty to be active for \( I + 1 \) periods: in the first \( I \) periods, agents make consumption saving decisions. In the last period of their lives, parents coexist with their offspring who are already in the first period of young adulthood. Parents decide how much to consume and transfer to their offspring. Transfers can be made in two ways: *inter vivos transfers* are received by the offspring during the coexistence period, and bequests are received at the beginning of the next period, after the parent dies. We show that the longer life cycle does not alter the main result, that is, bequests should be taxed. Furthermore, thanks to the coexistence period in the extended model, we are able to analyze parental inter vivos transfer behavior and establish the optimality of *taxing inter vivos transfers as well.*
Consider any calendar year $t$ in which there is a parent who is in the last period of his life. His preference over dynastic allocation is given by

$$V_t = u(c_t^o) + \gamma \left[ u(c_t) + \delta u(c_{t+1}) + \ldots + \delta^{I-1} u(c_{t+I-1}) + \delta^I V_{t+I} \right],$$

where $V_t$ represents the dynastic welfare of the parent who is in the last period of his life in period $t$ and $V_{t+I}$ represents that of the offspring in his terminal period, $t + I$. The term $c_t^o$ is the last period consumption of the parent and $c_t$ is the consumption level of the offspring who is at age 1 in period $t$. Observe that $c_t^o$ and $c_t$ occur in the same period. To keep aggregate labor supply constant across periods, we assume that only the offspring has one unit of time endowment in the period of coexistence. Clearly, this assumption is not material for any of our results.

There are two ‘natural’ ways to extent our baseline model to a multi-period setup, and both deliver same normative predictions qualitatively. First, one can envision agents as standard exponential discounters, perhaps with varying discount factors over the life cycle so that the discount factor between own consumption at age $i$ and $i + 1$ is $\beta_i \delta \leq \delta$. In this formulation, agents are time-consistent. Second, we can assume that agents face self-control problems and have time-inconsistent preferences. In this model, agents disagree with their future selves over intertemporal trade offs. That is, at any age $i$, people discount future with $\beta_i \delta$ and want their future selves to discount with $\delta$.

The time-consistent extension of the baseline model behaves quite similarly with the baseline model because in both models the only source of disagreement is the one that exists between different generations. Therefore, it is somewhat easier to see that the main results continue to hold in this extension. That is, as long as offspring are impatient from parents’ perspective in the period they receive transfers, it is optimal to tax parental transfers. This extension allows us to see that inter vivos as well as bequests should be taxed. Instead, in the rest of the section, we focus on the time-inconsistent extension where, in addition to intergenerational disagreement, people disagree with their future selves as well. We find that our results survive this double degree of disagreement as well.

We model time-inconsistent behavior following Laibson (1997)’s quasi-hyperbolic frame-
work. The offspring’s preference in period $i$ of his life is

$$u(c_{t+i-1}) + \beta_i \delta \left[ \sum_{j=i+1}^{I} \delta^{j-(i+1)} u(c_{t+j-1}) + \delta^{I-i} V_{t+I} \right], \quad \text{for } 1 \leq i \leq I - 1$$

$$u(c_{t+I-1}) + \beta_1 \delta V_{t+I}.$$ 

When $\beta_i = 1$ for all $i$, people are fully time-consistent at all ages; there is no present-bias problem. Whenever $\beta_i < 1$ for some $i$, people face present-bias problems at age $i$. Observe that we are extending the hyperbolic discounting model by allowing for the existence (and severity) of the present-bias problem, $\beta_i$, to depend on age.\[32\] We extend the standard model of present-bias in order to show that our transfer taxation results do not depend on how the degree of present-bias evolves over the life cycle. As we will see, the only assumption needed in order to establish optimality of transfer taxation is that the offspring has a present-bias problem in the period he receives the transfer.

Let $d_t$ and $b_{t+1}^o$ denote the inter vivos transfers and bequests made by the parent who is in his last period of life in period $t$. Let $b_{t+i}$ denote the offspring’s age $i$ saving level.

The parent, whose wealth level is $a_t = R_t b_t + w_t$, solves

$$V(a_t, Q_t) = \max_{b_t^{i+1},d_t} u(c_t^o) + \gamma \left[ \sum_{i=0}^{I-1} \delta^i u(c_{t+i}) + \delta^I V(a_{t+1}, Q_{t+1}) \right],$$

subject to the budget constraints.\[33\]

\[32\]If we were to take $\beta_i = \beta$ for all $i$, as previous papers have assumed, that would mean that the degree of present-bias problem is constant as people age, and all our results hold in this special case.

\[33\]The real interest and wage rates are given by marginal products of capital and labor as in the benchmark model. The only difference is for the period right after coexistence, total capital stock in the economy is equal to the sum of the offspring’s savings in the coexistence period and parental bequests

$$k_t = b_{t+1} + b_{t+1}^o.$$
and subject to the constraints defining the policy functions of future selves:

\[
\begin{align*}
    b_{t+1}(d_t, Q_t) &= \arg \max_{c_t} u(c_t) + \beta_1 \delta \left[ u(c_{t+1}) + \ldots + \delta^{t-1} u(c_{t+i-1}) + \delta^t V(a_{t+i}, Q_{t+i}) \right] \\
    &\quad \text{s.t.} \\
    b_{t+2}(d_t, b_{t+1}^o, Q_t) &= \arg \max_{c_{t+1}} u(c_{t+1}) + \beta_2 \delta \left[ u(c_{t+2}) + \ldots + \delta^{t-2} u(c_{t+i-1}) + \delta^{t-1} V(a_{t+i}, Q_{t+i}) \right] \\
    &\quad \text{s.t.} \\
    &\quad \ldots \\
    b_{t+i-1}(d_t, b_{t+i-1}^o, Q_t) &= \arg \max_{c_{t+i-1}} u(c_{t+i-1}) + \beta_i \delta \left[ u(c_{t+i-1}) + \delta V(a_{t+i}, Q_{t+i}) \right] \\
    b_{t+i}(d_t, b_{t+i}^o, Q_t) &= \arg \max_{c_{t+i}} u(c_{t+i}) + \beta_i \delta V(a_{t+i}, Q_{t+i})
\end{align*}
\]

To better understand the notation and the nature of policy functions in the planning problem, first observe that when an agent of age \(i\) in calendar year \(t + i - 1\) is deciding \(b_{t+i}\), he sees the direct effect of his choice on the next period self’s saving choice. Clearly, agent \(i\)’s saving choice indirectly affects his saving at age \(n > i\), \(b_{t+n}\), since \(b_{t+i}\) choice affects \(b_{t+i+1}\) choice, which then affects \(b_{t+i+2}\) choice, and so on. The dependence of \(b_{t+n}\) on \(b_{t+i}\) can be described by a (nested) function \(b_{t+n}(b_{t+n-1}(...b_{t+i+1}(b_{t+i}, Q_{t+i})...), Q_{t+n-1})\). Without creating a confusion, we can denote it as a single function of only \((b_{t+i}, Q_{t+i})\). Going backwards, the ‘initial’ choices in our recursion are \(d_t\) and \(b_{t+1}^o\). As a consequence, the vector of functions \([b_{t+1}(d_t, Q_t), b_{t+2}(d_t, b_{t+1}^o, Q_t), \ldots, b_{t+i}(d_t, b_{t+i-1}^o, Q_t)]\) denote the functions that describe how the offspring’s savings choices over the life cycle depend on parental transfers they receive, \((d_t, b_{t+1}^o)\). Observe that offspring’s saving at age 1 in year \(t\), denoted by \(b_{t+1}\), only depends on inter vivos transfers she receives in that period but not on the level of bequests, \(b_{t+i}^o\), since we assume that these people receive bequests only after their parents die, in year \(t + 1\).

We now derive a marginal condition that characterizes equilibrium inter vivos behavior assuming differentiability of the policy functions that describe offspring saving behavior at different ages.

Consider a parent’s problem under laissez-faire of choosing \(d_t\) and \(b_{t+1}^o\) subject to the flow budget constraints and the offspring’s policy functions. Let \(\frac{\partial b_{t+i}(d_t, b_{t+i}^o, Q_t)}{\partial d_t}\) represent how an increase in inter vivos transfers affects savings at age \(i\) calendar year \(t + i - 1\). For

---

Footnote 34: Again, the notation implicitly assumes policies are single valued. The usual caveat applies: whenever we have multiple solutions, the policies should be interpreted as selections from the policy correspondences.
notational simplicity, we will write this partial derivative as $\frac{\partial b_{t+1}}{\partial d_t}$ whenever doing so does not create a confusion. The parent’s first-order optimality condition with respect to the inter vivos decision is

$$u'(c_t^0) = \gamma \left( u'(c_t) \left[ 1 - \frac{\partial b_{t+1}}{\partial d_t} \right] + \sum_{i=1}^{l-1} \xi^i u'(c_{t+i}) \frac{\partial c_{t+i}}{\partial d_t} + \delta^i V_1(a_{t+1}, Q_{t+1}) R_{t+1} \frac{\partial b_{t+1}}{\partial d_t} \right),$$

(26)

where

$$\frac{\partial c_{t+i}}{\partial d_t} = \left[ R_{t+i} \frac{\partial b_{t+i}}{\partial d_t} - \frac{\partial b_{t+i+1}}{\partial d_t} \right]$$

(27)

and, clearly, the derivatives are evaluated at the equilibrium allocation and prices. Consider the problem of the offspring in the first period of his adult life, in period $t$. When choosing his level of savings, $b_{t+1}$, this agent faces the policy functions $(b_{t+2}(b_{t+1}), ..., b_{t+l}(b_{t+1}))$. Taking first-order condition for his period $t$ saving, we get

$$u'(c_t) = \beta_1 \delta \left\{ \sum_{i=1}^{l-1} \delta_i u'(c_{t+i}) \frac{\partial c_{t+i}}{\partial b_{t+1}} + \delta^l V_1(a_{t+1}, Q_{t+1}) R_{t+1} \frac{\partial b_{t+1}}{\partial b_{t+1}} \right\},$$

(28)

where, again, the derivatives are evaluated at the equilibrium allocation and prices. Using (28) in parental optimality condition for inter vivos, (26), and the fact that the only way $d_t$ affects decisions from $t + 1$ onward is through its effect on period $t$ decisions, meaning

$$\frac{\partial b_{t+i}}{\partial d_t} = \frac{\partial b_{t+i}}{\partial b_{t+1}} \frac{\partial b_{t+1}}{\partial d_t},$$

we get the following proposition which describes equilibrium parental inter vivos behavior under differentiability of policy functions.

**Proposition 9.** Suppose the policy functions that describe offspring behavior over the life cycle are differentiable. The equilibrium inter vivos behavior is characterized by

$$u'(c_t^0) = \gamma \left( u'(c_t) + \frac{\partial b_{t+1}}{\partial d_t} \frac{\partial b_{t+1}}{\partial d_t} u'(c_t) \left\{ -1 + \frac{1}{\beta_1} \right\} \right).$$

(29)

Condition (29) is analogous to (10), the optimality condition for bequests in the benchmark model under differentiability. The left-hand side is the marginal cost of increasing inter vivos transfers, which equals the utility loss from forgone parental consumption. The
first term on the right-hand side is the usual marginal benefit of increasing saving – the utility gain from increased consumption in the period during which returns to savings are received. The second term on the right-hand side summarizes how increasing inter vivos transfers affects parental welfare by affecting future consumption levels of the offspring. As long as $\beta_1 < 1$, meaning that the offspring faces present-bias problems in his first period of life, this term is positive: there is an additional marginal benefit of increasing transfers for the parent. Intuitively, the parent, being fully sophisticated, knows that from his perspective, self 1 is undersaving. So, parental welfare increases if the parent can make self 1 increase period 1 savings, which is possible by increasing inter vivos transfers since $\frac{\partial b_{t+1}(d_t)}{\partial d_t} > 0$.

