Loss Aversion and Poverty Traps

Miles T. Smith

November 29, 2025

Abstract

A common model of persistent poverty is that of poverty trap, which predict inescapable poverty and thus there is motivation for policy intervention. However, the persistent poverty literature often fails to identify poverty traps when looking for the key dynamics of traps within wealth panel data. In this paper, I develop an intergenerational model of a canonical poverty trap where agents have reference-dependent loss aversion over their bequest to the next generation. When agents would normally bequest less than their own initial endowment, they feel a loss (i.e. they feel like a failure) at giving their child a worse-off life than their own, causing poor agents to dramatically increase their bequest to avoid this felt loss. Downward wealth convergence dramatically slows or stagnates, and intergenerational wealth becomes more sensitive to shocks. This creates distributions that better align with data than most poverty trap models. Critically, I show this poverty trap model is often undetectable by common empirical methods. Failure to accurately identify the poverty trap them results in inefficient or welfare-reducing policy interventions.

1 Introduction

While the past century has brought about historically low levels of global poverty, a stubborn level persists despite well-intentioned and well-crafted interventions. Even the United States, one of the richest countries, still houses roughly 36.8 million people in poverty with pockets of abject poverty comparable to the poorest countries [Shrider and Creamer, 2023]. Despite some variance in that number, this has not meaningfully changed in the past 4 decades. Such poverty is not transitory. Most are poor for life [Benson et al., 2023].

One theory that attempts to explain persistent poverty and inequality is that of poverty traps. A poverty trap is a self-reinforcing mechanism whereby individuals below a certain wealth level (or income, asset, etc.) — an unstable steady state colloquially called the Micawber threshold — will only lose wealth every period and converge to a "poor" steady state; conversely, those above converge to a "rich" steady state within a multiple equilibrium trap. In these economies, an individual can do everything right, save every cent and maximize income, and still be resigned to poverty. Canonical examples of mechanical poverty traps range from rural farm work [Carter and Barrett, 2013], nutrition [Dasgupta and Ray, 1986], and geographical constrains [Jalan and Ravallion, 2002].

The burgeoning field of behavioral economics also attempts to explain the determinants of poverty. For example, Gennetian and Shafir [2015] offer a behavioral solution to the puzzle of why so many poor do not utilize welfare programs, pointing to behavioral theories like defaults and norms. They also explore how behaviors like present bias, while universal, are especially harmful to the poor. Other research has even generated behavioral poverty traps. For instance, poor agents might be overly risk averse, causing low investment into risky but high-return assets [Chivers, 2017, Barrett and Carter, 2013]. While behavioral economics offers insight into the persistence and creation of poverty, little work has been done to understand how behavioral traits can influence a mechanical poverty trap.

Despite the large theoretical literature on poverty traps, the empirical literature has struggles to find or can flatly reject the presence of a poverty trap with many contexts. Research from Kraay and McKenzie [2014] compile decades of research on trap identification and examine various panel data that might identify those vicious cycles. Their key concerns is that there is very rarely evidence of a Micawber threshold that bifurcates convergence paths; and there is little evidence that there is a vicious cycle at all. Aside from a few rare and specific cases, the authors conclude that poverty traps likely do not exist, especially in higher income economies. The debate is still divisive and while research continues, the consensus has largely resigned the existence of poverty trap to poor, rural economies.

However, there is a burgeoning literature within the broader social sciences that has not only brought poverty traps back into the academic discourse, but even points to their existence in the US. Recent sociological work consolidated by Desmond [2023] specifically points to the intersection of complex economic and social structures that, together, create and obfuscate poverty traps in America. And economic research is slowly catching up. For Black Americans, Derenoncourt et al. [2023] find that upward wealth convergence has stalled at low levels since the 1970s; in rural America, Allard [2019] find that poverty persists across generations and even the hardest workers are unable to escape. This motivates the question: is it possible that poverty traps can exist but the data fails to find evidence it exists or even rejects its existence?

To answer this question, I develop an intergenerational model of persistent poverty where agents can endogenously respond to the vicious cycle. Adapting the canonical framework from Banerjee and Newman [1993], I create an intergenerational model where each period an agents receives an initial wealth in the form of a bequest from their parent. They use this endowment as collateral for a loan which they invest in a capital market where, due to imperfect monitoring, an agent cannot always borrow their desired amount. After realizing capital gains, they chose an amount to bequest before exiting the economy at the end of the period. For agents with initially low wealth, they earn smaller incomes, and leave smaller bequests to their progeny, thereby passing on a worsening cycle of poverty. Agents internalize the effect of a vicious cycle by weighing the bequest they might give their child relative to the bequest they themselves enjoyed. I model this behavior using reference-dependent loss aversion first introduced by Kahneman and Tversky [1979] which states that agents overweigh perceived losses relative and will change their behavior to avoid them; here, agents anchor their expectation for the bequest based on their own initial wealth and will decrease consumption in order to increase their bequest towards that threshold. Intuitively, parents want to give their kids a life at least as good as they themselves had. This causes ranges of *sticky wealth* to emit across the poverty trap where agents consume less to keep their dynasty perpetually at the same higher wealth level, acting against the forces of the trap. Loss aversion creates a slowed and staled transition dynamics as well as greater wealth dispersion not predicted by other poverty trap models that obfuscate stable steady states and make intergenerational transition very sensitive to idiosyncratic shocks.

I find that the model presented in this paper fails to be identified using common empirical techniques even using rich and long panel data, corroborating the hypothesis that poverty traps can exist despite the conclusions of empirical analysis. Using simulations, I construct short and long panel data replicating what previous research has utilized. With very small shocks and no other noise to wealth transitions, I provide a best-case scenario for panel data. Then I employ two common benchmark tests. First, I attempt to empirically identify a Micawber threshold emulating the approach from Carter and Barrett [2013]. I find that the sticky intervals of wealth around the Micawber threshold obscure the ability of a threshold test to identify if transitions around a kink point are increasing or decreasing. Given certain parameters, the threshold test can actually predict a poverty trap is *not* present rather than simply a failure to predict a poverty trap. Second, I test for the S-shape curves using a cubic polynomial regression that attempts to identify the two stable and one unstable steady states which characterize the poverty trap model. Sensitive to the distribution, this test can potentially identify a trap, but it structurally mis-identifies the steady states and transition dynamics. Since the transition function from period-to-period flattens, the model predicts near-zero convergence to the identified stable steady states which is often argument against the existence of a poverty trap or actionable severity. Additionally, the sticky ranges of wealth mean around the steady states cause incorrect classification of the S-shape it identifies the poor steady state as much higher, the Micawber threshold as much lower, and the rich state as much higher. Together, this severely understates the scope of the trap and its severity, which can have significant policy implications.

Since even rich wealth panels and classic empirical identification methods produce unreliable detection methods, I formulate a new method to test for a poverty trap given loss averse preferences. I leverage reference-dependence by defining ranges in which random shocks to income can cause an agent from a lower wealth to earn as much as an agent from a higher wealth. By comparing their propensity to bequest, we can determine if there is a force of a vicious cycle, and where it ends. If there is a vicious cycle, an agent with a positive shock will code any bequest as a gain. But if an agent who makes s similar income without a shock, or an agent shocked negatively to that income level, has a different bequest propensity that is higher, then there is a vicious cycle. With many samples across the income-wealth distribution and over several periods, one could estimate the severity and scope of the trap based on when propensities from shocks differ and converge. Additionally, by estimating the observed half-life of individuals within the poverty trap, I provide a method to backward induct the half-life of the mechanical poverty trap.

A key consequence of failing to identify a poverty trap is that social planners might employ the wrong policy interventions, which I find can be welfare reducing. The typical poverty trap interventions is known as the "big push" where all agents below the Micawber threshold are given just enough extra income or wealth to be above it, and therefore enter a region of mobility. This big push, especially when targeted at the key poverty mechanism, is often the best policy. However, a social welfare planner who rejects or fails to identify a poverty trap will employ other policies typical in other contexts, like simple tax redistribution schemes. I show that when a benchmark flat-tax rebate scheme is employed within this model, the redistribution often lowers the rich tax base faster than it can elevate the poor tax base. The effect of even a small tax rate is often a larger loss of income than the gain of the rebate for those around the Micawber threshold, which increases the incentive for reneging since these agents' wealth decreases. This can cause the poverty trap to enlarge, and as more people become trapped, the current and future wealth distributions can quickly collapse. As I calculate, the optimal tax rate sits low, around 3%. I also compare this benchmark tax redistribution to a model without loss aversion, and find that while the optimal tax rate is about the same, the effects described above take place more intensely. Over long-run horizons, while the model without loss aversion collapses the economy quickly, the model with loss aversion sustains the wealth distribution for longer, allowing for some poverty alleviation to occur, but the welfare consequences remain uncertain since more parents must cut current consumption in lieu of taxes in order to maintain their dynasty's wealth.

Furthermore, I find even when a poverty trap is predicted the policies are inefficient and ineffective because the empirical tests fail to accurate estimate the extent of the poverty trap. When a social planner identifies a poverty trap, they employ a big push policy. Taxes, instead of being rebated equally, are given directly to the poor proportional to each agent's individual distance to the Micawber threshold. The big-push can also be funded by debt, which can be paid off in taxes to future generations, which itself has similar and slightly welfare-increasing consequences as the no-debt model. However, if a planner mis-identifies the threshold, which my model shows they might, a planner might employ too weak a big push policy, imposing longer-term financing through taxes as individuals fail to enter upward mobility and fall back into poverty. This prolongs the need for policy, making it inefficient. And it has a greater impact on efficiency impacts, decreasing the intergenerational transmission of wealth and therefore reducing the future tax base. If all individuals are taken to the edge of the sticky state, then every people some agents will be shocked back into the poverty trap. The stationary Markov distribution shows that not an insignificant measure of agents will always need to be helped back up, thereby enforcing a consistent tax that has the same effect on lending conditions as before. Loss aversion and the mis-identification of the poverty trap might explain why certain big-push policies have been seen as ineffective since they are too small.

The theoretical, empirical, and policy results of this paper are novel contributions to the fields of persistent poverty. To my knowledge, the identification and consequences of sticky regions from loss aversion has only vaguely been identified and discussed twice before. First, Blumberg and

Kremer [2014] discuss that loss aversion can create a poverty trap where there is a range poor stable steady states where agents are afraid to trade low-return safe assets for high-return, risky assets. But their analysis largely ends at the mention of possible existence, and does not identify possible sticky ranges across the wealth distribution. Additionally, my model does not generate persistent poverty through poor behavior, rather the endogenous behavior of the poor confounds the evidence, which has newly explored empirical and policy consequences. More recently, Bramoullé and Ghiglino [2022] use loss aversion to form a range of conspicuous consumption beliefs, but their paper is far outside the current literature, analysis and results.

