Poverty traps in Markov models of the evolution of wealth

Blume, Lawrence E.; Durlauf, Steven N.; Lukina, Aleksandra

Blume, Lawrence E.; Durlauf, Steven N.; Lukina, Aleksandra

Working Paper

Poverty traps in Markov models of theevolution of wealth

WZB Discussion Paper, No. SP II 2020-303

WZB
Wissenschaftszentrum Berlin
für Sozialforschung

Lawrence Blume
Steven Durlauf
Aleksandra Lukina

Poverty traps in Markov models of the evolution of wealth

Discussion Paper
SP II 2020–303
February 2020

Research Area
Markets and Choice
Research Unit
Economics of Change

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Affiliation of the authors:

Lawrence Blume, Cornell University and IHS Vienna

Steven Durlauf, Harris School of Public Policy, University of Chicago

Aleksandra Lukina, University of Chicago

Abstract

*Poverty traps in Markov models of the evolution of wealth**

Poverty trap models are dynamical systems with more than one attractor. Similar dynamical systems arise in optimal growth and macroeconomic models. These systems are often studied empirically by ad hoc methods relying on intuition from deterministic systems, such as looking for multiple peaks in the stationary distribution of states. We develop Markov wealth processes in which parents' investments in children stochastically determine children's wealth, and consequently their own investment choices. We show that, relative to a zero-shock process, some of the multiple attractors are less fragile than are others, and that their presence dominates the stationary behavior of the wealth distribution. Typically, mass accumulates around attractors. An only slightly stochastically perturbed deterministic system will have an invariant distribution which puts close to probability 1 on a single steady state rather than having significant mass distributed among several attractors. We also examine how policy effects the shape of the invariant distribution.

Keywords: Poverty traps, wealth distribution, wealth mobility

JEL classification: C6, D3

* Supported in part by the Wissenschaftszentrum Berlin für Sozialforschung and by a grant from the Carnegie Corporation of New York to the University of Wisconsin Center for Russia, East Europe, and Central Asia.

1 Introduction

A poverty trap is a "self-reinforcing mechanism which causes poverty to persist."[1] The persistence of poverty in aggregate data does not require the existence of traps that capture generations of families; poverty in aggregate data can be consistent with family mobility up and down the socioeconomic ladder. The degree of immobility is important, however, because intergenerational persistence requires different melioration strategies than does transitory poverty.

Our concern is with poverty-trap models in which not just the magnitude but the direction of motion depends upon initial conditions. Were dynamics deterministic, such models would exhibit multiple steady states. That more than one steady-state can be an attractor is the intuition behind the idea of a poverty trap.[2] Our concern in this paper is to examine how poverty traps appear in models with random shocks. Deterministic dynamics in poverty trap models are not ergodic; they have multiple attractors. Markov models of the evolution of the wealth distribution will typically be ergodic. The literature loosely associates poverty traps with multimodal invariant distributions of wealth. Often, these modes are near the attractors of a deterministic mean or zero-shock dynamic. We will show that this intuition is not always correct; typically there is a principal attractor around which most of the mass settles, with smaller blips in the vicinity of other attractors.

In deterministic poverty-trap models, policy can work in two ways: First, it can ameliorate low-level steady states, thereby creating a wealth distribution with poor and rich, but in which the poor are better off. Sometimes the shift in the dynamical system can be so great that the system passes through a bifurcation, and an attractor disappears. Second, it can shrink the basin of attraction of the low-level attractor. In deterministic models, almost regardless of starting point, any family ends up at one attractor or another. Where a family ends up depends upon which basin of attraction they start in. Policy can ameliorate low-level outcomes in stochastic models too. But unlike in deterministic models, in stochastic models there is continual motion; family fortune fluctuates from rich to poor and back again. Where in the distribution of wealth a family spends most of its time is described by the invariant distribution.

[1] Azariadis and Stachurski (2005).

[2] See surveys by Azariadis and Stachurski (2005) and Ghatak (2015) for descriptions of a variety of mechanisms that can lead to multiple steady-states.

In stochastic models, policy can also work by changing the shape of the invariant distribution, shifting mass from one attractor to another.

In the next section we introduce the model of family decisionmaking. Section 3 describes the resulting Markov wealth process. In section 4 we discuss the stationary distribution of wealth, the invariant distribution. In particular, we relate the shape of the distribution when noise is small to the behavior of the deterministic limit. In section 5 we apply the results of section 4 to describe the effects of some simple policy experiments on the shape of the invariant distribution. Section 6 concludes. Proofs are found in Appendix A. Appendix B. contains a technical discussion of the consequences of the non-convexities which arise naturally in our model.

2 The Model

The model contains overlapping generations of parents and children, one each per generation. We track the wealth of a single dynasty — an infinite sequence of one-parent, one-child families. Each adult produces one child. A parent allocates her wealth between a single consumption good and investment in her child. Following the benchmark Becker and Tomes (1979) model, we assume that a parent cares about the child's wealth, but not the child's utility.[3]

We will study the stationary distributions of Markov poverty trap models by considering not one model but a class of models parametrized by their large deviation rate functions. The behavior of these models in the low noise regime will be informative both about the shapes of invariant distributions in general and what can be learned from their deterministic counterparts. Our model is decribed by a structure $\langle U, F, C, \{\mu^\varepsilon\}\rangle$ for $\varepsilon \geq 0$. Parents' consumption sets are $\mathbb{R}_+ \times \mathcal{W}$, where $\mathcal{W} = [0, w^*]$ is a bounded interval.[4] An element (c_t, w_{t+1}) describes parent consumption today and the wealth of her adult child tomorrow. The parental utility function is $U : \mathbb{R}_+ \times \mathcal{W} \to \mathbb{R}$. The relationship between parent's investment today and child's wealth tomorrow is described by the function $F : \mathcal{K} \times \mathbb{R} \to \mathbb{R}_+$; $F(k_t, s_{t+1})$ gives the wealth of the adult at time $t + 1$ as a function of the amount k_t invested in her as

[3] The alternative is Loury's (1981) recursive utility model, which is not materially different for our purposes.

[4] The assumption of unboundedness of resources is technically difficult and economically unrealistic.

a child and a stochastic shock s_{t+1}. It maps parental investment and a shock into child's adult wealth. The set $\mathcal{K}$ contains the feasible investments. Since wealth is bounded, there is no loss of generality in bounding the set of feasible investments as well: $\mathcal{K} = [0, k^*]$. The cost of an investment k in a child is $C(k)$. The function $C : \mathcal{K} \to \mathbb{R}_+$ may be non-linear, and even non-convex, and will be subject to policy manipulations in Section 5. Finally, $\{\mu^\varepsilon\}$ is a family of shock distributions on $\mathbb{R}$. An instance of the model fixes a particular ε, and the shock process $\{s_t\}_{t\geq 1}$ is iid with respect to this distribution. These distributions vary continuously in the weak topology in the parameter ε, which controls the variance of the noise. The distribution μ^0 is point-mass at 0, and is our reference deterministic case. The four components U, F, C, and μ^ε define a Markov chain with stationary transition probabilities, which arises from chaining together solutions to one-period decision problems, the utility maximization problem for each period's parent.