To see the implication of Proposition 9 for Ramsey inter vivos taxation, first observe that in the Ramsey allocation we have

$$u'(c_{1t}^*) = \gamma u'(c_{1t}^*) \quad (30)$$

Now define the Ramsey inter vivos wedge as:

$$IW_t^* = -u'(c_{1t}^*) + \gamma \left( u'(c_{1t}^*) + \frac{\partial b_{t+1}(d_t^*, Q_t^*)}{\partial d_t} u'(c_{1t}^*) \left\{ -1 + \frac{1}{\beta_1} \right\} \right).$$

The Ramsey inter vivos wedge in period $t$ measures the distortion that the planner needs to create in the inter vivos decision of a parent in that period in order to make him choose the Ramsey level of inter vivos transfers. A positive (negative) $IW_t^*$ means that a parent would like to increase (decrease) his inter vivos transfers marginally above (below) the Ramsey level if there is no government intervention.

Corollary 10. Suppose the policy functions that describe offspring behavior over the life cycle are differentiable. Then, $IW_t^* > 0$ for all parenthood periods $t$.

Proof. Follows directly from (30), the definition of $IW_t^*$, and that $\frac{\partial b_{t+1}(d_t^*, Q_t^*)}{\partial d_t} > 0$, which follows from Lemma 12 in the Appendix. \qed

Corollary 10, which is analogous to Corollary 4 for the case of bequests in the main text, establishes that parents would increase their inter vivos transfers above the Ramsey level if there is no government intervention. One can further show that an analog to Proposition 5 also holds for the environment with multiperiod life cycle and coexistence: under the assumption that (Markov) equilibrium (with linear taxes) optimal policies are linear in
current wealth, there is a linear tax system that implements the Ramsey allocation, and in
this tax system, taxes on inter vivos is strictly positive.

The argument for positive bequest taxation in a multiperiod life-cycle environment is
identical to the case for inter vivos taxation and therefore will be omitted for the sake of
brevity. One can show optimality of bequest taxation by plugging the first-order optimality
condition of offspring for period 2 savings, $b_{t+2}$, into the first-order optimality condition of
bequests.

5.4 Long-Term Assets and Liquidity Constraints

The front-loading of consumption by offspring relative to what their parents prefer lies at
the heart of our transfer taxation results. A natural question then is: if the parents have
access to assets with more than one-period maturities, can they force their offspring into
the consumption patterns they want by carefully choosing the portfolio of these assets?
This is an important question because, if they can, then there would be no disagreement
in saving behavior from parent’s perspective. That would imply that the disagreement
wedge disappears which then would imply there is no need to tax transfers.

First, observe that if the offspring are not liquidity constrained, meaning that they can
borrow as much as they want within their natural borrowing limits, then the timing of
transfers cannot constrain their consumption patterns at all. In this case, the strategy of
using long-term assets is fruitless in disciplining offspring’s saving behavior, and we are
back at the benchmark environment without long-term assets: both bequests and inter
vivos transfers should be taxed.\footnote{For a formal analysis of this claim, see Pavoni and Yazici (2015) section 11.E.}
The real question, then, is what happens if parents have
access to long-term assets and children face liquidity constraints? This is the question we
pick up in this section. To sum up, we find that, as long as the parents do not have access
to a portfolio of assets that allow them to target transfers to each and every period of the
offspring’s life cycle, transfer taxation remains to be optimal.

We use the multi-period environment laid out in Section 5.3 for our analysis. To make
things simple, suppose people are not allowed to borrow at all. Suppose that in addition
to inter vivos transfers and bequests, parents can use an illiquid asset to transfer resources
directly to period $t + 2$. Let $b^0_{t+2}$ denote the amount of this illiquid bequest where the sub-
script refers to the period in which the child receives the bequest. Suppose the return to
this asset is $R_{t+1}R_{t+2}$. The budget constraints then are:

$$
\begin{align*}
    c_t^0 &= R_t b_t - d_t - b_{t+1}^0 - b_{t+2}^0, \\
    c_t &= d_t + w_t - b_{t+1}, \\
    c_{t+1} &= R_{t+1}b_{t+1} + w_{t+1} + R_{t+1}b_{t+1}^0 - b_{t+2}, \\
    c_{t+2} &= R_{t+2}b_{t+2} + w_{t+2} + R_{t+1}R_{t+2}b_{t+2}^0 - b_{t+3}, \\
    c_{t+i-1} &= R_{t+i-1}b_{t+i-1} + w_{t+i-1} - b_{t+i}, \text{ for } 4 \leq i \leq I.
\end{align*}
$$

Let $[b_{t+1}(d_t), b_{t+2}(d_t, b_{t+1}^0, b_{t+2}^0), ..., b_{t+1}(d_t, b_{t+1}^0, b_{t+2}^0)]$ denote the functions that describe how the offspring’s savings choices over the life cycle depend on parental transfers they receive, where we simplified notation by ignoring the dependence on the sequence of prices.

For the sake of argument, suppose that people do not have any labor income, meaning $w_t = 0$ for all $t$. In this case, one can show that parents use the transfers to keep offspring borrowing constrained in periods $t$ and $t + 1$. The intuition is simple. Suppose the offspring is not constrained in period $t$ in equilibrium. This means his optimal saving level $b_{t+1}$ is given by (28). Now, if the parent increases his bequests and decreases his inter vivos transfers in a way that keeps the total amount of transfers unchanged, the child will have to decrease his period $t$ savings to keep his preferred allocation. If the parent continues to backload transfers, there will be a point at which the child will deplete his savings completely and will become borrowing constrained. From this point onward, backloading of transfers strictly increases parental welfare, and the parent does this until the equilibrium saving behavior of the offspring is exactly in line with that of the parent:

$$
u'(c_t) = \delta \left\{ \sum_{i=1}^{I-1} \delta^{i-1} u'(c_{t+i}) \frac{\partial c_{t+i}}{\partial b_{t+1}} + \delta^{I-1} V_1(a_{t+1}, Q_{t+1}) R_{t+1} \frac{\partial b_{t+1}}{\partial b_{t+1}} \right\}, \tag{31}$$

where the term $\frac{\partial c_{t+i}}{\partial b_{t+1}}$ is defined analogous to (27). Similarly, the parent will use liquid and illiquid bequests to align offspring saving behavior with his own in period $t + 1$:

$$
u'(c_{t+1}) = \delta \left\{ \sum_{i=2}^{I-2} \delta^{i-2} u'(c_{t+i}) \frac{\partial c_{t+i}}{\partial b_{t+2}} + \delta^{I-2} V_1(a_{t+1}, Q_{t+1}) R_{t+1} \frac{\partial b_{t+1}}{\partial b_{t+2}} \right\}. \tag{32}$$

\[36\] Notice that a change in illiquid bequests affects the offspring’s saving decisions in periods $t + 1$ and $t + 2$, but not his saving decision in period $t$. This is due to our assumption that children learn the levels of both liquid and illiquid bequests in period $t$, after they make their period $t$ savings decision.
This implies that the parent does not have a motive to transfer too much and hence there is no need to distort transfers. For inter vivos transfers, this can be seen by plugging (31) into the parental optimality condition for inter vivos transfers, equation (26). For bequests, it can be shown by plugging (32) into the parental optimality condition for bequests.

We now argue that even in this case, it is optimal to tax parental transfers: not inter vivos or bequests but illiquid bequests. Notice that the parent does not have a long-term asset that pays in period \( t + 3 \). As a result, the offspring will save according to his present-bias problem in period \( t + 2 \) and choose \( b_{t+3} \) to satisfy

\[
u'(c_{t+2}) = \beta_3 \delta \left\{ \sum_{i=3}^{l-1} \delta^{i-3} u'(c_{t+i}) \frac{\partial c_{t+i}}{\partial b_{t+3}} + \delta^{l-3} V_1(a_{t+1}, Q_{t+1}) R_{t+1} \frac{\partial b_{t+1}}{\partial b_{t+3}} \right\}.
\]

(33)

Plugging (31), (32), and (33), in the parent’s optimality condition regarding illiquid bequests, we get:

\[
u'(c_i^o) = \gamma \left( R_{t+1} R_{t+2} \delta^2 u'(c_{t+2}) + \frac{\partial b_{t+3}}{\partial b_{t+2}} \delta u'(c_{t+2}) \left[ -1 + \frac{1}{\beta_3} \right] \right).
\]

(34)

The second term on the right-hand side is positive, which means that parents do too much illiquid bequeathing and therefore should be taxed.

Intuitively, since the parent cannot transfer resources to period \( t + 3 \) (the fourth period of the offspring’s life) directly, there is no way to discipline period \( t + 2 \) savings of the offspring via the timing of transfers. Therefore, the parent still bequests too much, this time through the illiquid asset, and hence, the optimal tax on illiquid bequests is positive. If there were a perfect set of long-term assets, \( \{b_i^o\}_{i=1}^l \), available to the parent, only in that case would the parent be able to control the offspring’s consumption completely and would not need to transfer too much relative to the Ramsey allocation through any of these assets. Only in that case would the optimality of transfer taxation break.

\[\text{For an exact derivation of (34), see Appendix C.}\]

\[\text{Remember that we assumed that the offspring have no labor income of their own, which – together with the no borrowing constraint – allowed parents to control the levels of the first two periods of consumption. When offspring receive labor income over their life cycle, then, depending on the shape of the wage profile over the life cycle, it might not be possible for parents to drive children to behave according to the Euler equations (31) and (32) even if parents have access to a perfect portfolio of assets. Alternatively, if the borrowing limits are not exactly zero but allow for some undersaving behavior for the offspring, then, depending on the wage profile over the life cycle, again parents may not be able to arrange children’s first and second period consumption according to (31) and (32). In these cases, parents will have motives to transfer too much through inter vivos and liquid bequests, and the optimal taxes on these transfers should again be positive.}\]
6 Conclusion

We study optimal taxation of parental transfers in a model where altruistic parents and their offspring disagree on intertemporal trade-offs. We prove that laissez-faire equilibrium is inefficient in a Pareto sense due to disagreement. We focus on the Ramsey allocation, the allocation that maximizes the welfare of the initial parent, as our benchmark Pareto efficient allocation. We show that, if offspring are impatient from parents’ perspective, then parents bequeath too much relative to Ramsey level, and hence, it is optimal to tax bequests. We then consider policies that target any point on the Pareto frontier of the economy. We find that intergenerational disagreement again call for a taxation of bequests. In this case, there is another component of bequest taxation which comes from intergenerational redistribution and this component always asks for a subsidy on bequests. The sign of the tax depends on which component dominates. We also consider an economy with income inequality across families in which government wants to redistribute from the rich to the poor. We find that if labor income taxes are distortionary and society puts direct weight on future agents, then bequest taxes are progressive.

Finally, we show that the optimality of positive taxes on transfers remains valid even when (i) we consider life cycles with arbitrarily long finite horizons, (ii) when parents can trade long maturity assets.

In Appendix D we also show that our result survives even when parents are naïve in the sense that they do not realize the existence of intergenerational disagreements.

PRICE EXTERNALITY ??

References


42
International Studies Program Working Paper 10-17, Andrew Young School of Policy Studies.


A Appendix: Proofs

A.1 Proof of Proposition 1

Before starting the proof, we demonstrate two preliminary lemmas which we use in the proof later on.

**Lemma 11.** The policy of the offspring \( b_{t+2}(\cdot, Q_{t+1}) \) is increasing in the amount of bequests received and the value function \( V(\cdot, Q_t) \) is strictly increasing in wealth.

**Proof.** It is easy to see that \( V(\cdot, Q_t) \) is strictly increasing. A higher amount of assets enlarges the constraint set of the parent with at least one allocation that strictly improves his welfare: the one in which he consumes all the extra wealth in period \( t \).

