Reference Dependence and Unobserved Poverty Traps

Miles T. Smith

December 15, 2025

Abstract

Decades of empirical research has found mixed results on the existence of poverty traps. In this paper, I revisit the question of their existence and propose an explanation for the mixed evidence using behavioral theory. I formulate a poverty trap using an imperfect credit market and wealth is transmitted to the next period through a bequest motive. When agents have a reference-dependent bequest preference, which I model through loss aversion, they change their bequest behavior in response to vicious cycles, which can slow and even stop downward transitions and makes the model sensitive to shocks. These dynamics are not accounted by standard poverty trap models and can make the trap invisible to empirical identification. Using Monte Carlo simulations, I confirm that that if reference-dependent loss aversion exists but is unaccounted for in the empirical specification then the poverty trap will not be identified. I analyze a tax-redistribution and "big push" policy, the different policies employed by different empirical results, formulate an optimal tax rate, and I find interventions can have welfare-reducing consequences if the poverty trap is mis-identified.

[CURRENT DRAFT]
[ONLINE APPENDIX]

1 Introduction

Poverty rates in the United States have remained high for many communities for decades [Benson et al., 2023]. One theory that attempts to explain persistent poverty are poverty traps: self-reinforcing mechanisms where those below a certain wealth threshold face economic forces that infringe on their ability to generate wealth. The evidence that poverty traps exist is mixed: while Kraay and McKenzie [2014] only find conclusive evidence for poverty traps in specific contexts, research by Carter and Barrett [2013] posit a greater prevalence of poverty traps but a dearth of reliable data. It remains an outstanding question as to why empirical data is so seldomly able to conclude the existence of poverty traps.

I reassess an intergenerational poverty trap framework while importantly allowing for reference-dependent bequest motives to understand how behavioral interactions might change the existing poverty dynamics and therefore observation data. My model borrows the framework from Banerjee and Newman [1993] where a mechanical poverty trap, separate from behavioral preferences, is generated within a credit market due to imperfect monitoring of loans. Agents begin their life with an initial wealth as a bequest from their parents. Those with low initial wealth are lent less, realize small incomes, and therefore bequeath less to their progeny which transmits poverty. The key contribution of my model is that agents' bequest preference is dependent on some reference point which I model using loss aversion. Loss aversion states that individuals feel a psychological loss much greater than an equivalent gain, and will change their behavior to avoid this felt loss. Agents' anchor their expectation for how much they should give based on their own initial wealth and will feel a loss if they give less: parents want to give their children a life *at least as good* as their *own* initial conditions. Agents will increase their bequest to avoid this loss.[1]

My model predicts a slower or entirely stagnant convergence to lower steady states than predicted by a model without loss aversion. Agents who normally give less than their own initial wealth — typically those within the vicious cycle of the poverty trap — increase their bequest, slowing downward convergence. If they can, an agent will increase their bequest until it exactly matches their own initial wealth, avoiding the loss with the least possible increase, which makes that wealth level *sticky*: dynasties at that initial wealth will bequest that initial wealth indefinitely. Intervals of sticky wealth levels, unstable steady states, emerge capturing large portions of the wealth distribution. These dynamics differ significantly from standard poverty trap models where all individuals below a threshold converge down to poverty. As Barrett and Carter [2013] posit, if poverty traps exist very few families would be near the unstable steady state in the data; in my model an interval of sticky wealth can emerge below the unstable state, clustering observations near that unsteady state opposite the crucial prediction researchers use to identify poverty traps. Sticky intervals also emerge near the stable steady states, rich and poor, such that dynasties above no longer converge to much poorer states.

I find that reference-dependence can make the poverty trap invisible to typical econometric empirical identification methods. I use Monte Carlo simulations of my model and, as a benchmark, my poverty trap without loss aversion to create wealth panel data. I run two common tests to identify the existence of a poverty trap: a threshold test that identifies the unstable steady state and an S-shape test that identifies the dynamics unique to a poverty trap. First, I find the threshold test fails to reject the null hypothesis that a poverty trap *does not exist*, so the existence of a poverty trap is inconclusive. I find the S-shape test is sensitive to the loss-averse parameters and if it does predict a trap, it often misidentifies the true weight of the vicious cycles and steady states. Without loss aversion, the empirical tests would otherwise accurately identify the trap. This paper also motivates researchers to revisit the definition of a poverty trap. If a poverty trap is defined as a threshold where the poor get poorer and the rich get richer — this is the most common definition researchers look for [Banerjee et al., 2019, Balboni et al., 2022, Banerjee, 2020] — this model shows empirics might not find those dynamics. But if poverty traps are defined as economic stagnation at multiple income/wealth levels, empirics might have more success identifying poverty traps.

If reference-dependent preferences exist, it could explain the mixed evidence. There is an abundance of literature that does support the theory of reference-dependent loss aversion, including related to both parental investment [Barone et al., 2021] and feelings within poverty [Yesuf and

[1] It is important to note that this behavior does not create a poverty trap, rather it allows agents to internalize the negative change in inter-period wealth and endogenously determine their bequest in response to that change.

Bluffstone, 2009]. This naturally begs two questions: if wealth panel data alone is unreliable how can researchers overcome those limitations and are there policy implications for failing to identify a trap? I offer a base step to answer these questions that reveals important insights. I plan to explore these questions using a more sophisticated models in the future to expand this analysis.

My model suggests adding consumption data might help identification, and predicts a potential quasi-experimental method to identify ranges of poverty traps. Over generations, while wealth or income might be steady, agents will sacrifice consumption to maintain bequests in response to negative shocks and headwind. Researchers looking to see a decline in wealth, especially when an agent positively or negatively shocked might not see it, but could identify if this is because of changes in consumption. Furthermore, a behavior uniquely rational in this model is that agents who are shocked to the same average income from different initial wealth levels will have different bequest propensities compared to each other and those un-shocked at the same income level. An agent whose income was shocked down will have a higher propensity to bequest than the agent shocked up to the same income level due to anchored preferences. This creates a micro quasi-experiment where, by comparing the bequest differences, we can estimate the headwind in that area of wealth levels by how much bequests need to change to maintain the anchors. By comparing many of these convergent income shocks, we could control for heterogeneity in consumption preference.

I show that empirical limitations can lead to two different policy interventions with potentially welfare-reducing effects. First, I assume a social planner is not able to identify a poverty trap and I formulate an optimal tax rate for a flat-tax rebate [Piketty and Saez, 2013]. Loss aversion makes the equity-efficiency trade-off sensitive to the distribution feeling a loss, and often lowers the optimal tax due to the agents' preference to preserve the wealth distribution and because richer agents who feel a loss have an outsized impact of both equity and efficiency. In the second analysis, a social planner identifies a trap so the optimal policy to take out debt to finance a "big-push": a demogrant large enough for poor individuals to enter a state of upward mobility. The implications of the empirical tests suggests planner's will underestimate the unstable threshold. Since the push is too small, the poor do not enter upward mobility and stochastic shocks and future distortionary taxes send many back into the vicious cycle which causes a continuous need for a planner to keep lifting part of the population out of poverty indefinitely, resulting in inefficient and ineffective policy. Future research would benefit by incorporating loss aversion into life-cycle models to study consumption and child spending or investment smoothing and models that allow for more tractable multi-period taxation.

I begin in Section 2 by setting up the economic model and discuss the poverty trap dynamics in Section 3. Then I discuss econometric implications in Section 4 and end with a policy analysis in Section 5.

2 Model

2.1 Environment

Suppose a continuum of agents where the population is normalized to 1. Each agent enters the economy with some initial wealth in the form of a bequest. Time is discrete with an infinite horizon where the distribution of the initial wealth in time t is $G_t(w)$, and every period represents a generation.

I assume that agents are economically active for one period. At the beginning of the period, they enter the economy with their initial wealth, earn an income, then choose how much of their income to leave as a bequest to their child. By the end of the period all economic activity has occurred in equilibrium and all agents leave a bequest to their child, consume the rest, then they exit the economy. In the next period, after a child receives a bequest, they become an agent in the economy. Since every agent is replaced by only one child the population size is constant across periods. While this set-up uses discrete time, the remainder of this section discusses the static equilibrium (the optimal policy function for a given period), and the period notation is dropped until necessary.

Agents have homogeneous, risk-neutral preferences over bequests and consumption characterized by the utility function in Equation (1):

$$U(b, c; \rho_b, \rho_c) = \underbrace{\gamma \ln(b) + (1-\gamma)\ln(c)}_{\substack{\text{Standard Cobb–Douglas} \\ \text{Preferences}}} + \underbrace{\eta\big[\gamma \cdot \upsilon\left(\ln(b)|\ln(\rho_b)\right) + (1-\gamma)\cdot \upsilon(\ln(c)|\ln(\rho_c))\big]}_{\text{Loss-Averse Preferences}}. \quad (1)$$

The first two terms represent standard Cobb-Douglas preferences over their bequest to their child, b (the "warm-glow" effect), and their consumption, c. The second two terms represent the loss-averse utility over their bequest and consumption. Loss aversion states that agents feel a "loss" greater than an equivalent "gain" relative to a reference point. Kőszegi and Rabin [2006] represent loss aversion using the gain-loss function $\upsilon(x|\rho)$ where utility input x has a reference point ρ. Agents gain utility $x - \rho$ when $x \geq \rho$ but lose utility $\lambda(x - \rho)$ when $x < \rho$, where $\lambda > 1$ is the loss-aversion coefficient. The value an agent places on loss-averse utility is $\eta > 0$. I define that agents are loss averse over their bequests and their consumption relative to the reference points ρ_b and ρ_c for bequests and consumption, respectively. [2]. One interpretation is that agents have an expectation for how much they *should* bequest to their child, and if they fail to meet that expectation they feel a loss (shame, guilt, compassion, etc.). This expectation could be driven by norms, personal preference, or some other factor. Regardless, agents will increase the proportion of their income they allocate for a bequest to avoid feeling this loss. There is also an expected amount of consumption they believe they should meet, potentially driven by conspicuous consumption or necessary nutrition.

[2]The gain-loss function in Utility Equation 1 is between the *utilities* of bequest and consumption, but practically this is equivalent to comparing the actual units since ln() is monotonically increasing on $\mathbb{R}_+$. For the purpose of this paper, I just refer to the comparison of the level amounts not the utility.

In this paper, I focus on the case where anchor expectations for bequests relative to an agent's own initial wealth, $\rho_b = w$, and I let their consumption always be coded as a gain, $\rho_c = 0^+$. In the most general version of the model, both could be differentially weighted and coded as a loss or gain, though this is outside the scope of this paper. For a discussion on the addition of losses over consumption and different anchors for bequests, see Appendix (A.1).

2.2 Production and Markets

Agents use their endowment to invest in a capital market, realize some return, then decide how much to consume and bequest. There is a single production technology that agents invest in and a lending market agents use to borrow capital for investment. Due to market imperfections, there is a possibility that that a borrower will renege on their loan and avoid repayment, which limits lending amounts.

All agents participate in the economy, using their initial wealth w as collateral to borrow k units of capital. Agents invest k and receive investment income $V(k)$. I assume $V(k)$ is increasing in k and that there exists a unique k^* that maximizes investment income, which is the "first-best" level of capital.[3] An agent who borrows $k < k^*$ borrows their "second-best" level of capital. There is also a fixed, exogenous global gross interest rate $r > 0$ which creates capitalized interest.

After an agent realizes their investment income, they choose whether to repay the loan or renege. Agents who repay their loan pay back with interest kr and get their collateral back with interest, wr. They are also gifted a lump-sum income supplement with expected value of T after repayment. We can think of this as a reward for good behavior. The expected value of repayment is therefore: $V(k) + wr - kr + T$.

Agents who renege on their loan forfeit their initial wealth and reward supplement to avoid repayment and might successfully run away with their investment gains. The probability of being caught is $\pi(k)$ increasing and concave with k: the bigger the loan, the easier it is to monitor. If caught, they receive some punishment F. These assumptions of $\pi(k)$ and F drives the lending dynamics. The expected value of reneging is $V(k) - \pi(k)F$. Borrowers will renege if the expected value of reneging is greater than the expected value of repayment: $V(k) - \pi(k)F > V(k) + wr - kr + T$, which can be rewritten as when the loan is sufficiently large compared to their wealth $k > w + \frac{\pi(k)F+T}{r}$. Intuitively, the poorer an agent is the less they have to lose by reneging. With perfect information, lenders know to only make loans where agents are indifferent between reneging and repayment, which is to offer a loan that satisfies the incentive-compatibility (IC) constraint: $k \leq w + \frac{\pi(k)F+T}{r}$.

The minimum initial wealth needed to borrow the first-best level of capital can be found by plugging in k^* in the IC and solving for w: $w^* = k^* - \frac{\pi(k^*)F+T}{r}$. Agents with $w \geq w^*$ only borrow k^*; Agents with $w < w^*$ borrow their second-best level of capital which depends on their initial wealth. Since the investment returns are increasing in k, an agents always want to borrow the maximum

[3] It is explored later that the production technology need not be concave. If $V(k)$ is concave, and the first-best level of capital is k^* that maximizes their investment income: $\frac{\partial}{\partial k}(V(k) - kr) = 0$ and $V'(k^*) = r$.

k they can below the first-best. The implicit loan function for $w < w^*$ is $k(w) = w + \frac{\pi(k(w))F+T}{r}$. As the probability of getting caught and the expected value of the reward increase, the lower the necessary wealth is needed to borrow k. For tractability, I assume there is a maximum level of capital that can be monitored by lenders is k_π, where $k^* < k_\pi$. This creates a slack region for those with $w \geq w^*$ where their maximizing level of capital is an interior solution, but those with $w < w^*$ have a IC constraint that emits a frontier solution. With these assumptions, this allows me to explicitly define an agent's income as a piecewise function around w^*:

$$I(w) = \begin{cases} V(k^*) - k^*r + wr + T & \text{if } w \geq w^* \\ V(k(w)) - k(w))r + T & \text{if } w < w^* \end{cases} \tag{2}$$

This income is split between bequests and consumption: $b + c = I(w)$, which is the agent's budget constraint.

2.3 Bequest Function

To understand the transmission of wealth from one generation to the next, we characterize the policy function for an agent as their optimal bequest to their progeny. Agents are already differentiated by those who can borrow the (1) first-best capital $w \geq w^*$ and those who (2) borrow their second-best $w < w^*$. Now, agents in *both* wealth branches face another bifurcation: (3) agents who feel a gain with their bequest $b \geq w$; and (4) agents who feel a loss with their bequest $b < w$.

