Loss Aversion and Poverty Traps

Miles T. Smith

November 29, 2025

Abstract

Decades of empirical research has found mixed results on whether poverty traps do and do not exist in many contexts. 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 a credit market. When agents have a reference-dependent bequest motive, they change their preferences in response to vicious cycles, can slow and even stop downward transitions which makes the model sensitive to income shocks. These dynamics are not accounted by previous poverty trap models and can make the trap invisible to empirical identification. I confirm that that if reference-dependence exists but is unaccounted for in the empirical specification then the poverty trap will not be identified by simulating the model and testing two common empirical methods. I analyze a tax-redistribution and "big push" policy, the different policies employed by different empirical results, formulate an implicit tax rate, and I find interventions can have welfare-reducing consequences if the poverty trap is mis-identified.

[CURRENT DRAFT]

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 is poverty traps: a 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] are unable to find conclusive evidence for poverty traps in many contexts explored in the theory literature, research by Carter and Barrett [2013] find data often insufficiently rich to identify poverty traps. 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 allow for reference-dependent bequest motives in order 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 bequest 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. The intuition is that 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 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, generating a new unstable steady state instead of converging towards a poorer stable steady state. Intervals of sticky wealth levels can emerge. This is a significantly different dynamic than other poverty trap models predict, 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. Sticky intervals also emerge near the stable steady states, rich and poor, such that dynasties above no longer converge to the stable states.

I find that reference-dependence can make the poverty trap invisible to typical econometric empirical identification methods. I use Monte Carlo simulations on my model and, as a benchmark, my model without loss aversion to create wealth panel data. I run two 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 dynamics and steady states. Without loss aversion, the empirical tests would otherwise accurately identify the trap.

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 Bluffstone, 2009]. This naturally begs two questions: if wealth (or income) panel data alone is unreliable how can researchers overcome those limitations and are there policy implications of reference-dependence? The model presented in this paper is constrained in answering these questions though I attempt a base step. I plan to explore these questions using a more sophisticated models and in the future to further 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

[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.

or income might be steady, agents will sacrifice consumption to maintain bequests in response to increased headwind (shocks) or vicious cycles. Shocks to individual income would reveal differences in propensity to consume to maintain a bequest amount since expectations are anchored based on an agents initial conditions. An agent whose income was shocked down will have a higher propensity to bequest than the agent shocked up to the same income level. 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 can control for heterogeneity in consumption preference. A potential dataset for this analysis is PSID, however there are other econometric issues that need to be addressed in order to isolate loss aversion and vicious cycles.

I offer a preliminary policy analysis by formulating an implicit function for the optimal tax rate to fund different interventions and show that empirical limitations can lead to sub-optimal policy with potentially welfare reducing effects. First, I assume a social planner is not able to identify a poverty trap and employ a flat tax on net income and equal rebate suggested by Mirrlees [1971] and use the framework from Piketty and Saez [2013] to formulate an implicit tax rate. There are two important forces that impact the equity-efficiency tradoff. (Equity) The optimal tax rate is sensitive to the distribution *of those feeling a loss*, rich and poor, as they will have outsized welfare losses (rich) or gains (poor) which can lower or raise the optimal tax and therefore equity. (Efficiency) While the optimal tax is often similar with and without loss averse preference, there is an efficiency gain since loss averse agents preserve the distribution in response to the tax, which allows for a higher tax rate. The second analysis explores if a social planner does identify a trap, the implications of the empirical tests suggests they might misidentify it. The optimal policy response to a poverty trap is for a planner to take on debt to fund a rebate sizable to lift everyone above the unstable steady state into mobility, and later tax individuals at these higher incomes to repay the debt. Since empirical tests grossly underestimate the unstable state, the lift is too small to get the poor population to enter mobility. This causes a vicious cycle for the planner, continuously applying a small push to keep raising poor families out of the poverty trap cycle when they re-enter due to small shocks and market distortions from a tax. Underestimating the unstable steady state there produces ineffient, ineffective and high deadweight loss policy than if a policy planner correctly identified the unstable steady state.

While this model gives useful insight into policy interventions, it lacks key features that would enrich the policy analysis, like richer life-cycle dynamics, heterogeniety in preferences, risk and insurance for income, and multi-period tax schedules.

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, 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 within a single 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):

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

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 ρ_b and ρ_c[2]. The 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.

In this paper, I 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

[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.

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 probability 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 global 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 these agents 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 to avoid repayment and might successfully run away with their investment gains. If they renege there is some probability $\pi(k)$ they are caught and receive some punishment F. I assume $\pi(k)$ is increasing and concave with k: the bigger the loan, the easier it is to monitor. 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 $k > w + \frac{\pi(k)F+T}{r}$. 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, and agents always want to borrow the maximum k they can. The implicit loan function for $w < w^*$ is $k(w) = w + \frac{\pi(k(w))F+T}{r}$. As the probability of catching renegers 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

[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 $V'(k^*) = r$.

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 admits 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 dependent. The amount of an agent's income spent on consumption or bequests is dependent on if they perceive their bequest as a gain or a loss.

2.3 Bequest Function

To understand the transmission of wealth from one generation to the next, we calculate the optimal bequest an agent will leave to their progeny by the end of the period. Agents are already differentiated by those who can borrow the first-best capital $w \geq w^*$ and those who borrow their second-best $w < w^*$. Now, agents in *both* wealth branches face another bifurcation: (1) agents who feel a gain with their bequest $b \geq w$; and (2) agents who feel a loss with their bequest $b < w$. While there is a natural intuition that second-best borrowers often feel a bequest loss, second best borrowers are permitted to feel a gain as equally as first-best borrowers can also feel a loss. So, an agent's bequest is determined both by their initial wealth, and therefore income, and their perception of their bequest. Since we already determined the different incomes for different initial wealth levels, we now turn to the perception of bequests.

I begin by assuming an agent codes their bequest as a gain, that is $b > w$. Then $\lambda = 1$, and we take the MRS of Utility Equation (1). While the η term is still in the MRS equation, it cancels out in the numerator and denominator: $c^* = \frac{(1-\gamma)(1+\eta)}{\gamma(1+\eta \cdot 1)} b^* = \frac{1-\gamma}{\gamma} b^*$. Plugging c^* into the Income Equation (2), we get our optimal bequest when the bequest is coded as a gain. I define this as the "baseline" bequest, $b_B(w)$ it is simply the standard Cobb-Douglas propensity to bequest with weight 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, an agent 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) and plugging it into Income Equation (2), we obtain a new propensity to bequeath: $\alpha = \frac{\gamma(1+\eta\lambda)}{1+\eta-\gamma \cdot \eta(1-\lambda)} > \gamma$. Thus, when an agent feels a loss, their increase their bequest up to (w).

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 $\alpha I(w) > \gamma I(w)$. This creates an open interval between the two bequest amounts. It is therefore possible that $\alpha I(w) > w > \gamma I(w)$. For ranges of wealth where this inequality is true, agents who bequest with propensity α will bequest more than their initial wealth, but bequeathing with propensity γ makes them feel a loss. In this case, a corner solution arises where the optimal bequest is a corner solution, exactly equal to the agents own initial wealth $b(w) = w$. 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). An interval of sticky wealth, S_j, are the ranges for which an agent with an initial wealth within this range will bequest exactly that initial wealth. It is important to note that *multiple* intervals of sticky wealth can emit across the wealth distribution depending on the function forms assigned to the economy. Specifically, sticky ranges can occur in regions of poverty *and wealthier regions.* Both the poor and the non-poor can bequest at a loss.