The monotonicity of \( b_{t+2}(\cdot, Q_{t+1}) \) is shown as follows. The definition of the offspring’s saving policy at \( b_{t+1} \) implies that, for \( \epsilon > 0 \), we have

\[
    u(R_{t+1}b_{t+1} + w_{t+1} - b_{t+2}(b_{t+1}, Q_{t+1})) - u(R_{t+1}b_{t+1} + w_{t+1} - b_{t+2}(b_{t+1} + \epsilon, Q_{t+1})) \geq \beta\delta[V(R_{t+2}b_{t+2}(b_{t+1} + \epsilon, Q_{t+1}) + w_{t+2}, Q_{t+2}) - V(R_{t+2}b_{t+2}(b_{t+1}, Q_{t+1}) + w_{t+2}, Q_{t+2})].
\]

Using the definition of the policy for the offspring at \( b_{t+1} + \epsilon \), we have

\[
    u(R_{t+1}(b_{t+1} + \epsilon) + w_{t+1} - b_{t+2}(b_{t+1}, Q_{t+1})) - u(R_{t+1}(b_{t+1} + \epsilon) + w_{t+1} - b_{t+2}(b_{t+1} + \epsilon, Q_{t+1})) \leq \beta\delta[R_{t+2}V(b_{t+2}(b_{t+1} + \epsilon, Q_{t+1}) + w_{t+2}, Q_{t+2}) - V(R_{t+2}b_{t+2}(b_{t+1}, Q_{t+1}) + w_{t+2}, Q_{t+2})].
\]

Combining (35) and (36), we get

\[
    u(R_{t+1}b_{t+1} + w_{t+1} - b_{t+2}(b_{t+1}, Q_{t+1})) - u(R_{t+1}b_{t+1} + w_{t+1} - b_{t+2}(b_{t+1} + \epsilon, Q_{t+1})) \geq u(R_{t+1}(b_{t+1} + \epsilon) + w_{t+1} - b_{t+2}(b_{t+1}, Q_{t+1})) - u(R_{t+1}(b_{t+1} + \epsilon) + w_{t+1} - b_{t+2}(b_{t+1} + \epsilon, Q_{t+1})).
\]

Assume for the sake of contradiction that \( b_{t+2}(b_{t+1} + \epsilon, Q_{t+1}) < b_{t+2}(b_{t+1}, Q_{t+1}) \). Combined with strict concavity of \( u \), this contradicts with (37). Thus, it must be that \( b_{t+2}(b_{t+1} + \epsilon, Q_{t+1}) \geq b_{t+2}(b_{t+1}, Q_{t+1}) \).

**Lemma 12.** Suppose the value function \( V(\cdot, Q_t) \) is differentiable. Then, \( b_{t+2}(\cdot, Q_{t+1}) \) is strictly monotone in the amount of bequests received.

**Proof.** Due to Inada assumption on the utility function, the solution to the offspring’s problem stated in (2) must be interior. A necessary condition for the optimality of offspring’s savings is then

\[
    u'(R_{t+1}b_{t+1} + w_{t+1} - b_{t+2}) = \beta\delta R_{t+1}v_1(R_{t+2}b_{t+2} + w_{t+2}, Q_{t+2})R_{t+2}.
\]
Now suppose the offspring receives a higher level of bequests from the parent, $b_{t+1} + x$, where $x > 0$. It would never be optimal for the offspring to use all of the increase in his wealth for current consumption since this does not satisfy his necessary condition for optimality:

$$u'(R_{t+1} (b_{t+1} + x) + w_{t+1} - b_{t+2}) < \beta \delta R_{t+1} V_1 (R_{t+2} b_{t+2} + w_{t+2}, Q_{t+2}) R_{t+2},$$

which follows from (38) and the strict concavity of the utility function.

**Proof. (Core Proof of Proposition)** In the proof of this proposition, we use monotonicity of $b_{t+2} (\cdot, Q_{t+1})$ and strict monotonicity of $V (\cdot, Q_t)$, which we establish in Lemma 11.

Assume for the sake of finding a contradiction that in equilibrium

$$- u' (R_t b_t + w_t - b_{t+1}) + \gamma R_{t+1} u' (R_{t+1} b_{t+1} + w_{t+1} - b_{t+2} (b_{t+1}, Q_{t+1})) > 0. \quad (39)$$

We want to show that there is a small positive $\epsilon > 0$ such that

$$u (R_t b_t + w_t - b_{t+1} - \epsilon) + \gamma \left[ u (R_{t+1} (b_{t+1} + \epsilon) + w_{t+1} - b_{t+2} (b_{t+1} + \epsilon, Q_{t+1})) + \delta V (R_{t+2} b_{t+2} (b_{t+1} + \epsilon, Q_{t+1}) + w_{t+2}, Q_{t+2}) \right]$$

$$> u (R_t b_t + w_t - b_{t+1}) + \gamma \left[ u (R_{t+1} b_{t+1} + w_{t+1} - b_{t+2} (b_{t+1}, Q_{t+1})) + \delta V (R_{t+2} b_{t+2} (b_{t+1}, Q_{t+1}) + w_{t+2}, Q_{t+2}) \right].$$

Since $u$ is a differentiable function, for $\epsilon$ sufficiently small, under the assumption (39), we have

$$u (R_t b_t + w_t - b_{t+1} - \epsilon) + \gamma u (R_{t+1} (b_{t+1} + \epsilon) + w_{t+1} - b_{t+2} (b_{t+1}, Q_{t+1}))$$

$$> u (R_t b_t + w_t - b_{t+1}) + \gamma u (R_{t+1} b_{t+1} + w_{t+1} - b_{t+2} (b_{t+1}, Q_{t+1})).$$

Then, we have

$$u (R_t b_t + w_t - b_{t+1} - \epsilon) + \gamma \left[ u (R_{t+1} (b_{t+1} + \epsilon) + w_{t+1} - b_{t+2} (b_{t+1} + \epsilon, Q_{t+1})) + \delta V (R_{t+2} b_{t+2} (b_{t+1} + \epsilon, Q_{t+1}) + w_{t+2}, Q_{t+2}) \right]$$

$$\geq u (R_t b_t + w_t - b_{t+1} - \epsilon) + \gamma \left[ u (R_{t+1} b_{t+1} + w_{t+1} - b_{t+2} (b_{t+1}, Q_{t+1})) + \delta V (R_{t+2} b_{t+2} (b_{t+1}, Q_{t+1}) + w_{t+2}, Q_{t+2}) \right] + (1 - \beta) \delta V (R_{t+2} b_{t+2} (b_{t+1} + \epsilon, Q_{t+1}) + w_{t+2}, Q_{t+2})$$

$$> u (R_t b_t + w_t - b_{t+1}) + \gamma \left[ u (R_{t+1} b_{t+1} + w_{t+1} - b_{t+2} (b_{t+1}, Q_{t+1})) + \delta V (R_{t+2} b_{t+2} (b_{t+1}, Q_{t+1}) + w_{t+2}, Q_{t+2}) \right] + (1 - \beta) \delta V (R_{t+2} b_{t+2} (b_{t+1}, Q_{t+1}) + w_{t+2}, Q_{t+2}),$$

where the first inequality follows from the definition of the policy $b_{t+2} (\cdot, Q_{t+1})$ and the monotonicity of both $b_{t+2} (\cdot, Q_{t+1})$ and $V (\cdot, Q_t)$, the second inequality follows from (39) and the differentiability of the utility function $u$ as discussed above, and the last one from the monotonicity of both
Proof of Proposition 2

We start the proof with a preliminary lemma.

**Lemma 13.** Let \( \{c_t, k_t\}_{t=0}^{\infty} \) and (interior) equilibrium allocation. Let \( c_{s+1} \) and \( c_{s+2} \) be the consumption of an agent in his two years of life (with \( c_{s+1} \) consumption when young, and \( c_{s+2} \) when parent), and \( b_{s+2} \) the agent’s level of saving from period \( s+1 \) to \( s+2 \), which equals in equilibrium the level of capital generated \( k_{s+2} \). Then, the following condition has to hold in equilibrium:

\[
-u'(c_{s+1}) + \beta \delta f'(k_{s+2}) u'(c_{s+2}) = 0.
\]

**Proof.** Recall that our equilibrium allocations are interior. The key point here is that the same individual is choosing both \( b_{s+2} \) and \( c_{s+2} \). In particular, the agent can always consider the usual Euler perturbation. The optimal choice of \( b_{s+2} \) must be such that there is no feasible deviation around \( b_{t+2} \) (keeping both \( b_{s+1} \) and \( b_{s+3} \) unchanged) that improves over \( b_{t+2} \). The agent optimal decision must be such that the problem

\[
\max_{\tilde{x} \in N} u\left( R_{s+1} b_{s+1} + w_{s+1} - b_{s+2} - \tilde{x}\right) + \delta \beta u\left( R_{s+2} (b_{s+2} + \tilde{x}) + w_{s+2} - b_{s+3}\right)
\]

where \( N \) is an open interval around zero, has solution \( \tilde{x}^* = 0 \). Since \( u(\cdot) \) is differentiable, and recalling that \( R_{s+2} = f'(k_{s+2}) \) and \( b_{s+2} = k_{s+2} \), the necessary condition for \( \tilde{x}^* = 0 \) to be optimal is the first order condition displayed in the statement of the lemma.

We will prove the statement for the case \( \beta < 1 \). The proof of the case of \( \beta > 1 \) is symmetric and will be sketched next. For notational simplicity, we will make the perturbation for consumption in periods 0,1, and 2. Define a perturbation by \( \tilde{c}_0 = c_0 - \varepsilon, \tilde{c}_1 = c_1 + f(k_1 + \varepsilon) - f(k_1) - \eta, \) and \( \tilde{c}_2 = c_2 + f(k_2 + \eta) - f(k_2) \).

**Proof.** (Core Proof of Proposition 2) We will show that there exist \( \varepsilon \) and \( \eta \) such that this perturbation makes both the initial parent and the agent who is born in period 1 strictly better off. The welfare of future generations will not be altered since the perturbation only involves periods 0,1, and 2. First, we show that the offspring’s welfare increases. We aim at showing that \( u(c_1 + f(k_1 + \varepsilon) - f(k_1) - f(k_2)) > u(c_1) \).
\( \eta + \delta \beta u(c_2 + f(k_2 + \eta) - f(k_2)) - [u(c_1) + \delta \beta u(c_2)] > 0. \) Letting \( \epsilon = \zeta \eta \) with \( \zeta > 0 \) to be chosen, the same condition is equivalent to itself divided by the positive number \( \eta \). So, for \( \eta > 0 \) we have

\[
u (c_1 + f(k_1 + \zeta \eta) - f(k_1) - \eta) + \delta \beta u(c_2 + f(k_2 + \eta) - f(k_2)) - [u(c_1) + \delta \beta u(c_2)] > 0
\]

\[
\iff \frac{\nu (c_1 + f(k_1 + \zeta \eta) - f(k_1) - \eta) + \delta \beta u(c_2 + f(k_2 + \eta) - f(k_2)) - [u(c_1) + \delta \beta u(c_2)]}{\eta} > 0.
\]

Since \( u(\cdot) \) and \( f(\cdot) \) are differentiable, we have:

\[
u (c_1 + f(k_1 + \zeta \eta) - f(k_1) - \eta) + \delta \beta u(c_2 + f(k_2 + \eta) - f(k_2)) - [u(c_1) + \delta \beta u(c_2)]
\]

\[
= \left[ (\zeta f'(k_1) - 1)u'(c_1) + \delta \beta f'(k_2)u'(c_2) \right] \eta + O(\eta) \eta,
\]

where, by definition of the remainder \( O(\cdot) \) (as the right derivative exists),

\[
\lim_{\eta \to 0^+, \eta > 0} O(\eta) = 0.
\]