First, I consider when an agent codes their bequest as a gain, that is $b > w$. The utility function in Equation (1) reduces to: $U(b > w) = \gamma \ln(b) + (1-\gamma) ln(c) + \gamma\ \eta[\ln(b) - \ln(w)] + (1-\gamma)\eta[\ln(c)]$, and we then calculate the MRS between bequests and consumption. While the η term is still in the utility equation, it cancels out in the numerator and denominator of the MRS: $c^* = \frac{(1-\gamma)(1+\eta)}{\gamma(1+\eta)} b^* = \frac{1-\gamma}{\gamma} b^*$. Plugging c^* into the Income Equation (2), we obtain our optimal bequest when coded as a gain. I define this as the "baseline" bequest, $b_B(w)$, since it is simply the standard Cobb-Douglas propensity to bequeath: γ. Thus, $\forall w$ such that $b_B(w) \geq w$, then

$$b_B(w) = \begin{cases} \gamma\big[V(k^*) - k^*r + wr + T\big] & \text{if } w \geq w^* \\ \gamma\big[V(k(w)) - k(w)r + wr + T\big] & \text{if } w < w^* \end{cases} \tag{3}$$

Next, I consider an agent who codes their bequest at a loss when their baseline bequest is less than their initial wealth, $b_B(w) < w$. In this case, $\lambda > 1$ and their propensity to bequest increases. Following the same exercise as before, taking the MRS of Utility Equation (1) now with a bequest loss and plugging it into Income Equation (2), we obtain a new propensity to bequeath: $\alpha = \frac{\gamma(1+\eta\lambda)}{1+\eta-\gamma\cdot\eta(1-\lambda)} > \gamma$.

Before we can define the full functional form for of the "loss" bequest function $b_L(w)$, we must first recognize that since $\alpha > \gamma$, then $\alpha I(w) > \gamma I(w)$. Importantly, in this open interval, it is possible that $\alpha I(w) > w > \gamma I(w)$. Here, agents who bequest $\alpha I(w) > w$ bequest so much they no longer feel a loss and would like to decrease the bequest. But if they drop it to the baseline $\gamma I(w) < w$, they feel a loss again. The optimal bequest is therefore a corner solution where an agent's bequest exactly equal to their own initial wealth $b_L(w) = w$. In other words, agents will increase their bequests until the minimum level they consider it a gain, which is their own initial wealth.

I define the wealth levels for which agents bequest exactly their own initial wealth as "sticky" in Definition (2.1). Here S_j described a describes a single interval of sticky wealth, and the collection of these intervals is S.

Definition 2.1 (Interval of Sticky Wealth)**.** The total set of "sticky" wealth levels is defined as

$$S = \{w \in G(w) : \alpha I(w) > w > \gamma I(w\}.$$

Set S is the union of S_j connected components (S_j an interval), $S_j \subset S$. So long as S_j is non-degenerate, then $S_j = [w_l^j, w_h^j]$.

Intuitively, sticky intervals emerge near where the baseline bequest crosses the 45-degree line, stable steady states. At these wealth levels, an agent is already close to giving a bequest match, and it takes only a small increase to avoid the loss. We can see this in Figure (1), were as the bequest function itself increases, the agents between the two crossing points of the 45-degree link give exactly their initial wealth, creating a sticky interval.

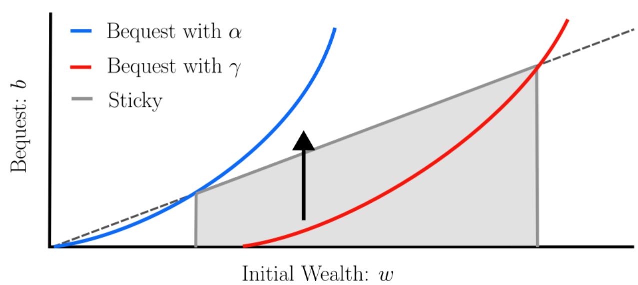


Figure 1: Bequest Changes and Sticky Interval

So, the bequest function at a loss is described in Simple terms in Equation (4). Agents poor and rich can all bequest at a loss, so $b_L(w)$ could be split by $I(w)$ and often is, so the notation remains vague.

$$b_L(w) = \begin{cases} w & \text{if } w \in S_j \\ \alpha I(w) & \text{if } w \notin S_j. \end{cases} \tag{4}$$

To write out the full, piecewise bequest mapping, $b(w)$, requires specifying the functional forms of the model as this determines pieces with loss-averse non-sticky ($b_L(w) < w$) and sticky bequests. I solve the policy function using different functional forms for the market economy in Appendix (A.2). To gain an intuitive understanding, below I present one toy version of the model.

2.3.1 Toy Model

Suppose a constant returns to investment, $V(k) = Rk$ where $R > r$ is the capital return rate. Let $\pi(k) = \pi k$, where $\pi \in (0, 1)$ and $F = V(k)$. The expected value of repayment is thus $k(R - r) + wr + T$ and the expected value of reneging is $kR - k^2 R\pi$. Setting these expectations equal to each other, we can explicitly solve for both the minimum wealth to borrow the first-best level of capital, w^*, and the second-best capital function, $k(w)$, as a convex function. The proofs for the toy model, including the algebra for w^* and $k(w)$ can be found in the appendix A.2.2.

Then, by solving for the optimal bequest while at a loss and at a gain, I can explicitly define the intervals for which sticky ranges emerge using the crossings of the gain and loss bequest functions. The total piecewise bequest function is there fore define in Proposition (1).

Proposition 1 (Loss-Averse Bequests with Convex Capital Gains)**.** *Given the functional forms of $V(k), \pi(k)$ and F, and the distribution of wealth as $G(w) \sim [\underline{w}, \bar{w}]$, we define the bequest function as:*

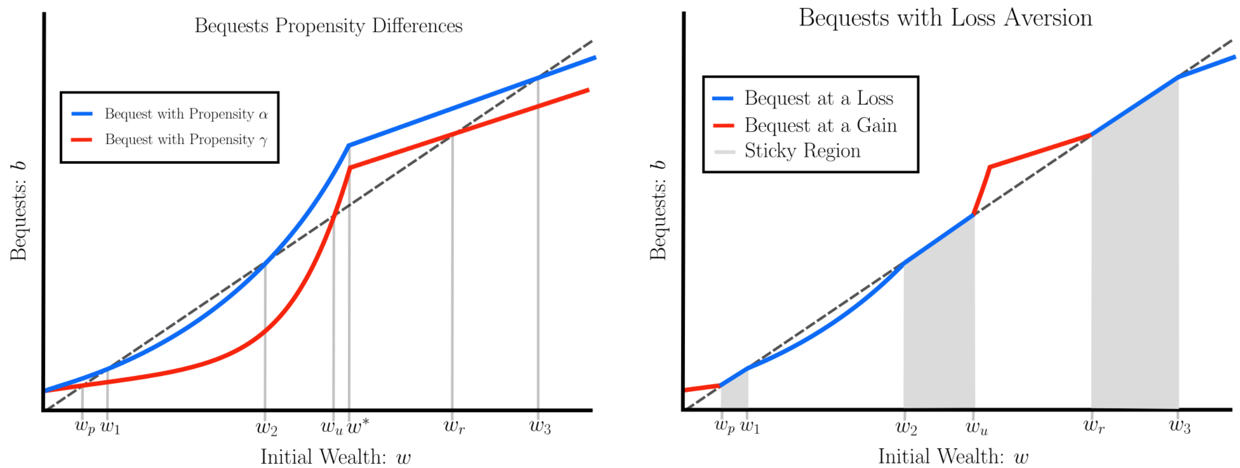
$$b(w) = \begin{cases} \alpha\left(k^*(R-r) + wr + T\right), & w_3 < w \leq \bar{w} \\ w, & w_r \leq w \leq w_3 \\ \gamma\left(k^*(R-r) + wr + T\right), & w^* \leq w \leq w_r \\ \gamma\left(k(w)(R-r) + wr + T\right), & w_u < w < w^* \\ w, & w_2 \leq w \leq w_u \\ \alpha\left(k(w)(R-r) + wr + T\right), & w_1 \leq w < w_2 \\ w & w_p \leq w < w_1 \\ \gamma\left(k(w)(R-r) + wr + T\right), & \underline{w} \leq w < w_p \end{cases}$$

where the explicitly defined kinks in the piecewise function are found in Appendix A.2.2. The sticky intervals of wealth are defined as $S = \{[w_p, w_1], [w_2, w_u], [w_r, w_]\}$. And $k(w) = \frac{r - \sqrt{r^2 - 4\pi R(wr+T)}}{2\pi R}$.

Proof. See Appendix A.2.2. □

While the above piecewise function looks complicated, ultimately the pieces simply describe when individuals are above or below w^* and when they do or do not feel a loss. Let's break it into steps. For those $w < w^*$, those directly below might earn a second-best income but still earn enough to feel a gain in the bequest without changing in from the norm, $w_u < w < w^*$. Those just below the unstable steady state w_u *do* feel a loss, but are close enough to giving a bequest that can let them feel a gain, $b = w$, in some range $w_2 \leq w \leq w_u$. But for agent just below that range, even increasing their bequest to α they cannot match their bequest, but they still bequest with propensity α and enter a vicious cycles in $w_1 \leq w \leq w_2$. Closer to the poor steady state, w_p, the threshold to match the bequest decreases:since agents start much poorer, it does not take much to feel a gain. This creates another sticky region, $w_p \leq w \leq w_1$. A similar dynamics plays out for those with $w \geq w^*$.

An illustration of this toy model can be seen in the Figure (2). In Figure (2a), the red line is the "baseline" bequest function that is similar to the frameworks in Banerjee and Newman [1993] and Banerjee and Newman [1994]. When agents feel a loss they give with propensity α, the blue line. In Figure 2a, when the blue line is above the 45-degree line while the red line is below, an agent is at a corner solution, a sticky wealth. Three new crossings emit near the baseline crossings: $w_1 > w_p$, $w_2 < w_u$, and $w_3 > w_r$. Those inequalities creates the sticky intervals in the toy model. The final function is shown in Figure (2b).



(a) Bequest Functions with Propensities α and γ.

(b) The Bequest Function in Proposition (1).

Figure 2: Bequests with Loss Aversion

When loss aversion increases – λ or η increase – the scope of sticky intervals also increases. Graphically, loss aversion rotates the bequest function counter-clockwise centered around the y-intercept since $b'_L(w) \geq b'_B(w)$. The stronger the behavioral response, the more individuals will match their exact initial wealth. The 45-degree lines becomes an absorbing line the function falls into. We can see this rotation in Figure 3, where $\lambda = 1$ is the baseline bequest curve and as λ increases to λ_1 then λ_2, so does the sticky regions and the slope of the bequest function all together.

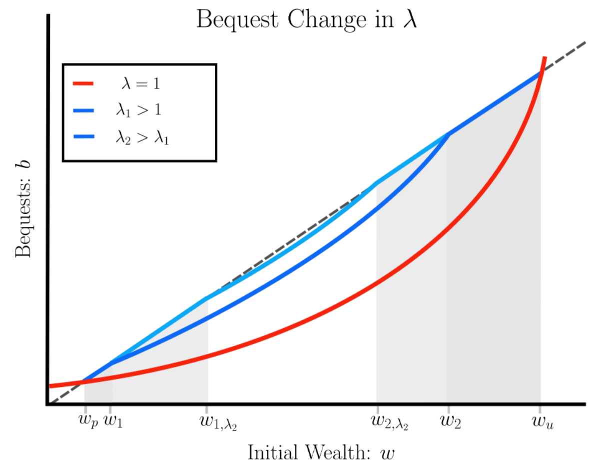


Figure 3: Loss-averse Bequest Function Change with Increase in λ.

3 Poverty Traps and Loss-Averse Dynamics

The bequest function has a naturally recursive nature, where the distributions of bequests in the current period is the wealth distribution in the next. To understand steady-state dynamics, I adopt the subscript t to relate the wealth in the current period w_t to the next period w_{t+1} where the bequest function is the mapping, $b(w_t) = w_{t+1}$. For this section, I show that my model can generate a poverty trap with dynamics very different from canonical poverty trap models.

General Poverty Traps

A poverty trap is defined as an economic mechanism that keeps agents below a certain threshold in poverty, unable to obtain a higher wealth or income. This threshold is an unstable steady state which is colloquially called the "Micawber" threshold. Agents with an initial wealth below the Micawber threshold will see their lineage converge towards a "poor" stable steady state, trapped within a vicious cycle. It is typically the case that agents above the Micawber threshold see their dynasty converge to a higher "rich" stable steady state in a virtuous cycle. Graphically, these typically looks like an "S-shaped" curve as seen in Figure (4). In this figure, the point U is the Micawber threshold. Agents with wealth w' in period t just less than U converge to the poor steady state P; agents with wealth w'' just greater than U converge to the richer steady state R. The transition function is this paper is the bequest function.

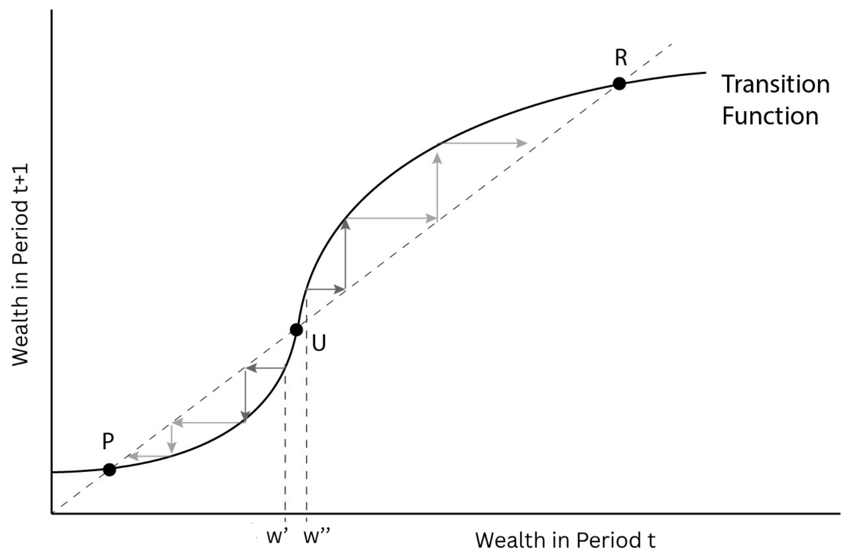


Figure 4: A General Diagram of a Poverty Trap and Wealth Transitions

The dramatic prediction of the canonical, S-shaped poverty trap is that over time, all those below the Micawber threshold converge to poverty and all those above converge to become rich. The end result is a binary, bimodal distribution, even accounting for some idiosyncratic shocks to income or wealth.