The decision problem of a date-t parent is to allocate wealth w_t between consumption c_t and investment k_t in her child so as to maximize her expected utility subject to a budget constraint. Let $V(c, k)$ denote parent expected utility from consuming c and investing k in the child. The optimization problem is:

$$\begin{aligned} &\max_{c,k} V(c, k) \equiv E_{\mu^\varepsilon} U\big(c, F(k, s)\big) \\ &s.t.\ c + C(k) \leq w,\ \ c \geq 0,\ \ k \in \mathcal{K}. \end{aligned} \tag{1}$$

From the solution to this problem, only the optimal investment policy is needed to describe the wealth dynamics:

$$\pi(w) = \{k \in \mathcal{K} : \text{for some } c \geq 0,\ (c, k) \text{ solves } (1)\}.$$

We assume the following about utility U and the production function F.

A.1. $U : \mathbb{R}_+ \times \mathcal{W} \to \mathbb{R}$ *is continuous, strictly increasing in* c*, and strictly concave in* c *for each* $w \in \mathcal{W}$.[5]

A.2. $U\big(c, F(k, s)\big)$ *is Lebesgue integrable with respect to the distributions* μ^ε *for any* c *and* k.

A.3. $C(k)$ *is continuous and strictly increasing in* k.

[5] Or, for computing examples, that it is Leontief.

A. 4. *$F(k, s)$ is non-negative, non-decreasing and continuous in k; and strictly increasing in s.*

A.5. *$U(c, w)$ has increasing differences in $c; w$.*

A.6. *The map $\varepsilon \to \mu^{\varepsilon}$ is continuous in the weak topology.*

Assumption 5 states that child's wealth and current consumption are complements.

Our assumptions guarantee the non-emptyness and upper hemi-continuity of the optimal policy correspondence π.

Theorem 1. *Suppose A.1–4 are satisfied. Then for all $w \geq 0$, $\pi(w) \neq \emptyset$. Furthermore, π is everywhere upper-hemicontinuous in parental wealth and ε. If, in addition, A.5 holds, then π is increasing in the following sense: If $w' > w''$ and if k' is optimal for w', k'' is optimal for w'', then $k' \geq k''$.*

The last part of the theorem demonstrates a strong form of normality for investment in children. This comparative statics result implies that π is usually singleton-valued.

Corollary 1. *If A.5 holds, the set $\{w \in \mathbb{R} : \#\pi(w) > 1\}$ is countable.*

We avoid strong convexity assumptions that would make π everywhere-continuous because some policies for poverty alleviation naturally introduce non-convexities. Nonetheless our exposition is greatly simplified by assuming throughout the body of the paper that π is a continuous function. In Appendix B we will extend our analysis to the non-convex case, assuming that the set of multivalued wealths is finite. Suffice it to say that our results will more or less continue to hold.

3 Markov Wealth Processes

3.1 The evolution of states

Figure 1 graphically depicts the stochastic process describing the evolution of wealth. The stochastic process $\{w_t\}_{t\geq 0}$ is Markov. A date-t parent with wealth w_t chooses k_t.

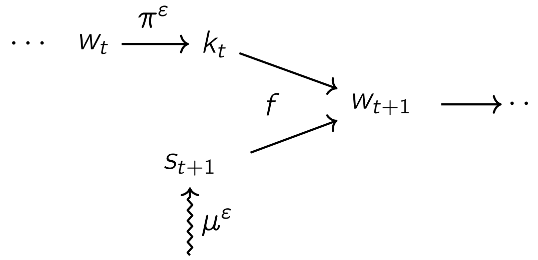


Figure 1: The stochastic evolution of wealth.

A random shock s_{t+1} is drawn from μ^ε. The investment technology produces w_{t+1} from these two variables.

We begin by describing the transition probability. We want to study how the magnitude of exogenous randomness effects the long-run behavior of the wealth process, so we introduce a family $\{\mu^\varepsilon\}$ of probability distributions that includes, at one extreme, a deterministic model. The following two assumptions describe the shock distribution.

B.1. *μ^ε has a density with respect to Lebesgue measure:*

$$\phi^\varepsilon(s) = Z(\varepsilon)\exp\left\{\frac{-h(s)}{\varepsilon}\right\}$$

where $Z(\varepsilon)$ is the normalizing constant such that $\phi^\varepsilon(s)$ integrates to 1 over s.

B.2. *$h(s)$ is non-negative; has a global minimum at 0 and $h(0) = 0$; is differentiable from the right and from the left at $s = 0$. For $s' > 0$, as $s \downarrow s'$, $h(s) - h(s') \sim \beta_1(s - s')^{\alpha_1}$. For $s' < 0$, as $s \uparrow s'$, $h(s) - h(s') \sim \beta_2(s - s')^{\alpha_2}$.*

One example of such a family is the class of Normal distributions with fixed mean μ and variance ε; take $h(s) = (s - \mu)^2/2$. Here $Z(\varepsilon) = 1/\sqrt{2\pi\varepsilon}$. In the next section we shall undertake computations with $h(s) = |s|$ and $Z(\varepsilon) = 1/2\varepsilon$, the family of Laplace distributions.

Finally, we assume that shocks are additive.

B.3. $F(k, s) = \max\{0, \min\{G(k) + s, w^*\}\}$ *where G is non-negative, non-decreasing, and continuous.*

This specializes assumption A.5 to an additive shock and a truncation. It takes a standard output function, adds an error term, and truncates if necessary at 0 and w^*. Thus the two endpoints of $\mathcal{W}$ will be atoms in the transition probability whose mass goes to 0 with ε. Note that assumption B.1 is significantly stronger then axiom A.6 because it implies variation-norm continuity for $\varepsilon > 0$.

The Markov wealth process $\{W_t^\varepsilon\}$ with initial condition W_0^ε is described by its transition probability

$$P^\varepsilon(w, A) = \mu^\varepsilon(F(\pi^\varepsilon(w), s) \in A). \tag{2}$$

For an interval $A = [0, w']$ and $\varepsilon > 0$,

$$P^\varepsilon(w, A) = \begin{cases} \displaystyle\int_{-\infty}^{w' - G(\pi^\varepsilon(w))} \phi^\varepsilon(s) ds & \text{if } w' < w^*, \\ 1 & \text{if } w' = w^*. \end{cases}$$

For any $w \in \mathcal{W}$ the transition probability has an atom at 0, an atom at w^*, and a density describing the process in between.