As a map of this paper, I begin in Section 2 by setting up the economic model. In section 3, I discuss the emergence and dynamics of a poverty trap within the model's framework. 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[1]. 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 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, $\rho_c = 0$. This allows me to examine how the transmission of wealth changes only as a result of preferences over bequests. 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. For a discussion on the addition of losses over consumption, see Appendix (...).

[1] 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.

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 L units of capital. Agents invest L and receive investment income $V(L)$. I assume $V(L)$ is increasing in L and that there exists a unique L^* that maximizes investment income, which is the "first-best" level of capital[2]. An agent who borrows $L < L^*$ borrows their "second-best" level of capital. There is also a global interest rate $r \geq 1$ 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: Lr; 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: $V(L) + wr - Lr + T$.

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 $\pi(L)$ they are caught and receive some punishment F. I assume $\pi(L)$ is increasing and concave with L: the bigger the loan, the easier it is to monitor. These assumptions of $\pi(L)$ and F drives the lending dynamics. The expected value of reneging is $V(L) - \pi(L)F$. Borrowers will renege if the expected value of reneging is greater than the expected value of repayment: $V(L) - \pi(L)F > V(L) + wr - Lr + T$, which can be rewritten as when the loan is sufficiently large $L > w + \frac{\pi(L)F+T}{r}$. 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: $L \leq w + \frac{\pi(L)F+T}{r}$.

The minimum initial wealth needed to borrow the first-best level of capital can be found by plugging in L^* in the IC and solving for w^*: $w^* = L^* - \frac{\pi(L^*)F+T}{r}$. Agents with $w \geq w^*$ only

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

borrow L^*; Agents with $w < w^*$ borrow their second-best level of capital which depends on their initial wealth. Since the investment returns are increasing in L, and agents always want to borrow the maximum L they can. The implicit loan function for $w < W^*$ is $L(w) = w + \frac{\pi(L(w))F+T}{r}$. As the probability of catching renegers and the expected value of the reward increase, the lower the necessary wealth is needed to borrow L. For tractability, I assume there is a maximum level of capital that can be monitored by lenders is L_π, where $L^* < L_\pi$. 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^*:

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

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

Each period reaches its own static equilibrium. Before the end of every period, all economic activity occurs in equilibrium – lenders and borrowers match, loans are made, incomes realized, and loans are repaid. At the end of every period, individuals maximize their Utility Function (1) over bequests and consumption subject to their budget constraint in Equation 2. Before exiting the economy, agents leave a bequest for their progeny. Then all current agents exit, and their children replace them in the economy with their parent's bequest as their initial wealth.

In the static equilibrium, we can explicitly define exactly what an agent bequeaths as a function of their initial wealth. Using the Utility Function (1), we can calculate the MRS between consumption and bequests to write consumption as a fraction of bequests. Substituting consumption c in the Income Function (2), we can write bequests explicitly as a piecewise function of initial wealth.

$$\frac{U_c}{U_b} = 1, \qquad b + c = I(w).$$

The MRS depends on if an agent is feeling a bequest loss of gain. If they feel a gain, the MRS is simply CObb-Douglas with weight γ; if they feel a loss, this augments their bequests propensity.

In order define the static equilibrium bequest and to understand the underlying mechanisms of the economy that drive the dynamics of the model, I first solve the equilibrium bequests when they are coded as a gain. Then, I solve the optimal bequest when coded as a loss.

2.3 Static Equilibria

Static Equilibrium without Loss Aversion.

To understand the mechanical effects of the market on an agent's optimal bequest, I first describe the bequest function for when agents code a bequest as a gain rather than a loss. When agents feel a gain, $\lambda = 1$ and when we take the MRS of Utility (1), the $(1 + \eta)$ components numerator and denominator cancel out, $c = \frac{(1-\gamma)(1+\eta)}{\gamma(1+\eta \cdot 1)} b = \frac{1-\gamma}{\gamma} b$. What results is the "baseline" bequest function where the propensity to bequest follows the standard Cobb-Douglass preference with bequest weight γ.

Definition 2.1 (Baseline (Cobb-Douglas) Bequests)**.** If $b(w) \geq \rho_b$, then the proportion of income reserved for bequests is the standard Cobb-Douglas preference γ. The bequest function can be written implicitly as a piecewise function of initial wealth w:

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

For illustrative purposes so that I can compare the effects on loss averse bequests on the underlying mechanisms of the economy, I let the reference point for bequests to be sufficiently small such that $\forall w$, $b(w) > \rho_b$. When agents always code a bequest as a gain, the model simplifies to a bequest structure that is similar to the model explored in Banerjee and Newman [1993] and closely aligns with their follow-up paper in Banerjee and Newman [1994], two examples of a canonical poverty trap.

Depending on the specification of the production function $V(L)$, the reneging probability $\pi(L)$, and the reneging punishment F, the baseline bequest phase diagrams can look very different. Still, a typical feature of this general bequest function is that there often exists an unstable steady state, where an agent with initial wealth just below this threshold will bequest less than their initial wealth and those just above bequest higher than their initial wealth.

Definition 2.2 (Unstable Steady State)**.** There exists a w_u if

- $b(w_u) = w_u$
- there exists a $\delta > 0$ such that $(b(w) - w) \cdot (w - w_u) > \delta$ for all $0 < \delta$. (i.e. $w < w_u$ will have $b(w) < w$ and $w > w_u$ will have $b(w) > w$.)

If $b(w)$ is continuously differentiable around w_u, then $b'(w_u) > 1$.

By definition, there is always at least one stable steady state where $b(w) = w$, when $w = 0$. Generally, the model will have a stable steady state in a lower-wealth area which I denote as w_p and another in a higher-wealth area as w_r. The wealth distribution paths to steady states will be discussed in Section (3), but for now it will suffice to define them.

Since the bequest function is sensitive to functional forms, as are the steady states, I present two examples of the production technology and monitoring function that characterize the range of potential functional forms the bequest function can take within this model. We will see that with different specifications of the economy, the mechanical effect on the bequest changes quite a bit.

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

$$b_B(w) = \begin{cases} \gamma\big[L^* \cdot (R - r) + wr + T\big] & \text{if } w \geq w^* \\ \gamma\big[wr + T\big] & \text{if } w < w^* \end{cases} \tag{4}$$

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

Proof. See appendix. □

In Proposition 1, 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 (2).

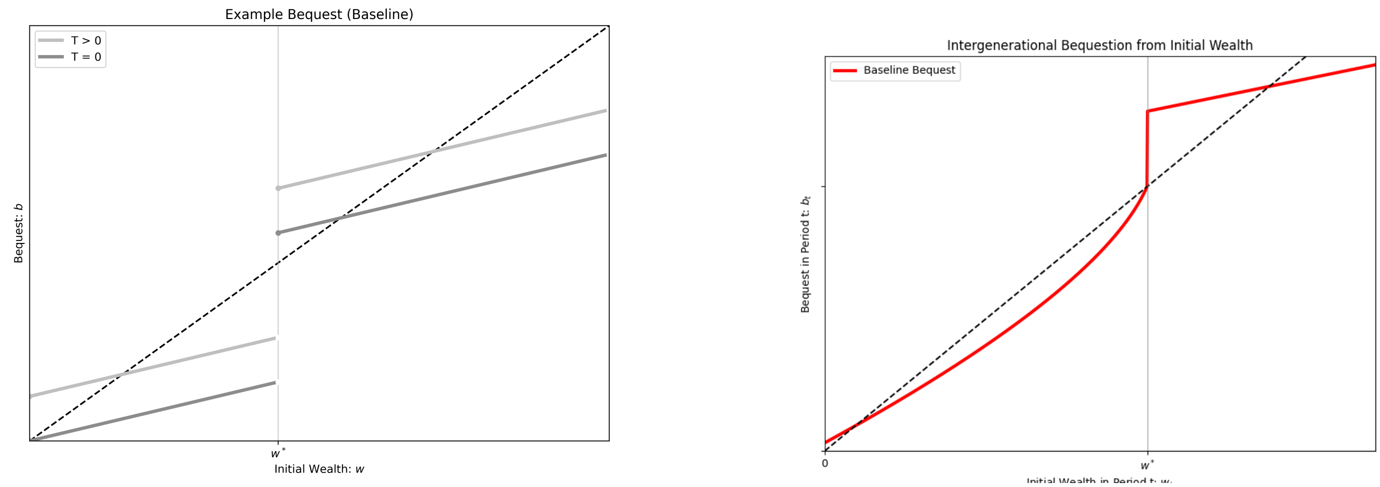
Proposition 2 (Bequest Function with a Convex Production Function)**.** *Let* $V(L) = R\min\{L, \bar{L}\}$ *with return rate* $r < R$. *Assume* $\pi(L) = \pi\min\{L, \bar{\bar{L}}\}$ *where* $\pi \in (0,1)$ *and* $\bar{L} < \bar{\bar{L}}$, *and a harsh punishment,* $F = V(L)$. *Then the baseline bequest function takes the form*

$$b_B(w) = \begin{cases} \gamma\big[L^* \cdot (R-r) + wr + T\big] & \text{if } w \geq w^* \\ \gamma\big[L(w) \cdot (R-r) + wr + T\big] & \text{if } w < w^* \end{cases} \tag{5}$$

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

Proof. See appendix. □

The specifications in Proposition (1) and (2) allow insight into the range of bequest functions possible within the model. The visualizations for these propositions are below. The bequest function in Proposition 1 closely replicates the bequest function in Banerjee and Newman [1993] and the bequest function in Proposition 2 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 1b and Figure 1a.



(a) Bequest Function from Propn (1)

(b) Bequest Function from Propn (2)

Figure 1: Example Baseline Bequest Figures

In Figure 1a we see a stark difference in the income and ability to bequest between agents just above and below the Micawber threshold. Meeting this threshold significantly increases an agent's income and therefore bequest, so much so that for agents with an initial wealth just above $w^* = w_u$ leave a bequest *higher* than their own initial wealth. Their children become richer than them. The opposite is true for those just below: those with initial wealth $w < w^*$ leave a substantially smaller bequest to their child.

In Figure 1b, the difference in bequest amount to the left and right of w_u is more subtle because $w_u < w^*$, meaning the unstable steady state is within one piece of the income function. It is rare that $w_u = w^*$, and in fact the unstable steady state can be much higher or lower than the wealth cut-off, which I explore quantitatively in Appendix (...). Here, wealth levels around the unstable steady can only borrow their second-best level of capital. Some individuals can borrow the second-best and still leave a larger bequest, but many who borrow their second-best leave a smaller bequest.