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 regions will appear near to crossings of the 45-degree line, if they exist. Since the baseline bequest function is already close to the 45-degree like, an α increase in the bequest propensity is likely to be higher than the 45-degree line. We would expect sticky intervals of wealth to emit near steady states, both stable and unstable. This will be discussed at length in the next section.

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.

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

And finally, we can write the total bequest function in Equation (5), which again is often split further by the income function which requires fitting functional forms to the model.

$$b(w) = \begin{cases} \gamma I(w) & \text{if } w \geq b_B(w) \\ b_L(w) & \text{if } w < b_B(w) \end{cases}. \tag{5}$$

Thus, the complete bequest mapping, b, for a given distribution and parameterization that defines the piecewise income function consolidates $b_L(w)$ and $b_B(w)$ into a single function. When agents bequest at a gain, they bequest $b_B(w) = \gamma I(w)$. When they bequest at a loss, $b_L(w) = \alpha I(w)$

or $b_L(w) = w$ when they can match their initial wealth. Intervals of sticky wealth where $b_L(w) = w$ emit near crossings of the 45-degree line. The explicit functional form for the bequest mapping for all initial wealths is very specific and sensitives to the functional forms and parameterizations of the model. For specific examples of the bequest function, I solve the policy function using different functional forms for the market economy in Appendix (A.2) including when loss aversion is not present, which returns the model to emulate the framework in Banerjee and Newman [1993]. 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.

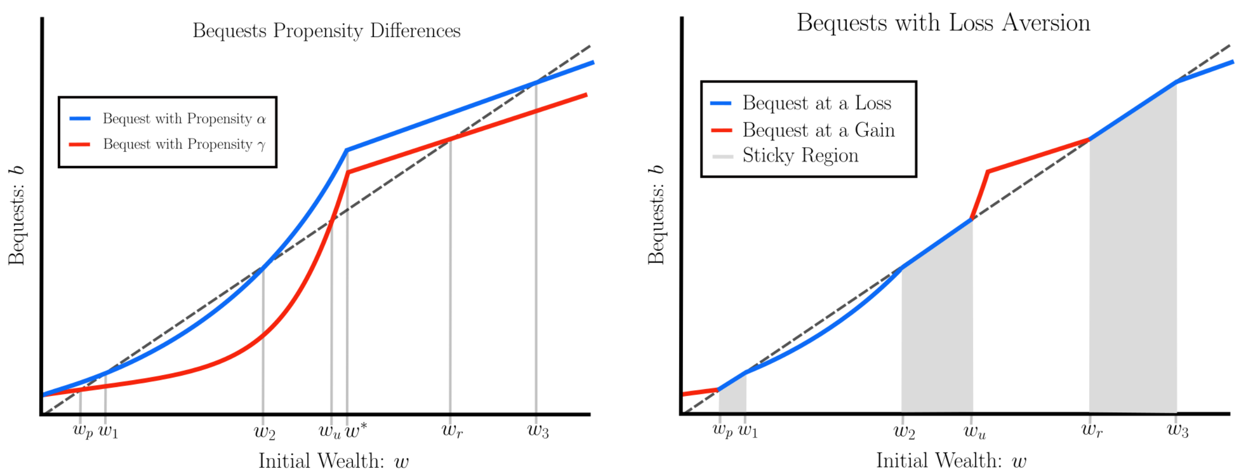
Then, by solving for the optimal bequest while at a loss and at a gain, I can explicitly define the intervals for which sticky ranges emerge using the crossings of the gain and loss bequest functions. The total piecewise bequest function is there fore define in Proposition (1).

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:*

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

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

While the above piecewise function looks cimplicated, ultiamtely the pieces simply describe when individuals are above or below w^* and when they do or do not feel a loss. An illustration of this toy model can be seen in the Figure (1). In Figure (1a), plotted in the red line is the bequest function *only* with propensity γ: this is the “baseline” bequest function that is similar to the model by Banerjee and Newman [1994], which is the common Cobb-Douglas bequest weight. When individuals feel a loss they give with propensity α. The bequest curve if all individuals gave α rather than γ is seen in the blue line, which graphically is a rotation counterclockwise around the y-intercept and a flattening of the curve. In Figure 1a, 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. When these two pieces are put together — the red line when it is above the 45-degree line, the blue line when it is below the 45-degree line, and the sticky ranges in-between – we arrive at the bequest curve in Figure (1b) that represents the bequest function in Proposition (1).



(a) Bequest Functions with Propensities α and γ.

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

Figure 1: Bequests with Loss Aversion

The important comparisons are the baseline bequest, the red line in Figure (1a), and the loss-averse bequest curve in Figure (1b). The baseline curve represents the bequest function when loss aversion is turned off, $\lambda = 1$ or $\eta = 0$, which returns the model to the framework similar to Banerjee and Newman [1993] and Banerjee and Newman [1994]. This is an important observation in order to understand the dramatic effect when loss aversion is present, which flattens the curve into the 45-degree line for all parts of the baseline function below the 45-degree line. The intergenerational consequences will be explored in-depth in the following section. For now, I focus on how loss aversion parameters affect the shape of the loss-averse bequest function compared to the baseline.

The sticky regions of wealth are determined by the crossings of the different bequest lines. In the baseline function, there are only three crossings: a poorer wealth stable steady state, w_p, an unstable steady state, w_u, and a richer stable steady state w_r. Around those points are the crossings from the α-loss bequest: $w_1 > w_p$, $w_2 < w_u$, and $w_3 > w_r$. Those inequalities creates the three sticky regions within this specific functional form of the model.

As parents feel the loss more intensely – $\lambda \uparrow$ – the scope of sticky intervals also increases. Graphically, the growing interval is because loss aversion rotates the baseline bequest counter-clockwise centered around the y-intercept. This is stronger than just a shift up for bequest, because when the bequest function is non-linear, neither is a change in the rate of the bequest when under a loss – here it is growing. Intuitively, the stronger the behavioral change, the more individuals will match their exact initial wealth. The 45-degree lines becomes an absorbing line the function falls into. We can this this rotation in Figure 2, where $\lambda = 1$ is the baseline bequest curve when a parent codes all bequests as a gain, and as λ increases to λ_1 then λ_2, so does the sticky regions and the slope of the bequest function all together.

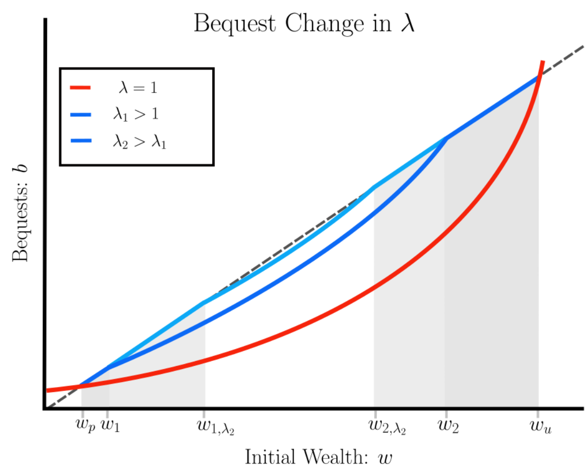


Figure 2: Loss-averse Bequest Function Change with Increase in λ.