The Markov processes for $\varepsilon > 0$ are strongly ergodic. This can be seen as a consequence of the assumed continuity of the π^ε. In fact, our construction guarantees that the transition probabilities satisfy Doeblin's condition for any selection from the optimal policy correspondence, and so this result will continue to hold when the continuity assumption is relaxed in Appendix B.

Theorem 2. *For all $\varepsilon > 0$ the Markov wealth process is strongly ergodic. The map from $\varepsilon > 0$ to its invariant distribution is continuous in the variation norm.*

3.2 Deterministic wealth dynamics

In the noiseless case of pointmass at 0 ($\varepsilon = 0$), wealth evolution is deterministic. It is governed by a difference equation whose right-hand side is the *parental Engel curve:*

$$w_{t+1} = G(\pi^0(w_t)) \equiv e^0(w_t). \tag{3}$$

A continuous Engel curve will have fixed points. Even if the parental Engel curve is not continuous, it will have fixed points because it is increasing. Some of the fixed points are attractors for the dynamical system of iterates (cobweb dynamics), $w_t = e^0(e^0(\cdots(e^0(w_0)\cdots) = e^{0t}(w_0)$. In the remainder of this section we describe the long-run behavior of the deterministic system. In section 4 we contrast this behavior with the behavior of a Markov chain driven by this system with small stochastic shocks.

Definition 1. *A fixed point w' is an* attractor *if there is an open interval $U \ni w'$ such that $\cap_{t\geq 0} e^{0t}(U) = \{w'\}$ and for all open $V \supset U$, $e^0(U) \subset V$.*

A discontinuous dynamical sytem may have no singleton attractors, but B.4 guarantees that e has at least one.

B.4. *$e^0(w)$ has only a finite number of fixed points.*

We sum this up in the following theorem:

Theorem 3. *There is a $w' \in \mathcal{W}$ such that $w' = e^0(w')$. If in addition e^0 satisfies B.4, then e^0 has at least one attractor.*

When $e(w)$ has more than one attractor, the smallest attractor is a *poverty trap*. Equally, it could be said that the highest attractor is an *affluence trap*.

4 Attractors and Invariant Distributions

Theorem 2 shows that for all $\varepsilon \geq 0$, an invariant distribution will exist. A sharp picture of the invariant distributions emerges when we study stochastic processes with low shock variance. We can think of these processes as stochastic perturbations of the deterministic system, parametrized by ε. All non-trivially random processes considered here have unique invariant distributions, unlike the deterministic system for which point mass at any attractor is invariant. Nonetheless, the multiple attractors play a special role in determining the shape of the invariant distribution which can be seen when ε is small.

As ε becomes small, the step-by-step transitions of the process $\{W^\varepsilon\}$ look increasingly like the deterministic evolution of e^0 . This suggests that the invariant distribution for small ε will be concentrated at attractors. But beyond this, some attractors are favored over others. Typically there will be a unique attractor around which the invariant distribution will concentrate. This fact suggests that in cross-section or in short panels it will be hard to find poverty traps by looking for multiple modes. Our analysis is for small ε, but while a larger ε might make the selection effect less strong, it also makes the overall picture more noisy, so it is not obvious that searching for multiple modes will pay off.

The key idea of the analysis is that there is a most likely path to take in traversing from one state to another, and as ε becomes small, the most-likely paths become infinitely more likely than other paths. So the problem simplifies to comparing the relative likelihood of traversing these paths. Describing this requires some apparatus. Define $A_T : \prod_{t=0}^{T-1} \mathcal{W}$, such that for path $\bar{w} \in \prod_{t=0}^{T-1} \mathcal{W}$,

$$A_T(\bar{w}) = \sum_{t=0}^{T-2} h(w_{t+1} - e^0(w_t)),$$

for $w', w'' \in \mathcal{W}$ define

$$B(w', w'') = \inf\{A_T(\bar{w}) : T \geq 1, w_0 = w', w_{T-1} = w''\},$$

and for any two attractors w_i, w_j of $e^0(w)$, define

$$B_{ij} = B(w_i, w_j).$$

For intuition, A_T can be thought of as the cost of following path $\bar{w}$ from w_0 to w_{T-1}. A zero-cost path follows an orbit of $e^0(w)$, while deviations from an orbit are charged a positive cost — the less likely the deviations, the higher the cost. Then $B(w', w'')$ is the cheapest cost (over all paths) of traversing from w' to w''. Finally, B_{ij} is the cost of travelling from fixed-point i to fixed-point j.

The next construction is closely related to a means of constructing an invariant distribution for a finite-state Markov chain. Let Γ_n denote the set of all directed trees, each of whose vertices corresponds to a distinct attractor, and whose root vertex corresponds to attractor n. Edges are directed so that each leaf has a path to the root. For each graph $g \in \Gamma_n$ define $B(g) = \sum_{j \to k \in g} B_{jk}$, and for each attractor w_n define $B(n) = \min_{g \in \Gamma_n} B(g)$. Finally, let $\mathcal{B} = \{n : B(n) = \min_j B(j)\}$. These can be thought of as the minimum-cost attractors. We have the following characterization theorem:

Theorem 4. *Let $K \subset \mathcal{W}$ be a closed set disjoint from $\{w_n : n \in \mathcal{B}\}$. Then for any sequence of invariant measures $\{\nu^\varepsilon\}$ of the transition probabilities P^ε with $\varepsilon \to 0$,*

$$\lim_{\varepsilon \to 0} \nu^\varepsilon(K) = 0.$$

If ν^* is a weak-topology accumulation point of a sequence of the ν^ε as $\varepsilon \to 0$, then $\operatorname{supp} \nu^* \subset \{w_n : n \in \mathcal{B}\}$.

Because of this theorem we will call the elements of $\mathcal{B}$ the *stochastically stable* attractors.

Since the set of probability measures on $\mathcal{W}$ is weakly compact, every sequence of measures with $\varepsilon \to 0$ will have a convergent subsequence, and all such limit measures have support on set of minimal-cost attractors. Typically there will be only one such attractor, say $\hat{w}$, and so all limit points are point-mass on that attractor. In this case, every sequence of invariant measures converges to the point-mass $\delta_{\hat{w}}$.

This theorem paints a picture of small-ε Markov processes. As ε becomes small, mass concentrates in neighborhoods of the set of attractors. The mass of the complementary areas, including fixed-points that are not attractors, converges to 0. This suggests the typical poverty-trap picture, of individuals being trapped near one or another of the various attractors. But more is true. Only the lowest-cost attractors matters, and typically there is only one of these in the systems we consider here. So, as ε becomes small, mass accumulates around only one attractor. This is not the multi-peak picture that we associate with poverty traps.