Static Equilibrium with Loss Aversion.

The introduction of loss aversion adds a level of complication, but rich results. While the static equilibria and results of the model hold for an arbitrarily defined ρ_b, I let $\rho_b = \rho_b(w) = w$. In my model, parent's anchor their belief of what they should give their child based on what they themselves received as their initial wealth. Put differently, parents want to give their child a better life, or at least as good a life, that they had. Failing to provide to at least this level makes a parent feel a loss. This allows us to explore a rich narrative of how parent's loss-averse preference affect the distribution of wealth and poverty. Other ways to define ρ_b are discussed in Appendix (A.2) but they do not change the underlyinng results of the model.

An agents feels a loss by a factor of λ when their baseline bequest (Equation 3) is less than their initial wealth, $b_B(w) < w$. Agents increase their propensity to bequeath to avoid this loss. This changes the share of consumption and bequests, $c = b\frac{(1-\gamma)(1+\eta)}{\gamma(1+\eta\lambda)}$. Plugging this into the income function in Equation (2), we get the propensity to bequeath when feeling a loss to be

$$b_L(w) = \alpha I(w) \quad \forall\ b_B(w) < \rho_b, \tag{6}$$

where $\alpha = \frac{\gamma(1+\eta\lambda)}{1+\eta-\gamma\cdot\eta(1-\lambda)}$.

Remark 1. Given $\lambda > 1$ and $\eta > 0$, then $\alpha > \gamma$ and $b_L(w) > b_B(w)$.

There is a careful stitching that must be done to incorporate the effect of loss aversion in the static equilibrium baseline bequest function. As we can see from Remark (1), there is a non-empty interval between $b_L(w)$ and $b_B(w)$. Importantly, when an agent's baseline bequest is *less* than w, it is possible their increased bequest is *greater* than initial wealth w, so within that non-empty interval between is the agent's initial wealth. In this situation, an agent is stuck between $b_B(w) < w < b_L(w)$, where they feel a loss at their baseline bequest, but they feel a gain if they bequest with propensity α. Here, the optimal bequest is a corner solution: $b_L(w) = w$. And there can be large ranges of wealth levels where an agent's optimal bequest is this corner solution. More formally, I define this as a range of "sticky" wealth in Definition (2.3).

Definition 2.3 (Ranges of Sticky Wealth)**.** Let $b_L(\cdot)$ denote the bequest rule under a loss and fix a branch $D_p \in \{[\underline{w}, w^*),\ [w^*, \bar{w}]\}$ *s.t.* $G(w) \sim [\underline{w}, \bar{w}]$. Define the fixed-point set $F = \{w \in D_p : b_L(w) = w\}$.

(i) Sticky interval. A closed interval $S_i = [\min(a,b), \max(a,b)] \subset D_p$ is a *sticky interval* if

(a) $b_L(w) = w\ \ \forall w \in [\min(a,b), \max(a,b)]$ (b) $\exists\, \varepsilon > 0 : \begin{cases} b_L(w) - w \ \geq 0 & \text{for } w \in (a-\varepsilon, a) \cap D_p, \\ b_L(w) - w \ \leq 0 & \text{for } w \in (b, b+\varepsilon) \cap D_p. \end{cases}$

(ii) Sticky Ranges of Wealth. The *sticky ranges on D_p* are defined as the set of sticky intervals $S = \{S_i\}$. Note that there can exist several sticky intervals on a D_p, one, or none.

The definition of a sticky ranges is broadly defined as a continuous and dense interval of unstable steady states where bequest at a wealth level just outside the interval is larger or smaller that w. These ranges appear in different parts of the wealth distribution and the ranges extend or contract depending on the functional forms and parameterization of the model. To gain an intuition for the economic conditions that determine how sticky ranges develop, I return to the examples outlined in the previous section and allow for loss-averse bequests in Equation (6). First, I examine the linear model in Proposition (3).

Proposition 3 (Loss-Averse Bequests with Linear Production Function)**.** *Given the assumptions in Proposition 1 and $\rho_b(w) = w$. For agent's with $b(w) < w$, they increase their bequest propensity to α defined then. Then, by Definition 2.3, we can see the set of sticky intervals on $G(w)$ is $\mathcal{S}(w < w^*) = \{[w_p, w_2]\}$,and $\mathcal{S}(w \geq w^*) = \{[w_r, w_4]\}$. We can explicitly define these regions: $w_p = \frac{\gamma}{1-\gamma r}\left(\underline{\underline{L}} \cdot (R-r) + T\right)$ and $w_2 = \frac{\alpha}{1-\alpha r}\left(\underline{\underline{L}} \cdot (R-r) + T\right)$. Further, $w_r = \frac{\gamma}{1-\gamma r}\left(L^* \cdot (R-r) + T\right)$ and $w_4 = \frac{\alpha}{1-\alpha r}\left(L^* \cdot (R-r) + T\right)$. Then the bequest function on is explicitly defined as*

$$b(w) = \begin{cases} \alpha\left(L^* \cdot (R-r) + wr + T\right), & w_4 < w \leq \bar{w} \\ w, & w_r \leq w \leq w_4 \\ \gamma\left(L^* \cdot (R-r) + wr + T\right), & w^* \leq w \leq w_r \\ \alpha\left(\underline{L} \cdot (R-r) + wr + T\right), & w_2 < w < w^* \\ w, & w_p \leq w_t \leq w_2 \\ \gamma\left(\underline{L} \cdot (R-r) + wr + T\right), & \underline{w} \leq w < w_p \end{cases} \tag{7}$$

and the $w^* = w_u$ remains the same as the baseline.

Proof. See Appendix. □

While the bequest function might seem complicated, graphically the function makes intuitive sense, as seen in Figure 2. The light-gray lines represent the bequest function of the baseline model without loss aversion and all agents simply bequeath γ share of their income. With loss-aversion, there are large ranges of sticky wealth just above *both* the poor and rich steady states on intervals $[w_1, w_2]$ and $[w_3, w_4]$. Those within their income function bracket above the sticky range now *all* bequest at a higher propensity, endowing the next generation with a much greater initial wealth. What is notable about this model it that sticky regions admit in multiple locations across the income distribution, which has not been previously modeled to my knowledge.

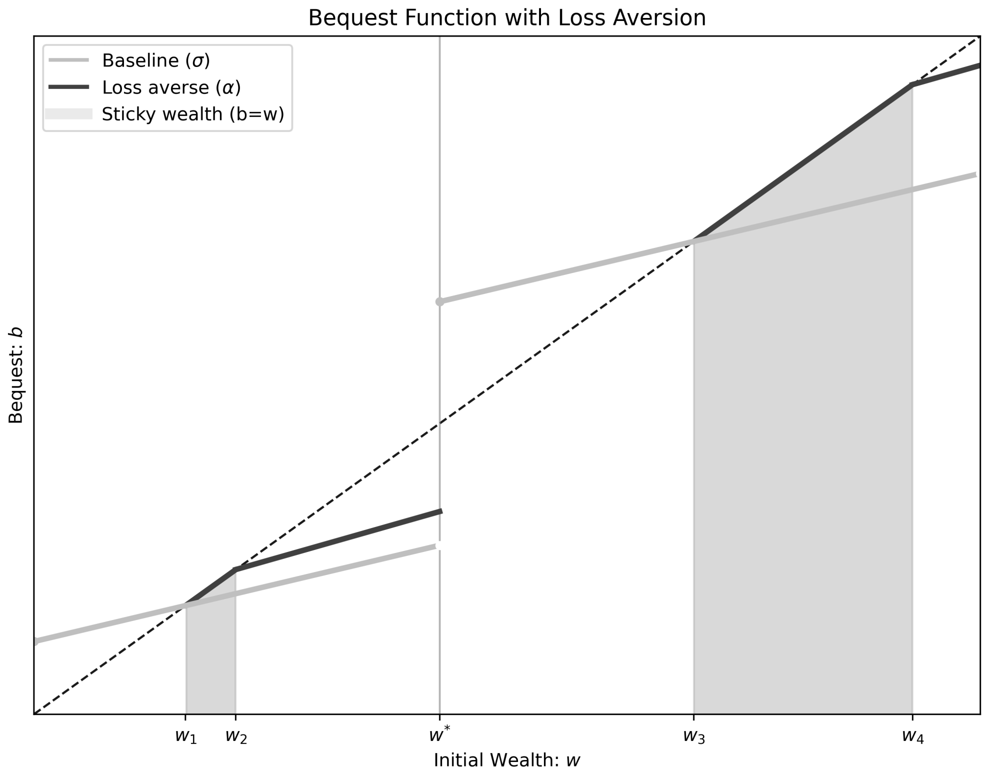


Figure 2: Bequest Function with Loss Aversion (Eq. 7) when $T > 0$.

This linear bequest function has a unique structure and is an example of an extreme version of the model. Allowing the loss-averse model to adopt richer functional forms like in Proposition 2 allows for the conditions for regions of sticky wealth to occur and where they occur. For the remainder of this paper, I focus on the more sophisticated version of the economy outlined in Proposition (2) and stitch it with loss averse preferences in Proposition (4) below.

Proposition 4 (Loss-Averse Bequest with Convex Production)**.** *Recall Proposition 2 and allow the same functional forms for $V(L)$, $\pi(L)$, and F and the basic assumptions. By Definition (2.3),*

$S(w < w^*) = \{[w_p, w_1], [w_3, w_u]\}$ *and* $S(w \geq w^*) = \{[w_4, w_5]\}$

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

where the explicitly defined kinks in the piecewise function are found in Appendix **??**. *When* $T = 0$, *the bequest function collapses for the bottom pieces:* $w_p = w_1 = 0$.

Proof. See appendix. □

Notice that when compared to linear loss-averse bequests in Proposition (3), the range of the poor sticky unstable steady states now includes a range directly below w_u, $\{[w_2, w_u]\}$. Mathematically, as the rate of bequests slows or accelerates near a crossing point with the 45-degree line, a sticky range can emerge. This is driven by the logarithmic utility function. Intuitively, agents are able to match their bequest when they earn enough,$\{[w_2, w_u]\}$ and $\{[w_r, w_3]\}$, or when their bequest is low enough, $\{[w_p, w_1]\}$. Agents between two sticky regions who feel a loss still bequeath more than without loss aversion.