Poverty Trap with Loss-Averse Bequests

The introduction of loss aversion over bequests significantly changes the canonical poverty trap model. Without changing the underlying mechanisms that create persistent poverty, loss aversion allows agents internalize the vicious cycle and respond to it by increasing their bequest, both shortening the range of wealth levels that converge down with the vicious cycle and also slowing those dynasties' convergence. Other dynasties in the sticky regions do not converge down at all, but stay entirely stagnate without any idiosyncratic shock. Loss aversion does not necessarily allow dynasties to themselves *overcome* or *eliminate* the mechanical poverty trap, but it does dramatically change its observable transition characteristics. The comparison between the poverty trap with loss-averse bequests and the canonical trap (baseline) can be seen in Figure (5) below. We can see that the baseline (red) curve resembles the S-shaped curve, but the loss averse model (blue) flattens the bequest transition function when below the 45-degree line.

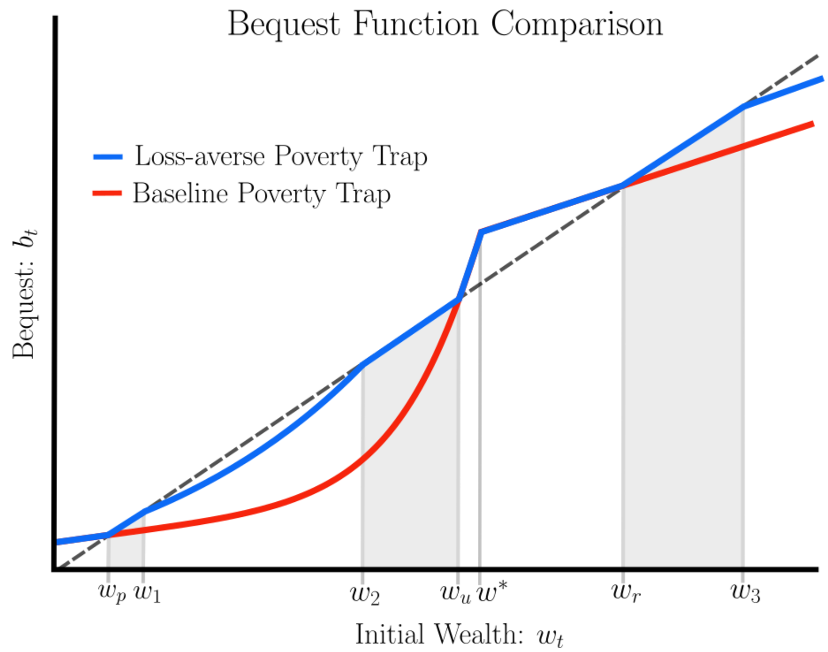


Figure 5: A poverty trap with loss averse preferences (this paper) and without (baseline)

First, loss aversion can change the observable Micawber threshold from the true Micawber threshold. The true, mechanical poverty trap threshold is at w_u, but with loss averse preferences, only agents below w_2 exhibit downward convergence. Agents between w_2 and w_u are still in the trap, but since they don't converge down, and a shock into this range would result in no downward transitions, if we identify the threshold as there start of a vicious cycle, we do not actually see a vicious cycle play out anymore in this region.

Second, downward convergence slowly significantly even grinding to a halt. The derivative of $b'_L(w) > b'_B(w)$ for all w and in the sticky regions is exactly equal to 1. As λ increases, convergence slows faster. In Figure (6), we see the transition of wealth given an agent with initial wealth w_0. In the loss-averse model, the initial wealth of an agent two generations (periods) later will be w_λ, which is modestly distant from w_0. But in the baseline poverty trap model, the initial wealth of an agent two generations (periods) later will be at w_γ, which is significantly further from w_0 than w_λ and already very close to the poor steady state. Furthermore, since w_λ will never converge to w_p but w_γ will, there will always be a difference in the long-run distribution.

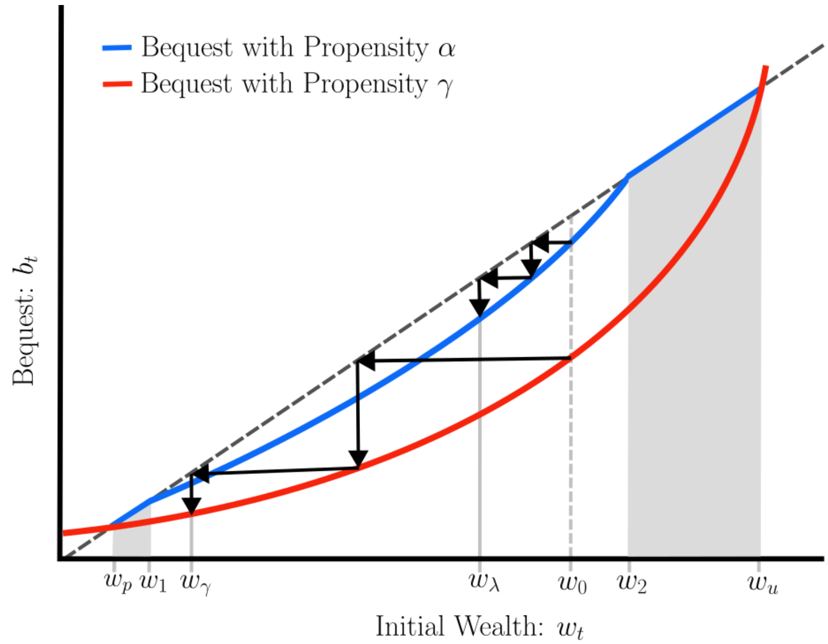


Figure 6: Differences of 2 transitions in initial wealth between baseline and loss-averse bequests.

Third, loss aversion predicts more realistic short- and long-run wealth distributions than other poverty trap models. After even one period, the distribution of initial wealth with loss averse bequests will have only slightly changes due to stickiness and slowness; but without loss aversion a researcher would expect a dramatically different distribution after one period, specifically a large divergence from the unstable steady state. [Barrett and Carter, 2013] explain that in panel data, we would expect few observations near the Micawber threshold since it is unstable. In my model, the sticky range emitting near the threshold pushes back against that postulation, predicting a permanent distribution near the threshold. In the long-run, sticky ranges and slow convergence preserves large parts of the initial distribution and does not predict the dire bimodal outcome that result from most poverty trap models.

We can explicitly compare the long-run distributions. Since b is time-invariant and continuous on a finite state space, the change in wealth follows a non-irreducible Markov process where we can formulate a limiting distribution assuming no idiosyncratic shocks.[4] By definition, poverty traps are non-ergodic. The stationary distributions simply depend on the initial distribution. Hence, in Proposition (1), we can explicitly define the long-run distribution given the current distribution $G_t(w)$ and the parameterization of the model. The complications in the new limiting distributions emit from those sticky ranges.

[4]If there are shocks associated within the transition function (i.e. to capital gains, interest, etc.) baseline models will still converge into the two basins of wealth: poor or rich. It is non-ergodic, but there is still a limiting distribution. But loss averse models with stochastic variance to income will increase the non-ergodic nature because the individual kinks in the piecewise functions become endogenously determined: there is a change in income but not in the anchor, which is different for wealth w and each shock, so the behavior is always different. Then sticky intervals are no longer totally absorbing states, they are perhaps more true to the word "sticky" where these some persistence in wealth in those regions will not necessarily capture a part of the distribution.

Lemma 1 (Limiting Wealth Distribution for Toy Model (Section 2.3.1))**.** *Let the bequest function take the parameterization of the toy model in Proposition (1). We know b is a direct mapping $b : G_t \mapsto G_{t+1}$. In regions outside the sticky regions of wealth, the bequest mapping is a contraction to the local basins: w_p, w_1, w_r, and w_3. Inside the sticky regions, a given wealth level w does not contract and stays at $G_0(w)$ indefinitely. The resulting stationary CDF is a mixed measure with point masses at the basins and a continuous distribution in sticky regions:*

$$G_\infty(w) = \begin{cases} 0, & w < w_p, \\ G_0(w), & w_p \leq w < w_1, \\ G_0(w_2), & w_1 \leq w < w_2, \\ G_0(w), & w_2 \leq w < w_u, \\ G_0(w_u), & w_u \leq w < w_r, \\ G_0(w), & w_r \leq w < w_3, \\ 1, & w \geq w_3. \end{cases} \tag{5}$$

As $\lambda \to 1$ or $\eta \to 0$, then $w_1 \to w_p$, $w_2 \to w_u$, and $w_3 \to w_r$, so $G_\infty = G_0(w_u)\delta(w_p) + [1 - G_0(w_u)]\delta(w_r)$ (baseline).

Fourth and finally, welfare remains ambiguous. Agents sacrifice their consumption to give a larger bequest. It is a philosophical debate if we prefer a world in which poorer parents must consume much less than they otherwise should to save their child from a worse poverty, even if it does make the wealth distribution better in the long-run. I leave this analysis to future research.

4 Econometric Implications

In this section, I show that a poverty trap with loss averse bequests can be invisible to empirical identification. First, I provide a general overview of the mixed evidence for poverty traps. I then formulate and test two different common empirical models to test for the presence of a poverty trap: a non-parametric and a parametric model. By comparing the results of the test on data sets generated from models with and without loss aversion, I show tests that otherwise would accurately confirm a poverty trap fail to reject the null hypothesis, H_0 : A poverty trap *does not exist.* I explain the reasons for each tests failure, and provide some motivation for further empirical research.

4.1 Benchmark Empirical Tests

Proving the existence of a poverty trap using income and wealth data has been difficult and divisive. Research by Kraay and McKenzie [2014] find it incredibly rare to empirically and definitively identify situations where all agents converge to a single state (poor) or multiple steady states (poor and non-poor) using income panel data. Carter and Barrett [2013] also acknowledge and attempt to rectify the short comings of income data by leveraging asset, or wealth, data that can better differentiate transitory from perpetual poverty. They also explain other limitations of panel data due to attrition and ability to capture income and wealth, and also that short panel data often do not contain enough periods that can allow the vicious cycle play out. But even when researchers have done their best to control for heterogeneity, use different empirical tests, and have sufficiently rich data, empirically identifying a poverty trap, or rather rejecting the null hypothesis a poverty trap does not exist, has yielded very mixed results.

The dynamics from loss averse bequests might provide an explanation. Due to sticky regions of wealth, researchers that predict convergence to poverty might not observe a decline in intergenerational wealth or income. Lineages stay close to the unstable steady state, which can confuse tests looking for one. And the slow transition dynamics in combination with sticky regions might require many generations to identify transitions and rich data due to the model's sensitivity to shocks. I preform a test to identify the unstable steady state, future research should explore how other tests, like the S-shape test, might be affected by my model's dynamics.

Threshold Test. The threshold test estimates a point (Micawber threshold) where the average change in wealth below tends to a lower state than the average change above. This is the most common empirical identification method for multiple equilibira poverty traps like in this model [Jalan and Ravallion, 2004, Carter and Barrett, 2013, Antman and McKenzie, 2007]. The true threshold is $c = w_u$ and the estimated threshold is $\hat{c}$. We represent the line of best fit for the expected change in wealth above and below the threshold for $w \geq \hat{c}$ as $g_h(w) = \mathbb{E}[w_{t+1} - w_t | w \geq \hat{c}]$ and $g_l(w) = \mathbb{E}[w_{t+1} - w_t | w \geq \hat{c}]$. These are the "drift equations" as the estimate the directions of change in wealth.[5]

I design the empirical estimation model in Equation (6) based on the framework developed by Hansen [2000]. The coefficients determine the drift equations $g_h(w) = \beta_0 + \beta_1 w$ and $g_l(w) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)w$. The null hypothesis is that multiple-equilibia poverty trap does not exist, and to reject requires (1) a bifurcation point where the two equations better represent the total change in wealth better than one line (different trajectories), so $\beta_2 = \beta_3 = 0$; and (2) that the lower drift is a negative change directly below the threshold, indicating convergence to a lower state, $\hat{g}_l(\hat{c}) < 0$.[6]

$$\text{(T1)} \quad \Delta w_{i,t+1} = \beta_0 + \beta_1 \, w_{i,t} + \beta_2 \, \mathbf{1}\{w_{i,t} < c\} + \beta_3 \big(w_{i,t} \cdot \mathbf{1}\{w_{i,t} < c\}\big) + \varepsilon_{i,t}. \tag{6}$$

I use a Monte Carlo simulation over 5 periods and 500,000 agents, significantly richer than even

[5] Monte Carlo simulations are especially powerful for ergodic Markov chains since more observations tends towards to true expected values of the underlying data generating process. However, this model is non-ergodic, so more simulations of the data generating process might produce better estimates of the expected value, drift, but will not necessarily tend towards a true stationary distribution.

[6] If $\hat{g}_l(\hat{c}) < 0$ then everyone is in the poverty trap. But a negative slope on the line of best fit just means on average, agents above are converging down, though still to a higher state than those below the threshold.

the best panel data previously used [Barrett and Carter, 2013]. I identify the threshold $\hat{c}$ using a grid search over a fine mesh of different c values and pick the $\hat{c}$ that minimizes the sum of squared errors for $\min_\beta ||Y - X(c)\beta||^2$.

The simulation results from Table (1) fail to reject the null for the model with a loss-averse bequest motive though, importantly, we can reject the null without loss-averse (baseline) preferences. In the data generated from the baseline poverty trap model, the threshold test accurately identifies the correct threshold, $\hat{c} = 0.8477$ and identifies that the lines of best fit for wealth levels below the threshold is quite different than those above, since $\hat{\beta}_3 = -0.795$ and $\hat{\beta}_4 = +0.5453$.[7] Finally, since $\hat{g}_l(\hat{c}) < 0$ and $\hat{g}_h(\hat{c}) > 0$, we know those below tend to a lower state and those above a higher. This is enough evidence to reject the null hypothesis. But my model shows when reference-dependent preferences are not accounted for, the threshold test fails to reject the null hypothesis. While $\hat{\beta}_3 = -0.2399$ and $\hat{\beta}_4 = +0.1782$, they are significantly closer to 0. Additional stochastic variance or less rich panel data might drive these estimates close to zero and become statistically insignificant, especially with strong loss aversion. Second, though there is evidence for a very weak difference in expected changed in wealth, the model predict *all* wealth levels grow on average. Since $\hat{g}_l(\hat{c}) = 0.0019$, the threshold test indicates a slow growth in wealth until it passes the threshold into a faster growth. This is driven by sticky regions of wealth staying close to the unstable threshold, no only flattening the predicted drift slope, but due to those individuals being sensitive to shocks and more easily able to get shocked over that threshold when close to it, the test fails to identify individuals will on average experience downward convergence.