3 Poverty Traps and Loss-Averse Dynamics

There is a naturally recursive nature to this model, 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}. The bequest function works as a mapping from wealth in one period to the next: $b(w_t) = w_{t+1}$. This has interesting properties. First, this implies the distribution of initial wealth is a Markov process. Second, the recursive nature allows us to investigate the novelties of the loss-averse bequest transition paths and steady state distributions.

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. For the context of my model, this means that agents with an initial wealth below the Micawber threshold will see their lineage converge towards a “poor” stable steady state every generation, trapped within a vicious cycle. It is typically the case that agents above the Micawber threshold see their family line converge to a higher “rich” stable steady state every generation in a virtuous cycle: poverty traps with multiple equilibria as opposed to just a poor single state. Graphically, these typically looks like an ”S-shaped” curve as seen in Figure (3). 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.

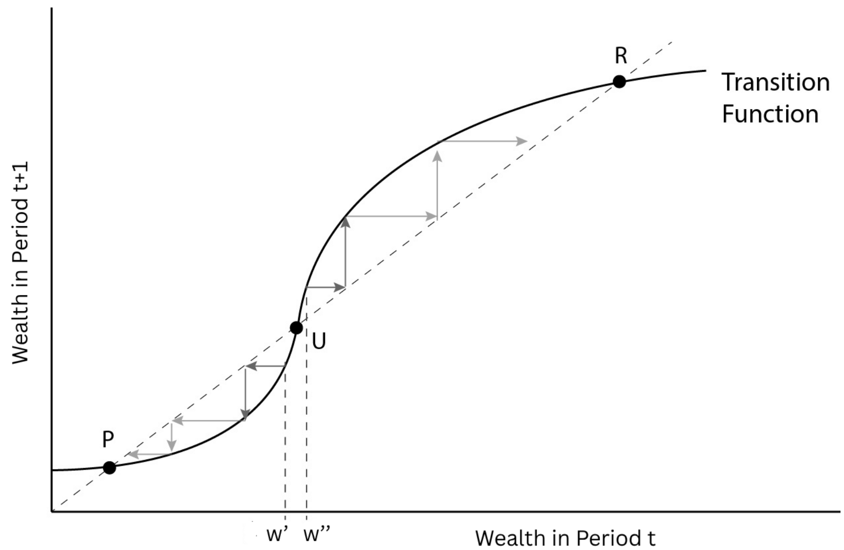


Figure 3: A General Diagram of a Poverty Trap and Wealth Transitions

As seen in Figure 3, the Micawber threshold is important to define and identify to understand who is trapped in poverty and the extent of the trap. We can define this threshold as a specification of the unstable steady state:

Definition 3.1 (Micawber threshold)**.** An unstable steady state w_u is known as a *Micawber threshold* if

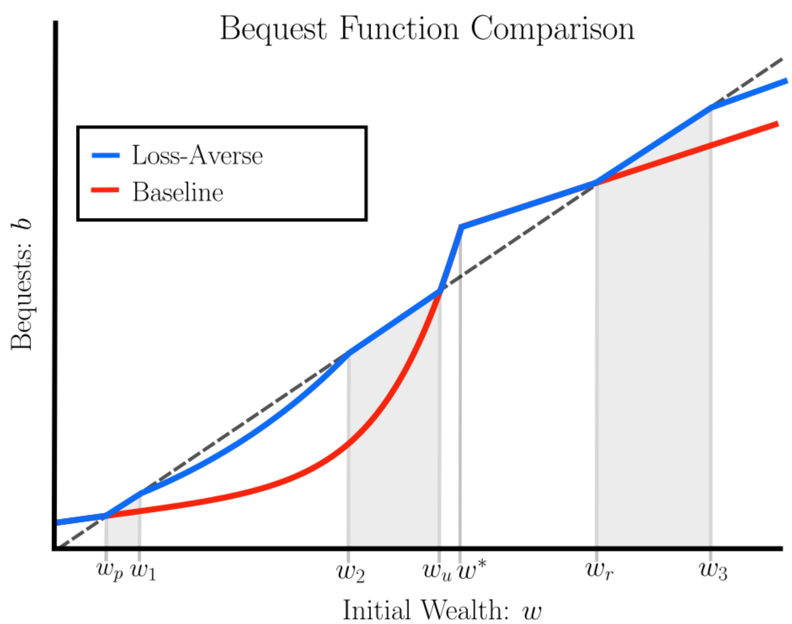
- $\forall w_t < w_u$, $\lim_{t\to\infty} w_t = w_p$, where w_p is locally stable
- $\forall w_t > w_u$, $\lim_{t\to\infty} w_t = w_r$, where w_r is locally stable

where $w_p, w_r \in G_t(w) \sim [\underline{w}, \bar{w}]$ are "poor" and "rich" stable steady states where all three steady states are time-invariant.

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 model without loss aversion ("baseline") can be seen in Figure (**??**) 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 making an S-shape difficult to identify.



First, loss aversion can change the observable Micawber threshold from the true Micawber threshold. Recalling the definition of the Micawber threshold (3.1), those below w_2 converge down instead of those below w_u which is the wealth level where families must combat the downward forces. As seen in Figure (**??**), the *true* Micawber threshold is much higher than would be the *observed* threshold.

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, both the sticky range of wealth increase and also the derivative of the bequest at a loss, meaning convergence slows. While this is true for those converging down to the rich and poor states, it is most obvious for those converging down into the poverty trap. In Figure (4), 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.

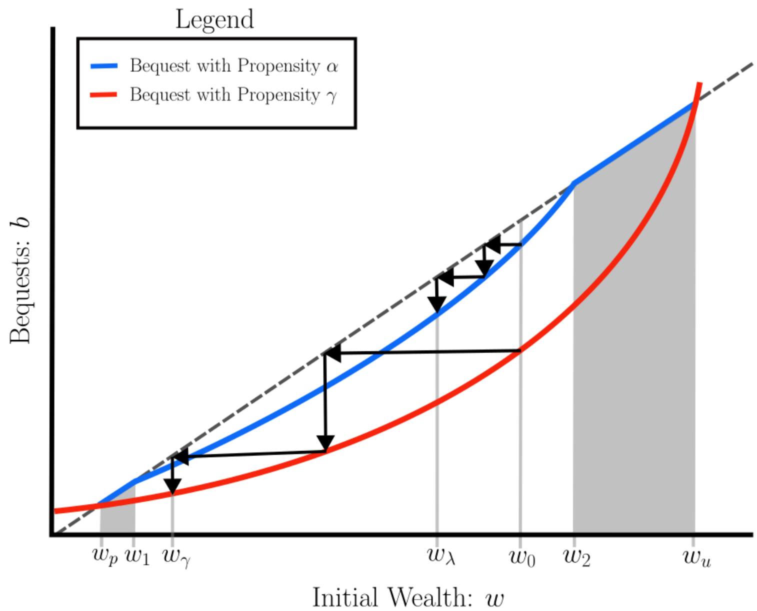


Figure 4: Differences of 2 transitions in initial wealth between baseline and loss-averse bequests.

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: w_p without loss aversion, w_1 with loss aversions. Visually, in Figure (4), we can tell convergence is fairly quick (in theoretical terms) for the baseline model, with a half-life of potentially just one period. The estimated half-life under loss aversion is much higher. There is a basic formula for calculating the half-life, and I adapt it to this model.

Corollary 1 (Bequest Half-Life)**.** *As defined in Proposition (4), 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.

