Intergenerational Mobility and Poverty Traps in the Markov Models of the Evolution of Wealth

Lawrence Blume*, Steven Durlauf**, Aleksandra Lukina***

April 26, 2023

This draft is incomplete. PLEASE DO NOT CIRCULATE. Entire sections and many of the figures are missing.

Abstract

This paper proposes a "stepping-stone" model of intergenerational mobility. We track the wealth of infinite sequences of one-parent, one-child families. A parent allocates her wealth between own consumption and investment in the child. A novel assumption of our work concerns the technology describing the production of wealth from parental investment. The stepping-stone technology reflects the idea that only discrete levels of parental investments allow the next generation to achieve an access to a certain quality of education, training, etc. We allow for unobserved heterogeneity in the effects of investment on children's wealth, and so we allow wealth to be random, conditional on the parental investment. Within the developed framework we study the difference between a zero-shock process (a deterministic system) and processes with small amounts of noise. We show that, relative to a zero-shock process, some of the multiple deterministic attractors are less fragile than the others, and that their presence dominates the stationary behavior of the wealth distribution. 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. Thus, poverty traps can arise naturally in stochastic models as they do in deterministic models, but their consequence is more subtle than simply a multi-peaked wealth distribution. Also, we analyze how perturbations of parameters of the model change the stationary wealth distribution. Finally, we offer some ideas related to measuring intergenerational mobility and analyzing the coevolution of inequality and mobility [in progress].

---

* Cornell University and IHS Vienna
** Harris School of Public Policy, University of Chicago
** Harris School of Public Policy, University of Chicago

1 Introduction

This paper proposes a stepping-stone model of intergenerational mobility, one in which there are human capital thresholds that must be met to obtain different levels of wealth. The logic of the model derives from the class mobility analyses of Becker and Tomes (1979) and Loury (1981) in that parental wealth constrains levels of human capital received by children and thus provides a mechanism to explain levels of mobility. By conceptualizing investments as discrete, this produces equilibrium laws of motions that can be characterized by Markov chains. What might appear to be a technical wrinkle will turn out to produce pathways to addressing a number of significant issues in the study of intergenerational dynamics. Substantively, we use this framework to analyze the nature of poverty and affluence traps in deterministic versus stochastic models.

According to Azariadis and Stachurski (2005) a poverty trap is usually defined as "any self-reinforcing mechanism which causes poverty to persist". However, the fact that the vast empirical evidence confirms that poverty is persistent does not yield poverty traps existence (Azariadis and Stachurski (2005)). On the contrary, we should first decide how to define them, justify why they exist, and only then we will be able to use them as an explanation of persistent poverty. It is also critical to make the difference between two different patterns of poverty (Carter and Barrett (2006)). In the first scenario the same individuals in the intragenerational case[1] (or the same dynasties in the intergenerational case[2]) stay poor period after period. The second scenario is characterized by moving over the social ladder during the individual (or dynasty) life course based on some random outcomes or on some deterministic mechanisms. In both situations we may observe the same fraction of poor, however, society described in the first case is much more polarized while in the second case individuals, on the contrary, "share the burden of poverty equally" (Carter and Barrett (2006)), and poverty is a transitory phenomenon. In the latter case there are less reasons for concern and government interventions. Thus, deep and multilateral understanding of mechanisms underlying poverty (and affluence) persistence seems to be fundamentally important. In particular, it is essential to provide some theoretical basis that may help to formulate the concept of poverty traps in the language of relative persistence of different wealth/income classes and

---

[1] When we track individual income/wealth/occupation changes, we deal with intragenerational mobility.

[2] Intergenerational mobility refers to any change in the social position of dynasty/family members that takes place from one generation to the next.

other properties of stochastic environment.

This paper proposes a new approach to modelling the mechanisms that determine intergenerational mobility and poverty traps in particular. Following the classic paper by Becker and Tomes (1979), we assume that parents invest in education of their children. Our model builds upon this idea by treating these investments as discrete, so the presented framework may be understood as a stepping-stone model of intergenerational mobility. We consider both deterministic and stochastic environments.