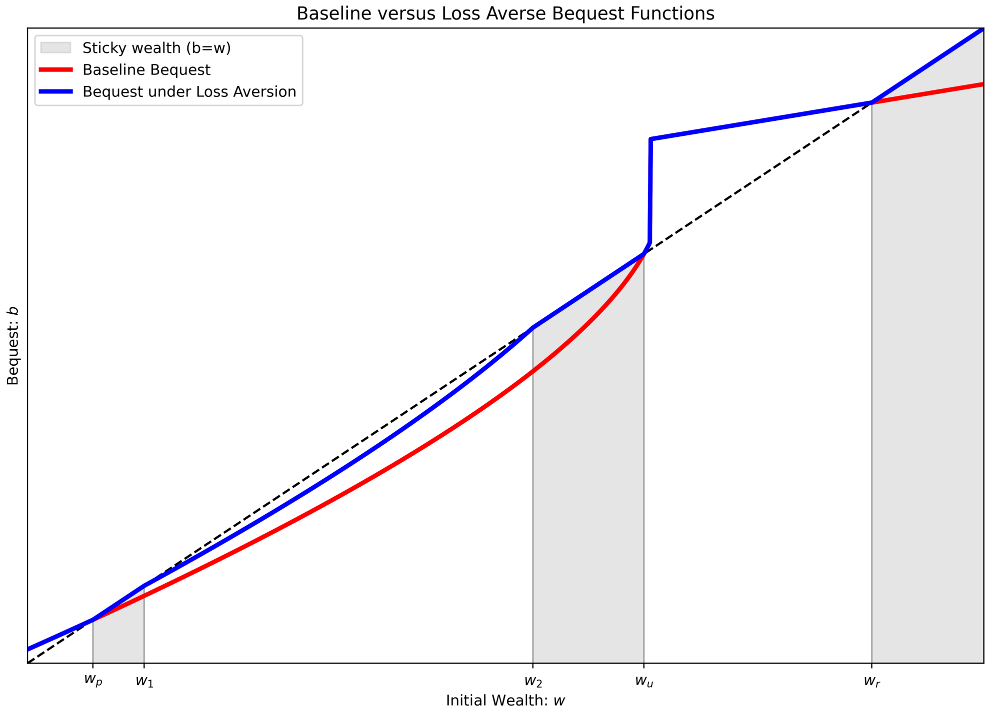
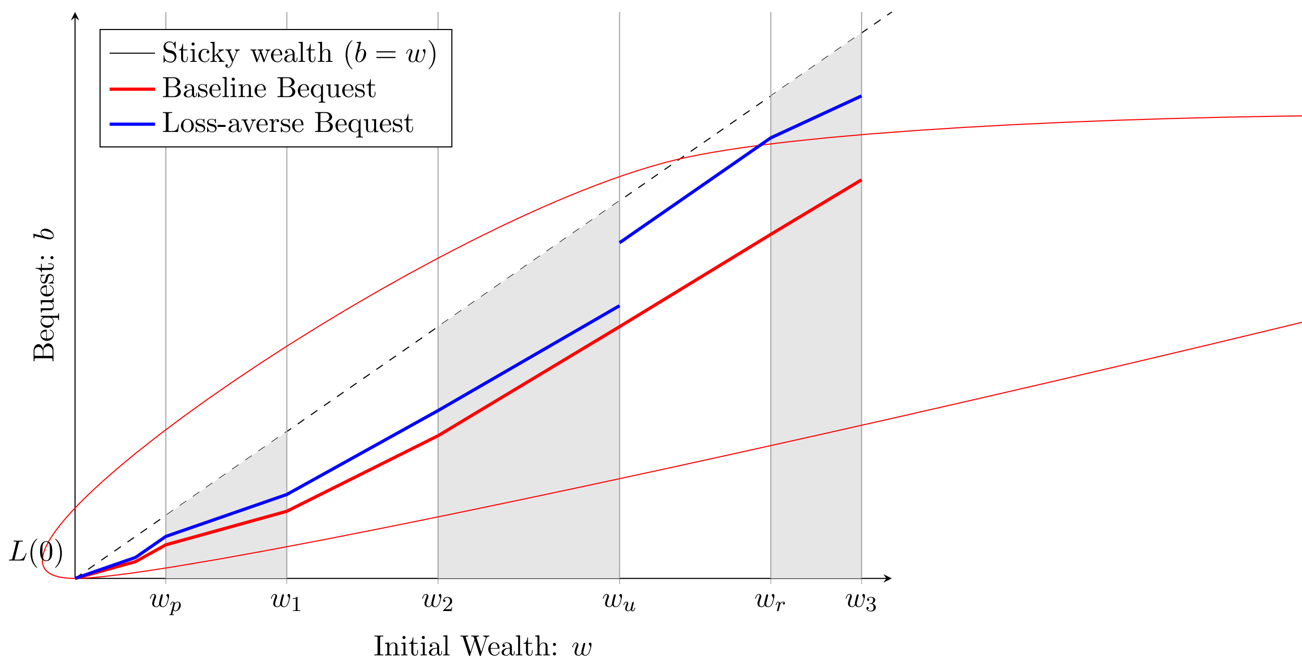


Figure 3: Bequest Function with Loss Aversion (Eq. 7) when $T > 0$.



As one could intuit, 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. We can this this rotation in Figure 4, where $\lambda = 1$ is the baseline bequest curve when a parent codes all bequests as a gain. To focus on just the sticky range beneath the unstable steady state, I let $T = 0$.

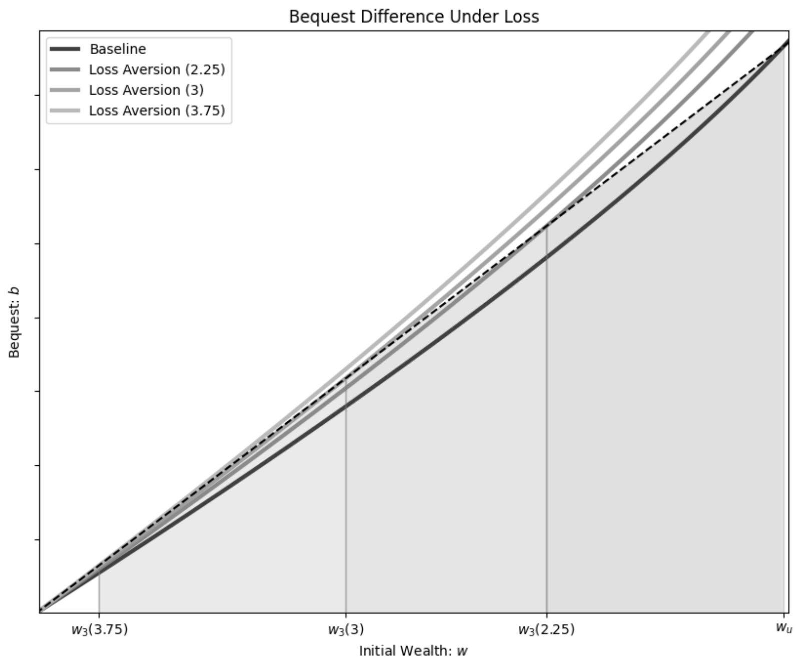


Figure 4: Bequest function for $b_L(w \ : \ w < w^*)$ change as λ increases when $T = 0$.

It is important here to note that generally, the measure of a sticky interval for the rich is greater than the measure of either sticky intervals for the poor. There are two mathematical sources for this. First, most simply, the rich have more income they can devote to avoiding the felt loss. Second, the the distance between the bequest and the reference point – and the difference in the utility for each – widens as income gross. Within context to this model, this means the rich feel a greater loss than the poor always. And with a greater means to avoid that, there admits a larger rich sticky range of wealth. Given a more complex functional form for the production technology where multiple equilibria could admit like in Barrett and Carter [2013], this is still still holds true by the very natural of the reference point and wealth. As we will see in the next section, this produces extremely stark differences in intergenerational dynamics compared to the baseline model.

Finally, it is important to note that the welfare implications of loss aversion within the static equilibrium are unclear: while an agent avoids the felt loss by increasing their bequest, they lose welfare by consuming less. For poor agents who increase their bequest, this results in a greater percent change from their previous consumption than for richer agents. When welfare is weighed off of only consumption, loss aversion results in a large welfare loss. When welfare is weighed off of the bequest – or reframed as the next generation's wealth distribution – loss aversion results in a large welfare gain. For the individual, the welfare consequences of loss aversion are unclear. The remainder of this paper will examine the welfare of society.

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 where below a certain wealth threshold, agents can only converge to a poor steady state. 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 (5). 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.

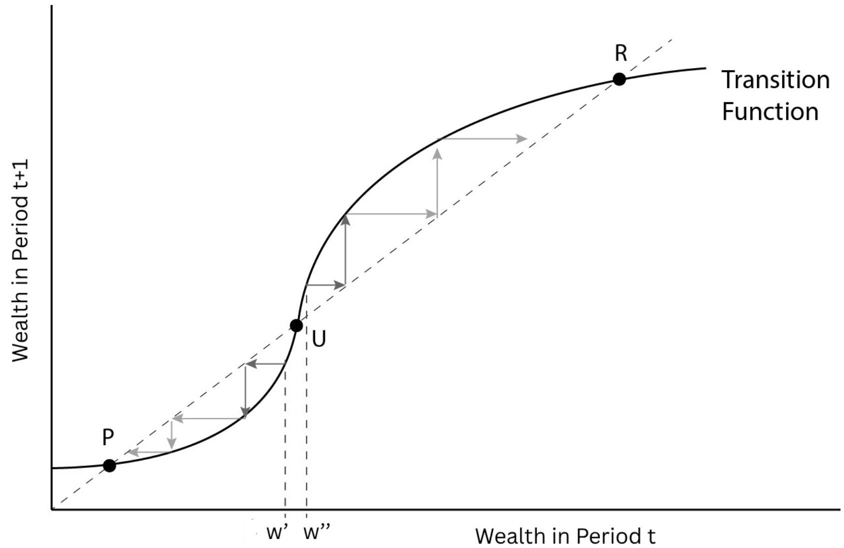


Figure 5: A General Diagram of a Poverty Trap and Wealth Transitions

As seen in Figure 5, 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 in Definition 2.2:

Definition 3.1 (Micawber threshold)**.** An unstable steady state w_u as defined in Definition 2.2 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 $w_p, w_r \in G \sim [\underline{w}, \bar{w}]$ are "poor" and "rich" stable steady states where all three steady states are time-invariant.

By defining the Micawber threshold and given at least one stable steady states, we can explicitly define the share of a given distribution that will be trapped in poverty in the long-run equilibrium. An important feature of this model – like other poverty trap models – is that the bequest function directly maps today's distribution of initial wealth into tomorrow's distribution of initial wealth: $b : G_t(w) \to G_{t+1}(w)$. This is because b does not depend on t for either the baseline or loss-averse model. Since b is time-invariant and continuous on a finite state space, then the baseline model follows a stationary Markov process[3]. This means the model is ergodic, and that by knowing the distribution in time $t = 0$, we will know the distribution in as $t \to \infty$.