One might wonder about the effects of this on convergence times: How long does it take for a family dynasty to have a "typical experience". How long does it take for the effects of the initial distribution of wealth to fade away. This is a difficult subject for models with state spaces that are not countable. In finite Markov chains it is known that convergence is always geometric beyond some point in time. Because our processes satisfy Doeblin's condition, they too will exhibit ultimately geometric convergence. But the interesting question is, how long until the geometric regime kicks in? Doeblin's condition gives an upper bound on the variation-norm distance from the time-t distribution to the invariant distribution of the form $k\rho^t$ with $\rho < 1$. The larger is k, the longer until the bound is informative. It is certainly clear that decreasing ε slows down convergence. Unfortunately the bound derived from Doeblin's condition is often not very good, and so we will not pursue this further.

5 Policy Experiments

5.1 The Deterministic Model

When $\varepsilon = 0$, the case of no shocks, comparative long-run dynamics are easily understood by examining the shape of the parental Engel curve. Attractors occur when the curve crosses the diagonal from above to below. For these fixed-points, policies that shift the curve up will increase the wealth of the attractor. At unstable fixed-points, the curve crosses the diagonal from below to above. Shifting up the parental Engel curve decreases the values of these fixed points.[6]

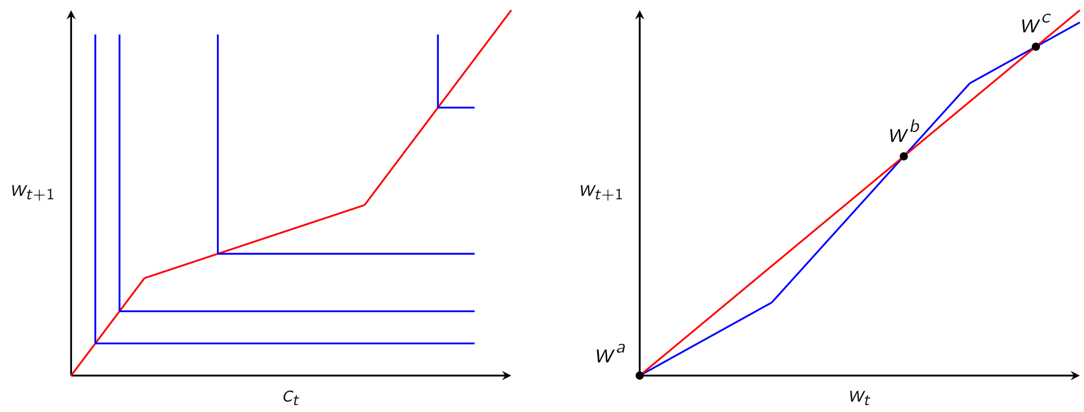
If preferences are homothetic and budget lines are linear, then Engel curves will be linear from 0 to w^*, and flat thereafter. There will be a single attractor, either at 0 or w^* depending on the slope of $e^0(w)$. Multiple attractors — poverty traps — can arise either from non-homotheticity or from non-linearity in the budget line. We will explore examples of both.

Imagine that at low wealth levels, parents' material needs are so great that they can afford investing only a small share of wealth in their children's capital accumulation. For wealthier parents this constraint lessens, but at high levels of capital investment the return to child wealth becomes small. Preferences are not homothetic.[7] Here we may expect to see multiple wealth attractors in deterministic dynamics. Figure 2 shows a highly stylized example: Preferences for $\varepsilon = 0$ are as shown, and $G(k) = k$. Figure 2a displays the expansion path and figure 2b the Engel curve. There are three fixed points, a stable poverty trap, a stable affluence trap, with an unstable fixed-point dividing their basins of attraction. The wealths of dynasties who begin to the left of the unstable fixed point converge to 0 while the wealths of those beginning to the right converge to w^*.

The long-run effects of policy are concerned with the levels of the two attrac-

[6] Non-generic behavior occurs when the curve has a fixed point where the curve touches but does not cross the diagonal. If the curve lies on and above the diagonal at this point, shifting up the curve eliminates the fixed point. If it lies on and below the diagonal, shifting it up bifurcates the fix-ed point into a lower unstable fixed point and a higher attractor.

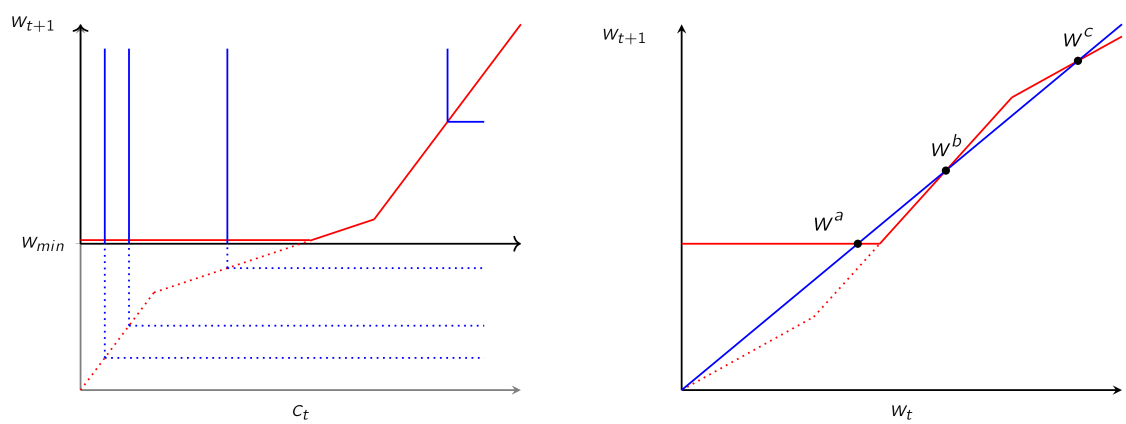
[7] For a discussion of the role of non-homothetic preferences in generating poverty traps, see Ghatak (2015).



(a) An indifference map (b) The Engel curve

Figure 2: Example preferences and Engel curve.

tors and the location of the fixed-point that separates them. Suppose, for instance, that every child is provided with a guaranteed wealth w_{min}. If the lower floor on child's wealth is w_{min}, the rectangle $\mathbb{R}_+ \times [0, w_{min}]$ is excised from the consumption set, and the results are as in figure 3. If the floor is below w^a, it has no long-run effect, raising it above w^a but below w^b increases the wealth of those in the poverty trap but has no effect on who is trapped and who is not. Finally, raising the floor above w^b eliminates the trap.



(a) An indifference map (b) The Engel curve

Figure 3: A basic guaranteed wealth.