Hence for \( \eta^* > 0 \) small enough we have

\[
\left[ (\zeta f'(k_1) - 1)u'(c_1) + \delta \beta f'(k_2)u'(c_2) \right] \eta^* + O(\eta^*) = \zeta f'(k_1)u'(c_1) + O(\eta^*) > 0,
\]

where we use Lemma 1.313\(^{39}\)

Next, we compute the change in welfare for the parent. Recall, \( \epsilon = \zeta \eta \). Set

\[
l := \max \left\{ u'(c_0) - \gamma \delta \beta f'(k_1)f'(k_2)u'(c_2), \ (1 - \beta) f'(k_2) \gamma \delta u'(c_2) \right\} > 0,
\]

\[
\zeta := \frac{(1 - \beta) \gamma \delta f'(k_2)u'(c_2)}{1 + l} \in (0, 1).
\]

Then, since \( u(\cdot) \) and \( f(\cdot) \) are differentiable, following the same argument as above, we can find a small enough \( \eta \) such that:

\[
u(c_0 - \zeta \eta) + \gamma u(c_1 + f(k_1 + \zeta \eta) - f(k_1) - \eta) + \gamma \delta u(c_2 + f(k_2 + \eta) - f(k_2)) - [u(c_0) + \gamma u(c_1) + \gamma \delta u(c_2)] \]

\[
\eta > 0
\]

if and only if

\[
-\zeta u'(c_0) + \gamma (\zeta f'(k_1) - 1)u'(c_1) + \gamma \delta f'(k_2)u'(c_2) > 0.
\]

\(^{39}\)To see in more detail why we can guarantee the existence of such \( \eta^* \) recall that the definition of \( O(\cdot) \) implies that for all \( \zeta > 0 \) we can find a \( \psi > 0 \) such that for all \( \eta < \psi \) we have \( |O(\eta^*)| < \zeta \). Thus, since \( \zeta f'(k_1)u'(c_1) > 0 \), we can find \( 0 < \zeta^* < \zeta f'(k_1)u'(c_1) \) such that for all \( 0 < \eta^* < \psi \) the requirement is satisfied.
But the latter inequality must be true as we have:

\[-ξu'(c_0) + γ(ξf'(k_1) - 1)u'(c_1) + γδf'(k_2)u'(c_2)\]

\[= -ξ [u'(c_0) - γβδf'(k_1)f'(k_2)u'(c_2)] + (1 - β)γδf'(k_2)u'(c_2) > 0.\]

The equality between the first and second rows uses Lemma 13 to replace \(u'(c_1)\) for \(βδf'(k_2)u'(c_2)\) and rearranging terms whereas the last inequality comes from our definition of \(ξ\) and \(l\) which directly implies \(ξ [u'(c_0) - γβδf'(k_1)f'(k_2)u'(c_2)] < (1 - β)γδf'(k_2)u'(c_2)\).

For the case \(β > 1\), define a perturbation by \(c_0 = c_0 - ε, c_1 = c_1 + f(k_1 + ε) - f(k_1) + η, \) and \(c_2 = c_2 + f(k_2 - η) - f(k_2).\) Following the same steps as above, for small \(η > 0,\) the change in offspring welfare again reduces to (disregarding the remainder \(O(η)\) since it goes to 0 as \(η\) goes to 0) \(u'(c_1)ξf'(k_1),\) which is strictly positive. The change in parental welfare reduces to:

\[-ξ [u'(c_0) - γβδf'(k_1)f'(k_2)u'(c_2)] + (β - 1)γδf'(k_2)u'(c_2) > 0,\]

which is strictly positive since both of the terms in this expression are positive and at least one of them is strictly positive. \(\square\)

### A.3 CEIS utility functions

**Proof of Proposition 6**

*Proof.* Given any joint sequence of prices and taxes \(Ψ,\) let

\[Γ_s(b) = R_s(1 - τ_s)b + w_s + T_s + G_s\]

be the net present value of wealth as of the beginning of period \(s\) of an agent who saved \(b\) units in the previous period (of course, we only consider prices and taxes such that this sum converges), where \(G_s\) denotes the net present value of wages and lump-sum taxes from period \(s + 1\) onwards

\[G_s = \sum_{m=1}^{∞} \frac{w_s+m + T_s+m}{\prod_{n=1}^{m} R_{s+n}(1 - τ_{s+n})} = \frac{w_{s+1} + T_{s+1} + G_{s+1}}{R_{s+1}(1 - τ_{s+1})}.\]

We are going to construct an equilibrium in which agents’ policies are linear in the current net present value of wealth. We do so in three steps.

**Step 1.**

We first guess that the value function of the parent has the form

\[V(R_s(1 - τ_s)b + w_s + T_s, Ψ_s) = \hat{V}(Γ_s(b), Ψ_s),\]
where $\hat{V}$ is homogeneous of degree $1 - \rho \leq 1$ in period $s$ net present value of wealth, that is

$$\hat{V}(\lambda \Gamma_s(b), \Psi_s) = \lambda^{1-\rho} \hat{V}(\Gamma_s(b), \Psi_s), \quad \forall \lambda > 0.$$  

**Step 2.**

We now show that, given this guess about the value function, the consumption policy of period $s$ is linear in $\Gamma_s(b)$ for each period $s$. In Step 3, we will verify that this policy indeed generates a value function that has homogeneity of degree $1 - \rho$ in $\Gamma_s(b)$.

We proceed by backward induction. Consider the problem of an offspring in period $s+1$.

**Claim 1.**

$$\hat{c}_{s+1} \text{ solves max } u(c_{s+1}) + \beta \delta \hat{V}(\Gamma_{s+2}(b_{s+2}), \Psi_{s+2}) \text{ s.t. } c_{s+1} + b_{s+2} = \Gamma_{s+1}(b_{s+1}) - G_{s+1} \text{ if and only if }$$

$$\lambda \hat{c}_{s+1} \text{ solves max } u(c_{s+1}) + \beta \delta \hat{V}(\Gamma_{s+2}(b_{s+2}), \Psi_{s+2}) \text{ s.t. } c_{s+1} + b_{s+2} = \lambda \Gamma_{s+1}(b_{s+1}) - G_{s+1}.$$  

First we show that budget constraint is homogenous of degree 1 in consumption and net present value of wealth. To do so, let $\Gamma_{s+1}(\hat{b}_{s+1})$, be the period $s+1$ net present value of wealth when $\Gamma_s(b_s)$ is period $s$ net present value of wealth and period $s$ consumption choice is $\hat{c}_s$. Now we show that when period $s$ wealth is $\lambda \Gamma_s(b_s)$ and the agent consumes $\lambda \hat{c}_s$ in period $s$, then period $s+1$ net present value of wealth will be $\lambda \Gamma_{s+1}(b_{s+1})$. First, observe that period $s$ saving in the latter case is given by

$$[\lambda \Gamma_s(b_s) - G_s - \lambda \hat{c}_s].$$  

Plugging this value in the definition of net present value of wealth for period $s+1$, we get:

$$R_{s+1}(1 - \tau_{s+1}) \left[ \lambda \Gamma_s(b_s) - G_s - \lambda \hat{c}_s \right] + w_{s+1} + T_{s+1} + G_{s+1} = \lambda \Gamma_s(b_s) R_{s+1}(1 - \tau_{s+1}) - \lambda \hat{c}_s R_{s+1}(1 - \tau_{s+1}) - G_s R_{s+1}(1 - \tau_{s+1}) + T_{s+1} + G_{s+1}$$

$$= \lambda \left[ \Gamma_s(b_{s-1}) R_s(1 - \tau_s) - \hat{c}_s R_s(1 - \tau_s) \right]$$

Suppose $\hat{c}_{s+1}$ solves

$$\text{max } u(c_{s+1}) + \beta \delta \hat{V}(\Gamma_{s+2}(b_{s+2}), \Psi_{s+2})$$

s.t. $c_{s+1} + b_{s+2} = \Gamma_{s+1}(b_{s+1}) - G_{s+1}$, and for the sake of contradiction suppose $\lambda \hat{c}_{s+1}$ does not solve

$$\text{sup } u(c_{s+1}) + \beta \delta \hat{V}(\Gamma_{s+2}(b_{s+2}), \Psi_{s+2})$$

s.t. $c_{s+1} + b_{s+2} = \lambda \Gamma_{s+1}(b_{s+1}) - G_{s+1}$. Let $v^*$ be the solution of the previous supremum problem and
define
\[ \kappa := v^* - \left[ u(\lambda \hat{c}_{s+1}) + \beta \delta \hat{V}(\lambda \Gamma_{s+2}(\hat{b}_{s+2}), \Psi_{s+2}) \right] > 0. \]

Under our assumptions, for each \( \varepsilon > 0 \) there is a feasible \( \hat{c}^{\varepsilon}_{s+1} \) such that
\[ u(\hat{c}^{\varepsilon}_{s+1}) + \beta \delta \hat{V}(\Gamma_{s+2}(\hat{b}^{\varepsilon}_{s+2}), \Psi_{s+2}) > u(\lambda \hat{c}_{s+1}) + \beta \delta \hat{V}(\lambda \Gamma_{s+2}(\hat{b}_{s+2}), \Psi_{s+2}) + \kappa - \varepsilon, \]

where \( \hat{b}^{\varepsilon}_{s+2} \) is adjusted so as to maintain feasibility. By homogeneity of the utility and value functions, for all \( \lambda > 0 \) the previous statement is equivalent to
\[ u \left( \frac{\hat{c}^{\varepsilon}_{s+1}}{\lambda} \right) + \beta \delta \hat{V} \left( \frac{\Gamma_{s+2}(\hat{b}^{\varepsilon}_{s+2})}{\lambda}, \Psi_{s+2} \right) > u(\hat{c}_{s+1}) + \beta \delta \hat{V} \left( \Gamma_{s+2}(\hat{b}_{s+2}), \Psi_{s+2} \right) + \frac{\kappa - \varepsilon}{\lambda}. \]

We also know that if \( \hat{c}^{\varepsilon}_{s+1} \) and \( \Gamma_{s+2}(\hat{b}^{\varepsilon}_{s+2}) \) are feasible in the problem of the agent facing wealth \( \lambda \Gamma_{s+1}(b_{s+1}) \), so are \( \frac{\hat{c}^{\varepsilon}_{s+1}}{\lambda} \) and \( \frac{\Gamma_{s+2}(\hat{b}^{\varepsilon}_{s+2})}{\lambda} \) in the problem of the agent facing \( \Gamma_{s+1}(b_{s+1}) \). Setting \( \varepsilon^* = \frac{\kappa}{2} > 0 \), we obtain a contradiction since \( \frac{\hat{c}^{\varepsilon^*}_{s+1}}{\lambda} \) and \( \frac{\Gamma_{s+2}(\hat{b}^{\varepsilon^*}_{s+2})}{\lambda} \) are feasible and give strictly higher utility to the agent’s problem in year \( s \). The converse of the claim is shown symmetrically.

Now, we prove the second step in the backward induction.