Proposition 5 (Stationary Wealth Distribution in a Canonical Poverty Trap)**.** *Given the baseline bequest function in Equation (3), b_B is a direct mapping to and from the state space $W \in [\underline{w}, \bar{w}]$, so that $b_B(G_t(W)) \mapsto G_{t+1}(W)$. By definition of the Micawber threshold (3.1), the bequest function is a contraction between $[\underline{w}, w_u)$ and $(w_u, \bar{w}]$. Then, we can explicitly define the stationary distribution of the model as*

$$G_\infty = G_0(w_u)\,\delta_{w_p} \;+\; G_0(w_u)\,\delta_{w_r}.$$

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

In Figure (1), we can use Proposition (5) to play out the bequest transition function as $t \to \infty$. In Figure (1a), individuals just to the poorer than $w_u = w^*$ bequest at a loss every period, where eventually a generation will bequest w_p, where the dynasty will forever remain. The predictions of this model are sensitive to the parameterization which defines w_u and the scope of those trapped are, of course, sensitive to the initial distribution of wealth $G_0(W)$. THis is also true for the convex bequest function in Figure (1b).

The dramatic prophecy of Proposition (5) where the economy will stratify into only the rich (or simply rich-*er*) and the abject poor is seldom proven in observable data. Even in the poorest regions, [Kraay and McKenzie, 2014] find there is not much evidence for a stark Micawber threshold and two steady state equilibria. Using long-to-short sized panel data, they also claim there is no observable convergence down to a poorer state. What I find, however, is that without changing the underlying economic forces, loss aversion creates a wealth dispersion, intergenerational elasticity of wealth, and observable steady states that are much more realistic and might explain the limitations of data to prove a poverty trap.

Poverty Trap with Loss-Averse Bequests

The introduction of loss aversion over bequests significantly disrupts the canonical poverty trap

[3] 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 (A.2), 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.

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, dampening and observationally changing the dynamics of the poverty trap to no longer look like one just from observable characteristic of intergenerational wealth. Still, loss aversion does not necessarily allow dynasties to *overcome* or *eliminate* the poverty trap.

First, loss aversion can change the observable Micawber threshold from the true Micawber threshold. The conditions for when this divergence occurs in outlined in Lemma (1). Since loss aversion does not generate the poverty trap, the scope of the vicious cycle still extends to w_u. But, as seen in Figure (3), agents with wealth below w_2 are observed to lose initial wealth every generation. To a observer of data naive to the potential effects of loss aversion on intergenerations wealth, they might predict a *lower* Micawber threshold than is true.

Lemma 1 (Change in Observable Micawber Threshold)**.** *I define w_m^0 is the observable Micawber threshold and w_m^t is the true Micawber threshold. Given $\lambda > 1$ and $\eta > 0$, and $b(w_t) < w$, then $w_m^o \leq w_m^t$. And if $b'(w_m) > 1$, then $w_m^o < w_m^t$.*

Often, too, is that agents converge to a higher poor and rich steady state at the edge of the ranges of sticky wealth. Agents near poverty (and higher wealth levels) have some dispersion due to these sticky ranges. And individuals in both Figure (2) and Figure (3) would converge to $w_1 > w_p$. For agents with initial wealth in a sticky state $\mathcal{S} = \{[w_p, w_1], [w_2, w_u], [w_r, w_3]\}$, agents would not converge to a lower steady state at all, which goes against the conventional intuition of the poverty trap. A naive observer of the intergeneration wealth data might predict a much *higher* poor steady state than is true or a lack of vicious cycle for the sufficiently poor.

If loss aversion is high enough, the dynamics of the poverty trap can entirely disappear from observable data. This is clear from Figure (4): as lambda increases, the scope of the ranges of sticky wealth also increase. If there are two ranges of sticky wealth, the expansion of both ranges can overlap, creating no observable poverty trap and an intergenerational elasticity of 1 across the wealth distribution for $[w_p, w_u]$. Even if there is one sticky range, high loss aversion (η or λ), can also take over the entire vicious cycle. Again, it's important to note that this doesn't change the underlying fact a poverty trap exists, it just makes it nearly unobservable.

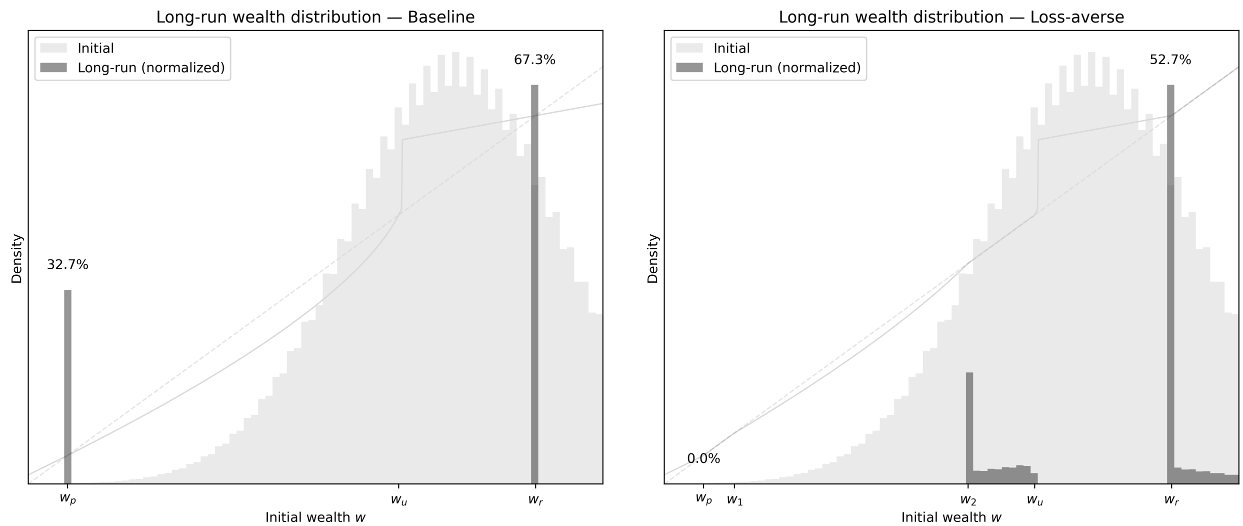
The collection of these long-run distribution effects can be summarized by the stationary distribution with loss-averse preferences, the conditions and results of which defined in Proposition (6). Instead of two steady states, the stationary distribution must sum across the wealth levels within the ranges of unstable steady states and only range of wealth that convergences downward to a poor steady state are those poorer than the new observable, weaker Micawber threshold.

Proposition 6 (Stationary Wealth Distribution with Loss Averse Bequests)**.** *Given the same assumptions in Proposition (5) and agents has loss-averse preferences over bequests, then the stationary distribution can be defined as the dynasties that converge or that are contained in the sticky ranges of wealth:*

$$G_\infty = G_0(w_p)\,\delta_{w_p}\ + ...$$

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

Sensitive to the initial wealth distribution, the results of the updated stationary wealth distribution in Proposition (**??**) can have dramatic long-run results. In Figure (6b), I show the stationary wealth distribution predicted by the baseline bequest function in Proposition (2) and the loss-averse bequest function in Proposition (4). As an example of how stark the long-run distirbutions can be, I parameterize $G_0(W)$ as a normal distribution centered within the area of mobility (w_u to w_r). In the baseline bequest function, the stationary distribution admits that nearly 1/3 of the population will converge to abject poverty and 2/3 to a higher wealth. In the loss averse model, nearly 0% of agents converge to abject poverty. Agents are stuck in the higher range of sticky wealth, unable to enter the mobility region but also not falling deeper into poverty.



(a) Long-run Wealth Distribution from Bequest Function from Propn (2)

(b) Long-run Wealth Distribution from Bequest Function from Propn (4)

Figure 6: Example Initial and Final wealth distribution from Bequest Functions in Figure (3)

Figures (6) show a greater dispersion of wealth in the loss averse model than predicted by the baseline bequest function. In order to obtain a richer wealth dispersion typical poverty trap models must assume different economic mechanic steps, like in Banerjee and Newman [1993] which have different occupations. Without having to impose new or complex economic structures, loss aversion generates a short- and long-run wealth dispersion much more similar to what a researcher would observe in data.

Loss-averse preferences in poverty traps not only *changes* the stationary distribution, it also *slows* the transition itself. The math is fairly intuitive. An $\alpha > \gamma$ increase in bequest propensity means $b'_L(w)$ is closer to 1 (outside sticky regions, when it is 1) than $b'_b(w)$. As $\alpha \uparrow$, $b_L(w) \to 1$. If $b_L(w) < 1$ (w not in a sticky range), at any given initial wealth level $w_0 < w_2$, significantly more periods must pass to reach a lower wealth level than the baseline model predicts. Put differently, after k periods, the wealth level of a dynasty that feels loss averse over bequests will be higher than one that does feel the loss. This slows the transition function and can be seen graphically in Figure (7). After 5 periods, the initial wealth w_0 travels to w_λ with the loss averse model and to w_γ without loss aversion. These states are dramatically different. w_γ has traveled nearly 85% the distance to the higher poor steady state at w_2. But w_λ has only traveled about a third of that distance. Of course, the baseline bequest function will have that agent at w_0 converge to w_p and the difference speeds toward convergence will tends to infinity since w_0 will never converge to w_p under loss aversion.