Table 1: T1: Conventional threshold regression for downward convergence

Model	$\hat{c}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{g}_l(\hat{c})$	$\hat{g}_h(\hat{c})$
Loss-averse	0.8418	+0.2381	-0.1734	-0.2399	+0.1782	0.0019	0.0929
Baseline	0.8477	+0.7928	-0.6850	-0.7952	+0.5453	-0.12	0.212

Notes: All coefficients are statistically significant within the 99th percentile. "Slope below" is the local slope of $\mathbb{E}[\Delta w \mid w]$ for $w < \hat{c}$. In the loss-averse run, the slope below $\hat{c}$ is near zero (sticky).

I conduct a robustness check in Appendix (A.5).

S-shaped Curve. Another popular estimation uses a cubic polynomial function to identify an S-shaped curve particular to a multiple-equilibria poverty trap, like in Banerjee et al. [2019]. This estimates 3 points where the bequest function crosses the 45-degree line, as seen in Equation (7). A poverty trap exists if (a) there are three roots, and (b) if the slopes for the lowest (β_1) and highest (β_3) are negative and the middle slope (β_2) is positive. This implies that the middle root is an unsteady state, the rate of change around is increasing, while the rate of change is decreasing into the roots. Therefore, the null hypothesis is that there are no three distinct roots and the slope

[7] Note, $\hat{\beta}_2 < 0$ and $\hat{\beta}_3 > 0$ only means the intercept and slope of the expected change in wealth is different than the slope above. The slope of those below could be increasing or greater than the slope of the expected change below, but the two lines will still result in one higher and one lower steady state, which is what the next part of the test shows.

at root 2 is non-positive and the other roots non-negative, $H_0 : 0 \nless \beta_1 \nless \beta_3$ and $\beta_2 \geq 0$ and $H_1 : \beta_2 < 0 < \beta_1 < \beta_3$.

$$\text{(S2)} \quad \Delta w_{i,t+1} = \beta_0 + \beta_1 \, w_{i,t} + \beta_2 \, w_{i,t}^2 + \beta_3 \, w_{i,t}^3 + \varepsilon_{i,t}. \tag{7}$$

The results in Table (2) indicate the parametric model can still identify a poverty trap with significant statistics, but importantly it is misidentified. The roots for the poor (w_p), non-poor (w_r), and the Micawber threshold (w_u) exist, but are not the true steady states as can be identified in the baseline. As predicted by my model, the estimated (observed) poor and rich steady states are higher than is true at .1285 and 1.4267 due to the sticky ranges of wealth halting downward convergences. Most notably, the Micawber threshold has decreased significantly from 0.8454 (true) to 0.4599 (observed) due to the sticky range that emits below the unstable state, as seen in Figure (5). Again, I drop notation for statistical significance because of the simple data generated process.

Table 2: S2: S-shaped regression

S-shaped regression coefficients						
Model	Root 1	(slope)	Root 2	(slope)	Root 3	(slope)
Loss-averse	0.1285	(S, −0.063)	0.4599	(U, +0.047)	1.4267	(S, −0.184)
Baseline	0.1058	(S, −0.623)	0.8454	(U, +0.187)	1.1624	(S, −0.267)

Notes: All coefficients are statistically significant within the 99th percentile.

Another notable result from this exercise is that the slopes for all steady states decrease fairly significantly. This is due to the "flattening" of the curve loss averse bequests create. If loss aversion were stronger, or the poverty trap weaker, it is entirely possible with some stochastic variance to inter-period wealth that those signs converge to zero or even flip, becoming statistically insignificant. The S-shaped model is very sensitive to the underlying bequest function generating the data. Given the best-case scenario of data these exercises use, a poverty trap is still difficult to identify and potentially unidentifiable using the S-shaped empirical model.

4.2 Insights for New Analysis

I turn to discuss ways in which researchers might better identify poverty traps if agents are loss averse.

Consumption Data. Consumption data might give the full picture of what has been sacrificed to maintain the bequest in the presence of income shocks. As Carter and Barrett [2013] point out, income transitions alone are very flawed in their abilities to disaggregated transitory poverty from a trap; but wealth data too is limited in explaining what downward forces an agent might be experiencing just to maintain a current wealth level. They note that a negative income shock that does not change underlying wealth in the next period (measured in assets) is evidence there might not be a trap. However in my model, this is not true—anchored beliefs preserve bequests through a decrease in consumption. Consumption is pro-cyclical with shocks and first changed to avoid drops in the bequest, as seen in Figure (7).

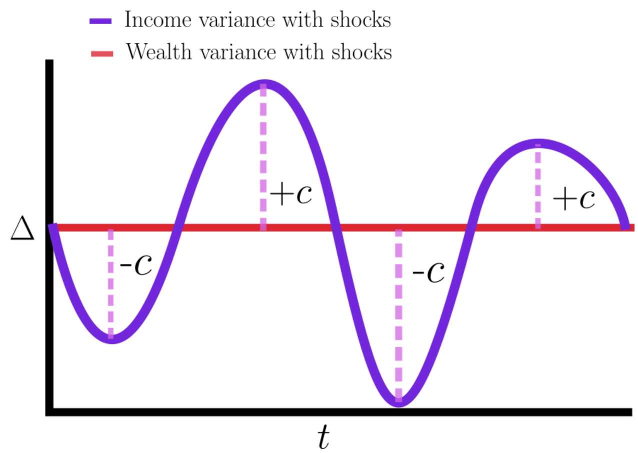


Figure 7: Consumption change in response to income shocks with anchored bequest expectations.

Consumption data often reveals a "tick": those who are very poor have a higher propensity to consume than those a little less poor [Brewer et al., 2017]. This model offers a potential solution to that puzzle. Agents at or below w_p will consume with propensity $1-\gamma$, but all agents above w_p will consume with a lower propensity until $1-\alpha$. Not only is the propensity to consume lower, the very poor might consume much more levels than the slightly less poor, too. Given the limitations of panel data being too short or giving mixed evidence for a poverty trap's existence, incorporating consumption data and controlling for anchored bequests might circumnavigate those shortcomings.

My model is limited to only described permanent income shocks and not temporary shocks or shocks within a life cycle. If we want to understand how loss aversion might affect consumption or bequest smoothing across a multi-period lifetime, this would allow better insight into identifying poverty traps within panel data that covers decades rather than generations. Second, the definition of a "bequest" in the model is abstract, and depending on how it is defined would change what we would look for in the data. If bequests are defined as an investment (i.e. college savings, dowry), which fits my model, then researchers would be looking for different observations in the data than if bequests are treated as an expenditure (i.e tutoring, books). As an investment, temporary income shocks might decrease bequests as agents smooth then raise it later. As an expenditure, temporary income shocks might negatively change consumption (i.e. consuming less, or consuming less quality goods) before bequests. Future research should consider how bequests and shocks should be defined and how to find those in the data, like PSID.

Uniquely Rational Propensity Behavior. Barrett and Carter [2013] encourage researchers to identify behaviors in the data that are only rational within a poverty trap. For example, agents who experience the hardships of poverty might shorten their time horizons as to not suffer the expectation of future poverty. My paper connects with this hypothesis: the behavior of agents is endogenous to their presence in the poverty trap, and we can identify the wealth levels where that behavior occurs.

Two agents shocked to an income level between their expected income levels will have different propensities potentially due to their initial condition anchor. Idiosyncratic shocks to capital gains (or, generally, income) will change the realized income from the expected income but it will not change the anchor specific to that initial wealth level. If there was no poverty trap or outside a poverty trap, an agent who experiences a *positive* income shock would have the same propensity to bequeath as an individual who experienced no shock but earned the same income. Only within a poverty trap with strong economic headwind will the agent with a positive shock bequeath less than an agent with no shock at the same income level. Conversely, a richer agent shocked down, regardless of a poverty trap or not, will increase the bequest propensity under loss aversion. But an agent who did not receive a shock at the same lower income level will have a similarly high bequest propensity within a poverty trap: both feel a strong headwind relative to their initial condition. The richer agent shocked down will almost always bequest with α, and the agent without a shock earning the same income in the poverty trap will have a propensity near α but always higher than γ. Individuals within and outside of the poverty trap can have bequest propensity γ or α depending on the direction of the shock and their initial condition. We see this in the figure below.

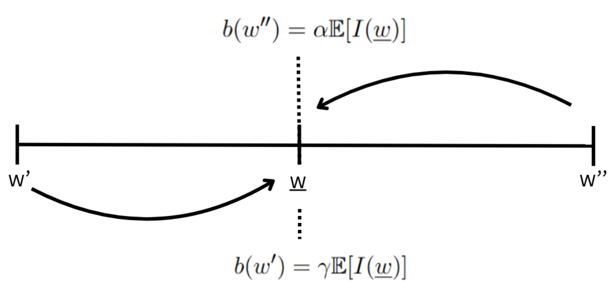


Figure 8: Diagram of bequest shocks that synchronize two agents' income but not their preferences.

Both agents feel a shock and though their income is the same, the initially richer agent increases their bequest while the initially poorer agent decreases their bequest. Both *might* be different from the agent that did not receive a shock. Outside a poverty trap, the middle income person would have propensity γ. Inside a poverty trap, the middle income agent will have propensity $\leq \alpha$. Identifying what the income propensity of an agent at that convergent income in combination with the other two propensities, could identify if there is a poverty trap.

In the data, we could identify a case of an income-averaging shock which creates a small quasi-experiment, where by aggregating many cases allows us to estimate the average effect of an upward/downward shock on bequest propensity while controlling for heterogeneity. This requires three data: consumption, income, and initial conditions. Within context to this model, initial conditions are initial wealth, though in other contexts could be identified by correlating variables, like education.

As described before, future research will have to disseminate consumption smoothing, life-cycle effects, and hetergenous preferences, but more sophisticated models and using these quasi-experiements in aggregate could help better measure economic headwind and identify when it is a trap.

5 Policy Interventions

Before defining a social welfare planner and their policy response, it is important here to note that the welfare consequences of loss-averse bequests are ambiguous. As discussed in Sections (2) and (3), agents sacrifice their present consumption to offer a greater bequest when trapped in a vicious cycle. When welfare is weighed from present consumption, loss aversion negatively affects aggregate welfare, especially for the poor who have higher marginal utility from consumption. When welfare is weighed from the bequest – or rather, future wealth distributions – then loss aversion positively affects aggregate welfare. A policy planner that cares about both present consumption and future wealth distributions balances these opposing forces in addition to the standard equity-efficiency tradeoffs.

A "big push" policy is the typical policy response to a poverty trap in which a individualized demogrant given to all individuals below the (observed) unstable steady state such that their income is equal to the income of an agent immediately above the threshold [Kraay and McKenzie, 2014, Barrett and Carter, 2013]. A social planner will only enact this policy if there is a poverty trap; in the world without a poverty trap where poverty might be slow but ultimately transitory, there is not necessarily a welfare gain by rushing the upward transition. But in the presence of a poverty trap, there are large welfare gains (aggregate output, utilitarian social welfare, etc.) to get everyone in the state of mobility and then let natural economic mechanisms of mobility handle the rest. However, a social planner must first identify the presence of a trap to know if this policy is necessary.

As discussion in Section (4.1), under loss-averse preferences the poverty trap can be almost unidentifiable. If a policy planner fails to identify a poverty trap, they will rely on the classic public policy redistribution policies which could be totally ineffective. Even if it is identified, the correct the Micawber threshold is likely negatively under-estimated, then a targeted big push policy would be inefficient or ineffective.

In this section, I used the toy model from Proposition (1) and evaluate the disparate impacts of two policies: a tax-rebate redistribution (given a poverty trap is not identified) and a big push policy (given a poverty trap is identified). The policy analysis in this model is limited, but gives us insight into the effects on bequests from a distortionary tax, primary drivers of the equity-efficiency tradeoff, and long-run outcomes for both policies. I provide insights from a one-period intervention, and leave it to future research to study path-dependent policy and inter-period taxes. Future research should also use more sophisticated models to provide a normative policy analysis.

5.1 Benchmark Policy: Flat-Tax Rebate

Suppose that given empirical tests of the integerantional wealth and income data, *a poverty trap is not identified*. The typical social planners response, then a typical policy response would be a basic tax-redistribution scheme.

Let the economy and agent preferences be defined by the toy model in Proposition (1). Agents not face a flat-tax rate τ^* on new investment income $(1-\tau)k(R-r)$ following the Mirrlees [1971]. Agents then receive a demogrant which we let be T. Agents who do not renege realize an expected income net of capital profits with the rebate: $(1-\tau)kR - kr + wr + T$. Agents who renege forfeit the rebate and their collateral, but leave without needing to repay the loan or the tax, leading to an expected income: $[1-\pi k]kR$. While the tax rate decreases the take-home pay for repaying borrowers, the good-behavior supplement increases their income. This creates a trade-off for borrowers that affects the lenders new incentive-compatibility constraint:

$$(1-\tau)kR - kr + w \cdot r + T \geq [1-\pi k] \cdot (1-\tau)kR \tag{8}$$

Rearranging Equation (19), lenders will only administer a small enough such that agents are indifferent to repayment and reneging, where they choose repayment in the static equilibrium. To borrow the first-best level of capital k^*, an agent must have $w^* = \frac{k^*(R\tau+r)-(k^*)^2(\pi R)-wr-T}{r}$. Intuitively, as the tax rate goes up, so does the required wealth for first-best borrowers $k(\tau R + r) > kr$ but goes down as the rebate goes up $T > 0$. Intuitively, as the tax goes up the rebate increases the lowest poverty stable steady state, but the rebate might be small compared to the loss in returns for higher wealth individuals $T - k^* R\tau < 0$. . For non-first best borrowers, the lending function for those with $w < w^* =$

$$k(w,\tau,T) = \frac{\tau(R-r)}{2\pi R} + \frac{r - \sqrt{(\tau(R-r)+r)^2 - 4\pi R(wr+T)}}{2\pi R} \tag{9}$$

Mathematically, while $T > 0$ is an upward shift in the return function, the tax τ causes a clockwise rotation around the y-intercept. Intuitively, while it increases the wealth of agents, the loss of returns also hurt agents borrowing the second-best level of capital. The new bequest function incorporates these trade offs. Solving the MRS given the Utility Function (1) and the income function in Equation (20), the bequest function when agents feel a gain is:

$$I(w,\tau) = \begin{cases} k^* \cdot (1-\tau)(R-r) + w \cdot r + T, & \text{if } w \geq w^* \\ k(w,\tau) \cdot (1-\tau)(R-r) + w \cdot r + T, & \text{if } w < w^* \end{cases} \tag{10}$$

By combining the above income in Equation (21) with the utility function in Equation (1), we can rederive the updated bequest function. Instead of rewriting the long bequest function here, note that the only change from Proposition (1) is that net capital gains are taxed, resulting a decrease in the marginal returns of investment, also reducing the optimal capital k^*.