Claim 2.

\[(\hat{c}_s, \hat{c}_{s+1}) \text{ solves } \max u(c_s) + \gamma \left[ u(c_{s+1}) + \delta \hat{V}(\Gamma_{s+2}(b_{s+2}), \Psi_{s+2}) \right] \]
\[ \text{s.t. (1) } c_s + b_{s+1} = \Gamma_s(b_s) - G_s \text{ and (2) } c_{s+1} \text{ solves agent’s problem in year } s + 1 \]
\[ \text{if and only if} \]
\[ (\lambda \hat{c}_s, \lambda \hat{c}_{s+1}) \text{ solves } \max u(c_s) + \gamma \left[ u(c_{s+1}) + \delta \hat{V}(\Gamma_{s+2}(b_{s+2}), \Psi_{s+2}) \right] \]
\[ \text{s.t. (3) } c_s + b_{s+1} = \lambda \Gamma_s(b_s) - G_s \text{ and (4) } c_{s+1} \text{ solves agent’s problem in year } s + 1. \]

Suppose for the sake of contradiction that \((\hat{c}_s, \hat{c}_{s+1})\) solves the corresponding problem but \((\lambda \hat{c}_s, \lambda \hat{c}_{s+1})\) does not. Since the proof follows the same principle as that in the proof of Claim 1, to save notation, we now propose the proof assuming existence of a solution for both problems. This assumption is not needed as we have shown above. Then, there exists a \((\hat{c}_s, \hat{c}_{s+1})\) such that
\[ u(\hat{c}_s) + \gamma \left[ u(\hat{c}_{s+1}) + \delta \hat{V}(\Gamma_{s+2}(\hat{b}_{s+2}), \Psi_{s+2}) \right] > u(\lambda \hat{c}_s) + \gamma \left[ u(\lambda \hat{c}_{s+1}) + \delta \hat{V}(\lambda \Gamma_{s+2}(\hat{b}_{s+2}), \Psi_{s+2}) \right], \]
and (3) and (4) are satisfied. By homogeneity of utility and value functions, we have
\[ u \left( \frac{\hat{c}_s}{\lambda} \right) + \gamma \left[ u \left( \frac{\hat{c}_{s+1}}{\lambda} \right) + \delta \hat{V} \left( \frac{\Gamma_{s+2}(\hat{b}_{s+2})}{\lambda}, \Psi_{s+2} \right) \right] > u(\hat{c}_s) + \gamma \left[ u(\hat{c}_{s+1}) + \delta \hat{V}(\Gamma_{s+2}(\hat{b}_{s+2}), \Psi_{s+2}) \right]. \]
Furthermore, as we have shown in the first step of the induction, if \( \bar{c}_{s+1} \) solves the problem in year \( s + 1 \) under \( \Gamma_{s+1}(b_{s+1}) \), then \( \frac{\bar{c}_{s+1}}{\lambda} \) solves the same problem under \( \Gamma_{s+1}(b_{s+1}) \). This means that \( \left( \frac{c_s}{\lambda}, \frac{\bar{c}_{s+1}}{\lambda} \right) \) is in the constraint set of the agent’s problem in year \( s \), which combined with the fact that it gives strictly higher welfare than the equilibrium allocation implies a contradiction.

Step 3.

Now, we verify that, under consumption policies that are linear in the current wealth, the value function is in fact homogeneous of degree 1 in \( \Gamma \), as assumed:

\[
V(\lambda \Gamma_s(b), \Psi_s) = u(\lambda \hat{c}_s) + \gamma \left[ u(\lambda \hat{c}_{s+1}) + \delta V(\lambda \Gamma_{s+2}(\hat{b}_{s+2}), \Psi_{s+2}) \right] = \lambda^{1-\rho} V(\Gamma_s(b), \Psi_s).
\]

To complete the proof, observe that we have shown that consumption defined as a function of net present value of wealth, prices, and taxes, denote it by \( c_s(\Gamma_s(b), \Psi_s) \), satisfies the following homogeneity of degree one in wealth:

\[
c_s(\lambda \Gamma_s(b), \Psi_s) = \lambda c_s(\Gamma_s(b), \Psi_s).
\]

In particular this implies

\[
c_s(\Gamma_s(b), \Psi_s) = \Gamma_s(b) c_s(1, \Psi_s),
\]

which means consumption is a linear function of wealth with a constant multiplier of \( c_s(1, \Psi_s) \).

Using the homogeneity of the value function, we can show that the value function has the following simple form as a function of wealth:

\[
V(\Gamma_s(b), \Psi_s) = \Gamma_s(b)^{1-\rho} V(1, \Psi_s).
\]

This ends the proof. Note in particular, that since we are driving conditions for an equilibrium, no verification stage is needed.

A.4 The Logarithmic Case

Proof of Proposition 7

Proof. The proof proceeds in two steps. In the first step, for logarithmic utility, we compute closed form solution for equilibrium linear consumption policy as a function of net present value of wealth.
and tax-price sequence. In the second step, we use the consumption policy computed in step one to compute \( M^\ast_{t+2} \) and plug that in the tax formula in Proposition 5.

Step 1. We use guess and verify method to compute value and policy functions. First, remember from the proof of Proposition 3 that, given any joint sequence of taxes and prices \( \Psi \), we can write the parent’s value function as a function of his current net present value of wealth \( \tilde{V}(\Gamma_t(b_t), \Psi_t) \), where \( \Gamma_t(b_t) \) represents the current net present value of wealth of a parent who saved \( b_t \) units during his young adulthood in period \( t-1 \). (Observe that in fact \( \Gamma_t(b_t) \) also depends on the tax-price sequence \( \Psi_t \), however, we omit to make this dependence explicit in order to ease notation). Now, we guess that the value function has the following form:

\[
\tilde{V}(\Gamma_t(b_t), \Psi_t) = D \log(\Gamma_t(b_t)) + B(\Psi_t),
\]

where \( D \) is the constant of the parent’s value function.

By assumption we are interested in equilibria where policies are linear in net present value of wealth. Therefore, let consumption in period \( t \) under wealth \( \Gamma_t(b_t) \) be given by

\[
c_t(\Gamma_t(b_t), \Psi_t) = C_t(\Psi_t)\Gamma_t(b_t),
\]

where \( C_t(\Psi_t) \) is the fraction of wealth consumed by agent under \( \Psi_t \). In what follows, we omit the dependence of \( C_t \) on \( \Psi_t \) in order to ease notation. Using linearity of the policy functions, we can rewrite the parent’s problem as:

\[
\tilde{V}(\Gamma_t(b_t), \Psi_t) = \max_{C_t} u(C_t\Gamma_t(b_t)) + \gamma[u(C_{t+1}\Gamma_{t+1}(b_{t+1})) + \delta \tilde{V}(\Gamma_{t+2}(b_{t+2}), \Psi_{t+2})] \quad (40)
\]

\[
s.t.
\]

\[
u'(C_{t+1}\Gamma_{t+1}(b_{t+1})) = \delta \beta \tilde{V}_1(\Gamma_{t+2}(b_{t+2}), \Psi_{t+2}) R_{t+2}(1 - \tau_{t+2}). \quad (41)
\]

Note that net present value of wealth in two consecutive periods are linked as follows:

\[
\Gamma_{t+1}(b_{t+1}) = R_{t+1}(1 - \tau_{t+1})b_{t+1} + w_{t+1} + T_{t+1} + G_{t+1}
\]

\[
= R_{t+1}(1 - \tau_{t+1})[R_t(1 - \tau_t)b_t + w_t + T_t - C_t\Gamma_t(b_t)] + w_{t+1} + T_{t+1} + G_{t+1}
\]

\[
= R_{t+1}(1 - \tau_{t+1})\left[R_t(1 - \tau_t)b_t + w_t + T_t - C_t\Gamma_t(b_t) + \frac{w_{t+1} + T_{t+1} + G_{t+1}}{R_t(1 - \tau_t)}\right]
\]

\[
= R_{t+1}(1 - \tau_{t+1})[R_t(1 - \tau_t)b_t + T_t - C_t\Gamma_t(b_t) + G_t]
\]

\[
= R_{t+1}(1 - \tau_{t+1})\Gamma_t(b_t)[1 - C_t]. \quad (42)
\]

Plugging the value function guess in the constraint of the planning problem, (41), and using \( u(\cdot) = \log \), we get:
\[(C_{t+1} \Gamma_{t+1} (b_{t+1}))^{-1} = \frac{\delta \beta R_{t+2} (1 - \tau_{t+2}) D}{\Gamma_{t+2} (b_{t+2})},\]

where the second equality follows from the relationship between consecutive wealth levels that we just established. This implies

\[
(C_{t+1})^{-1} = \frac{\delta \beta D}{1 - C_{t+1}}
\]
or

\[
C_{t+1}(D) = \frac{1}{1 + \delta \beta D}.
\]

Taking first-order condition with respect to \(C_t\) in the parent’s problem and using again (42), we get:

\[
C_t(D) = \frac{1}{1 + \gamma (1 + \delta D)}.
\]

Now verify the value function to compute \(D\):

\[
D \log(\Gamma_t (b_t)) + B(\Psi_t) = \log (C_t(D) \Gamma_t (b_t)) + \gamma [\log (C_{t+1}(D) \Gamma_{t+1} (b_{t+1})) + \delta \{D \log (\Gamma_{t+2} (b_{t+2})) + B(\Psi_{t+2})]\},
\]

which, using (42) and comparing the coefficients of \(\log(\Gamma_t (b_t))\) on both sides of the above equation, implies

\[
D = \frac{1 + \gamma}{1 - \gamma \delta}.
\]

**Step 2.** Remember that \(b_{t+2}(b_{t+1}, \Psi_{t+1}) = R_{t+1}(1 - \tau_{t+1})b_{t+1} + w_{t+1} + T_{t+1} - C_{t+1} \Gamma_{t+1}(b_{t+1}).\) Thus, we get

\[
M_{t+2}^* \equiv \frac{db_{t+2}(b_{t+1}, \Psi_{t+1})}{db_{t+1}} = R_{t+1}(1 - \tau_{t+1}) - C_{t+1}R_{t+1}(1 - \tau_{t+1})
\]

\[
= R_{t+1}(1 - \tau_{t+1}) \frac{\delta \beta D}{1 + \delta \beta D}.
\]

Plugging this \(M_{t+2}^*\) value in the tax formula in Proposition 5 and doing some simple manipulation, we get the result.

\[\square\]

**Proof of Proposition 8** We first prove Proposition 14 below which is of independent interest and is used for our computations in Table 2.
Proposition 14. The Ramsey allocation Pareto improves over laissez-faire market equilibrium allocation if and only if

\[
\log \left( \frac{(1 - \gamma \delta)(1 + \gamma - \delta + \beta \delta)}{(1 + \gamma)(1 - \gamma \delta \beta - \delta + \beta \delta)} \right) + \log A + \log(1 - \delta + \beta \delta)B - \log(\beta)C \geq 0, \tag{44}
\]

where

\[
A = 1 + \frac{\beta \delta}{1 - \gamma \delta}(1 + \gamma),
\]
\[
B = 1 + \frac{\beta \delta}{1 - \gamma \delta}(1 + \gamma + \gamma),
\]
\[
C = \frac{\beta \delta}{1 - \gamma \delta}(1 + \gamma).
\]

Proof. Step 1. We first compute the Ramsey allocation. The euler equations that characterize Ramsey equilibrium are: for \(t = 0, 2, 4, \ldots\)

\[
\begin{align*}
&u'(c_t^*) = \gamma R_{t+1}u'(c_{t+1}^*), \\
&u'(c_{t+1}^*) = \delta R_{t+2}u'(c_{t+2}^*).
\end{align*}
\]

Since utility is logarithmic, this implies that for \(t = 0, 2, 4, \ldots\)

\[
\begin{align*}
c_{t+1}^* &= c_t^*x_{t+1}, \\
c_{t+2}^* &= c_{t+1}^*y_{t+2},
\end{align*}
\]

where:

\[
\begin{align*}
x_{t+1} &= \gamma R_{t+1}, \\
y_{t+2} &= \delta R_{t+2}.
\end{align*}
\]

Equivalently, we can write

\[
\begin{align*}
c_1^* &= c_0^*x_1, \\
c_2^* &= c_0^*x_1y_2, \\
c_3^* &= c_0^*x_1y_2x_3, \\
c_4^* &= c_0^*x_1y_2x_3y_4, \\
&\quad \ldots
\end{align*}
\]