Third, the short- and long-run distributions will be much richer with loss averse bequets. In the short-run, after even one period, the distribution of intial wealth under loss averse bequests will only be slightly different than the firs period; but without loss aversion a researcher would expect a dramatically different distribution after one period, specifically a large divergence from the unstable steady state. But in the long-run, the sticky ranges of wealth permanently preserve the distributions that started within them and maintain higher dispersion of the wealth distribution.

We can explicitly compare the long-run distributions since b is time-invariant and continuous on a finite state space, and therefore the model follows a stationary Markov process[4]. Since the model is ergodic, knowing the distribution in time $t = 0$ means we will know the distribution as $t \to \infty$. Hence, in Proposition (2), we can explicitly define the long-run distribution given the current distribution $G_t(w)$ and the parameterization of the model.

Proposition 2 (Stationary Wealth Distribution for Toy Model (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 sticky regions, the transition is already stationary.*

$$\begin{aligned} G_\infty = {} & G_0(w_p)\,[\delta_{w_p} - 1] \; + \; [G_0(w_2) - G_0(w_1)][1 - \delta(w_1)] \\ & + [G_0(w_r) - G_0(w_u)][\delta(w_r) - 1] \; + \; [1 - G_0(w_3)][1 - \delta(w_3)] \end{aligned} \tag{6}$$

Where $\delta(\cdot)$ is the dirac-delta function for a given steady state. 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).

Proof. See Appendix A.3 □

My model predicts a much richer steady state distribution than normally predicts by baseline, two-equilibria poverty trap models. In Equation (6), they key transitions are only the distributions between sticky regions: $[w_1, w_2]$ and $[w_u, w_r]$. The rest of the distribution is preserved. When these ranges are small, most of the distributional changes would only be from stochastic shocks. Without loss aversion, a researcher would predict many individuals clustered around the two stable steady states since researchers often predict data is in the long run [Barrett and Carter, 2013]. In my model, even if data observed the stable steady state, we predict there would still be a lack of clustering in the stable steady states and a strong preservation of the initial wealth distribution. Since

Fourth and finally, welfare remains ambiguous. Agents sacrifice their consumption to give a larger bequest. In the long-run, agents in a poverty trap will enjoy greater consumption than if they converge to the poor steady state. However, it is a philosophical debate if we prefer a world in which poorer parents must consume much less than they otherwise should in order 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]There are formulations of this model with non-ergodic Markov process. One example from Banerjee and Newman [1993] where occupational choice depends on the wealth distribution. The introduction of loss aversion in their model would likely corroborate the findings in this paper though perhaps with interesting long-run subtitles. Another formulation, as discussed in Appendix (**??**), is to make the bequest anchor distributionally dependent: $\rho_b(w, G_t(w))$. As the distribution of wealth changes, so does the population's expectations for their child's outcomes. It would be worthwhile for future research to study these avenues.

4 Econometric Implications

In this section, I show that a poverty trap with loss averse bequest preference 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. If a person is poor now, but they have high-return assets, they are unlikely to be poor forever. Conversely, those with high income now might not sustain their non-poverty status without any wealth. These authors also explore how the panel data itself is often limited, not only due to attrition and ability to capture income and wealth, but also that short panel data often do not contain enough periods that can allow the vicious cycle play out, potentially confounded by preference heterogeneity and averaging of meso-level dynamics. 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.

A potential explanation could be that empirical models fail to incorporate reference-dependent preferences. When a poverty trap is present, but empirical models do not account for reference-dependent bequests (which can be applied to a broader context), I show that identification methods can fail to reject the null hypothesis that a trap does not exist.

Threshold Test. The threshold test estimates a point (threshold) where the average change in wealth below is substantively different than the average change above. Or, intuitively, that two lines of best fit better match the data than one line, separating discontinuously at a threshold. This is the most common empirical identification for poverty traps — Jalan and Ravallion [2004], Carter and Barrett [2013], Antman and McKenzie [2007] — estimating the Micawber threshold. As Barrett and Carter [2013], describe, this is the most robust way to identify the potential of a poverty trap given limited panel data, as it does not assume or rely on the transition paths taking on a specific functional form. The tradeoff is that it does not estimate the richer dynamics of a transition function. This is still the most powerful tool to researchers as it indicates a bifurcation in paths: if the average change below the threshold is negative and the average change above is positive, this is the crucial prediction of the poverty trap.

Using a Monte Carlo simulation, I compare the efficacy of the threshold test to identify a poverty trap in the loss averse model and as a benchmark test estimate the existence of a poverty trap in the model without loss aversion. The true threshold is c, and the estimated threshold is $\hat{c}$. We represent the line of best fit for the average wealth change as $g_h(w) = w_{t+1} - w_t = \Delta w_{t+1}$ for wealth levels higher and as $g_l(w) = w_{t+1} - w_t = \Delta w_{t+1}$ for wealth levels lower than $\hat{c}$.

I design the empirical estimation model in Equation (7) based on the framework developed by Hansen [2000]. Here, β_0 and $\beta_0 + \beta_3$ represent the intercepts of the the higher and lower average lines, respectively. The coefficient β_1 represents the rate of the average change for the distribution above c; similarly, $\beta_1 + \beta_3$ are the rate of change below. We can define the discontinuous average wealth change lines as $g_h(w) = \beta_0 + \beta_1 w$ and $g_l(w) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)w$. The null hypothesis is that a poverty trap does not exist, which is true if *one line of best fit* estimates the model better than *two* around a threshold and that *those just below the threshold experience a negative change and those just above a positive change*, implying that there is no Micawber threshold where those above get richer and those below get poorer. In the model, we represent the null as H_0 : (no unstable threshold at $\hat{c}$) $\beta_1 + \beta_3 = 0$ and (bifurcating paths) $\hat{g}_l(\hat{c}) \geq 0,\ \hat{g}(\hat{c}) > 0$.

$$\text{(T1)} \quad \Delta w_{i,t+1} = \beta_0 + \beta_1 \, w_{i,t} + \beta_2 \, \mathbf{1}\{w_{i,t} < c\} + \beta_3 \big(w_{i,t} \cdot \mathbf{1}\{w_{i,t} < c\}\big) + \varepsilon_{i,t}. \tag{7}$$

We 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 $\min_\beta ||Y - X(c)\beta||^2$. I play out the simulation for 5 periods, and 500,000 agents, which is both a longer panel with more agents than most panel data used for this test Carter and Barrett [2013].

The simulation results from Table (1) fail to reject the null for the model with loss-averse bequest motive thought importantly we can reject the null without loss-averse (baseline) preferences. This means without loss aversion, the data generating function should allow for identification, but with the empirical specification is insufficient. In Panel A, with the baseline, we can first reject the null because the change in wealth just below the estimated threshold negative, $\beta_1 + \beta_3 = -0.1397$: the change in wealth for those below the threshold is expected to be negative. With loss aversion, the estimated slope is weakly positive though close to zero, 0.0048: the expected change in wealth is positive and small. The slope being close to zero is due to the flattening of the transition function. While in this model the results are all statistically significant due to the simple data generating process, which I show later, it is important to note that any additional noise in the data might easily flip the sign of the estimated slope when agents are loss averse and the empirical model does not control for that. In addition, the estimated threshold with loss averse preferences ($\hat{c} = 0.8418$) is marginally lower than without ($\hat{c} = 0.8477$), which is expected given the sticky range lowers the observed Micawber threshold.