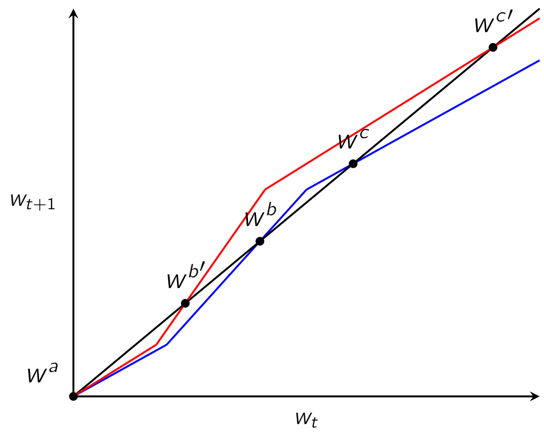


Figure 4: An investment subsidy.

Consider now a straightforward fixed-rate child-investment subsidy. The price of a unit of child wealth falls, and if child wealth is a normal good, the Engel curve $e^0(w)$ will shift up as shown in figure 4. Pinning 0 investment at 0 wealth, w^b shifts down and w^c shifts up. The poverty trap does not disappear, and the long-run wealth of those families in the trap does not change, but now fewer families will be trapped. As w^b slides down, the set of initial conditions leading to the trap shrinks.

If the subsidy cuts off beyond a certain investment level, the budget set will be convex with a piecewise-linear budget line. The effect is to raise the lower part of the parental Engel curve, introduce a flat where the subsidy tails off, and then to join the pre-policy Engel curve for the highest wealths. Again the effect is to pull down the separating point if the subsidy is extensive enough; otherwise it has no long-run effect.

The role of 0 in these examples should not be overemphasized. We might imagine that children are born with a certain innate wealth $\underline{w} > 0$. The Engel curve will start above 0 and therefore first cross the diagonal from above to below, creating an attractor. The picture is no different from Figure 3. In this case, the subsidy will shift up the steady-state wealth in of the low attractor and, as before, pull downward the dividers between basins of attraction.

The effect of these policies on inequality is ambiguous. Imagine an Engel curve with a single attractor at $\underline{w}$. A subsidy could, by shifting up the Engel curve, create an afluence trap, so that even though the post-policy $\underline{w}$ is higher, there is a still higher $\bar{w}$ which attracts some families. By any measure, wealth inequality has increased.

5.2 The Stochastic Model

We saw in the previous subsection that policy effects result in an upward shift of the parental Engel curve, whose effects can be easily calculated. When $\varepsilon > 0$, policies again work through shifting up the Engel curve, but the consequences can no longer be calculated by observing the shifting of the fixed points. Nonetheless, we can see in general that the effect of such policies is to shift up the distribution of wealth.

Theorem 5. *Suppose that F is strictly increasing in k. Let $e^{\varepsilon}(w)$ and $e'^{\varepsilon}(w)$ be two parental Engel curves, with invariant distributions μ^{ε} and μ'^{ε}, respectively. If parental Engel curve $e^{\varepsilon}(w)$ is everywhere at least as big as parental Engel curve $e'^{\varepsilon}(w)$, and if on some open set of wealths $e^{\varepsilon}(w) > e'^{\varepsilon}(w)$, then μ^{ε} dominates μ'^{ε} in the first-order stochastic dominance relation.*

The motion of a Markov process provides an additional channel for policy to work that is not apparent in deterministic models. Consider again an upward shift in a parental Engel curve with two attractors separated by an unstable fixed point w', and suppose that $\mathcal{B} = \{\underline{w}\}$. The effect of the shift is to decrease w', increasing the size of the basin of attraction of the large attractor $\bar{w}$ and shrinking that of the small attractor $\underline{w}$. The locations of the two attractors may also increase, further magnifying the changes in size. The effect of this under any shock distribution h is to decrease the cost $B(\underline{w}, \bar{w})$ of moving from $\underline{w}$ to $\bar{w}$, and to increase the cost $B(\bar{w}, \underline{w})$ of moving from $\bar{w}$ to $\underline{w}$. A large enough policy will switch the ranking of these costs. When moving up becomes cheaper than moving down, $\mathcal{B} = \{\bar{w}\}$, and the invariant distribution shifts from most mass near $\underline{w}$ to most mass near $\bar{w}$. When ε is large — variance is high — this transition is gradual. When ε is small, the transition will be quite sharp. These policy effects look like phase transitions.

6 Conclusion

We have worked out a rather stripped-down Becker-Tomes style Markov model of the evolution of wealth in order to examine its stationary distribution. We have shown that poverty traps can arise naturally in these models as they do in deterministic models, but their consequence is more subtle than simply a multi-peaked wealth distribution. We have also worked out the effects of some commonly discussed policies for both deterministic and stochastic models.

The poverty trap concept is closely related to — in fact, a piece of — the concept of resilience as it appears in the study of ecological dynamics and systems engineering. There are two aspects of resilience that recur repeatedly in these literatures: recovery time, how long it takes an out-of-equilibrium system to return to an equilibrium state; and what we can call domain resilience, how much of a shock can a system sustain before flipping towards another equilibrium.[8] In deterministic systems where steady states occur as wells and the system flows downhill, recovery time corresponds to the depth of a well and domain resilience corresponds to the width of wells. These two concepts combine to determine the cost of an attractor, and so they shape the distribution of wealth. There is a tradeoff, however, between the two ideas. If, for instance, the h is the absolute value function, giving rise to the family of Laplace distributions, only width matters. The optimal path from attractor w_i to attractor w_j is always to move in one step to the edge of any basin of attraction in which the direction of motion is away from w_j (and ride for free through those parts where the direction of motion is in the right direction). On the other hand, if h is a high-degree even polynomial, depth is important for determining cost. Those attractors that are less resilient in the ecological dynamics sense have low mass in the stationary wealth distribution; they are easier to escape from, and harder to return to, than more resilient states. This suggests a connection between inequality and immobility that we will explore in a subsequent paper.

The empirical relevance of poverty traps is contested in the literature. Kraay and McKenzie (2014) argues against the importance of poverty traps. Ravallion (2015, pp. 1226-7) provides a more nuanced view. We have seen that the absence

[8]Gunderson (2000) refers to these two concepts as engineering and ecological stability. He errs, however, in assuming that engineering resilience, recovery time, is applicable only to systems with a single, globally stable state, and more generally in assuming that resilience is applicable only to systems in which attractors are single states.

of clear multiple peaks in cross-section does not indicate the absence of a trap. A first step towards the econometric investigation of poverty traps is a model of the dynamics that generate them. Further study of properties of Markov processes related to stationary measures, such as first-passage times and exit times suggests that current measures of immobility can be improved upon. We will explore this issue in a subsequent paper.

References

Azariadis, Costas and John Stachurski. 2005. "Poverty Traps." In *Handbook of Economic Growth, vol. 1*, vol. 1, chap. 5. Elsevier.

Becker, Gary S. and Nigel Tomes. 1979. "An equilibrium theory of the distribution of income and intergenerational mobility." *Journal of Political Economy* 87 (6):1153–1189.