Aggregate capital in the economy can be written as K, which is simply the mass of those borrowing first-best and then the expected value of those earning their individualized second-best level of capital, $K = \int_{w^*}^{\bar{w}} k^* g(w)\partial w + \int_{\underline{w}}^{w^*} k(w,\tau)g(w)\partial w$. This can be reduced and utilized in Equation (22):

$$K(\tau) = k^* \cdot [1 - G(w^*)] \; + \; \mathbb{E}[k(w,\tau) \mid w \in [\underline{w}, w^*]] \tag{11}$$

Assume a myopic social welfare planner who chooses an optimal tax rate τ^* for a one-period flat-tax and rebate to all agents that maximizes welfare. Since the mass of workers is normalized to 1, $T = \tau K(R-r)$. I let social welfare be utilitarian weighted, and over an agent's Cobb-Douglas utility for simplicity. While this is a one-period tax policy, since the bequest function is a direct mapping of next period's welfare, the planner already internalizes the effect of the tax rate on the next period's wealth distribution. Using the aggregate capital function in Equation (22), we can define the social planner's problem:

$$\mathbb{W} = \int_i g(w^i) \cdot U^i(w^i,\tau)\partial w^i = \sum_{(a,b)} \int_a^b g(w^i) \cdot U^i_j(w^i,\tau)\partial w^i \tag{12}$$

$$(a,b) = \{[\underline{w}, w_p], [w_p, w_1], [w_1, w_2], [w_2, w_u], ...\}$$
$$\text{s.t. } \tau K(R-r) \geq 0$$

Given the social planner's welfare function in Equation, we use the envelope theorem to write an implicit function for the optimal tax using the framework from Piketty and Saez [2013] in Proposition (2)

Proposition 2 (Implicit Optimal Tax)**.**

$$\tau * = \frac{1 - \overline{\mathbf{G}}}{1 - \overline{\mathbf{G}} + e}$$

where $\overline{\mathbf{G}} = \frac{1}{K}\mathbf{G}$*, and* $\mathbf{G}$ *is the normalized social marginal weight:*

$$\mathbf{G} = \frac{\sum \Big(\mathbb{E}[I^{j,k}_\tau \mid b^*(w) \geq w]\Big) + \frac{1-\gamma}{1-\alpha}\sum\Big(\mathbb{E}[I^{i,j}_\tau \mid b^*(w^i) < w^i]\Big)}{\sum\left(\mathbb{E}[(I^{j,k})^{-1} \mid b^*(w) \geq w]\right) + \frac{1-\gamma}{1-\alpha}\sum\left(\mathbb{E}[(I^{j,k})^{-1} \mid b^*(w^i) < w^i]\right)}.$$

I define $e = \frac{1-\tau}{K}\frac{\partial K}{\partial(1-\tau)}$ *as the elasticity of aggregate capital to net-of-tax rate,* $1-\tau$*.*

Proof. See Appendix (A.6). □

In Proposition (2), loss aversion affects this optimal tax rate by (1) up-weighting those feeling a loss (equity) and (2) preserving the wealth distribution (efficiency). Both poor and rich agents can feel a loss, and their utility is unweighted by $\frac{1-\gamma}{1-\alpha}$ when social planner's *do not* care about loss averse utility. When they do, this weight becomes stronger: $\frac{(1+\eta)(1-\gamma)}{\eta(1-\alpha)} > \frac{1-\gamma}{1-\alpha}$. Assuming this is a world in which a policy planner doesn't identify a poverty trap, it is reasonable to assume they also did not account for loss averse preferences. The standard equity-efficiency trade off without loss aversion is simply that taxing today, while redistributing income to increase equality, changes the policy function for agents, reducing the wealth bequeathed to the next generation and the ability of a social planner to tax later, causing a dead weight loss and a less efficient economy.

Those trade-off blow up in loss aversion. First, the distribution of those feeling a loss will change the optimal tax by equity alone. If many are poor and feeling a loss, this will increase the optimal tax compared to the baseline model. Conversely, if many of the rich feel a loss, this will decrease the tax compared to the baseline, though in both scenarios, the wealth distribution is the same. Additionally, it is possible that one very rich agent who feels a loss might decrease the optimal tax more than much poorer agents who feel less of a loss in pure monetary terms.

This also changes the efficiency. Naturally, loss aversion preserves the wealth distribution, allows higher tax rates to have a muted negative effect on the wealth distribution loss in the next generation. However, a small increase in a poor agents income is likely to just be eaten rather than passed one, since they consume everything after matching the bequest and the rich will always bequeath less when taxed. This means the tax policy has a smaller positive impact on the next wealth distribution, drawing down the optimal tax. In Figure (9), the change in outcome when agents are loss averse is shown in the decrease in optimal tax given a normal distribution centered within the range of mobility.

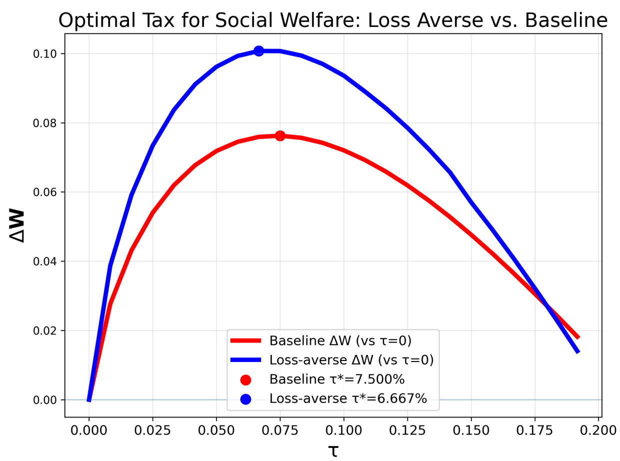


Figure 9: Optimal tax rates for loss-averse and baseline poverty trap models.

The result of the tax-redistribution in the poverty trap model is often severe and can result in welfare reducing outcomes, as seen in Figure (10). The tax distorts the capital market outcomes, since very quickly the tax burden is greater than the demogrant gain, the incentive to repay decreases. This causes necessary wealth to borrow the first-best level of capital, w^*, to increase, capturing more wealth levels in the poverty trap. While the poor steady state increases, if the tax policy was turned off the next period, you would simply have a larger share stuck in the poverty trap than if the planner implemented no tax policy at all. Similarly, first-best borrowers realize smaller incomes and reduce their bequest dramatically. While preserving the wealth distribution some, the next period distribution will be substantially effected, causing welfare-reducing long-run results.

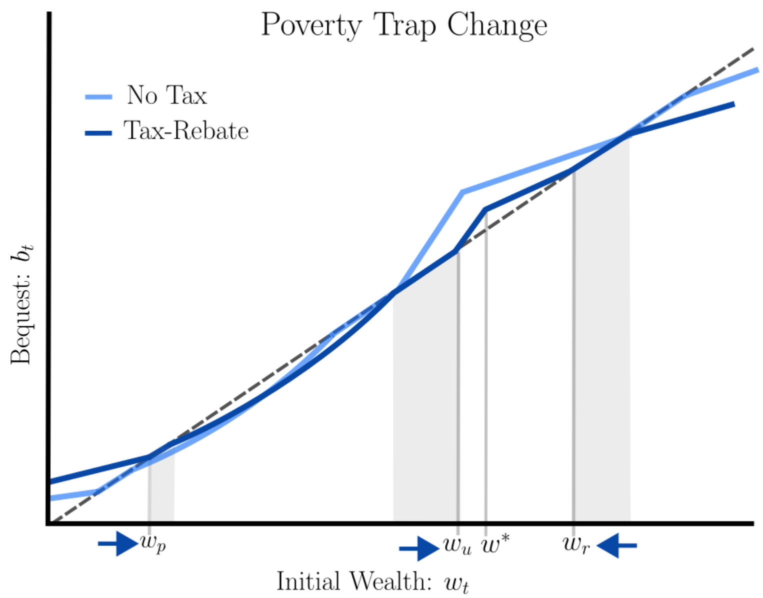


Figure 10: Loss averse bequest change with redist. policy

5.2 Big-Push Policy

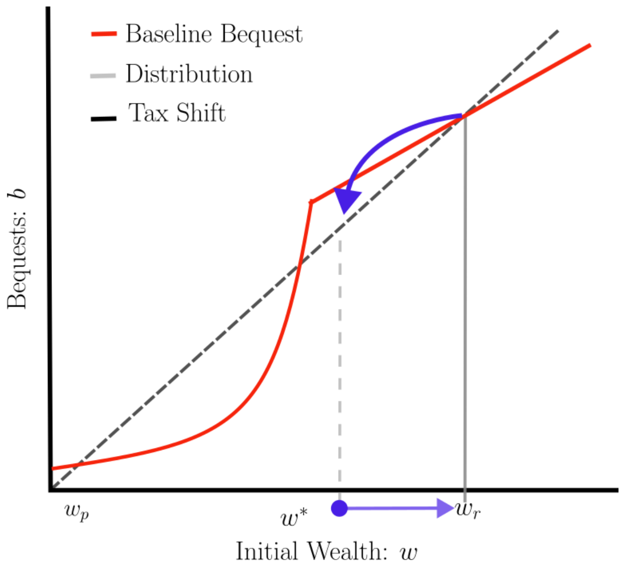
Suppose a poverty trap is identified but the stable steady states and transition dynamics are misidentified, as shown to be probably by the S-shaped test in Section (4.1). A planner uses targeted "big-push" transfers to agents below the estimated threshold, which in this model would push agents to $w_2 < w_u$, in the sticky region.

I study the simplest, most optimistic case. Suppose the policy is financed through debt (bonds or sovereign) to be repaid much later. We could characterize the total demogrant spending as $D = I(w_m)G_0(w_m) - \int_{\underline{w}}^{w_m} I(w^i)g_0(w^i)\partial w^i$ where w^i is wealth of agent i and $w_m \in \{w_2, w_u\}$, then each eligible agent receives a personalized demogrant $d(w^i)$.

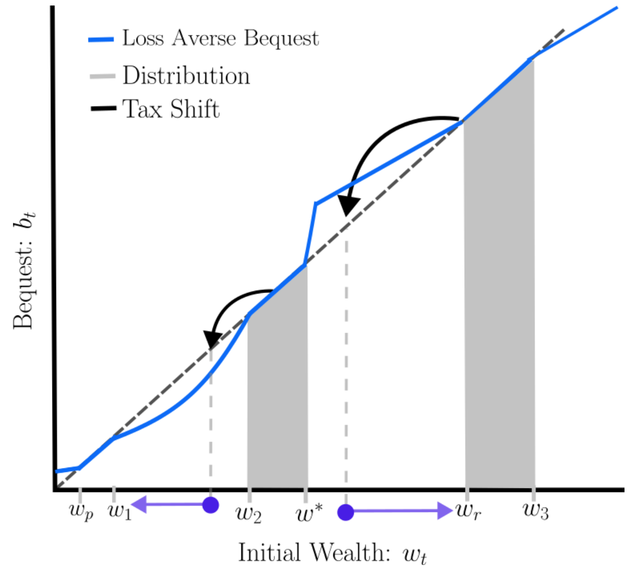
Let the debt be repaid through a flat net-income tax when the distribution reaches equilibrium (for simplicity) and that generations know they will not be taxed until then.[8]. In the baseline model, agents are accurate lifted above w_u and all agents have wealth w_r. In the loss averse model, a large measure of the population is stuck between w_2 and w_u.

Accounting the market distortions from the tax, we can view a tax as a negative "shift" in agents' income that changes (often lowers) what they bequest. In Figure (11a), the unit mass of agents at w_r are pushed down, but eventually rises back to w_r. In Figure (11b), agents held in the sticky regions are pushed down but now fall into the vicious cycle down to w_1. Even those still in sticky wealth do not recover higher wealth.

[8] I lose Ricardian equivalence by assuming that future generations *do not* expect to be taxed. Since any tax policy distorts the capital lending market equilibirum, lowering the Micawber threshold than in the first period, the optimal policy is path dependent and quickly complicates. This is not necessary to calculate to understand the basic intuition I present in this paper, but recommend it for future research.



(a) Income changes from tax with Baseline Preferences



(b) Income changes from tax with Loss Averse Preferences

A tax or negative shocks will through agents back into worsening poverty traps, causing continued big-push policies in perpetuity, making this policy hugely ineffective. Instead of a large, one-time deadweight loss from a tax, a policy planner will impose a smaller tax burden but the continued need will accumulate, and the long-run output is likely to be much lower, causing greater inefficiencies.

6 Conclusion

In this paper, I formulate a model of a poverty trap where agents have reference-dependent bequest preferences. Due to the non-linear bequest motive, the dynamics of the resulting observable poverty traps are significantly different than previous models. I show the empirical methods used to reveal the presence of a poverty trap will fail to reject the null hypothesis that they do not exist if these preferences are present and not accounted for by researchers. I then go on to provide different policy analysis based on two typical empirical failings, and show these policies can be welfare reducing or ineffective.

There are important next steps in this line of research. First, this model assumes the reference point for bequests is dependent only on the initial wealth level but it would be important to understand how the distribution of wealth, wealth inequality, might shape a parent's expectation for how much they should give their kid: the reference-point is distribution-dependent. Next, it would be worthwhile to use this micro-foundation to inform structural models parameterized using actual data to better measure the potential effects and policy solutions. Researchers might also leverage the insights of this model to revisit previous contexts where poverty traps might exist.