Table 1: T1: Conventional threshold regression for downward convergence

Panel A: Threshold regression coefficients

Model	$\hat{c}$	β_0	β_1	β_2	β_3	Slope below (β_1+β_3)
Loss-averse	0.8418	+0.2381	-0.1734	-0.2399	+0.1782	+ 0.0048
Baseline	0.8477	+0.7928	-0.6850	-0.7952	+0.5453	-0.1397

Panel B: Drift equations and values at $\hat{c}$

Model	Drift equations	At $\hat{c}$
Loss-averse	$g_l(w) = -0.0016 + 0.0041\,w$ $g_h(w) = 0.2398 - 0.1746\,w$	$g_l(\hat{c}) = 0.0019$ $g_h(\hat{c}) = 0.0929$
Baseline	$g_l(w) = -0.0025 - 0.1392\,w$ $g_h(w) = 0.8477 - 0.6832\,w$	$g_l(\hat{c}) = -0.12$ $g_h(\hat{c}) = 0.212$

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).

In Panel B, we find no reliable bifurcation paths below and above the estimated threshold with the loss averse model. Since $\hat{g}_l(\hat{c}) > 0$ and $\hat{g}_h(\hat{c}) > 0$, this implies a different in slopes, where above an individual increases their rate of wealth accumulation, but only that individuals below the threshold get wealthier more slowly. In the baseline model, we would estimate a clear bifurcation in immediate changes in wealth below and above the threshold, which would finally allow us to reject the null completely. Notice again the coefficients for the loss-averse model are closer to zero, implying that noise make flip the signs of the coefficient and the estimations might be insignificant. I conduct a robustness check in Appendix (A.4).

S-shaped Curve. Another popular estimation uses a cubic polynomial function parametric estimation model 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 inter-period change in wealth crosses the 45-degree line, as seen in Equation (8). 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 at root 2 is non-positive and the other roots non-negative, $H_0 : 0 \not< \beta_1 \not< \beta_3$ and $\beta_2 \geq 0$ and $H_1 : \beta_2 < 0 < \beta_1 < \beta_3$.

$$\text{(S2)} \quad \Delta w_{i,t+1} = \beta_0 + \beta_1\, w_{i,t} + \beta_2\, w_{i,t}^2 + \beta_3\, w_{i,t}^3 + \varepsilon_{i,t}. \tag{8}$$

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 (**??**). 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 and 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 discuss three potential strategies to identify the poverty trap: supplementing wealth and income data with consumption data, using the half-life formula to reverse-engineer the true half-life using the observed half-life, and by identifying a behavior rational only within a trap.

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. Together, a researcher might view a stochastic income shock and unchanged interperiod wealth levels, and miss the consumption sacrifice after the shock it took to maintain that wealth level. I model this behavior below in Figure (5).

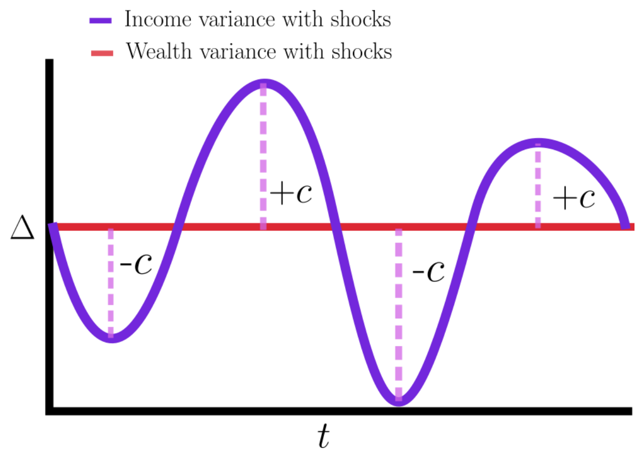


Figure 5: Consumption change in response to income shocks with anchored bequest expectations.

In Figure (5), this describe a *permanent* shock to income. While consumption drops, this is to maintain a bequest that does not noticeably change. In essence, consumption will rise and fall much more sensitively to permanent income shocks than the bequest. Outside a poverty trap, permanent income shocks should effect both bequests and income fairly equally, but if it doe not, this could be due to anchored bequest expectations and without accounting for that, a researcher might fail to identify the poverty trap. 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 permanent income shocks and not temporary income shocks; furthermore, 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 identify the distance between a parents realized outcome and their expected outcome, and measure if the investment noticeably changed from their parent's own received investment. The data required is different than if bequests were instead defined more as a present expenditure (i.e. tutoring lessons, books). In this case, temporary income shocks might negatively change consumption (i.e. consuming less, or consuming less quality goods) far more than impact those human capital expenditures. Outside a poverty trap, permanent There is more discussion to be had over datasets and definitions, but a good future research avenue would be to use PSID and OLG life-cycle models to tease out potential results in the data.

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 (assuming shocks are i.i.d. centered around 0) 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 under a poverty trap, where agents caught within it feel a loss and bequest higher, 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: they should both feel a deep entrenchment, fighting a similar headwind in a poverty trap. 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.

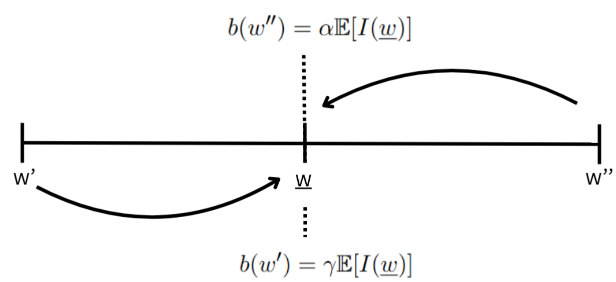


Figure 6: Diagram of bequest shocks that synchronize two agents' income but not their preferences.

Both agents feel a shock and though their income is the same, the initial richer agent increase 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 $\leq \alpha$. Looking at counter-cyclical relationships between income shocks and bequest propensities will identify the presence of economic headwind and the existence of loss averse preferences. 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.

There are obstacles to this approach that future research should explore. First, one would have to differentiate the effects of loss aversion from consumption smoothing, life-cycle effects, and heterogeneous preferences. At a national aggregate, this might be impossible. But poverty traps are context-dependent, and when investigating specific communities, this might be more feasible. Additionally, some features of the model are difficult to quantify. Initial wealth, conditions, for one. But consumption itself might change based on the quality of consumption first, or bequests might be low due to threshold effects (if an agent knows their child will receive a full ride then saving for their college might be moot).

In the Survey for Consumer Finance, for example, one might look at how an individual spends their money when given a windfall, the options being over consumption, savings, and paying off debt. If agents are loss averse and within the poverty trap, we'd expect to see investment in areas that support their children. If individuals were in the range of mobility, they might save this money (in the context of the model, this is more about consumption because it is not a direct present investment in their child), and in the poverty trap individuals might consume or pay-off debt (implying an immediate reaction to worsening income and investment in their child's current condition rather the agent's future consumption).

Half-lives. Using the Corollary (1), if a researcher can estimated 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 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

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 in order to know if this policy is necessary.