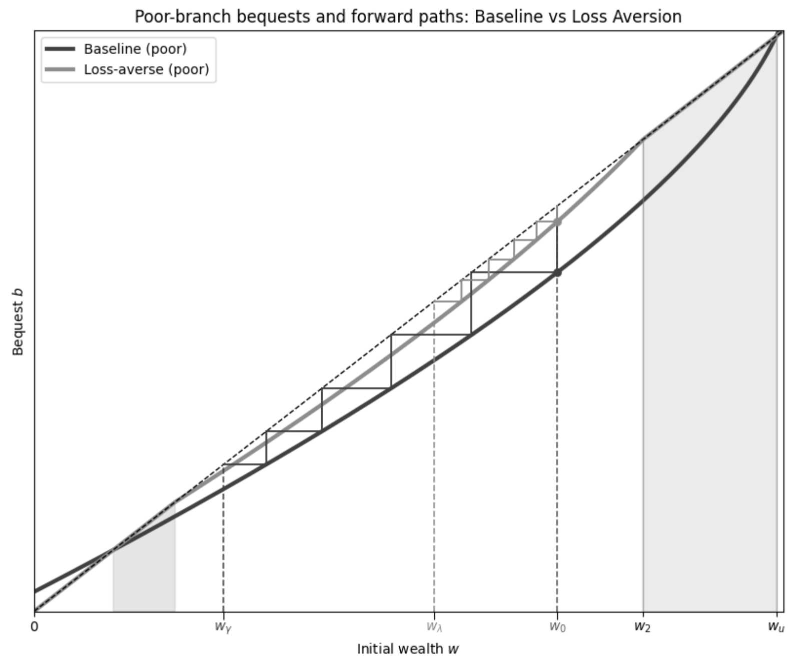


Figure 7: 5-period transition in initial wealth

This can behavior can be formalized for potential empirical and simulation investigation in Corollary (1). The half-life of a given w_0 is simply number of periods it takes to be half-way to between w_0 and w_p (or the converging steady state). Visually, in Figure (7), we can tell it is fairly quick (in theoretical terms) for the baseline model, with a half-life of about 3 periods. The estimated half-life under loss aversion is about 10 periods. 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 (3), for $w_t < w_2$, define the half-life to the poor steady state, w_p, as $H_i = \frac{\ln(1/2)}{\ln(b_i'(\underline{w})}$ where i refers to the baseline model B or the loss-averse model L. Then,*

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

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

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.*

In Corollary (1), the half-life of the mechanical poverty trap can be implicitly solved for numerically given the exogenous parameters are known and the half-life of the observed transition data can be estimated. The half-life under loss aversion goes up as the ratio of between the propensity at a loss and at a gain goes up, or when the baseline half-life increases.

4 Econometric Implications

In this section, I explore how loss aversion can help explain why data limitations and empirical methods often fail to identify poverty traps, and offer unique a method to test the presence of a poverty trap by identifying behaviors only rational and consistent in the presence of a poverty trap. First, I provide an overview of the discourse on the data necessary to identify poverty traps and the empirical models researchers have employed. Then, I offer simulations that reveal how loss averse preference can cause the murky data and failure of empirical models though a poverty trap is present using two different empirical methods. Due to the ranges of sticky wealth and slower convergence, loss aversion makes identifying the poverty trap dynamics — which are what researchers look for — unreliable. However, both the kink-point characteristics of loss aversion and the reference-dependent bequests allow us to utilize shocks to examine differences in bequest and consumption propensity. In the presence of a trap without loss aversion and no trap, indiviudals from a higher and lower income shocked to an income in the middle should have the same propensity to bequest. But if the initially richer agent develops a stronger propensity and the poorer a lower propensity, this behavior is only possible in the context of both loss aversion and vicious cycles.

4.1 Benchmark Empirical Tests

Proving the existence of a poverty trap using income and wealth data has been difficult and divisive. The seminal refutation of the existence of most poverty trap hypotheses by Kraay and McKenzie [2014] find it incredibly rare to empirically uncover 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 data richness, but also that short panel data often do not contain enough wealth and income transition periods that can definitely show the vicious cycle play out, especially with preference heterogeneity and large enough stochastic errors.

Even if there is ample and diverse panel data, empirical models might still struggle to find evidence of a trap due to the identification strategy. The most basic strategy is called the threshold test. As Barrett and Carter [2013] discuss, the threshold test empirically identifies the existence of a threshold such that the average change below is negative and above is positive, which could indicate that the threshold is a Micawber frontier. Still, while this test is a more general framework, it is difficult to identify the dynamics beyond a threshold presence. Before this test was widely accepted, researchers relied on testing for the S-shape curve like in Figure (5) which painted a better picture of the inter-period dynamics. For example, Banerjee et al. [2019] empirically identify a potential S-shaped return function from smaller to larger micro-finance loans, where those that cannot gain access to large enough loans cannot purchase the high-investment production technology. But, as discussed by both Kraay and McKenzie [2014], both approaches are seldom conclusive and more often than not, these tests inform researchers a poverty trap does not exist.

The remainder of this section tests the two common identification strategies in the loss averse model and without, showing how these empirical strategies — given very rich and sufficient data — can identify poverty traps but fail to when loss aversion is present.

Threshold Test. Using a Monte Carlo simulation, I compare the efficacy of the threshold test to identify a poverty trap in the loss averse model and without. I find that loss aversion significant distorts the identification of divergent trends this test would otherwise predict. The basic idea of the threshold test is that there are two lines of the average change in wealth discontinuous around a threshold c that better estimates the model's overall change in wealth than a single line. We can represent these lines of average wealth change as the lines higher than c as $g_h(w) = w_{t+1} - w_t$ and lower as $g_l(w) = w_{t+1} - w_t$. I design the empirical model in Equation (8) based on the framework developed by Hansen [2000]. Here, β_0 represents 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 $g_l(w) = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)w$. 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, representing 5 generations.

$$\text{(T1)} \quad \Delta w_{i,t+1} = \beta_0 + \beta_1\, w_{i,t} + \beta_2\, \mathbf{1}\{w_{i,t} < c\} + \beta_3\big(w_{i,t} \cdot \mathbf{1}\{w_{i,t} < c\}\big) + \varepsilon_{i,t}. \tag{8}$$

Following Barrett and Carter [2013], to identify the presence of a poverty trap using the threshold requires examining the behavior around the estimated threshold, $\hat{c}$. If wealth change for the lower trend is negative the the threshold, $g_l(\hat{c}) < 0$, and the change for the higher level is positive $g_l(\hat{c}) > 0$, then we can show a bifurcation between these two trends. To identify if this is a Micawber threshold where those lower and higher converge to different steady states, we would need to see the slope of the lower line trend downward: $(\beta_1 + \beta_3) < 0$. This implies that not only is there a bifurcation point, but that those below the bifurcation point trend to a lower state. If the slope for the higher line, β_1, is positive, this is evidence for a two-state poverty trap. But the sign of β_1 is less important. This entire estimation process is sensitive to the distribution. If nearly the entire distribution is clumped within the area of mobility, it becomes harder to identify a poverty trap. The estimated $\hat{c}$

The simulation results from Table (3) corroborates the hypothesis that loss aversion might explain the obfuscation of a poverty trap in the data due to strong endogenous, counter-cyclical decision making. It is important to note that the data in this simulation is best-case scenario, where intergenerational initial wealth is free of any heterogeneity and shocks are relatively small. The time horizon is also very large compared to most panel data past papers which rely on 1-2 generations or 3-5 time periods [Kraay and McKenzie, 2014].

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)
Baseline	0.8477	+0.7928	-0.6850	-0.7952	+0.5453	-0.1397
Loss-averse	0.8418	+0.2381	-0.1734	-0.2399	+0.1782	+ 0.0048

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: "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 A, we find a stark decrease in the coefficient estimates under loss aversion which emit from the slow and sticky convergence. In both the model with loss aversion and without, the threshold is about the same. However, across all coefficients, the loss averse model predicts less than half of the value of the baseline, though typically preserves the sign. This means that the loss averse model predicts a much less severe trap than what actually exists: the change in inter-period wealth is much slower under loss aversion. This is exactly the result of Corollary (1), in which the half-life significantly slows (as do all transitions), and to observe a transition from a higher wealth to a lower wealth can over double the number of periods than in the model without loss aversion. In panel data, this means that to identify a poverty trap, shorter panels might not tell us much. Especially with more noise and less fine and granular wealth data, the loss averse coefficients will decline even further potentially indicating no or even positive change around the threshold. Importantly, the test with loss aversion also fails to identify diverging trends. Under loss aversion, the average change in wealth below $\hat{c}$ is positive, $\beta_1 + \beta_3 = +0.0048$. This implies that overall, those below have an increase in their inter-period wealth. Even if it was not positive, it is also very close to 0, implying little to no change. In a poverty trap, we would expect this to be strongly negative, as we do in the baseline case. This alone does not confirm nor reject a poverty trap.

What allows a researcher to fail to reject the null hypothesis (that a poverty trap doesn't exist) is that under loss aversion in Panel B, the lower change in wealth line at the cut-off is greater than 0, $g_l(\hat{c}) > 0$, implying that this is not a Micawber threshold. Since it is positive, and close to 0, this empirical findings would predict simply very slow convergence out of poverty, where at that threshold they grow much faster, rather than downward convergence. Especially since the slope of the average change for g_l is positive and close to 0, a researcher could confidently fail to reject the null. Recalling Corollary (1), If loss aversion was present, the same threshold would not only predict a bifurcation, but that one trajectory is negative and the other positive, which would confirm a poverty trap.

I employ a robustness check of the identified thresholds using bootstrapping, which identifies the $\hat{c}$ that maximizes the F-stat for a given guess of c. This reveals that both previously identified thresholds are statically significant within the 99.9975 percentile; we can be quite confident these are correct. While this is a somewhat trivial exercise in a Monte Carlo simulation, 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 2: T1 threshold existence (sup-F across c; wild bootstrap p)

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

What is critically important is that a researcher might confidently and accurately predict the presence of a threshold, and confidently predict there is no poverty trap. These sticky ranges and slow convergence can strongly maintain the wealth distribution. Especially in the region of the poverty trap, an empirical technique might find inconclusive evidence of negative or any change in wealth. Above the poverty trap, they might identify upward mobility and some downward rich convergence. These results are sensitive to the initial distribution of wealth, but the efficacy constraints from the influence of loss averse preferences remain the same.

S-shaped Curve. An alternative method, while much harder to get sufficient data to identify, allows researchers to more confidently identify an S-shaped curve if present. In my empirical estimation model, I fit the data to a parametric cubic polynomial to estimate 3 points where the inter-period change in wealth crosses the 45-degree line, as seen in Equation (9). This model was popularized by Antman and McKenzie [2007] and is the framework for this current analysis. This model identifies different roots of the model, where a w^* is a root is $\Delta(w^*) = 0$. To evaluate stability, we examine if the derivative of Equation (9) is positive or negative. If $\Delta'(w^*) = \beta_1 + 2\beta_2 w^* + 3\beta_3 (w^*)^2 < 0$, it is stable and if it is > 0 it is unstable (as can visually be determined by Figure (1b)). I run a basic OLS regression on my model, but shorter panels, income data, non-parametric data as well as other confounding data issues require different empirical approaches as described in Antman and McKenzie [2007].

$$\text{(S2)} \quad \Delta w_{i,t+1} = \beta_0 + \beta_1\, w_{i,t} + \beta_2\, w_{i,t}^2 + \beta_3\, w_{i,t}^3 + \varepsilon_{i,t}. \tag{9}$$

To confidently reject the null hypothesis a poverty trap does not exist, we would expect to find two stable steady states, one poorer and one richer, and in the middle a unstable steady state, the Micawber threshold. As we see in Table (**??**), the cubic fit does find evidence of a poverty trap but it is much less severe than without loss aversion. The roots for the poor (w_p), non-poor (w_r), and the Micawber threshold (w_u) exists, but are not the true steady states as can be identified in the baseline. As predicted by my model the poor and rich steady state is 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 (the true threshold) to 0.4599 (the observable threshold), results that follow from from Proposition (**??**).