Using this in the budget constraint, we get

\[
c_0^*(1 + \frac{x_1}{R_1} + \frac{x_1y_2}{R_1R_2} + \frac{x_1y_2x_3}{R_1R_2R_3} + \frac{x_1y_2x_3y_4}{R_1R_2R_3R_4} + \ldots) = R_0b_0 + w_0 + \frac{w_1}{R_1} + \frac{w_2}{R_1R_2} + \ldots \equiv \Gamma,
\]
which after plugging in values of $x_{t+1}$ and $y_{t+2}$ gives:

$$c_0^* (1 + \gamma + \gamma \delta + \gamma^2 \delta + \gamma^2 \delta^2 + ...) = \Gamma,$$

implying

$$c_0^* = \frac{\Gamma (1 - \gamma \delta)}{1 + \gamma}.$$

**Step 2.** Next, we compute the laissez-faire equilibrium allocation, which we denote below by hat. The euler equations that characterize equilibrium are: for $t = 0, 2, 4, ...$

$$u'(\hat{c}_t) = \frac{\gamma R}{1 - \delta + \beta \delta} u'(\hat{c}_{t+1}),$$

$$u'(\hat{c}_{t+1}) = \beta \delta R u'(\hat{c}_{t+2}).$$

Since utility is logarithmic, this implies that for $t = 0, 2, 4, ...$

$$\hat{c}_{t+1} = \hat{c}_t \hat{x}_{t+1},$$

$$\hat{c}_{t+2} = \hat{c}_{t+1} \hat{y}_{t+1},$$

where

$$\hat{x}_{t+1} = \frac{\gamma R}{1 - \delta + \beta \delta},$$

$$\hat{y}_{t+2} = \beta \delta R.$$

Following what we did while computing the Ramsey allocation, we find that

$$\hat{c}_0 = \Gamma \frac{1 - \delta + \beta \delta - \beta \gamma \delta}{1 - \delta + \beta \delta + \gamma}.$$

**Step 3.** Now we compare the welfare of all generations under the two allocations and find the restrictions on parameters under which Ramsey improves welfare for all.

Notice that the Ramsey allocation strictly increases initial old’s welfare over laissez-faire since by definition Ramsey allocation is the unique allocation maximizes initial old’s welfare. We need to find when it improves over laissez-faire for the rest of the agents.

Let’s first check agent who is born in period 1. Denote his utility under Ramsey allocation by $U_{RA}^1$ and that under laissez-faire equilibrium by $U_{LF}^1$:

$$U_{RA}^1 = \log(c_0^* x_1) + \beta \delta \log(c_0^* x_1 y_2) + \gamma \log(c_0^* x_1 y_2 x_3) + \gamma \delta \log(c_0^* x_1 y_2 x_3 y_4) + ...$$

$$U_{LF}^1 = \log(\hat{c}_0 \hat{x}_1) + \beta \delta \log(\hat{c}_0 \hat{x}_1 \hat{y}_2) + \gamma \log(\hat{c}_0 \hat{x}_1 \hat{y}_2 \hat{x}_3) + \gamma \delta \log(\hat{c}_0 \hat{x}_1 \hat{y}_2 \hat{x}_3 \hat{y}_4) + ...$$
Agent 1 is better off under Ramsey allocation if and only if $U_{RA}^1 \geq U_{LF}^1$. But this difference can be computed as follows:

$$U_{RA}^1 - U_{LF}^1 = [\log(c_0^*) - \log(\hat{c}_0)]A + [\log(\gamma) - \log(\frac{\gamma}{1-\delta + \beta \delta})]B + [\log(\delta) - \log(\beta \delta)]C,$$

where

$$A = 1 + \beta \delta[1 + \gamma + \gamma \delta + \gamma^2 \delta + \gamma^2 \delta^2 + ...],$$
$$B = 1 + \beta \delta[1 + 2 \gamma + 2 \gamma \delta + 3 \gamma^2 \delta + 3 \gamma^2 \delta^2 + ...],$$
$$C = \beta \delta[1 + \gamma + 2 \gamma \delta + 2 \gamma^2 \delta + 3 \gamma^2 \delta^2 + ...].$$

With some algebra, we can show that

$$A = 1 + \frac{\beta \delta}{1 - \gamma \delta}(1 + \gamma),$$
$$B = 1 + \frac{\beta \delta}{1 - \gamma \delta}(\frac{1 + \gamma}{1 - \gamma \delta} + \gamma),$$
$$C = \frac{\beta \delta}{1 - \gamma \delta}(\frac{1 + \gamma}{1 - \gamma \delta}).$$

Plugging in the values of $c_0^*$ and $\hat{c}_0$, condition (45) becomes

$$U_{RA}^1 - U_{LF}^1 = A \log D + B \log(1 - \delta + \beta \delta) - C \log \beta,$$

where

$$D = \frac{(1 - \gamma \delta)(1 + \gamma)}{(1 + \gamma)(1 - \gamma \delta + \beta \delta)}.$$ 

Since neither the Ramsey nor the laissez-faire allocations are assumed to be stationary, the fact that agent 1 is better off under Ramsey allocation does not imply that future generations are better off too. Now, we show that condition (46) guarantees that agents born in future generations are better off under Ramsey as well. Let us first look at the agent born in period 3.

$$U_{RA}^3 = \log(c_0^* x_1 y_2 x_3) + \beta \delta[\log(c_0^* x_1 y_2 x_3 y_4) + \gamma \log(c_0^* x_1 y_2 x_3 y_4 x_5) + ...]$$
$$= U_{RA}^1 + \log(y_2 x_3) + \beta \delta[\log(x_3 y_4) + \gamma \log(x_4 y_5) + ...].$$

Similarly,

$$U_{LF}^3 = U_{LF}^1 + \log(\hat{y}_2 \hat{x}_3) + \beta \delta[\log(\hat{x}_3 \hat{y}_4) + \gamma \log(\hat{x}_4 \hat{y}_5) + ...].$$
After some algebra, we reach:

\[
U_{RA}^3 - U_{LF}^3 = U_{RA}^1 - U_{LF}^1 + \left[ \log(\gamma\delta) - \log\left(\frac{\gamma\delta\beta}{1 - \delta + \beta\delta}\right) \right] A
\]

\[
= U_{RA}^1 - U_{LF}^1 - \log\left(\frac{\beta}{1 - \delta + \beta\delta}\right) A
\]

\[
> U^1(x, y) - U^1(\hat{x}, \hat{y}),
\]

since \( \frac{\beta}{1 - \delta + \beta\delta} < 1 \). Thus, we have shown that condition (46) guarantees that agent 3 is better off under Ramsey allocation.

In a similar way, one can show that for any \( t \) odd, \( U_t(x, y) - U_t(\hat{x}, \hat{y}) > U^1(x, y) - U^1(\hat{x}, \hat{y}) \)
ending the proof.

\[ \Box \]

**Core Proof of Proposition 8**

*Proof.* When \( \gamma = \delta \), then, the condition for Ramsey allocation to be Pareto improving becomes

\[
\log(1 - \delta + \beta\delta) \left[ 1 + \frac{\beta\delta}{1 - \delta \left(1 + \frac{\delta}{1 - \delta^2}\right)} \right] - \log(\beta) \frac{\beta\delta}{1 - \delta \left(1 - \delta^2\right)} \geq 0.
\]

At no \( \beta \in (0, 1) \), a limit exists for the above expression as \( \delta \to 1 \). But the left and right hand limits exist. Since we are interested in the value as \( \delta \) approaches 1 from below, the relevant limit is the left hand one, which is

\[
\lim_{\delta \to 1} \log\left(\frac{1}{1 - \delta + \beta\delta}\right) \left[ 1 + \frac{\beta\delta}{1 - \delta \left(1 + \frac{\delta}{1 - \delta^2}\right)} \right] + \log(\beta) \frac{\beta\delta}{1 - \delta \left(1 - \delta^2\right)} = \infty.
\]

Thus, we have proved the proposition.

\[ \Box \]

**B Horizontal Redistribution: Implementation**

We now show that the efficient allocation can be implemented by inducing parents to bequest an amount to the kids that increases with \( s \), subsiding savings of the kids and taxing or subsidizing bequests in a progressive way. Whether bequests are subsidized or taxed again depends on the level of the direct weight \( \lambda \) the planner assigns to future generations (exactly as in Section 5.1). In the implementation exercise, parent’s budget constraint is:

\[
c_0 + b_1 = R_0b_0 + y_0 - T^b(b_1) + T^y(y_0),
\]
where labor income taxes $T^y_0$ will be function of the parent’s declaration and - by the taxation principle - can always be written in terms of the (observable) period zero income $y_0$. The function $T^b(\cdot)$ represents the bequest tax that is allowed to depend non-linearly on the level of bequests $b_1$ and for notational simplicity we assumed to be paid by the parent. The offspring budget constraints are represented by

$$
c_1 + b_2 = R_1(1 - \tau_1(y_0))b_1 + y_1 + T^y_1(y_0);
\quad c_2 = R_2(1 - \tau_2(y_0))b_2 + y_2 + T^y_2(y_0),
$$

where, for $t = 1, 2$, $\tau_t$, and $T^y_t$ represent wealth (or inheritance) taxes and labor income taxes, respectively. As indicated, both labor and wealth taxes are allowed to depend on the initial income level $y_0$ of the dynasty. In the above, both $T^b(\cdot)$ and $\tau^*_1$ are bequest taxes and we can obviously dispense for one of the two. We will denote the tax scheme that implement the efficient allocation with $\hat{T}^b(\cdot), \hat{T}^y_t(\cdot)$ for $t = 0, 1, 2$; and $\hat{\tau}_2(\cdot)$ (we set $\hat{\tau}_1 \equiv 0$)

**Proposition 15.** Assume CEIS preferences and $\beta < 1$. Then there is vector of taxes function on income $\hat{T}^y_t(\cdot)$, $t = 0, 1, 2$, on bequests $\hat{T}^b(\cdot)$, and offspring wealth $\hat{\tau}_2(\cdot)$ that implement the Mirrlees allocation. Moreover, if $\lambda > 0$ then $\hat{\tau}_2(\cdot)$ is constant in $y_0$, and the bequest tax function $\hat{T}^b(\cdot)$ is linear in $b_1$. (ii) If $\lambda < 0$ then $\hat{\tau}_2(\cdot)$ decreases with $y_0$ and the bequest tax is progressive, that is, $\hat{T}^b(\cdot)$ convex whenever $\rho \leq 1$ (including log).

**Proof.** We start with deriving the savings wedge for the offspring. The optimality condition for the planner when looking at the perturbation $z^s_0(s) = z^s_0(s), z^s_1(s) = z^s_1(s) + \epsilon$, and $z^s_2(s) = z^s_2(s) - \delta$, is

$$
\lambda(1 - \beta) - \frac{\mu_1}{u'(c^*_1(s))} + \frac{\mu_1}{\delta R^*_2 u'(c^*_2(s))} = 0,
$$

which can be rearranged as

$$
\frac{\beta - 1}{\beta} \left( \frac{\lambda u'(c^*_1(s))}{\mu_1} - 1 \right) = \frac{u'(c^*_1(s))}{\delta R^*_2 u'(c^*_2(s))} - 1.
$$

Consider now, the problem of a parent endowed with asset level $b_0(= K_0)$, skill level $s$, and declar-

---

[40] As discussed in [Farhi and Werning, 2010], it is easy to see that for each of this tax we can set a tax on the offspring (inheritance tax) with the same characteristics.
increases with skill $\sigma$:
\[
V(\sigma|s) := \max_{\{c_t, b_t\}_{t=0}^\infty} u(c_0) - \sigma \left( \frac{y_0(\sigma)}{w_0s} \right) + \gamma [u(c_1) + \delta u(c_2)]
\]
\[
s.t. \quad c_0 + b_1 = R_0 b_0 + y_0(\sigma) - T^b(1) b_1 + T_0^y(\sigma);
\]
\[
b_2 \in \arg \max_{\hat{b}} u(\hat{c}_1) + \delta \beta u(\hat{c}_2);
\]
\[
s.t. \quad \hat{c}_1 + \hat{b} = R_1 (1 - \tau_1(\sigma)) b_1 + y_1 + T^y_1(\sigma);
\]
\[
\hat{c}_2 = R_2 (1 - \tau_2(\sigma)) \hat{b} + y_2 + T^y_2(\sigma).
\]

**Definition.** We say that a system of taxes implements the Mirrlees allocation $\{c^*_t(s)\}_{t=0,1,2}, y^*_0(s)$, if for each $s$ we have $V(s|s) \geq V(\sigma|s)$ for all $\sigma$ and for each $s$ the solution to $V(s|s)$ is $\{c^*_t(s)\}_{t=0,1,2}, y^*_0(s)$.