As discussion in Section (4), 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 [Mirrlees, 1971], which could be totally ineffective. Even if it is identified, the correct specification of the Micawber threshold and the steady states is still difficult to correctly estimate. If a policy planner misidentifies the true ranges of the poverty trap (namely w_u), then even 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:

$$(1-\tau)kR - kr + w \cdot r + T \geq [1-\pi k] \cdot (1-\tau)kR \tag{9}$$

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 $w < w^* =$

$$k(w, \tau, T) = \frac{\tau(R-r)}{2\pi R} + \frac{r - \sqrt{(\tau(R-r)+r)^2 - 4\pi R(wr+T)}}{2\pi R} \tag{10}$$

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:

$$I(w, \tau) = \begin{cases} k^* \cdot (1-\tau)(R-r) + w \cdot r + T, & \text{if } w \geq w^* \\ k(w, \tau) \cdot (1-\tau)(R-r) + w \cdot r + T, & \text{if } w < w^* \end{cases} \tag{11}$$

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):

$$K(\tau) = k^* \cdot [1 - G(w^*)] \; + \; \mathbb{E}[k(w, \tau) \mid w \in [\underline{w}, w^*]] \tag{12}$$

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:

$$\mathbb{W} = \int_i g(w^i) \cdot U^i(w^i, \tau) \partial w^i = \sum_{(a,b)} \int_a^b g(w^i) \cdot U^i_j(w^i, \tau) \partial w^i \tag{13}$$

$$(a, b) = \{[\underline{w}, w_p], [w_p, w_1], [w_1, w_2], [w_2, w_u], ...\}$$

$$\text{s.t. } \tau K(R-r) \geq 0$$

Given the social planner's welfare function in Equation, we can be write an implicit function for the optimal tax using the framework from Piketty and Saez [2013] in Proposition (3)

Proposition 3 (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:*

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

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.5). □

In Proposition (3), the outcome from the equity-efficiency trade-off is very sensitive to the distribution. The mechanical effect to the tax is the loss in net capital income and the gain from the rebate. For agents with wealth higher than the rebate, this is a negative mechanical effect; for those above it is positive. Since bequest and consumption utilities are logarithmic, the social planner might still implement a tax to support those with very income.

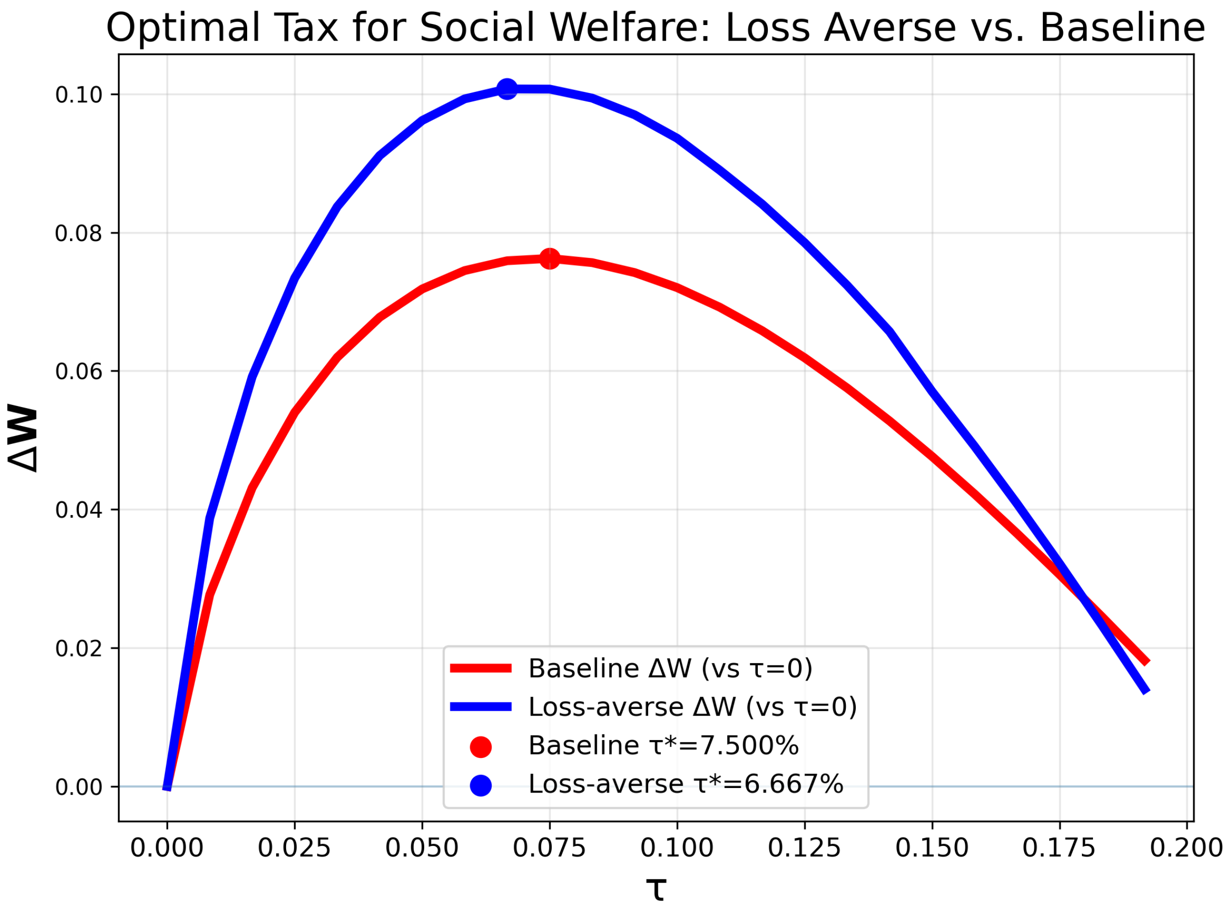
Loss aversion effects this optimal tax rate (1) up weighting those feeling a loss (equity) and (2) preserving the wealth distribution (efficiency). The more poor (rich) agents who feel a loss, the stronger (weaker) force for redistribution since that loss is weighted very highly.

by up-weighting those feeling a loss, and the optimal tax is therefore most sensitive to the distribution that feels a loss, rich and poor. As seen in **G**, those that give an optimal bequest less than or equal to their initial wealth have their marginal utility up-weighted by $\frac{1-\gamma}{1-\alpha} > 1$. Even if a social planner did not take into account loss averse utility, the optimal share of income allocated for consumption is still different for these individuals, which alone emits this weight. If loss averse utility is included, this is unweighted even more to $\frac{(1+\eta)(1-\gamma)}{\eta(1-\alpha)} > \frac{1-\gamma}{1-\alpha}$. 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. While loss aversion up weights those feeling a loss, if a greater part of the initial wealth distribution falls above w_r, i.e richer people feeling a loss, then this reduces the optimal tax rate since their mechanical effect is negative.

Though it is still optimal to borrow their first- or second-best level of capital, the first-best borrowing threshold the follows from the IC constraint (19) likely changes. If the tax is net-negative around w^*, then the new w^* will be higher than without the tax. The behavioral response from the tax creates an efficiency trade-off. Those who experience a net-negative from the tax will change their bequest, keeping more for current consumption and reducing the wealth passed on to the next generation, ultimately reducing the taxable revenue in the next period.

In Figure (**??**), the change in outcome when agents are loss averse is shown in the decrease in optimal tax.

ultimately the distribution itself determines the optimal tax. As seen in Figure (**??**), the change in the bequest with and without loss aversion hardly changes the optimal tax. This means a social planner cannot ex-post determine if there was a poverty trap simply based on elasticities of capital revenue or estimated optimal tax.



The result of the tax-redistribtion in the poverty trap model is often severe and can result in welfare reducing outcomes, as seen in Figure (7).