7 References

References

C Benson, A Bishaw, and B Glassman. Persistent poverty: Identifying areas with long-term high poverty., 2023.

Aart Kraay and David McKenzie. Do poverty traps exist? assessing the evidence. *Journal of Economic Perspectives*, 28(3):127–148, 2014.

Michael R Carter and Christopher B Barrett. The economics of poverty traps and persistent poverty: An asset-based approach. In *Understanding and Reducing Persistent Poverty in Africa*, pages 12–33. Routledge, 2013.

Abhijit V Banerjee and Andrew F Newman. Occupational choice and the process of development. *Journal of political economy*, 101(2):274–298, 1993.

Christopher B Barrett and Michael R Carter. The economics of poverty traps and persistent poverty: Empirical and policy implications. *The Journal of Development Studies*, 49(7):976–990, 2013.

Abhijit Banerjee, Emily Breza, Esther Duflo, and Cynthia Kinnan. Can microfinance unlock a poverty trap for some entrepreneurs? Technical report, National Bureau of Economic Research, 2019.

Clare Balboni, Oriana Bandiera, Robin Burgess, Maitreesh Ghatak, and Anton Heil. Why do people stay poor? *The Quarterly Journal of Economics*, 137(2):785–844, 2022.

Abhijit Vinayak Banerjee. Field experiments and the practice of economics. *American Economic Review*, 110(7):1937–1951, 2020.

Carlo Barone, Katherin Barg, and Mathieu Ichou. Relative risk aversion models: How plausible are their assumptions? *Rationality and Society*, 33(2):143–175, 2021.

Mahmud Yesuf and Randall A Bluffstone. Poverty, risk aversion, and path dependence in low-income countries: Experimental evidence from ethiopia. *American Journal of Agricultural Economics*, 91(4):1022–1037, 2009.

Thomas Piketty and Emmanuel Saez. Optimal labor income taxation. *Handbook of public economics*, 5:391–474, 2013.

Botond Kőszegi and Matthew Rabin. A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165, 2006.

Abhijit V Banerjee and Andrew F Newman. Poverty, incentives, and development. *The American Economic Review*, 84(2):211–215, 1994.

Jyotsna Jalan and Martin Ravallion. Household income dynamics in rural china. *Insurance against poverty*, pages 108–124, 2004.

Francisca Antman and David McKenzie. Poverty traps and nonlinear income dynamics with measurement error and individual heterogeneity. *The journal of development studies*, 43(6):1057–1083, 2007.

Bruce E Hansen. Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603, 2000.

Mike Brewer, Ben Etheridge, and Cormac O'dea. Why are households that report the lowest incomes so well-off?, 2017.

James A Mirrlees. An exploration in the theory of optimum income taxation. *The review of economic studies*, 38(2):175–208, 1971.

Liam C Malloy. Loss aversion, education, and intergenerational mobility. *Education Economics*, 23 (3):318–337, 2015.

Rebecca Ryan, Ariel Kalil, Mesmin Destin, Marlis Schneider, David Silverman, and Ivan Hernandez. Parents' beliefs in the. *University of Chicago, Becker Friedman Institute for Economics Working Paper*, (2024-123), 2024.

Jonathan D Fisher, David S Johnson, Timothy M Smeeding, and Jeffrey P Thompson. Estimating the marginal propensity to consume using the distributions of income, consumption, and wealth. *Journal of Macroeconomics*, 65:103218, 2020.

A Appendix

A.1 Reference-Dependence Anchors

In the most general version of this model λ should be differentiated into λ_m^b (λ specific to losses over bequests) and λ_m^c (λ specific to losses over consumption) to examine how preference and behaviors change when agents perceive a loss to neither, one, or both consumption and bequests. However, for the purposes of this paper, I restrict consumption to always be coded as a gain ($c \geq 0$). This allows me to examine how the transmission of wealth changes only as a result of preferences over bequests, with reference. I still include loss-averse preferences over consumption as it allows the marginal rate of substitution between consumption and bequests to be the same as Cobb-Douglas preferences when agents do not feel a loss over bequests, which is a useful assumption.

In this paper I pin the relative loss-averse anchor as an agent's own initial wealth. There is an abundance of evidence to support this parameterization of ρ_b. For example, **?** find that parents anchor their expectations of their child's future education attainment compared to their own. And Barone et al. [2021] find that empirical evidence that social class and loss aversion heavily affects parent's investments in their children, including parents that are well-off and well-educated. Across the wealth and income spectrum, parents anchor their expectations for their child's future outcomes based off their own, and will invest more to make sure their children meet that expectation.

There is also supporting evidence that loss-averse anchors are distributionally-dependent for different behaviors and products. For example, Malloy [2015] finds that consumption anchors are dependent on the overall expectation and observation of other's consumption. So, as the average consumption increases, so too would the reference point. Another good alternative for a distribution-dependent anchor point is on average wealth, an anchoring that teases at the "American dream". In Ryan et al. [2024], there's evidence that parent's anchor their beliefs in what their children should received based on the average wealth of the economy. In times of growth and mobility, this expectation rises; in times of stagnation, the expectation lowers. This would create interesting dynamics that should be explored.

In my model, the assumption that loss averse reference points are only dependent on one's wealth levels which though it is non-ergodic, still allows for an explicit limiting distribution to be defined. If references are anchored in the distribution, then the analysis significantly complicates. This is an important and empirically backed alternative, and further research should explore the effects of non-ergodic processes due to distributional-dependent references in poverty models.

A.2 Toy Models

A.2.1 Linear Production Technology

Let's assume a linear production technology, $V(k^*) = k^*R$ as the first-best level of capital and $V(k) = \underline{k}R$ for $k < k^*$ as the second-best, and R is the return rate $r < R$. The monitoring function is piecewise: $\pi(k) = 0$ if $k < \bar{k}$ and $\pi(k) = \pi$ if $k \geq \bar{k}$. The incentive-compatibility condition is therefore: $k(R-r)+wr+T \geq kR - kR\pi$. Rearranging after plugging in k^*, the first-best cut-off is $w^* = k^* - \frac{k^* R\pi + T}{r}$. Income is therefore piecewise:

$$I(w) = \begin{cases} k^*(R-r)+wr+T & \text{if} w \geq w^* \\ \underline{k}(R-r)+wr+T & \text{if} w < w^* \end{cases}.$$

To calculate the bequest, a poor or rich agents that feels a gain will bequest $\gamma I(w)$. If they feel a loss, they increase their bequest to $\alpha I(w)$. In the either the poor or rich branches wealth, $w < w^*$ or $w \geq w^*$, we can calculate the sticky ranges of wealth explicitly by formulating the crossings of the 45-degree line at the different shares of income, $\gamma I(w)$ or $\alpha I(w)$. Piecing these together, we get the bequest function in Proposition (**??**)

Proposition 3 (Loss-Averse Bequests with Linear Production Function)**.** *Given the assumptions previously defined, we can explicitly define the kinks in the piecewise bequest function:* $w_p = \frac{\gamma}{1-\gamma r}\left(\underline{\underline{k}} \cdot (R-r)+T\right)$ *and* $w_1 = \frac{\alpha}{1-\alpha r}\left(\underline{\underline{k}} \cdot (R-r)+T\right)$*. Further,* $w_r = \frac{\gamma}{1-\gamma r}\left(k^* \cdot (R-r)+T\right)$ *and* $w_3 = \frac{\alpha}{1-\alpha r}\left(k^* \cdot (R-r)+T\right)$*. Then, by Definition 2.1,* $\mathcal{S}(w < w^*) = \{[w_p, w_2], [w_r, w_4]\}$*. Thus,*

$$b(w) = \begin{cases} \alpha\left(k^* \cdot (R-r)+wr+T\right), & w_3 < w \leq \bar{w} \\ w, & w_r \leq w \leq w_3 \\ \gamma\left(k^* \cdot (R-r)+wr+T\right), & w^* \leq w \leq w_r \\ \alpha\left(\underline{k} \cdot (R-r)+wr+T\right), & w_1 < w < w^* \\ w, & w_p \leq w_t \leq w_1 \\ \gamma\left(\underline{k} \cdot (R-r)+wr+T\right), & \underline{w} \leq w < w_p \end{cases} \tag{13}$$

and the $w^* = w_u$*.*

Proof. To calculate the poor steady state, we calculate when the gain bequest crosses the 45 degree line: $w = \gamma[\underline{k}(R-r)+wr+T]$. Rearranging, we obtain $w_p = \frac{\gamma}{1-\gamma r}\left(\underline{\underline{k}} \cdot (R-r)+T\right)$. The poor branch bequest at a loss crosses the 45-degree lilne with propensity α, $w_2 = \frac{\alpha}{1-\alpha r}\left(\underline{\underline{k}} \cdot (R-r)+T\right)$, and everything between w_p to w_1 is sticky, and everything above w_1 is a bequest with α propensity. The same exercise is done for the rich branch.

Since $0 \leq \underline{k} < k^*$, there is a discontinuous jump in the income function and therefore the bequest function. Because the change of the income function is constant, $I'(w) = r$, then $\alpha r < 1$ and the income function increases the distance from the 45-degree line indefinitely, so no upper poor sticky interval exists. The jump in income at w^* is also w_u if k^* sufficiently high, which I assume.

□

The function is numerically solved and visualized in Figure 12. The light-gray lines represent a bequest at propensity γ, and the dark-gray lines with propensity $> \gamma$ up to α. The loss averse bequest function follows the light-gray line until w_p transitioning to the loss averse bequest of the dark-gray line.

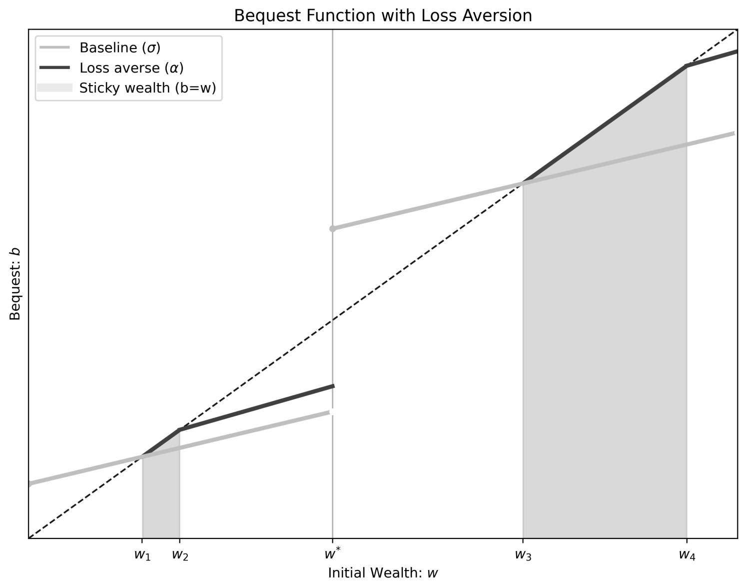


Figure 12: Bequest Function with Loss Aversion (Eq. 13) when $T > 0$.

A.2.2 Convex Production Technology

Given the set-up of Section (2.3.1), we define the incentive comparability condition as: $k(R - r) + wr + T \geq kR - k^2R\pi$. Plugging in k^*, we rearrange and solve for the first-best borrowers: $w^* = \underline{k} - \frac{\pi R\underline{k}^2}{r}$. For those with $w < w^*$, we rearrange the IC as solve for $k(w)$, which is

$$k(w) = \frac{r \pm \sqrt{r^2 - 4\pi R(wr + T)}}{2\pi R}.$$

To calculate the sticky regions and to fit the pieces of the bequest function together, we calculate the crossings on the 45-degree line. For gains for the poor, the smaller and larger root of $k(w)$ represent w_p and w_u. For losses for the poor, the smaller and larger root represent the range of the vicious cycle, and in-between the vicious cycle and the gain crossings are sticky intervals. Generally,w e calculate this as:

$$\textbf{Crossings on } [0, w^*): \quad a_\theta = 4\pi^2R^2(\theta r - 1)^2, \; b_\theta = 8\pi^2R^2\theta(\theta r - 1)T + 4\pi Rr\,\theta(R - r)(\theta R - 1),$$

$$c_\theta = 4\theta^2\pi RT\big(\pi RT + R(R - r)\big).$$

$$w_{\theta,\pm} = \frac{-b_\theta \pm \sqrt{b_\theta^2 - 4a_\theta c_\theta}}{2a_\theta}, w_p = \min\{w_{\gamma,-}, w_{\gamma,+}\},$$

$$w_u = \max\{w_{\gamma,-}, w_{\gamma,+}\}, \quad w_1 = \min\{w_{\alpha,-}, w_{\alpha,+}\}, \quad w_2 = \max\{w_{\alpha,-}, w_{\alpha,+}\}.$$

$$\textbf{Crossings on } [w^*, \infty): \quad w_r = \frac{\gamma\,[k^*(R-r)+T]}{1-\gamma r}, \qquad w_3 = \frac{\alpha\,[k^*(R-r)+T]}{1-\alpha r}.$$

We are careful because depending on the parameter, these sticky regions can collapse, expand, or entirely diminish. A sticky region can consume the entire range of w_p to w_u, and the general form written above attempts to accommodate the many situations. Assuming nice parameters, we get the result in Proposition (1) and can numerically solve and visualize in Figure (13)

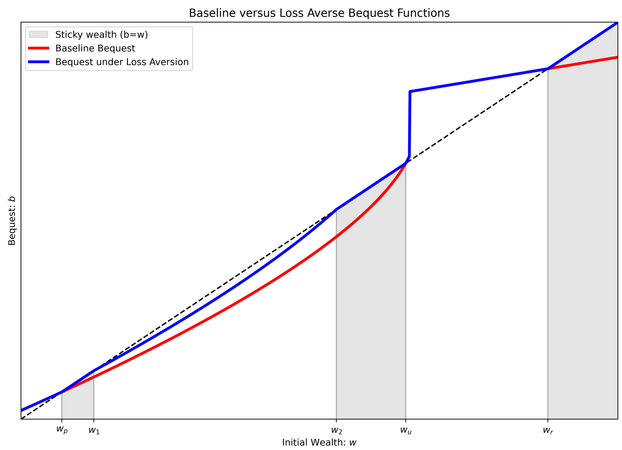


Figure 13: Numerically solved bequest function with convex bequests, $T > 0$.