The core of the implementation exercise follows the same logic as in [Kocherlakota (2005)](#). We will show that for all $s$ a dynasty whose parent declares $\sigma$ will find optimal to choose the allocation prescribed by the efficient allocation $c^*_t(\sigma)$, $t = 0, 1, 2$ (and, of course, the parent is forced to generate income $y^*_0(\sigma)$). One this key fact has been shown, the inequality $V(s|s) \geq V(\sigma|s)$ is a direct consequence of the incentive compatibility property of the Mirrleesian allocation.

**Offspring Policy with CEIS preferences.** Since the offspring problem is strictly concave, if taxes will be set at their optimal values, the offspring will choose the second best allocation if budget feasible. It hence suffices to choose labor taxes that guarantee\(^{41}\)
\[
c^*_1(s) + \frac{c^*_2(s)}{R_2(1 - \hat{\tau}_2(s))} = y_1 + \frac{T^y_1(s)}{R_2(1 - \hat{\tau}_2(s))} + \frac{y_2 + T^y_2(s)}{R_2(1 - \hat{\tau}_2(s))} + R_1 b^*_1(s).
\]

In the above, the value $R^*_1 b^*_1(s)$ as yet to be determined, but it is pretty clear that the presence of state contingent labor income taxes gives enough degrees of freedom to generate feasibility of the Mirrlees allocation for the offspring. Now, recall that from Proposition\(^{4}\) - with CEIS utility - the policy of the kid is linear and increasing in wealth. We now show how its depends on taxes in our example\(^{42}\).

Let $M_T := \frac{\partial b_2(\cdot; R_1(1 - \tau_1), R_2(1 - \tau_2))}{\partial b_1}$ be the derivative of the policy when facing taxes $\tau_1, \tau_2$ and prices

\(^{41}\)We will abuse notation and use $\hat{\tau}_1$ and so on even for taxes defined on $s$ instead on $y_0$. Then since $y_0(s)$ increases with $s$ there will be a non-to-one relationship between our taxes and that in the proposition.

\(^{42}\)Note that the kid problem is as above, with the budget linear in $b_1$ so the proof of linearity is immediate.
$R_1, R_2$. From the Euler equation and the budget constraint, the policy of the offspring must solve

$$u'(c_1) = \delta \beta R_1 (1 - \tau_2) u'(c_2)$$

$$\iff \quad (R_1 (1 - \tau_1) b_1 + y_1 + T_1^y - (A_\tau + M_\tau b_1))^{-\rho} = \delta \beta R_2 (1 - \tau_2) (R_2 (1 - \tau_2) (A_\tau + M_\tau b_1) + y_2 + T_2^y)^{-\rho}$$

$$\iff \quad (R_1 (1 - \tau_1) b_1 + y_1 + T_1^y - (A_\tau + M_\tau b_1)) = (\delta \beta R_2 (1 - \tau_2))^\frac{1}{\rho} (R_2 (1 - \tau_2) (A_\tau + M_\tau b_1) + y_2 + T_2^y),$$

where $A_\tau$ and $M_\tau$ are the constants defining the policy and $\frac{1}{\rho}$ is the intertemporal elasticity of substitution. Since the above relationship must hold for all $b_1$, taking the derivative with respect to $b_1$ on both sides, for all $\tau_1, \tau_2, R_1, R_2$ we have

$$M_\tau = \frac{(1 - \tau_1) R_1}{(\delta \beta)^{-\frac{1}{\rho}} R_2 (1 - \tau_2)^{1-\frac{1}{\rho}} + 1}. \quad (49)$$

Recalling equation (10) we can derive the first order condition for the parent in presence of taxes:

$$\frac{u'(c_0(s)) [1 + \frac{\partial T^b(b_1(s))}{\partial b_1}]}{\gamma R_1 u''(c_1(s))} = 1 - \tau_1(s) + \frac{M_\tau(s)}{R_1} \left[ \frac{1}{\beta} - 1 \right]. \quad (50)$$

If we divide both sides by $1 - \tau_1 \neq 0$, we have

$$\frac{u'(c_0(s)) [1 + \frac{\partial T^b(b_1(s))}{\partial b_1}]}{\gamma R_1 u''(c_1(s))} \frac{1}{1 - \tau_1(s)} = 1 + \frac{M_\tau(s)}{(1 - \tau_1(s)) R_1} \left[ \frac{1}{\beta} - 1 \right].$$

If we set $\tilde{\tau}_1(s) = 0$ for all $s$ as required by the proposition, using (49), we have

$$\frac{u'(c_0(s)) [1 + \frac{\partial T^b(b_1(s))}{\partial b_1}]}{\gamma R_1 u''(c_1(s))} - 1 = \frac{1}{\beta} - 1 \quad \frac{1}{(\delta \beta)^{-\frac{1}{\rho}} R_2 (1 - \tilde{\tau}_2(s))^{1-\frac{1}{\rho}} + 1}.$$

which can be rearranged as:

$$\frac{u'(c_0(s))}{\gamma R_1 u''(c_1(s))} - 1 = - \frac{u'(c_0(s))}{\gamma R_1 u''(c_1(s))} \frac{\partial T^b(b_1(s))}{\partial b_1} + \frac{1}{\beta} - 1 \quad \frac{1}{(\delta \beta)^{-\frac{1}{\rho}} R_2 (1 - \tilde{\tau}_2(s))^{1-\frac{1}{\rho}} + 1}.$$

Matching this with the planner optimality conditions we have

$$\frac{\lambda}{\mu_0 \gamma} u'(c_0^*(s)) = - \frac{u'(c_0^*(s))}{\gamma R_1^2 u''(c_1^*(s))} \frac{\partial T^b(b_1^*(s))}{\partial b_1} + \frac{1}{\beta} - 1 \quad \frac{1}{(\delta \beta)^{-\frac{1}{\rho}} R_2^2 (1 - \tilde{\tau}_2(s))^{1-\frac{1}{\rho}} + 1}. \quad (51)$$
(i) \( \lambda = 0 \) (Ramsey). From condition \([18]\) when \( \lambda = 0 \) the offspring wedge is independent from \( s \). In particular, if for all \( s \), we set \( 1 - \hat{\tau}_2(s) \equiv \frac{1}{\beta} \) we exactly match the offspring wedge. If we set \( \lambda = 0 \) in \((51)\) and hence also note that \( \frac{u'(c^*_1(s))}{\gamma R_1 u'(c_1(s))} \equiv 1 \) for all \( s \), we obtain

\[
\frac{\partial \hat{T}^b(b_1^*(s))}{\partial b_1} = \frac{1}{\beta} - \frac{1}{(\delta \beta)^{\frac{1}{\gamma}} (R_2^* (1 - \hat{\tau}_2(s)))^{1 - \frac{1}{\gamma}} + 1} = \frac{M_+(s)}{R_1^*} \left[ \frac{1}{\beta} - 1 \right].
\]

Note that since \( M_+(s) \) is independent from \( s \) while \( b_1^*(s) \) in creases with \( s \) the function \( \hat{T}^b(\cdot) \) is linear in this case. In fact, this expression is familiar to us as it corresponds exactly to the bequest tax of Proposition \([5]\) in the baseline model without horizontal heterogeneity.

Finally, note that the actual level of skills does not enter into the intertemporal budget constraint nor into the Euler equations of the individuals. This implies that for all \( s \), a parent declaring \( \sigma \) (possibly different from \( s \)) will find it optimal to choose the Mirrleese consumption plan \( c_1^*(\sigma) \), \( t = 0, 1, 2 \). Since the Mirrleese allocation is incentive compatible the parents will actually tell the truth. As we said, there is a continuum of labor taxes generating the net present value of income. One possibility is hence that each parent bequests the same level \( b_1^* = K_1 \) to his kids. A particularly appealing allocation is the one that uses \( T^y_0(s) \) as a main distributional tool and \( \hat{T}_1^y(s) \), and \( \hat{T}_2^y(s) \) to only rebate lump sum the wealth taxes paid. More precisely, we will implement the Mirrlees allocation by setting for \( t = 1, 2 \), \( \hat{T}_t^y(s) = \hat{\tau}_tb_t^*(s) \), where \( b_1^*(s) \) and \( b_2^*(s) \) are the bequest label on the efficiency allocation under such taxes. Then \( \hat{T}_0^y(s) \) is set so that to have the intertemporal budget constraint of the dynasty to hold exactly.

(ii) \( \lambda > 0 \). In this case, we set labor income taxes with the same rule as above, namely, for \( t = 1, 2 \), and all \( s \), \( \hat{T}_t^y(s) = \hat{\tau}_t(s)b_t^*(s) \). So we can concentrate on the tax rates.

Equation \((51)\) can be rearranged to get:

\[
\frac{1}{\beta} - \frac{1}{(\delta \beta)^{\frac{1}{\gamma}} (R_2^* (1 - \hat{\tau}_2(s)))^{1 - \frac{1}{\gamma}} + 1} = \frac{\gamma R_1 u'(c_1^*(s))}{\mu_0 u'(c_1^*(s))} - \frac{\lambda R_1^*}{\mu_0} u'(c_1^*(s)) = \frac{\partial \hat{T}^b(b_1^*(s))}{\partial b_1}.
\]

Also recall that when \( \lambda > 0 \) the wedge on the offspring is state dependent. We again use \( \hat{\tau}_2 \) to correct for this wedge. We set for all \( s \)

\[
\hat{\tau}_2(s) = \frac{1 - \beta}{\beta} \left( \frac{\lambda u'(c_1^*(s))}{\mu_1} - 1 \right) = 1 - \frac{u'(c_1^*(s))}{\delta \beta R_2^* u'(c_2^*(s))}.
\]

Note that since \( c_1^*(s) \) increases with \( s \) and \( 0 < \beta < 1 \), \( \hat{\tau}_2(s) \) decreases with \( s \). Given that the tax on offspring savings is state dependent, the offspring policy will be state dependent as well, but again linear. This is so since preference safe CEIS and for each given \( s \), wealth taxes are linear.

Condition \((52)\) is crucial in determining the curvature of \( \hat{T}^b(\cdot) \). Recall first that from the planner
first order conditions \( \frac{u'(c_0(s))}{\gamma R_1 u'(c_1(s))} \) decreases with \( s \) since \( c_0^*(\cdot) \) is increasing. As a consequence, its inverse \( \gamma R_1 u'(c_0^*(s)) \) increases. In addition, we just saw that \( 1 - \hat{\tau}_2(s) \) increases. Finally, \( \frac{\lambda R_1}{R_0} u'(c_1^*(s)) \) decreases with \( s \) as \( c_1^*(\cdot) \) is increasing. So for \( \rho \leq 1 \) everything in the left hand side of (52) increases with \( s \). If we show that \( c_1^*(s) \) increases with \( s \), we get a convex \( \hat{\tau}^I \) as claimed. So we need to show that in the Mirrlese allocation \( c_1^*(s) \) increases with \( s \). From the offspring budget constraint at the efficient allocation and taxes, and recalling that \( \hat{\tau}^I(s) = \tau_1(s) b_1^*(s) \), for all \( s \) we have: \( c_1^*(s) + b_2^*(s) = R_1 b_1^*(s) + y_1 \) and \( c_2^*(s) = R_2 b_2^*(s) + y_2 \), which combined together generates

\[
c_1^*(s) + \frac{c_2^*(s)}{R_2} = y_1 + \frac{u_2}{R_2} + R_1 b_1^*(s)
\]

Since, \( c_1^*(s) \) and \( c_2^*(s) \) is increasing in \( s \), then \( b_1^*(s) \) must also increase with \( s \). In fact, the value of \( b_1^*(s) \) is fully pinned down from this equation for all \( s \). Finally, we need to set the period zero labor tax \( \hat{\tau}^I_0 \). For each \( s \) given \( c_0^*(s), b_1^*(s), y_0^*(s) \) and \( K_0 \), we set \( \hat{\tau}^I_0(s) = c_0^*(s) + b_1^*(s) + \hat{\tau}^b(b_1^*(s)) - R_0 K_0 - y_0^*(s) \). As argued above, since agents will choose this allocation for each declaration \( \sigma = s \) no matter what is their true level of skills, incentive compatibility implies the parent will always tell the truth.