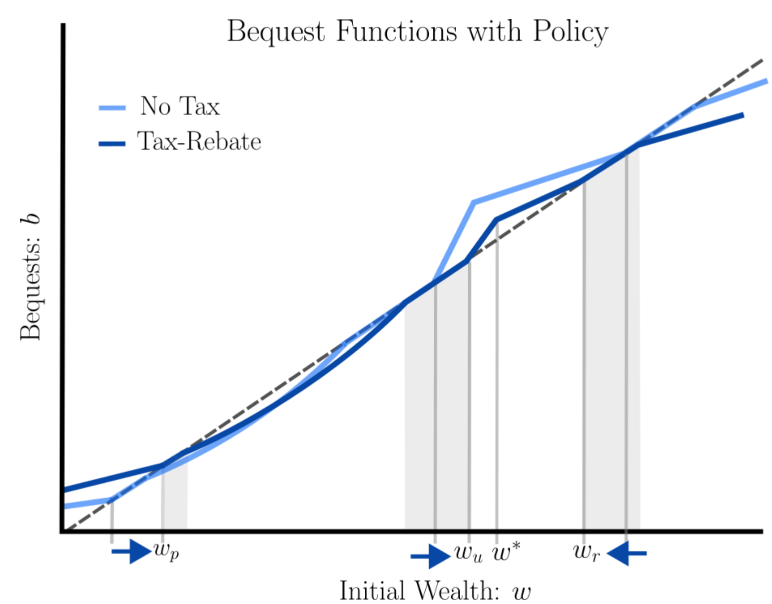


Figure 7: Loss averse bequest with redist. policy

Note that due to the efficiency effects and the dramatic bequest reduction from a large part of the distribution, as the time horizon of the social planner increases, so does the tax rate. If a policy planner fails to identify the poverty trap, an insufficient or welfare-reducing outcome might occur in the long run.

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 he S-shaped test in Section (**??**). This has policy consequences, which make the big push policy inefficient and ineffective.

To study the simplest possible case, I assume a social planner uses sovereign debt (alternatively, bonds if agents buy them as their bequest) in order to finance a big-push policy to be repaid through income taxes later on. The sovereign debt is equal to the demogrant each individual should receive to make an expected income equivalent to the expected income of an agent just above the empirically identified Micawber, w_m:

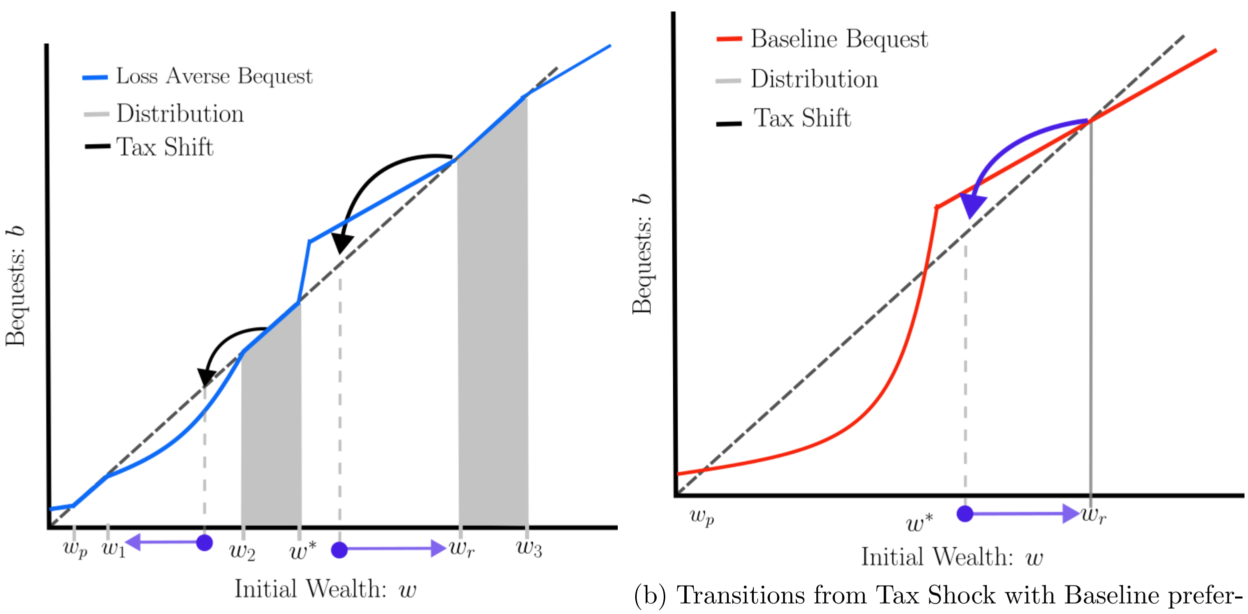
$$D = I(w_m)G_0(w_m) - \int_{\underline{w}}^{w_m} I(w^i)g_0(w^i)\partial w^i$$

where w^i is the initial wealth of individual i. An agents will receive an individualized demogrant d_i and so $D = \int_{\underline{w}}^{w_m} d(w^i)g_0(w^i)\partial w^i$.

Based on the S-shape identification results in Section (), a social planner might underestimate the big push when a poverty trap is identified. Under loss aversion, a planner will borrow less but only because it is spending less than it should. Demogrants will shift the distirbution up to w_2 but these individuals will not grow in income and infact are sensitive to falling back in due to small shocks. From Proposition 2, the big push long-run distribution will eliminate the mass of agents that might fall into w_p. In the loss averse model, the distribution crowds around the sticky states, both the interval lower than w_u and the one higher than w_r. In the baseline, everyone crowds around w_r.

If the debt bares no interest and has no repayment deadline (or long time-horizon), the social planner will levy a tax policy to repay the debt when the wealth distribution has reached the stable steady state.[5]. In the baseline model, since the entire measure of the wealth distirbution is at w_r, a lump-sum tax or flat tax rate is mathematically equivolent. In the loss averse model, these are different, but would emit similar effects. Both policies distort capital lending equilibirum and change the bequest transition function in a similar was as previously seen in Figure ().

Assuming market distortions have taken place, we can view a tax as a negative "shift" in someone's income that changes where in the bequest transition function they are, even when the tax distortions disappear. In Figure (8a), individuals initially in the sticky ranges are shifted down the transition function. Those above w_r might shift down below, but would rise towards w_r in the long run. Those in the lower sticky interval, $[w_2, w_u]$ might get shocked below w_2 but now enter a vicious cycle (both due to distortions of the market and also the negative income shift). If the big push was correctly estimated, all individuals would crowd around w_r or the sticky interval. When agents are shocked down, they will re-enter the area of mobility and overtime rise back up, as can be seen in Figure (8b).



(a) Transitions from tax with Loss Averse Preferences

(b) Transitions from Tax Shock with Baseline preferences

Due to repayment, a social planner will observe a part of the distribution fall back towards

[5]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.

poverty. While the initial cost of the big push was smaller because the big push was smaller, the planner will have to pay for this big push again, entering a cycle of pushing people back up only for them to fall back down because of small shocks or distortions from taxes. This is not just ineffective, it is a less efficient and more expensive policy. The long-run aggregate capital output will also be less: a large mass of agents who should produce capital gains at w_r are stuck earning less in the lower sticky interval. The deadweight loss of many small distortionary taxes and smaller aggregate output is much higher than if the threshold was correctly estimated.

6 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.

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.

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

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.

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.

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.

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.