A.2.3 Collapsing to Canonical Models

Proposition 4 (Bequest Function with a Linear Production Function). *Let $\pi(k) = 0$ if $k < \bar{k}$ and $\pi(k) = \pi$ for $k \geq \bar{k}$ where $0 < \pi \leq 1$, and $F > 0$. Furthermore, I let $V(k) = Rk$ where R is the return rate $r < R$ and $\bar{k} < k^*$. Then $V(k(w)) = 0$ and $V(k^*) = RL^*$. Then the baseline bequest function functions takes the form*

$$b_B(w) = \begin{cases} \gamma\big[k^* \cdot (R-r) + wr + T\big] & \text{if } w \geq w^* \\ \gamma\big[wr + T\big] & \text{if } w < w^* \end{cases} \tag{14}$$

where $w^* = w_u = k^* - \frac{\pi F + T}{r}$. The stable steady states are $w_p = \frac{\gamma T}{1-\gamma r}$ and $w_r = \frac{\gamma(k^*(R-r)+T)}{1-\gamma r}$.

Proof. See appendix. □

In Proposition 4, I assume the most basic structure this model can take and is a useful abstraction. Like in Banerjee and Newman [1993], when the model increases complexity in occupational choice – potentially in risky versus non-risky investment production technologies – the piecewise bequest function can quickly complicate, and a linear technology and monitoring function ensure a tractable model. Since this paper is concerned with the bequest dynamics, the majority of my analysis uses an example of a more sophisticated and realistic production technology and monitoring function outlined in Proposition (5).

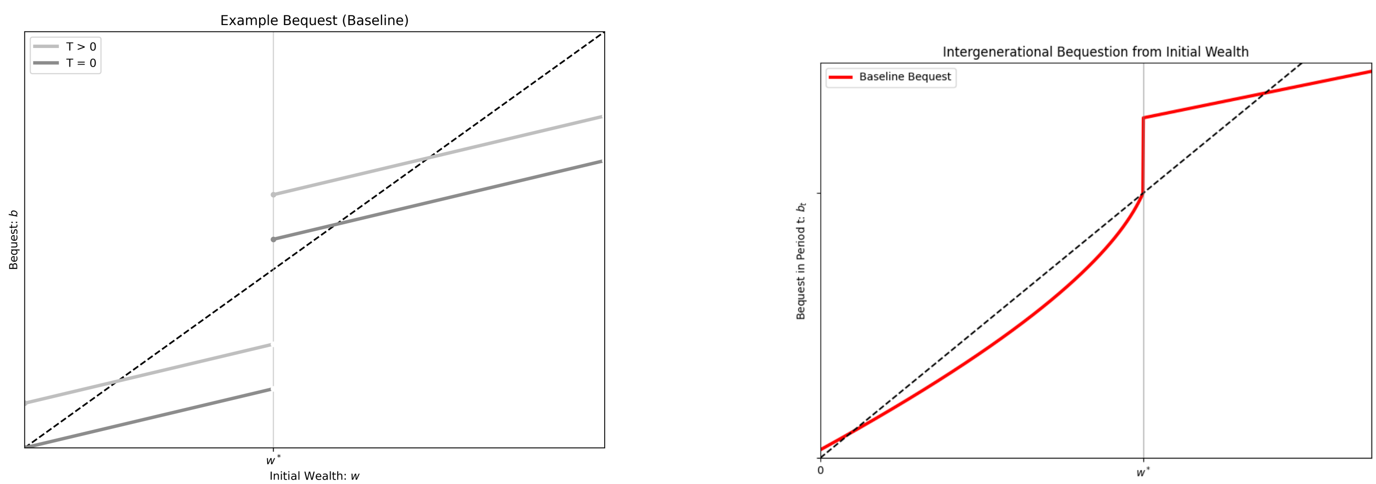
Proposition 5 (Bequest Function with a Convex Production Function)**.** *Let* $V(k) = R\min\{k, \bar{k}\}$ *with return rate* $r < R$*. Assume* $\pi(k) = \pi\min\{k, \bar{\bar{k}}\}$ *where* $\pi \in (0,1)$ *and* $\bar{k} < \bar{\bar{k}}$*, and a harsh punishment,* $F = V(k)$*. Then the baseline bequest function takes the form*

$$b_B(w) = \begin{cases} \gamma\big[k^* \cdot (R-r) + wr + T\big] & \text{if } w \geq w^* \\ \gamma\big[k(w) \cdot (R-r) + wr + T\big] & \text{if } w < w^* \end{cases} \tag{15}$$

where $k^* = \bar{k}$*,* $w^* = k^* - \frac{\pi R(k^*)^2 + T}{r}$ *and* $k(w) = \frac{r - \sqrt{r^2 - 4\pi R(wr+T)}}{2\pi R}$*, and* $w_u = \frac{-b+\sqrt{b^2-4ac}}{2a}$*. Further,* $w_p = \frac{-b-\sqrt{b^2-4ac}}{2a}$ *and* $w_r = \frac{\gamma\left[k^*(R-r)+T\right]}{1-\gamma r}$.
The coefficents are explicitly defined as $a = 4\pi^2 R^2 (\gamma r - 1)^2$*,* $b = 8\pi^2 R^2 \gamma(\gamma r - 1)\, T + 4\pi R\, r\, \gamma(R-r)\,(\gamma R - 1)$*, and* $c = 4\,\gamma^2\,\pi R T\Big(\pi R T + R(R-r)\Big)$

Proof. See appendix. ☐

The specifications in Proposition (4) and (5) allow insight into the range of bequest functions possible within the model. The visualizations for these propositions are below. The bequest function in Proposition 4 closely replicates the bequest function in Banerjee and Newman [1993] and the bequest function in Proposition 5 closely replicates the bequest function in Banerjee and Newman [1994]. We can visualize the mechanical effects of the market on an agent's bequest in Figure 14b and Figure 14a.



(a) Bequest Function from Propn (4)

(b) Bequest Function from Propn (5)

Figure 14: Example Baseline Bequest Figures

A.3 Stationary Distrbution

$$G_\infty = G_0(w_p)\delta(w_p) + [G(w_1) - G(w_p)] + [G(w_2) - G(w_1)]\delta(w_1) \\ + [G(w_u) - G(w_2)] + [G(w_r) - G(w_u)]\delta(w_r) + [G(w_3) - G(w_r)] + [1 - G(w_3)]\delta(w_3).$$

We can simply rearrange this equation to obtain the consolidated stationary distribution.

$$g_\infty(w) = g_0(w)\,\mathbf{1}\{w \in [w_p, w_1] \cup [w_2, w_u] \cup [w_r, w_3]\} + \pi_p\,\delta(w - w_p) + \pi_1\,\delta(w - w_1) + \pi_r\,\delta(w - w_r) + \pi_3\,\delta(w - w_3) \tag{16}$$

A.4 Empirical Insights, Half-lives

Half-lives. The intergenerational wealth transmission speed can be formalized for simulations and potential empirical analysis in Corollary (1). The half-life of a given w_0 is simply number of periods it takes to converge half-way to the corresponding stable steady state, as seen visually in Figure (6).

Corollary 1 (Bequest Half-Life)**.** *As defined in Proposition (3), for $w_t < w_2$, define the half-life to the poor steady state, w_p, as $H_i = \frac{\ln(1/2)}{\ln(b_i'(\underline{w})}$ where i refers to the baseline model B or the loss-averse model L. Then,*

$$H_L = \frac{-\ln(2)}{\ln(\frac{\alpha}{\gamma}) + \ln(b'_B(w_t))} > \frac{-\ln(2)}{\ln(b'_B(w_t))} = H_B.$$

$$H_B = H_L(\frac{\ln(2)}{\ln(2) + \ln(\frac{\alpha}{\gamma})H_L})$$

when $0 < \frac{\alpha}{\gamma}\frac{\ln(2)}{HL_B} < 1$, which is true for $w \in [w_1, w_2]$. Otherwise, the $H_L \to \infty$ without some idiosyncratic shock.

Using the Corollary (1), if a researcher can estimate the observed half-life in panel data, this model allows them to then calculate the underlying true half-life. This would require estimating two additional components: the propensity to bequest when not feeling a loss and the propensity at a loss. Since this is hard to capture in wealth data, consumption data would allow insight into those different bequest propensities, which is well-analyzed [Fisher et al., 2020]. It is beyond the scope of this paper to do this, but could be a fruitful avenue for research

A.5 T1 Robustness Check

I employ a robustness check of the identified thresholds in the loss-averse and baseline models using bootstrapping, which identifies the $\hat{c}$ that maximizes the F-stat for a given guess of c in Table 3. While this shows confidence in the 99.75 percentile, this is a somewhat trivial exercise in a Monte Carlo simulation since the data generating function is simple with clear i.i.d stochastic errors. Note that the sup-F for loss aversion, while still high, is about 1/4 the value of the model without loss aversion. As data becomes noisier and less rich, the F-stat for the best $\hat{c}$ will decrease, and the loss averse model might produce a low F-stat even if there is a poverty trap.

Table 3: T1 threshold existence (sup-F across c; bootstraping p)

Model	$\hat{c}$ (sup-F)	sup-F	p-value
Loss-averse	0.8418	107464.9	0.0025
Baseline	0.8477	408392.8	0.0025

Due to the sticky ranges of wealth and the slower convergence towards steady states, the loss averse model shows that even the reliable threshold test can fail to reject the null hypothesis when it otherwise should in the baseline.

A.6 Proofs for Optimal Tax in single-period flat tax rebate policy

Model Set-up

To begin, note that each individual will maximize utility to their updated budget constraint, where *net capital gains* is taxed at rate τ, and they receive some lump-sum rebate, T:

$$\max_{b^i} U(c^i, b^i) = (1-\gamma)\ln(c^i) + \gamma\ln(b^i) + \eta\cdot\gamma\nu(\ln(b)|\ln(w)) + \eta\cdot(1-\gamma)\nu(\ln(c)|\ln(0^+)) \tag{17}$$

$$\text{s.t.} \quad c^i + b^i = (1-\tau)k(w^i,\tau)(R-r) + w^i\cdot r + T \tag{18}$$

$$c^i, b^i \geq 0$$

This creates a trade-off for borrowers that affects the lenders' new incentive-compatibility constraint:

$$(1-\tau)k(R-r) + w\cdot r + T = [1-\pi k]\cdot Rk \tag{19}$$

Rearranging Equation (19), lenders will only administer a small enough such that agents are indifferent to repayment and reneging, where they choose repayment in the static equilibrium. To borrow the first-best level of capital k^*, an agent must have $w^* = \frac{k^*(\tau(R-r)+r)-(k^*)^2(\pi R)-T}{r}$. Intuitively, as the tax rate goes up, so does the required wealth for first-best borrowers $k(\tau(R-r)+r) > kr$ but goes down as the rebate goes up $T > 0$. Intuitively, as the tax goes up the rebate increases the lowest poverty stable steady state, but the rebate might be small compared to the loss in returns for higher wealth individuals $T - k^*R\tau < 0$. For non-first best borrowers, the lending function for those with $w < w^*$ is

$$k(w,\tau,T) = \frac{\tau(R-r)}{2\pi R} + \frac{r - \sqrt{(\tau(R-r)+r)^2 - 4\pi R(wr+T)}}{2\pi R} \tag{20}$$

Mathematically, while $T > 0$ is an upward shift in the return function, the tax τ causes a clockwise rotation around the y-intercept. Intuitively, while it increases the wealth of agents, the loss of returns also hurt agents borrowing the second-best level of capital. The new bequest function incorporates these trade offs. Solving the MRS given the Utility Function (17) and the income function in Equation (20), the bequest function when agents feel a gain is:

$$I(w,\tau) = \begin{cases} k^* \cdot (1-\tau)(R-r) + w \cdot r + T, & \text{if } w \geq w^* \\ k(w,\tau) \cdot (1-\tau)(R-r) + w \cdot r + T, & \text{if } w < w^* \end{cases} \tag{21}$$

By combining the above income in Equation (21) with the utility function in Equation (17), we can rederive the updated bequest function. Instead of rewriting the long bequest function here, note that the only change from Proposition (1) is that net-income is taxed, and the good-behavior T benefit is the demogrant.

To calculate the demogrant, I first describe the total capital in the economy as K, which is simply the mass of those borrowing first-best and then the expected value of those earning their individualized second-best level of capital, $K = \int_{w^*}^{\bar{w}} k^* g(w)\partial w + \int_{\underline{w}}^{w^*} k(w,\tau)g(w)\partial w$. This can be reduced and utilized in Equation (22):

$$K = k^* \cdot [1 - G(w^*)] \; + \; \mathbb{E}[k(w,\tau) \mid w \in [\underline{w}, w^*]] \tag{22}$$

Given aggregate capital in Equation (22), the rebate each agent receives in $T = \tau K(R-r)$, since $G(w)$ is normalized to one.

Solve the Model (without loss aversion).