C The Optimality Condition for Illiquid Bequests

In this section, we explicitly derive equation \( \langle 34 \rangle \) in the main text, which is the optimality condition for illiquid bequests. The first-order condition for illiquid bequests is given by

\[
(b_{t+2}^o : u'(c_{t+1}^o)) = \gamma \delta \left( u'(c_{t+1}^o) \left[ -\frac{db_{t+3}}{db_{t+2}} \right] + \delta u'(c_{t+2}) \left[ R_{t+2} \frac{db_{t+3}}{db_{t+2}} + R_{t+1} R_{t+2} - \frac{db_{t+3}}{db_{t+2}} \right] \right)
\]

Observe that a change in illiquid bequests affects \( b_{t+3} \) both directly and indirectly through its effect on \( b_{t+2} \):

\[
\frac{db_{t+3}}{db_{t+2}} = \frac{db_{t+3}}{db_{t+2}} + \frac{db_{t+3} db_{t+2}}{db_{t+2} db_{t+2}}.
\]

Now, plugging \( \langle 53 \rangle \) into the parent’s first-order condition for illiquid bequests and rearranging, we get

\[
u'(c_{t+1}^o) = \gamma \left( R_{t+1} R_{t+2} \delta^2 u'(c_{t+2}) + \frac{db_{t+2}}{db_{t+2}} \Delta_{t+1} + \frac{db_{t+3}}{db_{t+2}} \Delta_{t+2} \right),
\]

\[\text{Of course, this ‘social’ discounting without taxes only occurs at the efficient allocation, in other terms, this is not a definition of the budget constraint. As described above, the offsprings perceives a different (subsidized) discounting of future income and consumption.}\]
where

\[
\Delta_{t+1} = \left[ -u'(c_{t+1}) + \delta \left( \sum_{i=2}^{t+1} \delta^{i-2} u'(c_{t+i}) \frac{\partial c_{t+i}}{\partial b_{t+2}} + \delta^{i-2} V_1(a_{t+1}, Q_{t+1}) R_{t+1} \frac{\partial b_{t+1}}{\partial b_{t+2}} \right) \right], \quad (55)
\]

\[
\Delta_{t+2} = \left[ -u'(c_{t+2}) + \delta \left( \sum_{i=3}^{t+1} \delta^{i-3} u'(c_{t+i}) \frac{\partial c_{t+i}}{\partial b_{t+3}} + \delta^{i-3} V_1(a_{t+1}, Q_{t+1}) R_{t+1} \frac{\partial b_{t+1}}{\partial b_{t+3}} \right) \right]. \quad (56)
\]

Using (32) in (55) and the offspring’s optimality condition for period \( t + 2 \) savings, (33), in (56), we get

\[
\Delta_{t+1} = 0, \\
\Delta_{t+2} = \delta u'(c_{t+2}) \left( -1 + \frac{1}{\beta_3} \right).
\]

Pluggin these back into (54), we reach equation (34) in the text. □

### D Naïvete

In the main body of the paper, we assume that parents are sophisticated in the sense that they are aware of the fact that their descendants disagree with them about how much to save for their old age. In this section, we analyze whether the optimality of transfer taxation depends on the assumption of sophistication. To do so, we assume that agents are naïve, meaning that they believe (incorrectly) that their descendants fully agree with themselves regarding intertemporal allocation of resources. We focus on the Ramsey wedge but the results can be readily extended to other Pareto efficient allocations following the approach in Section 5.1. We show that even naïve parents bequeath too much in equilibrium relative to the Ramsey level and, hence, it is optimal to tax parental transfers.\(^{44}\)

The problem of a naïve parent in period \( t \) is given by:

\[
\bar{V} (a_t, Q_t) = \max_{b_{t+1} \geq -B(Q_{t+1})} u(c_t) + \gamma [u(\bar{c}_{t+1}(b_{t+1}, Q_{t+1})) + \delta \bar{V} (\bar{a}_{t+2}(b_{t+1}, Q_{t+1}), Q_{t+2})],
\]

subject to the budget constraints and the definition of wealth

\[
c_t = a_t - b_{t+1}, \\
\bar{c}_{t+1}(b_{t+1}, Q_{t+1}) = R_{t+1} b_{t+1} + w_{t+1} - \bar{b}_{t+2}(b_{t+1}, Q_{t+1}), \\
\bar{a}_{t+2}(b_{t+1}, Q_{t+1}) := R_{t+2} \bar{b}_{t+2}(b_{t+1}, Q_{t+1}) + w_{t+2},
\]

\(^{44}\)It might be worthwhile noting that the spirit of government intervention in this section is in some sense ‘paternalistic,’ in that it is aiming to correct a bias emerging from bounded rationality. No ‘paternalistic’ policy has been proposed so far in the paper.
together with the condition defining the policy of the offspring:

\[ \bar{b}_{t+2}(b_{t+1}, Q_{t+1}) = \arg \max_{b_{t+2} \geq B(Q_{t+2})} u \left( R_{t+1} b_{t+1} + w_{t+1} - \bar{b}_{t+2} \right) + \delta \bar{V} \left( R_{t+2} \bar{b}_{t+2} + w_{t+2}, Q_{t+2} \right). \] (57)

Observe that this problem is identical to the problem of the sophisticated parent, except that in the condition defining the policy of the offspring, the offspring discounts future with \( \delta \) instead of \( \beta \delta \). This reflects the assumption that parents naively believe that their descendants are in full harmony with them regarding intertemporal preferences. We denote the naive parents’ value function by \( \bar{V} \left( a_t, Q_t \right) \). The naive value function gives the value of the solution to a standard dynamic programming problem with exponential discounters. It is well-known that since \( u \) is concave, this value function is concave as well. Notice that the saving policy function of the offspring given in (57) is incorrect; it represents the naive belief of the parent about the saving policy of the offspring. To distinguish the parent’s belief about the policy and the actual policy of the offspring in equilibrium, we denote the former by \( \bar{b}_{t+2}(\cdot) \). Similarly, \( \bar{c}_{t+1}(\cdot) \) and \( \bar{a}_{t+2}(\cdot) \) refer to the naive belief of the parent about the offspring’s period \( t + 1 \) consumption and beginning of period \( t + 2 \) wealth policies.

Now, we consider the true saving behavior of the offspring in equilibrium. The offspring solves

\[ \max_{\tilde{b}_{t+2} \geq -B(Q_{t+2})} u \left( R_{t+1} b_{t+1} + w_{t+1} - \tilde{b}_{t+2} \right) + \beta \delta \bar{V} \left( R_{t+2} \tilde{b}_{t+2} + w_{t+2}, Q_{t+2} \right). \]

Observe that the naive offspring also faces the naive value function \( \bar{V} \) since his naive beliefs about how his descendants will allocate consumption over time is in line with that of his parent.

We are now ready to establish the optimality of bequest tax. The parent naively believes that the offspring solves (57), and hence, the following first-order condition describes offspring saving behavior

\[ u'(\bar{c}_{t+1}(b_{t+1}, Q_{t+1})) = \delta R_{t+2} \bar{V}_1(\bar{a}_{t+2}(b_{t+1}, Q_{t+1}), Q_{t+2}). \] (58)

Taking the first-order optimality condition with respect to bequests in the naive parent’s problem and substituting (58) in, we find that the optimality condition for bequests in equilibrium is given by:

\[ u'(c_t) = \gamma R_{t+1} u'(\bar{c}_{t+1}(b_{t+1}, Q_{t+1})). \]

We therefore define the naive Ramsey bequest wedge as follows:

\[ \overline{BW}_t = 1 - \frac{u'(c_t^*)}{\gamma R_{t+1} u'(\bar{c}_{t+1}(b_{t+1}^*, Q_{t+1}^*)^*}, \]

where, \( \bar{c}_{t+1}(b_{t+1}^*, Q_{t+1}^*) \) corresponds to what the naive parent believes the offspring will choose if he receives \( b_{t+1}^* \) as bequests and faces the price sequence implied by the Ramsey allocation, \( Q_{t+1}^* \).
We want to show that the naive bequest wedge is strictly positive, meaning given the Ramsey allocation, the naive parent would like to increase bequests. Remember that in the Ramsey allocation, we have

\[-u'(c_t^*) + \gamma R_{t+1}^* u'(c_{t+1}^*) = 0.\]

Pluggin this condition in the definition of the naive bequest wedge implies

\[\frac{BW_t}{u'} = 1 - \frac{u'(c_{t+1}^*)}{u' \left( \hat{c}_{t+1}(b_{t+1}^*, Q_{t+1}^*) \right)}. \quad (59)\]

Therefore, in order to establish the sign of \(BW_t\), we need to compare the Ramsey level of consumption of the offspring, \(c_{t+1}^*\), with the parent’s belief about how much the offspring will consume: \(\hat{c}_{t+1}(b_{t+1}^*, Q_{t+1}^*)\).

Now, if the offspring chooses the Ramsey allocation in equilibrium, say thanks to the tax \(\tau_{t+2}^*\), then his first-order condition reads

\[u'(c_{t+1}^*) = \beta \delta R_{t+2}^*(1 - \tau_{t+2}^*) \bar{V}_1(a_{t+2}, Q_{t+2}^*).\]

The parent - on the other hand - believes the offspring will behave according to

\[u'(\hat{c}_{t+1}(b_{t+1}^*, Q_{t+1}^*)) = \delta R_{t+2}^*(1 - \tau_{t+2}^*) \bar{V}_1(\hat{a}_{t+2}(b_{t+1}^*, Q_{t+1}^*), Q_{t+2}^*).\]

Concavity of \(u\) and \(\bar{V}\) then imply that

\[\hat{c}_{t+1}(b_{t+1}^*, Q_{t+1}^*) < c_{t+1}^*.\]

This ordering is intuitive: an agent with a lower discount factor, \(\beta \delta\), tends to consume more in the current period than an agent with a higher discount factor, \(\delta\), and the parent naively believes that the offspring has the higher discount factor. Therefore, we have the following result:

**Proposition 16.** Assume agents are naive and \(\beta < 1\), then the naive Ramsey bequest wedge is positive.

Therefore, we conclude that even when agents are naive, parents bequeath too much relative to the Ramsey level, and hence, it is optimal to tax bequests. The intuition for the bequest tax result when agents are naive is different from the case with sophisticated agents though. In the naive case, parents bequeath too much relative to the Ramsey level not because they want to compensate for their offspring’s undersavings but because their perceived marginal return to bequeathing, \(u'(\hat{c}_{t+1}(b_{t+1}^*, Q_{t+1}^*))\), is larger than the actual marginal return, \(u'(c_{t+1}^*)\). One can further show that, if we restrict attention to linear equilibria, then we can implement the Ramsey allocation via taxes, and the optimal tax on bequests is given by the naive bequest wedge in (59).