Ghatak, Maitreesh. 2015. "Theories of Poverty Traps and Anti-Poverty Policies." *World Bank Economic Review* 29:S77–S105.

Gunderson, Lance H. 2000. "Ecological resilience — in theory and application." *Annual Review of Ecological Systems* 31:425–39.

Kifer, Yuri. 1990. "A discrete-time version of the Wenzell-Friedlin theory." *Annals of Probability* 18 (4):1676–92.

Kraay, Aart and David McKenzie. 2014. "Do poverty traps exist? Assessing the evidence." *Journal of Economic Perspectives* 28 (3):127–48.

Loury, Glenn C. 1981. "Intergenerational transfers and the distribution of earnings." *Econometrica* 49 (4):843–867.

Meyn, Sean P. and Richard L. Tweedie. 1993. *Markov Chains and Stochastic Stability*. Springer-Verlag.

Ravallion, Martin. 2015. "The idea of antipoverty policy." In *Handbook of Income Distribution*, vol. 2. Elsevier, 1967–2061.

Appendix A. Proofs

Proof of Theorem 1. Our assumptions 1, 3 and 4 imply that $V(c,k)$ is usc and strictly concave in c for each k. With assumption A.6, the budget correspondence will be continuous. That π is non-empty valued and uhc follows from the maximum theorem.

Assumptions A.4 and A.5 imply that V is supermodular. Let $w' > w''$, $k' \in \pi(w')$, and $k'' \in \pi(w'')$. Suppose that the statement of the theorem is false, that is, $k' < k''$. Since $k'' \in \pi(w'')$, $w'' - C(k'') \geq 0$, and consequently, $w' - C(k'') > 0$ which means that the investment level k'' is feasible for w'. Since k' is optimal for w', $V(w'-k',k') \geq V(w'-k'',k'')$. On the other hand, $C(k') < C(k'')$, so k' is feasible for w''. Since k'' is optimal for w'', $V(w''-k'',k'') \geq V(w''-k',k')$. Combining these two inequalities, we get

$$V(w'-k',k') - V(w''-k',k') \geq V(w'-k'',k'') - V(w''-k'',k'').$$

Next, $w'-k'' > w''-k''$ and $k'' > k'$, so supermodularity of V implies that

$$V(w'-k'',k'') - V(w''-k'',k'') \geq V(w'-k'',k') - V(w''-k'',k').$$

Putting these two equations together,

$$V(w'-k',k') - V(w''-k',k') \geq V(w'-k'',k') - V(w''-k'',k').$$

Equivalently,

$$V(w'-k',k') - V(w'-k'',k') \geq V(w''-k',k') - V(w''-k'',k').$$

This contradicts the strict concavity of V in c. □

Also note that assumption A.6 implies that μ^ε is weakly continuous in $\varepsilon \geq 0$. Consequently, the Maximum Theorem implies as well that π is uhc in ε.

Proof of Corollary 1. If $\pi^\varepsilon(w')$ is multivalued, then $\min\{k : k \in \pi(w')\}$ and $\max\{k : k \in \pi(w')\}$ span an open interval. Similarly if $\pi^\varepsilon(w'')$ is multivalued for some $w'' \neq w'$. The monotonicity property of Theorem 1 implies that these two open intervals are disjoint. The proof is completed by recalling that $\mathbb{R}_+$ contains only a countable number of disjoint open intervals. □

Proof of Theorem 2. For all $w \in \mathcal{W}$ the transition probability $P^\varepsilon(w, \cdot)$ is absolutely continuous with respect to the measure ψ that puts point mass on 0 and on w^*, and is Lebesgue measure on $(0, w^*)$. Thus for each $w \in \mathcal{W}$, $P^\varepsilon(w, \cdot)$ has a density with respect to ψ. Each density can be described by a triple (a, b, f_w^ε) where a and b are non-negative weights on 0 and w^*, and f_w^ε is a density on $(0, w^*)$. It follows from assumption B.1 that $a \geq \int_{-\infty}^{-w^*} \phi^\varepsilon(s)ds \equiv a_\psi$, $b \geq \int_{w^*}^{\infty} \phi^\varepsilon(s)ds \equiv b_\psi$, and $f_w^\varepsilon(z) \geq \min_{\{z \in \mathcal{W}\}}\{\phi^\varepsilon(z), \phi^\varepsilon(z - w^*)\} \geq \min\{\{\phi^\varepsilon(w^*), \phi^\varepsilon(-w^*)\} \equiv c_\psi$, and all these bounds are strictly positive. It follows that if $\psi(A) > \gamma > 0$, then

$$P^\varepsilon(w, A) \geq a_\psi \psi(A \cap \{0\}) + b_\psi \psi(A \cap \{w^*\}) + \gamma c_\psi,$$

which immediately gives Doeblin's condition. (See Meyn and Tweedie (1993).) The chain is clearly irreducible and aperiodic by virtue of the positivity of the densities, so it is uniformly ergodic.

To show variation-norm continuity, note that the correspondence is weakly uhc. Choose a sequence $\{\varepsilon_n\}$ with limit $\varepsilon^* > 0$. Notice that the map $z \to (a^\varepsilon(z), b^\varepsilon(z), f_w^\varepsilon(z))$ is jointly continuous in z and ε on $\mathcal{W} \times \mathbb{R}_{++}$. Consequently, each μ^ε has a density $\xi^\varepsilon(z)$ with respect to ψ, and $\xi^{\varepsilon_n}(z)$ converges pointwise to $\xi^{\varepsilon_n}(z)$. The conclusion now follows from Scheffé's Theorem. □

Proof of Theorem 3. The existence of fixed points is guaranteed by Tarski's Fixed-Point Theorem since $\mathcal{W}$ is a compact lattice and e^0 is increasing.

Let w' denote the largest fixed point of e^0, whose existence is guaranteed by assumption B.4. Suppose that no fixed point $w'' < w'$ is an attractor, and that w' is not an attractor. Then $e^0(w) > w$ for all $w > w'$. Choose w'' strictly between w' and w^*. e^0 maps the interval $[w'', w^*]$ into itself, and so it has a fixed point, contradicting the hypothesis that w' is the largest fixed point. □

Proof of Theorem 4. The first part of the Theorem is an application of Kifer's (1990) Theorem 3.1.[9] The next lemma states that his assumption 1.1 is satisfied.

Lemma 1. *For any open set $U \in \mathcal{W}$,*

$$\lim_{\varepsilon \to 0} \varepsilon \log P^\varepsilon(w, U) = -\inf_{v \in U} h(v - e^0(w))$$

uniformly in w.

[9]See in particular the discussion on page 1678 and the paragraph following Kifer's equation 1.6.