The general social planner problem is below. However, I am first going to solve for the case when $\eta = 0$, individuals don't feel loss averse. This will allow me to then easily flesh out the SWF for loss averse preferences.

$$\mathbb{W} = \int_i g(w^i) \cdot U^i(w^i,\tau)\partial w^i = \sum_{(a,b)} \int_a^b g(w^i) \cdot U^i_j(w^i,\tau)\partial w^i \tag{23}$$
$$(a,b) = \{[\underline{w}, w_p], [w_p, w_1], [w_1, w_2], [w_2, w_u], ...\}$$
$$\text{s.t. } \tau K(R-r) \geq 0$$

A social planner only considering the cobb-douglas utility for non-loss averse agents will solve the following equation for τ^*:

$$\frac{\partial SWF}{\partial \tau} = \frac{\partial}{\partial \tau}\left(\int_{w^*}^{\bar{w}} U(w^i,\tau)g(w^i)\partial w^i + \int_{\underline{w}}^{w^*} U(w^i,\tau)g(w^i)\partial w^i\right) = 0.$$

To solve this, we begin by solving for the partial derivative of utility. Since agents optimize their utility over bequests, $\frac{\partial U(b^*)}{\partial b} = 0$, then using the envelope theorem as described, we see

$$\begin{aligned}
\frac{\partial U}{\partial \tau} &= \frac{\partial(\gamma \ln(b^*))}{\partial b} \cdot \frac{\partial b^*}{\partial \tau} + \frac{\partial((1-\gamma)\ln(I-b^*))}{\partial b} \cdot \left(\frac{\partial I}{\partial \tau} - \frac{\partial b^*}{\partial \tau}\right) \\
&= \left(\frac{\partial(\gamma \ln(b^*))}{\partial b} - \frac{\partial((1-\gamma)\ln(I-b^*))}{\partial b}\right) \cdot \frac{\partial b^*}{\partial \tau} + \frac{\partial((1-\gamma)\ln(I-b^*))}{\partial b} \cdot \frac{\partial I}{\partial \tau} \\
&= 0 \cdot \frac{\partial b^*}{\partial \tau} + \frac{1-\gamma}{I-b^*} \cdot \frac{\partial I}{\partial \tau} \\
&= \frac{1-\gamma}{(1-\gamma)I(w,\tau)} \cdot \frac{\partial I}{\partial \tau} \\
U_\tau &= \frac{1}{I(w,\tau)} \frac{\partial I}{\partial \tau}
\end{aligned}$$

Let $L_\tau(w,\tau)$ denote $\partial k(w,\tau,T)/\partial \tau$ as implied by (20), and $K_\tau = \int_{\underline{w}}^{w^*} L_\tau(w,\tau) g(w_i) \partial w_i$. This allows us to take the derivative of each part of the income function in equation (21):

$$\begin{aligned}
I_\tau(w < w^*) &= (R-r) \cdot ((1-\tau)L_\tau(w,\tau) - k(w,\tau)) + (R-r) \cdot (K + \tau K_\tau) \\
&= (R-r) \cdot ((1-\tau)L_\tau(w,\tau) - k(w,\tau)) + (R-r) \cdot (K - \tau \frac{\partial K}{\partial(1-\tau)})
\end{aligned}$$

$$\begin{aligned}
I_\tau(w \geq w^*) &= (R-r) \cdot (-k^*) + (R-r) \cdot (K + \tau K_\tau) \\
&= (R-r) \cdot (-k^*) + (R-r) \cdot (K - \tau \frac{\partial K}{\partial(1-\tau)})
\end{aligned}$$

Intuitively, the first part of the income derivative simply says a marginal increase in the tax will increase the total tax revenue which has a positive influence on income, but decrease the total capital gain. The second parts thus describe that as the marginal tax increases, so too do the intensive marginal effects of income. For small $k(w,\tau)$, an increase in the tax has a positive effect with diminishing marginal returns. Eventually this switches. Both describe the mechanical and reaction effects of the tax on income.

Now, we can take the derivative of the welfare function. Given the Leibniz integration rule, we can simply move the τ derivative into the integrals. We do not need to differentiate $\frac{\partial w^*}{\partial \tau}$ since the measure of agents at that wealth is zero. However, it can easily be derived for an exercise.

$$\frac{\partial W}{\partial \tau} = \int_{w^*}^{\bar{w}} U_\tau(w^i,\tau)\cdot g(w^i)\partial w^i + \int_{\underline{w}}^{w^*} U_\tau(w^i,\tau)\cdot g(w^i)\partial w^i = 0$$

$$\begin{aligned} 0 &= \int_{w^*}^{\bar{w}} (R-r)\frac{-k^* + [K - \tau\frac{\partial K}{\partial(1-\tau)}]}{I(w^i,\tau)}\cdot g(w^i)\partial w^i \\ &\quad + \int_{\underline{w}}^{w^*} (R-r)\frac{(1-\tau)L_\tau(w,\tau) - k(w,\tau) + [K - \tau\frac{\partial K}{\partial(1-\tau)}]}{I(w^i,\tau)}\cdot g(w^i)\partial w^i \end{aligned}$$

$$\begin{aligned} 0 &= \int_{w^*}^{\bar{w}} \frac{-k^*}{I(w^i,\tau)}\cdot g(w^i)\partial w^i + \int_{\underline{w}}^{w^*} \frac{(1-\tau)L_\tau(w,\tau) - k(w,\tau)}{I(w^i,\tau)}\cdot g(w^i)\partial w^i \\ &\quad + \int_{w^*}^{\bar{w}} \frac{K - \tau\frac{\partial K}{\partial(1-\tau)}}{I(w^i,\tau)}\cdot g(w^i)\partial w^i + \int_{\underline{w}}^{w^*} \frac{K - \tau\frac{\partial K}{\partial(1-\tau)}}{I(w^i,\tau)}\cdot g(w^i)\partial w^i \end{aligned}$$

$$\begin{aligned} 0 &= \int_{w^*}^{\bar{w}} \frac{-k^*}{I(w^i,\tau)}\cdot g(w^i)\partial w^i + \int_{\underline{w}}^{w^*} \frac{(1-\tau)L_\tau(w,\tau) - k(w,\tau)}{I(w^i,\tau)}\cdot g(w^i)\partial w^i \\ &\quad + \left[K - \tau\frac{\partial K}{\partial(1-\tau)}\right]\left(\int_{w^*}^{\bar{w}} \frac{1}{I(w^i,\tau)}\cdot g(w^i)\partial w^i + \int_{\underline{w}}^{w^*} \frac{1}{I(w^i,\tau)}\cdot g(w^i)\partial w^i\right) \end{aligned}$$

$$\begin{aligned} &\left[K - \tau\frac{\partial K}{\partial(1-\tau)}\right]\left(\int_{w^*}^{\bar{w}} \frac{1}{I(w^i,\tau)}\cdot g(w^i)\partial w^i + \int_{\underline{w}}^{w^*} \frac{1}{I(w^i,\tau)}\cdot g(w^i)\partial w^i\right) \\ &\qquad = \int_{w^*}^{\bar{w}} \frac{k^*}{I(w^i,\tau)}\cdot g(w^i)\partial w^i + \int_{\underline{w}}^{w^*} \frac{k(w,\tau) - (1-\tau)L_\tau(w,\tau)}{I(w^i,\tau)}\cdot g(w^i)\partial w^i. \end{aligned}$$

We can define $e = \frac{1-\tau}{K}\frac{\partial K}{\partial(1-\tau)}$. Then, we can simplify the integrals in expectation. With a specification of $G(W)$, we can explicitly solve for τ^*. But with this general form, I offer the implicit solution.

$$\begin{aligned} &K\left[1 - \frac{\tau}{1-\tau}e\right]\left(\mathbb{E}\left[\frac{1}{I(w^i,\tau)}\,\middle|\, w^i \in [w^*,\bar{w}]\right] + \mathbb{E}\left[\frac{1}{I(w^i,\tau)}\,\middle|\, w^i \in [\underline{w},w^*]\right]\right) \\ &\qquad = \int_{w^*}^{\bar{w}} \frac{k^*}{I(w^i,\tau)}\cdot g(w^i)\partial w^i + \int_{\underline{w}}^{w^*} \frac{k(w,\tau) - (1-\tau)L_\tau(w,\tau)}{I(w^i,\tau)}\cdot g(w^i)\partial w^i \end{aligned}$$

$$K\left[1 - \frac{\tau}{1-\tau}e\right] = \frac{\mathbb{E}\left[\frac{k^*}{I(w^i,\tau)}\,\middle|\, w^i \in [w^*,\bar{w}]\right] + \mathbb{E}\left[\frac{k(w,\tau)-(1-\tau)L_\tau(w,\tau)}{I(w^i,\tau)}\,\middle|\, w^i \in [\underline{w},w^*]\right]}{\mathbb{E}\left[\frac{1}{I(w^i,\tau)}\,\middle|\, w^i \in [w^*,\bar{w}]\right] + \mathbb{E}\left[\frac{1}{I(w^i,\tau)}\,\middle|\, w^i \in [\underline{w},w^*]\right]}.$$

Let G be defined as the right-side of the equation, and $\bar{G} = \frac{G}{K}$

$$K\left[1 - \frac{\tau}{1-\tau}e\right] = G$$

Then we can implicitly solve for the welfare-maximizing tax rate,

$$\tau^* = \frac{1-\bar{G}}{1-\bar{G}+e}.$$

This is the same implicit τ^*.

Solve the Model (with loss aversion).

Now I allow for loss avserion using the same set up as Equation (23). Much like the additional complications loss aversion creates to the bequestion function as seen in Proposition (1), the solution the social welfare problem will be similar the model without loss aversion, only this time there are several more integrals branches instead of just those above and below w^*. These integrals follow the different branches in the loss averse bequest function. As outlined in Equation (23), we just sum the integrals where those roots exist.

To begin, using the envelop theorem, we can determine that for agents feeling a *gain*, their utility derivative with respect to the tax rate is

$$\frac{\partial U(b^* \geq w)}{\partial \tau} = \frac{1-\gamma}{(1-\gamma)I(w,\tau)}\frac{\partial I(w,\tau)}{\partial \tau} + \frac{\eta(1-\gamma)}{(1-\gamma)I(w,\tau)}\frac{\partial I(w,\tau)}{\partial \tau}$$

$$= \frac{1+\eta}{I(w,\tau)}\frac{\partial I(w,\tau)}{\partial \tau}$$

The term $(1+\eta)$ increases the welfare, but if all agents were to feel a gain, this term would cancel out of the equation $\frac{\partial SWF}{\partial \tau} = 0$, and we return to the world without loss aversion at all. Additionally, if we didn't cancel the term, this would not change the maximizing tax rate, just the value of $SWF(\tau^*)$ though this is not in-and-of itself comparable to other welfare results.

For agents experiencing a loss, their utility derivative is

$$\frac{\partial U(b^* < w)}{\partial \tau} = \frac{1-\gamma}{(1-\alpha)I(w,\tau)}\frac{\partial I(w,\tau)}{\partial \tau} + \frac{\eta(1-\gamma)}{(1-\alpha)I(w,\tau)}\frac{\partial I(w,\tau)}{\partial \tau}$$

$$= \frac{(1+\eta)(1-\gamma)}{(1-\alpha)I(w,\tau)}\frac{\partial I(w,\tau)}{\partial \tau}$$

As λ or η increase, so does $U_\tau(b^*, \tau)$ since $\frac{1-\gamma}{1-\alpha} > 1$. For $\lambda > 1$ and $\eta > 0$, then marginal effect of a tax rate will be greater under a loss than without. For agents with $w \geq w^*$, this means a increase in a tax has a greater *negative* impact. For agents with $w < w^*$, this means an increase in a tax has a greater *positive* impact. Thus, the optimal tax becomes more sensitive to the distribution. If we add one more person to the distribution with initial wealth $w < w^*$ and who feels a loss, then the tax rate will increase compared to the world without loss aversion; conversely, if we add one more person with $w \geq w^*$ who feels a loss, the tax decreases

Since income is not dependent on loss averse preferences, the derivative doesn't change. However, since welfare function must account for the different piece of possible bequest branches, which are also utility branches. I will write out the general derivative below, then provide the final expected value since it follow a similar albeit more droning process as the model without loss aversion. There are two camps of individuals that *do not* feel a loss: those poorer than w_p converging up and those in the range of mobility, $[w_u, w_r]$. Those that do feel a loss are those in the trap (w_p, w_u) and those above w_r, from $(w_r, \bar{w}]$. In each camp, there can be first-best and second-best borrowers. Due to the envelope theorem, the bequest amount is not necessary to consider here when the planner maximizes, so we aren't concerned with the sticky regions.

$$\frac{\partial W}{\partial \tau} = \int_{w_r}^{\bar{w}} U_\tau^L(w^i,\tau) \cdot g(w^i) \partial w^i + \int_{w^*}^{w_r} U_\tau^G(w^i,\tau) \cdot g(w^i) \partial w^i$$
$$+ \int_{w_u}^{w^*} U_\tau^G(w^i,\tau) \cdot g(w^i) \partial w^i + \int_{\underline{w}}^{w_u} U_\tau^L(w^i,\tau) \cdot g(w^i) \partial w^i = 0$$

Following this, we can

$$\left[K - \tau \frac{\partial K}{\partial(1-\tau)}\right] \cdot [\frac{1-\gamma}{1-\alpha} \cdot \mathbb{E}[\frac{1}{I^*(w^i,\tau)} | w^i \in (w^*, \bar{w}] + \mathbb{E}[\frac{1}{I^*(w^i,\tau)} | w^i \in (w_r, \bar{w}]$$
$$\mathbb{E}[\frac{1}{I(w^i,\tau)} | w^i \in (w^*, \bar{w}] + \frac{1-\gamma}{1-\alpha} \cdot \mathbb{E}[\frac{1}{I(w^i,\tau)} | w^i \in (w_r, \bar{w}]]$$

$$= [\frac{1-\gamma}{1-\alpha} \cdot \mathbb{E}[\frac{k^*}{I^*(w^i,\tau)} | w^i \in (w^*, \bar{w}] + \mathbb{E}[\frac{k^*}{I^*(w^i,\tau)} | w^i \in (w_r, \bar{w}]$$
$$\mathbb{E}[\frac{k(w^i,\tau) - (1-\tau)L_\tau(w^i,\tau)}{I(w^i,\tau)} | w^i \in (w^*, \bar{w}] + \frac{1-\gamma}{1-\alpha} \cdot \mathbb{E}[\frac{k(w^i,\tau) - (1-\tau)L_\tau(w^i,\tau)}{I(w^i,\tau)} | w^i \in (w_r, \bar{w}]]$$

While the equation looks very messy and complicated, ultimately this is just a sum of expected values for across those making first-best and second-best, and then those that feel a gain and those feel a loss, so 4 integrals and expected values. For those feeling a gain, regardless of income level, they are waited as in the baseline. For those feeling a loss, they are upweighted.

$$K - \tau \frac{\partial K}{\partial(1-\tau)} = \frac{\sum_j \mathbb{E}_j[I_\tau(w^i,\tau) \mid b^* \geq w^i] + \frac{1-\gamma}{1-\alpha}\mathbb{E}_j[I_\tau(w^i,\tau) \mid b^* < w^i]}{\sum_j \mathbb{E}_j[\frac{1}{I(w^i,\tau)} \mid b^* \geq w^i] + \frac{1-\gamma}{1-\alpha}\mathbb{E}_j[\frac{1}{I(w^i,\tau)}) \mid b^* < w^i]}$$

where j indicates a sum over the expected values when $w < w^*$ and $w \geq w^*$. In the LHS, the denominator sums the expected values by their loss/gain weight of the recipricol of their branch's income. The numerator sums the expected values of the change of income given a marginal increase in the tax. Whether the numerator or denominator is larger or smaller is entirely dependent on the distribution and the rebate/tax. For wealth individuals, a marginal increase has a negative effect on the tax, and quickly shrinks the numerator faster than the denominator. For poorer individuals, an increase in tax reduces their income, but the rebate *might* be a stronger benefit. Let G equal the RHS.

$$\tau^* = \frac{1 - G_L}{1 - G_L + e}. \tag{24}$$