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 also allows the Markov process to be ergodic, and we can explicitly define the long-run distribution only dependent on the initial distribution. If references are anchored in the distribution, then the process is non-ergodic, and 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 4 (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,*

$$b(w) = \begin{cases} \alpha\left(k^* \cdot (R-r) + wr + T\right), & w_3 < w \leq \bar{w} \\ w, & w_r \leq w \leq w_3 \\ \gamma\left(k^* \cdot (R-r) + wr + T\right), & w^* \leq w \leq w_r \\ \alpha\left(\underline{k} \cdot (R-r) + wr + T\right), & w_1 < w < w^* \\ w, & w_p \leq w_t \leq w_1 \\ \gamma\left(\underline{k} \cdot (R-r) + wr + T\right), & \underline{w} \leq w < w_p \end{cases} \tag{14}$$

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 9. 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.

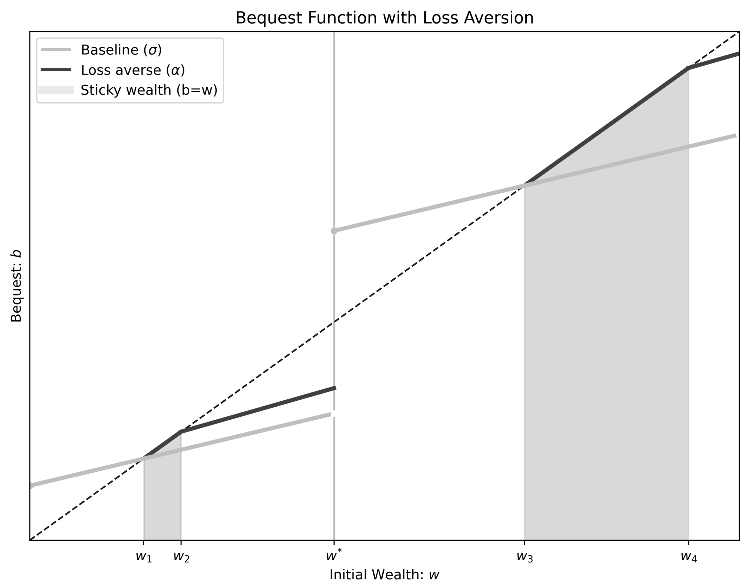


Figure 9: Bequest Function with Loss Aversion (Eq. 14) when $T > 0$.

A.2.2 Convex Production Technology

Given the set-up of Section (2.3.1), we define the incentive comparability condition as: $k(R - r) + wr + T \geq kR - k^2 R\pi$. 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:

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

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

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

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

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

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 (10)

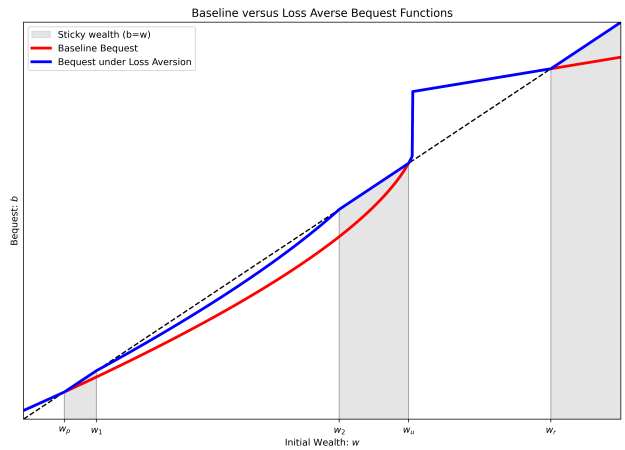


Figure 10: Numerically solved bequest function with convex bequests, $T > 0$.

A.2.3 Collapsing to Canonical Models

Proposition 5 (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*

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

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 5, 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 (6).

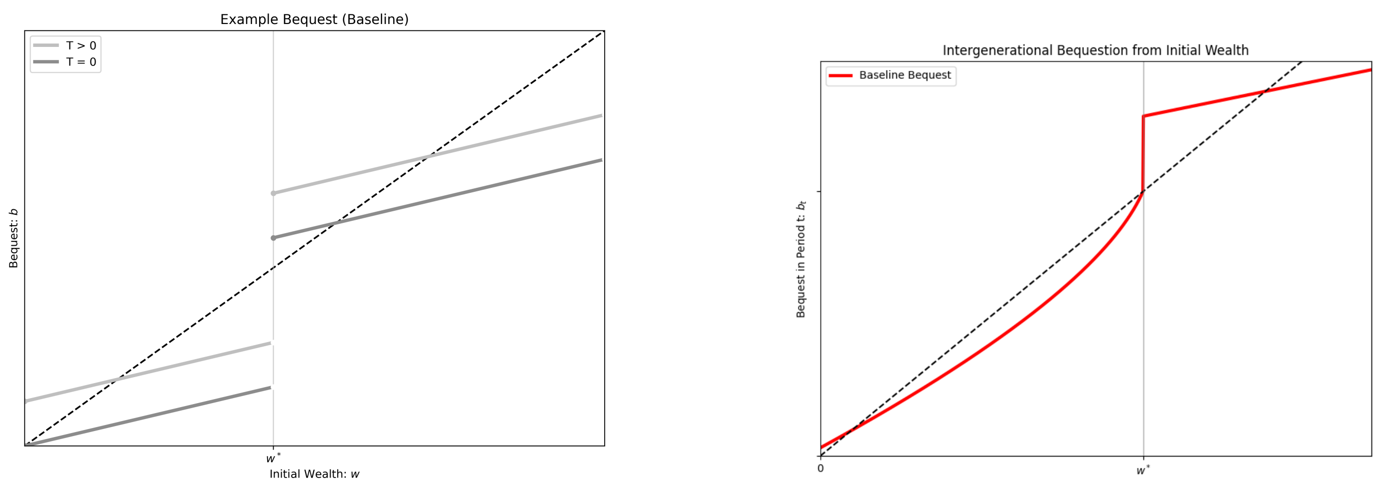
Proposition 6 (Bequest Function with a Convex Production Function)**.** *Let $V(k) = R\min\{k, \bar{k}\}$ with return rate $r < R$. Assume $\pi(k) = \pi\min\{k, \bar{\bar{k}}\}$ where $\pi \in (0, 1)$ and $\bar{k} < \bar{\bar{k}}$, and a harsh* *punishment,* $F = V(k)$. *Then the baseline bequest function takes the form*

$$b_B(w) = \begin{cases} \gamma\Big[k^* \cdot (R-r) + wr + T\Big] & \text{if } w \geq w^* \\ \gamma\Big[k(w) \cdot (R-r) + wr + T\Big] & \text{if } w < w^* \end{cases} \tag{16}$$

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 (5) and (6) allow insight into the range of bequest functions possible within the model. The visualizations for these propositions are below. The bequest function in Proposition 5 closely replicates the bequest function in Banerjee and Newman [1993] and the bequest function in Proposition 6 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 11b and Figure 11a.



(a) Bequest Function from Propn (5)

(b) Bequest Function from Propn (6)

Figure 11: Example Baseline Bequest Figures

A.3 Stationary Distrbution

$$\begin{aligned} G_\infty = {} & G_0(w_p)\delta(w_p) + [G(w_1) - G(w_p)] + [G(w_2) - G(w_1)]\delta(w_1) \\ & + [G(w_u) - G(w_2)] + [G(w_r) - G(w_u)]\delta(w_r) + [G(w_3) - G(w_r)] + [1 - G(w_3)]\delta(w_3). \end{aligned}$$

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

A.4 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.5 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.

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

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

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

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

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

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.

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

Following this, we can

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

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

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.

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

We can visualize the social planner's trade off for the curren