Proof of Lemma 1. This follows from Laplace's method of integration. It suffices to show the result on an open interval $U = (a, b) \subset \mathcal{W}$.

$$\begin{aligned} P^\varepsilon(w, U) &= Z(\varepsilon) \int_a^b \exp - \frac{h(s - e^\varepsilon(w))}{\varepsilon} ds \\ &= Z(\varepsilon) \int_{a - e^\varepsilon(w)}^{b - e^\varepsilon(w)} \exp - \frac{h(z)}{\varepsilon} dz, \end{aligned}$$

and the second integral is in the form required by Laplace. Choose an $\varepsilon_0 > 0$. Since $e^\varepsilon(w)$ is jointly continuous on $[\varepsilon_0, 0] \times \mathcal{W}$, the functions $e^\varepsilon(w)$ converge uniformly to $e^0(w)$ on $\mathcal{W}$ for any sequence $\varepsilon \to 0$. Therefore, for any $\delta > 0$ there is an $\varepsilon' > 0$ such that for all $\varepsilon < \varepsilon'$, $|e^\varepsilon(w) - e^0(w)| < \delta$. For such ε,

$$Z(\varepsilon) \int_{a - e^0(w) + \delta}^{b - e^0(w) - \delta} \exp - \frac{h(z)}{\varepsilon} dz < P^\varepsilon(w, U) < Z(\varepsilon) \int_{a - e^0(w) - \delta}^{b - e^0(w) + \delta} \exp - \frac{h(z)}{\varepsilon} dz.$$

If $a > e^0(w)$, then for small enough $\delta > 0$, Laplace's method shows that as ε goes to 0, the lower and upper bounds are asymptotically approximated by

$$\frac{Z(\varepsilon)}{\alpha_1} \Gamma\left(\frac{1}{\alpha_1}\right) \left(\frac{\varepsilon}{\beta_1}\right)^{1/\alpha_1} \exp - \frac{h(a - e^0(w) + \delta)}{\varepsilon}$$

and

$$\frac{Z(\varepsilon)}{\alpha_1} \Gamma\left(\frac{1}{\alpha_1}\right) \left(\frac{\varepsilon}{\beta_1}\right)^{1/\alpha_1} \exp - \frac{h(a - e^0(w) - \delta)}{\varepsilon}$$

respectively. Laplace's method gives us that $\lim_{\varepsilon \to 0} \varepsilon \log Z(\varepsilon) = 0$, and so

$$-h(a - e^0 - \delta) < \lim_{\varepsilon \to 0} \varepsilon \log P^\varepsilon(w, U) < -h(a - e^0 + \delta).$$

Since δ is arbitrary and h is continuous, $\lim_{\varepsilon \to 0} \varepsilon \log P^\varepsilon(w, U) = \inf_{s \in U} h(s - e^0(w))$.

If $b < e^0(w)$, the same argument works with b replacing a and α_2 and β_2 replacing, respectively, α_1 and β_1. If $a < e^0(w) < b$, the same method shows that $P^\varepsilon(w, U) \sim 1$, and so $\varepsilon \log P^\varepsilon(w, U) \to 0 = h(0) = -\inf_{s \in U} h(s - e^0(w))$. $\square$

For the second part of the Theorem, choose an open set O whose closure does not contain $\mathcal{B}$, and a sequence $\varepsilon_n \to 0$ such that the corresponding ν^n converge weakly to ν^0. From the Portmanteau Theorem, $0 = \liminf \nu^n(O) \geq \nu^0(O) \geq 0$, so O^c has ν^0-measure 1. The closed set $\mathcal{B}$ is the intersection of a countable number of such closed sets $O^c \supsetneq \mathcal{B}$, and so $\nu^0(B) = 1$. □

Proof of Theorem 5. Let $e'(w)$ and $e''(w)$ denote two parental Engel curves derived from utility maximization problems with the same $\varepsilon > 0$. (ε superscripts will be henceforth omitted.) Suppose that on some set A of positive measure with respect to ψ, $e'(w) > e''(w)$. Let $(S, \mathcal{S}, \tilde{\psi})$ denote the measure space of *sequences* of shocks s_t with the Borel σ-field, and $\tilde{\psi}$ the product measure of iid draws from density ψ.

Define the Markov process on $\mathcal{W} \times \mathcal{W}$ by starting from the same initial wealth w_0, the first component evolving acording to e' and the second according to e'', both with the same shock sequence $\bar{s} \in S$. The transition probability has the property that the projection onto the first component is the wealth process derived from e', the projection onto the second component is the wealth process derived from e'', and the set $\Delta = \{w', w'' : w' \geq w''\}$ is reached from any (w_0', w_0'') with probability 1. The Markov process on pairs has an invariant distribution $\tilde{\mu}$ whose projection onto each component is the invariant distribution of the corresponding wealth process, and whose support is all of Δ. Let f be any increasing function. Then

$$\int_{\mathcal{W}} f(w')d\mu' - \int_{\mathcal{W}} f(w'')d\mu'' = \int_{\mathcal{W}\times\mathcal{W}} \big(f(w') - f(w'')\big)d\tilde{\mu} = \int_{\Delta} \big(f(w') - f(w'')\big)d\tilde{\mu} \geq 0.$$

Consequently, μ' is at least as big as μ'' in the first-order stochastically dominance relation.

For the last part of the theorem, let A denote the open set of wealths on which $e'(w) > e''(w)$, and let f denote a function which is increasing on A and constant on A^c. Without loss of generality we can assume A is contained in the interior of $\mathcal{W}$. Extend f from $\mathcal{W}$ to $\mathbb{R}$ by letting $f(w) = f(w^*)$ for $w > w^*$ and $f(w) = 0$ for $w < 0$. By doing so we can avoid specially notating the point masses at 0 and w^*.

Then

$$\int_{\mathcal{W}} f(w')d\mu' - \int_{\mathcal{W}} f(w'')d\mu'' = \int_{\Delta} \big(f(w') - f(w'')\big) d\tilde{\mu} = \int_{\Delta} \big(f(w') - f(w'')\big) d\tilde{\mu}$$
$$= \int_{\Delta} \left(\int_{\mathcal{W}} f(z')P'(w', dz') - \int_{\mathcal{W}} f(z'')P''(w'', dz'') \right) d\tilde{\mu}$$
$$= \int_{\Delta} \int_{\mathbb{R}} \left(f\big(g(e'(w')) + s\big) - f\big(g(e''(w'') + s)\big) \right) \phi(s) ds \, d\tilde{\mu}$$

The integral on the last line is at least 0 for all $(w', w'') \in \Delta$ and s. Furthermore for each w'' there is a set of positive ϕ-measure such that $g(e''(w'') + s) \in A$. Then

$$f\big(g(e'(w')) + s\big) - f\big(g(e''(w'')) + s\big) \geq f\big(g(e'(w'')) + s\big) - f\big(g(e'(w'')) + s\big) > 0$$

on some cube inside of $\mathcal{W} \times A$, and non-negative elsewhere. Since this cube has positive measure, $\int_{\mathcal{W}} f(w')d\mu' - \int_{\mathcal{W}} f(w'')d\mu'' > 0$, and so μ' strictly stochastically dominates μ''. □