Table 3: S2: S-shaped regression

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

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 New Methods

In their follow up paper, Barrett and Carter [2013] encourage researchers to identify behaviors in the data that are only rational within a poverty trap. The constraints of panel data and the limitations of empirical models might be circumnavigated by identifying behavior that is 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. Alternatively, the traumas of impoverished areas, due to high rates of crime or warfare, might cause the poor to be overly risk averse. Their paper teased at the underlying statement of mine: the behavior of agents is endogenous to their presence in the poverty trap, and we can identify that behavior. In this section, I offer a unique behavioral trait and identification technique using the loss-averse model of bequests.

The main empirical identification contribution comes from the effect of reference-dependent bequests. Since agents are reference-dependent over their *initial wealth*, idiosyncratic shocks to their income do not change that reference point. Inside the poverty trap, there is a unique behavior. A richer parents who receives a lower income than expected due to a shock, regardless of where on the distribution they are, will feel loss averse. A poorer parent who received a positive income shock will always feel a gain. Inside the poverty trap, we expect all agents will feel loss averse. But simply sampling who in the poverty trap range feels a loss without any other context does not allow us to separate loss aversion in response to a poverty trap compared to heterogenity in preference or risk aversion. But if we look at individuals who are shocked to earn the same income, their propensities will always be different within a sticky range of wealth.

The main case is when a richer agent is shocked down, a poorer agent is shocked up, and they earn similar incomes. We can see this in Figure (8). If agents did not feel loss averse, then their propensities should be the same. If agents feel loss averse, but are not in a poverty trap, then we'd expect the propensity to bequath between someone who started out at $\underline{w}$ to be nearly the same as the positively-shocked individual at w' (they both feel a gain) while w'' bequeath with a very high propensity. If the agents are in a poverty trap and feel loss averse, then agents at both $\underline{w}$ and w'' will bequest at a higher propensity than w'.

$\underline{w}$ will bequest with propensity α, and w' with γ.

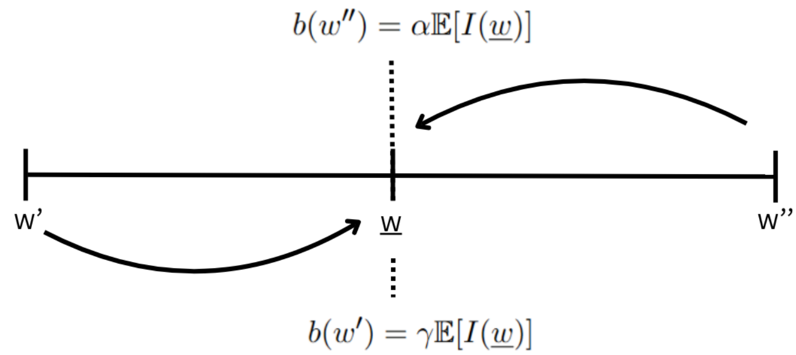


Figure 8: Diagram of bequest shocks that synchronize two agents' income but not their preferences.

In the data, we can idetify a tight range and classify agents who had an upward shock, a downward shock, and nearly no shock then compare the bequest (or consumption) propensity differences. by finding many instances where two agents are shocked up and down into a small range of each other, we can compare the aggregate changes in propensity to bequest and consume. In areas where the propensity of the richer agent better matches the propensity of those who start and stay in that range ($\underline{w}$) and the poorer agent has a lower bequest propensity, there might be cause for concern. This implies that agents at $\underline{w}$ are already feeling a downward current from the vicious cycle and if they have a similar propensity as someone already shocked down. But if an agent has similar propensities as someone shock up to them, then they aren't feeling a downward pull. In the range of upward mobility, which is the null hypothesis of poverty traps, even richer agents shocked down might not need to change their bequest propensity to still match or exceed their threshold. Essentially, the reference dependence allows us — using three data points — to potentially find where vicious cycles might be occurs, evidencing a potential poverty trap.

In the data, if we collect these small intervals of shock-income convergence and aggregate the comparisons over time, we could potentially identify the scope of the poverty trap. If there are ranges where vicious cycles indicators occur, and both a certain threshold they stop mattering, according to the model this would be because these agents cross the Micawber threshold: only agents pushed very far down would change their bequest propensity.

Furthermore, the loss aversion shows that it is not just necessary to look at the transmission or accumulation of wealth, but also it might be useful to look at present *consumption.* In this model, there is a difference between an agent who has very little wealth but had a nice income shock, versus an agent who has a lot of wealth but no (or negative) income shock. When agents are loss averse to bequests, we'd expect that positive (negative) shock would increase (decrease) their propensity to consume. This assumption that consumption and bequest propensities rise and fall countercycically means we should be able to see that reaction in the data. If wealth and income data is spare, but consumption data is plentiful, consumption changes in propensity might offer a decent stand-in. 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).

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 bequest alone is 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 will need to balance these opposing forces.

The typical policy response to a poverty trap is a "big push" where a benefactor endows agents below the Micawber threshold with just enough wealth subsidies to push them above and into a state of mobility. This policy response faces significant limitations. First, this policy relies on identification of a poverty trap and the true Micawber threshold. 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, 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 reliably estimate. If a policy planner fails to correctly identify the true ranges of the poverty trap, then even a targeted big push policy would be inefficient or ineffective.

Second, if we assume these policies do not rely on foreign aid but instead an economies own resources to alleviate poverty, then the poverty trap and the optimal policy are extremely sensitive to one another. Foreign aid is an exogenous, large lump-sum gift and all a social planner needs to do is allocate those resources efficiently, hopefully target the mechanical root of the poverty trap. But if agents are taxed, then the policy planner's income is endogenously determined and is very sensitive to both the current wealth/income distribution, and the bequest function. As will be discussed in this section, even targetted policy plans could fail and social planners within a poverty trap must be very delicate.

In this section, I used the toy model from Proposition (4) and I begin with a benchmark social welfare function and policy response using a universal flat-tax and rebate scheme to simulate the effects when a poverty trap is not detected. While this policy itself is an unrealistic response from a policy planner, it allows us to understand how empirical identification and different policy levers effect present and long-run outcomes. After establishing the benchmark policy results, I transition to more realistic intervention policies and compare the limitations and benefits of each policy.

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 (4). Agents not face a flat-tax rate τ^* on new investment income $(1-\tau)L(R-r)$ 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: $(1-\tau)LR - Lr + wr + T$. 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: $[1-\pi L]LR$. 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:

$$(1-\tau)LR - Lr + w \cdot r + T \geq [1-\pi L] \cdot (1-\tau)LR \tag{10}$$

Rearranging Equation (17), 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 L^*, an agent must have $w^* = \frac{L^*(R\tau+r)-(L^*)^2(\pi R)-wr-T}{r}$. Intuitively, as the tax rate goes up, so does the required wealth for first-best borrowers $L(\tau R + r) > Lr$ 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 $T - L^*R\tau < 0$. . For non-first best borrowers, the lending function for those with $w < w^* =$

$$L(w,\tau,T) = \frac{\tau(R-r)}{2\pi R} + \frac{r - \sqrt{(\tau(R-r)+r)^2 - 4\pi R(wr+T)}}{2\pi R} \tag{11}$$

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 (18), the bequest function when agents feel a gain is:

$$I(w,\tau) = \begin{cases} L^* \cdot (1-\tau)(R-r) + w \cdot r + T, & \text{if } w \geq w^* \\ L(w,\tau) \cdot (1-\tau)(R-r) + w \cdot r + T, & \text{if } w < w^* \end{cases} \tag{12}$$

By combining the above income in Equation (19) 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 (4) is that net capital gains are taxed, resulting a decrease in the marginal returns of investment, also reducing the optimal capital L^*.

Aggregate capital in the economy can be written as $\mathbb{L}$, 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, $\mathbf{L} = \int_{w^*}^{\bar{w}} L^* g(w)\partial w + \int_{\underline{w}}^{w^*} L(w,\tau) g(w)\partial w$. This can be reduced and utilized in Equation (20):

$$\mathbb{L} = L^* \cdot [1 - G(w^*)] \; + \; \mathbb{E}[L(w,\tau) \mid w \in [\underline{w, w^*}]] \tag{13}$$

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, $T = \tau\mathbb{L}(R - r)$. 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 (20), we can define the social planner's problem:

$$\mathbb{W} = \int_i g(w^i) \cdot U^i(w^i,\tau)\partial w^i = \sum_{(a,b)} \int_a^b g(w^i) \cdot U^i_j(w^i,\tau)\partial w^i \tag{14}$$

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

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

Given the social planner's welfare function in Equation, we can be write an implicit function for the optimal tax using the framework fromPiketty and Saez [2013]. Before defining the

in Proposition (7)

Proposition 7 (Implicit Optimal Tax)**.**

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

where $\bar{G} = \frac{1}{\mathbb{L}} G$*, and* G *is the normalized social marginal weight:*

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

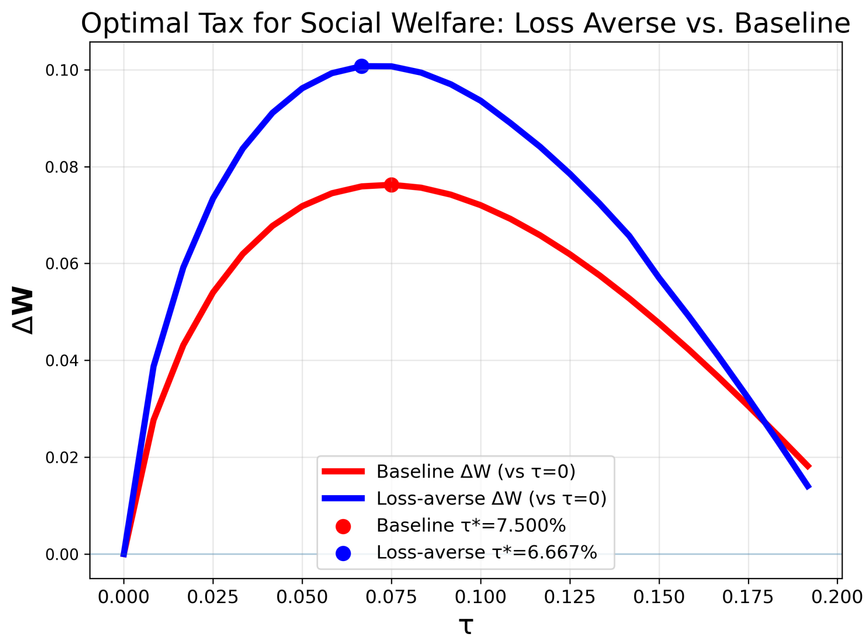
I define $e = \frac{1-\tau}{\mathbb{L}}$