Appendix B. Discontinuous Parental Engel Curves

Some policies that support investment in children can lead to discontinuous Engel curves. A flat-rate subsidy for low levels of child investment that cuts out beyond some level introduces a kink in the budget line but leaves the cost function $C(k)$, and therefore the budget set, convex. (Its effect will be to introduce a flat in the Engel curve.) But a flat-rate subsidy that kicks in beyond a certain level, such as a subsidy for college education, introduces a non-convexity into the cost function, a kinked budget line that kinks the "wrong way", and consequently a non-convex budget set. A few moments at a blackboard will convince the reader that a discontinuity in the Engel curve will occur at that wealth level at which optima occur on both sides of the kink. The conclusions of Theorem 1 still hold, and a consequence of upper hemicontinuity is that at a discontinuity, the optimal policy function must be multivalued. The argument in the proof of Corollary 1 applies here as well, and so the number of discontinuities must be countable. We will assume

B.5. *The number of wealths at which the optimal policy correspondence is multivalued is finite, and includes neither 0 nor w^*.*

Theorem 4 is the only result that relies on continuity. Our technique for dealing with discontinuities is to "smooth" the optimal policy correspondence; to replace selections from it with a continuous function which is uhc and differs from the original only in arbitrarily small neighborhoods of discontinuities. We show in Theorem 6 that choosing a good continuous approximation to the optimal policy correspondence gives a Markov process whose invariant distribution is close to that belonging to a process derived from any selection from the optimal policy correspondence.

The first step is to choose a selection from the optimal policy correspondence. Let D denote the set of discontinuities $\{w_1, \dots, w_d\}$. We construct Markov processes by choosing a selection from this correspondence and proceeding as in Section 3. Two selections can disagree only on D, in the interior of $\mathcal{W}$, so the invariant distribution will not depend upon the choice of selection. We suppress notation of $\varepsilon > 0$ for this appendix, and let $\tilde{\pi}(w)$ denote the selection from the optimal policy correspondence which takes on the largest optimal investment at each $w \in D$. Choose $\delta > 0$ to be less than w_1 and $\min_{i=2,\dots,d} w_i - w_{i-1}$.

$$\pi(w) = \begin{cases} \dfrac{1}{w - (w_i - \delta)} \displaystyle\int_{w_i - \delta}^{w} \tilde{\pi}(v)dv & \text{if } w_i - \delta < w < w_i,\ w_i \in D, \\ \dfrac{1}{w_i + \delta - w} \displaystyle\int_{2w - w_i - \delta}^{w} \tilde{\pi}(v)dv & \text{if } w_i \leq w < w_i + \delta,\ w_i \in D, \\ \tilde{\pi}(w) & \text{otherwise.} \end{cases}$$

This operation smooths $\tilde{\pi}$ at w by averaging over nearby values. The δ-smoothed Engel curve will be continuous and increasing. It equals the optimal investment except in each interval $w_i \pm \delta$ around the discontinuity w_i. This averaging uses a backwards-looking window to ensure budget feasibility at low wealths; otherwise a parent with wealth 0 could choose a positive investment level for the child. Other choices are, of course, possible. This operation introduces at most one fixed point for each discontinuity, and at these fixed points $\tilde{\pi}$ crosses the diagonal from below to above, so no new attractors are added. Let P^δ denote the transition probability for the Markov process with the δ-smoothed Engel curve, and ν_δ its invariant measure. Our justification for smoothing is the following theorem:

Theorem 6. *The invariant measures ν_δ converge weakly to ν_0 in variation norm.*

To reiterate, a smoothed approximation to a selection $\tilde{\pi}$ with a small window-diameter δ gives a good approximation to the invariant distribution of the unsmoothed

Markov process. Consequently, the shape result of Theorem 4 is informative about the shape of the unsmoothed process as well.

Proof. The invariant measures ν_δ for the P^δ have densities with respect to ψ. To prove 2 we must show that for any sequence $\delta \to 0$ the densities converge to the density for ν_0 ψ-almost surely. Choose a sequence of invariant distributions $\{\nu_\delta\}$ with weak limit ν. First we show that ν is absolutely continuous with respect to ψ. Each density can be written $(a_\delta, b_\delta, f_\delta(w))$. For weak convergence to hold, the two number sequences $\{a_\delta\}$ and $\{b_\delta\}$ must converge to limits a and b. Now consider the sequence $\{f_\delta\}$, and choose a set N with $\psi(N) = 0$. Choose a decreasing sequence of open sets N_n with limit N, such that $0 < \psi(N_n) < 1/n$. Let $q_n(w)$ denote a continuous function bounded by 0 and 1, that is 1 on N and 0 on N_n^c. Then for all $\delta \geq 0$,

$$0 \leq \mu_\delta(N) \leq \int q_n(w) d\mu_\delta \tag{4}$$

and for $\delta > 0$,

$$\int q_n(w) d\mu_\delta = \int q_n(w) f_\delta(w) d\psi \leq \int q_n(w) d\psi < 1/n. \tag{5}$$

The first three inequalities hold for all δ including 0. It follows from (5) and weak convergence of the ν_δ that $\int q_n(w) d\mu_0 \leq 1/n$, and then from (4) that $\mu_0(N) \leq 1/n$. Since n is arbitrary, $\mu_0(N) = 0$. Thus μ_0 has a density f_0 with respect to ψ. If f_δ does not converge to f_0 pointwise ψ-almost surely, there is a continuous function q for which $\int q f_\delta d\psi \not\to \int q f_0 d\psi$, which contradicts weak convergence.

Finally, we need to show that μ_0 (with density f_0) is invariant for P^0. The transition probability densities $p^\delta(x, y)$ are bounded, and p^δ converges pointwise to p_0^ε ψ-almost surely. (B.1 puts the finite set of discontinuities, where pointwise convergence fails, in the interior of $\mathcal{W}$.) Then

$$f_0(y) = \lim_\delta f_\delta(y) = \lim_\delta \int p^\delta(x, y) f_\delta(x) d\psi = \int p^0(x, y) f_0(x) d\zeta,$$

the last equality by the dominated convergence theorem. $\square$

Discussion Papers of the Research Area Markets and Choice 2020

Research Unit: **Economics of Change**

Anselm Hager and Justin Valasek Refugees and social capital: Evidence from Northern Lebanon	SP II 2020-301
Maja Adena and Anselm Hager Does online fundraising increase charitable giving? A nation-wide field experiment on Facebook	SP II 2020-302
Lawrence Blume, Steven Durlauf, and Aleksandra Lukina Poverty Traps in Markov Models of the Evolution of Wealth	SP II 2020-303