Proof. See Appendix (**??**) $\square$

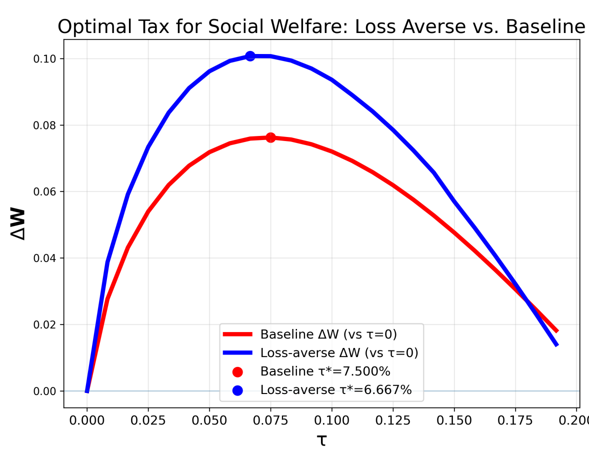
In Proposition (7), the net outcome from the equity-efficiency trade-off is very sensitive to the distribution. The mechanical effect to the tax is the loss in net capital income and the gain from the rebate. For agents with wealth higher than the rebate, this is a negative mechanical effect; for those above it is positive. Since utilities is logarithmic, the social planner might still implement a tax to support those with very low wealth and income. 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 (17) 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.

Loss aversion effects this optimal tax rate by up-weighting those feeling a loss, though this is dependent on which part of the distribution more feels a loss. As seen in G, those that give and 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.

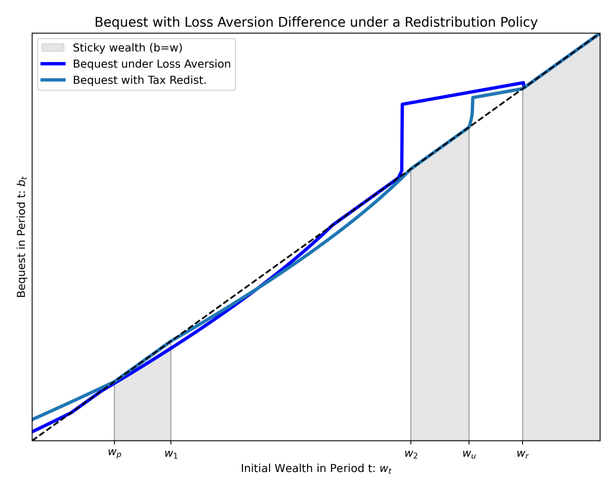
Though loss aversion changes the weights of the distribution, ultimately the distribution itself determines the optimal tax. As seen in Figure (**??**), the change in the bequest with and without loss aversion hardly changes the optimal tax. This means a social planner cannot ex-post determine if there 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



(a) Baseline bequest with redist. policy



(b) 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

Now let's assume a social planner is able to identify a poverty trap. Through the example in S-shaped identification results in Section (5), the steady states might be misidentified, which can cause insufficent policy. Since the typical poverty trap policy is a big bush, pushing agents above $w_2 < w_u$ does not actually push these agents into an upward cycle of mobility as we might expect. This might explain why Kraay and McKenzie [2014] did not find small interventions from various studies sufficient to save many of the poor from persistent poverty. If reference-dependent behavior reduces the observed Micawber threshold, it reduces the target a policy planner is aiming to push agents over. While this keeps the tax rate lower in the present period, it requires a higher continued tax in future periods in order to consistently save people as they keep falling back into the trap due to sensitivity to shocks.

5.3 Alternative Policy Interventions

I describe in qualitative terms the potential effects of alternative policy interventions funded through the tax rate. For each type, unless otherwise noted, I assume all individuals are taxed and rebated a lump-sum to allow for discussion on a policy's unique attributes. Mixing these policies would generate combined or heightened effects discussed in each.

The "Big Push".

As described at length by researchers of poverty traps (e.g., [cite barret]), the typical prescribed to poverty traps is that of a "big push": a heterogeneous income supplement, $L(w_t)$ that raises those below the Micawber threshold just above it $(w_t + L(w_t) - \epsilon > w^*)$. Then, without further policy intervention, all those previously trapped in the vicious cycle enter a virtuous wealth cycle.

However, the ability of a social planner to lift individuals in poverty above the Micawber threshold is constrained. For one, the distribution of taxable income matters. For distributions concentrated closer to poverty, even a 100% tax might fail to lift everyone above the Micawber threshold and in-fact could worsen the impact of the poverty trap. This is because the rebate is small compared to net-taxed income, negatively effecting the threshold to obtain the first-best level of capital. Time horizons also matter, because if a social welfare planner fails to get everyone out of the poverty trap, worsening the poverty trap itself, this lowers the

In the presence of loss aversion, a social planner might tax too low, bringing individuals to the observed Micawber threshold (first sticky steady state) rather the true Micawber threshold. Subsequent wealth distributions would contain a greater wealth dispersion and greater poverty than the big push predicts.

Targeted Transfers.

A targetted transfer would specifically target those in poverty, which depends on the poverty treshold. If a social planner can observe the presence of a poverty trap, they would simply

Rawlsian Weights.

6 References

References

Emily A Shrider and John Creamer. Poverty in the united states: 2022. *US Census Bureau*, 12, 2023.

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

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.

Partha Dasgupta and Debraj Ray. Inequality as a determinant of malnutrition and unemployment: Theory. *The Economic Journal*, 96(384):1011–1034, 1986.

Jyotsna Jalan and Martin Ravallion. Geographic poverty traps? a micro model of consumption growth in rural china. *Journal of applied econometrics*, 17(4):329–346, 2002.

Lisa A Gennetian and Eldar Shafir. The persistence of poverty in the context of financial instability: A behavioral perspective. *Journal of Policy Analysis and Management*, 34(4):904–936, 2015.

David Chivers. Success, survive or escape? aspirations and poverty traps. *Journal of Economic Behavior & Organization*, 143:116–132, 2017.

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.

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

Matthew Desmond. Poverty, by america, 2023.

Ellora Derenoncourt, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick. Changes in the distribution of black and white wealth since the us civil war. *Journal of Economic Perspectives*, 37 (4):71–89, 2023.

Scott W. Allard. Poverty traps. *Pathways*, 2019.

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

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979. URL `http://www.jstor.org/stable/1914185`.

Seth Blumberg and Michael Kremer. Dynamic loss aversion, growth, and development. *Unpublished manuscript*, 2014.

Yann Bramoullé and Christian Ghiglino. Loss aversion and conspicuous consumption in networks. 2022.

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.

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.

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

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.

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.

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

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

[TO BE COMPLETED]

A.1 Proposition (4)

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

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

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

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

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

A.2 Reference-Dependence Anchors

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.3 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:

$$\max_{b^i} U(c^i, b^i) = (1-\gamma)\ln(c^i) + \gamma \ln(b^i) + \eta \cdot \gamma\nu(\ln(b)|\ln(w)) + \eta \cdot (1-\gamma)\nu(\ln(c)|\ln(0^+)) \tag{15}$$

$$\text{s.t.} \quad c^i + b^i = (1-\tau)L(w^i, \tau)(R-r) + w^i \cdot r + T \tag{16}$$

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

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

$$(1-\tau)L(R-r) + w \cdot r + T = [1 - \pi L] \cdot RL \tag{17}$$

Rearranging Equation (17), 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 L^*, an agent must have $w^* = \frac{L^*(\tau(R-r)+r)-(L^*)^2(\pi R)-T}{r}$. Intuitively, as the tax rate goes up, so does the required wealth for first-best borrowers $L(\tau(R-r)+r) > Lr$ 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 $T - L^*R\tau < 0$. For non-first best borrowers, the lending function for those with $w < w^*$ is

$$L(w, \tau, T) = \frac{\tau(R-r)}{2\pi R} + \frac{r - \sqrt{(\tau(R-r)+r)^2 - 4\pi R(wr+T)}}{2\pi R} \tag{18}$$

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 (15) and the income function in Equation (18), the bequest function when agents feel a gain is:

$$I(w, \tau) = \begin{cases} L^* \cdot (1-\tau)(R-r) + w \cdot r + T, & \text{if } w \geq w^* \\ L(w, \tau) \cdot (1-\tau)(R-r) + w \cdot r + T, & \text{if } w < w^* \end{cases} \tag{19}$$

By combining the above income in Equation (19) with the utility function in Equation (15), we can rederive the updated bequest function. Instead of rewriting the long bequest function here, note that the only change from Proposition (4) 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 $\mathbb{L}$, 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, $\mathbf{L} = \int_{w^*}^{\bar{w}} L^* g(w) \partial w + \int_{\underline{w}}^{w^*} L(w,\tau) g(w) \partial w$. This can be reduced and utilized in Equation (20):

$$\mathbb{L} = L^* \cdot [1 - G(w^*)] \; + \; \mathbb{E}[L(w,\tau) \mid w \in [\underline{w}, w^*]] \tag{20}$$

Given aggregate capital in Equation (20), the rebate each agent receives in $T = \tau \mathbf{L}(R - r)$, 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.

$$\mathbb{W} = \int_i g(w^i) \cdot U^i(w^i, \tau) \partial w^i = \sum_{(a,b)} \int_a^b g(w^i) \cdot U^i_j(w^i, \tau) \partial w^i \tag{21}$$

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

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

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 $\partial L(w,\tau,T)/\partial\tau$ as implied by (18), and $\mathbb{L}_\tau = \int_{\underline{w}}^{w^*} L_\tau(w,\tau) g(w_i) \partial w_i$. This allows us to take the derivative of each part of the income function in equation (19):

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

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

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 $L(w,\tau)$, 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.

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

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

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

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

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

We can define $e = \frac{1-\tau}{\mathbb{L}}\frac{\partial \mathbb{L}}{\partial(1-\tau)}$. 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.

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

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

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

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

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 (21). Much like the additional complications loss aversion creates to the bequestion function as seen in Proposition (4), 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 (21), 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.

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

Following this, we can

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

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

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.

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

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.

$$\tau^* = \frac{1 - G_L}{1 - G_L + e}. \tag{22}$$