The introduction of discreteness in educational investment matters at two levels. At a technical level, we are able to characterize equilibrium sequences of wealth levels across generations. In turn, a number of features of these trajectories are examined. As such, we provide explicit microfoundations for the use of Markov chains, as opposed to linear regressions used for the intergenerational elasticity (IGE) estimation[3], to model the mobility relationship.

The characterization of Markov chain descriptions of mobility turns out to lead to interesting issues relating to the measurement of mobility. Specifically, when one moves from deterministic to stochastic environments, phenomena such as poverty or affluence traps, which are natural in the deterministic world, disappear as long as all possible wealth levels communicate. Traps do not generalize in a straightforward way, which leads to proposing some new measures of mobility and persistence that are appropriate for stepping-stone environments.

Mechanisms leading to poverty traps at the individual or the family level can be split into two large groups: those that can be explained by external frictions (such as market failure) and those that are "due to behaviour under extreme scarcity in the absence of any frictions" (Ghatak (2015)). The latter story is perfectly covered by our modelling. Individuals struggling for everyday survival spend most of their income on their own consumption and have no chance to invest in their children enough. Let us underline that this is not necessarily the story about monetary investments, it may also concern lack of time and attention span caused by a huge amount of working hours etc. At this stage we do not introduce external frictions into our framework,

---

[3] $\log y_{i,t+1} = \alpha + \beta \log y_{i,t} + \varepsilon_{i,t+1}$, where $y_{i,t+1}$ is child's income (or wealth, or earnings, or any other characteristics of interest), $y_{i,t}$ is parental income, α "captures the trend in average incomes across generations" (Corak, 2013), β is the IGE (of income, or wealth, or earnings), $\varepsilon_{i,t+1}$ is an error term, i denotes family index, t – time period.

but the fundamental ideas of our analysis are quite general and we expect them to carry over to more complex models that cover other mechanisms leading to poverty.

This paper suggests that parental investments determine child prospects in isolation from investments of other parents, i.e. it is assumed that the demand for each skill level is infinitely elastic. In a following paper we are also going to allow the effect of a given parental investment to depend on investments of others.

2 The Stepping-Stone Model and its Properties

2.1 Basic Assumptions

In this paper we track the wealth of infinite sequences of one-parent, one-child families, that we call "dynasties". Each member of the dynasty is active for one period and produces one child. A parent allocates her wealth between her own consumption and investment in her child. For simplicity we consider one good economy. Following the benchmark Becker-Tomes model (Becker and Tomes (1979)), we assume that a parent cares only about wealth of the child and not about her welfare, i. e. not about child's future utility (as in another classic paper Loury (1981); such assumption technically leads to dynamic programming problems). We regard this assumption as a reasonable one because it seems plausible that when a parent makes investment decisions she does not have enough inputs to predict what a child's future utility might be (child's actual preferences are formed much later than these decisions are made). A novel assumption of our work concerns the technology describing the production of wealth from parental investment. While Becker and Tomes (1979) consider the linear dependence of child's wealth on parental investment we study what we call the "stepping-stone technology", reflecting the idea that only discrete levels of parental investments allow the next generation to achieve an access to certain quality of education, training, etc. We allow for unobserved heterogeneity in the effects of investment on children's wealth, and so we allow wealth to be random, conditional on the parental investment. The stochastic shock that affects child's outcome does not have an explicit interpretation. We can use it to discuss, for instance, macroeconomic shocks. We can also use it to discuss the distribution of unobservables, such as children's genes – thus contributing to the "nature versus nurture" discussion (how important is the genetic component of poverty versus how the child is raised[4]) by modeling the evolution of the stochastic shock as the outcome of genetic and epigenetic processes[5].

We start with the general formulation of one-period problem and then we move to infinite horizon framework and the stepping-stone technology. Though the main purpose of the paper is to analyze the dynamic stochastic stepping-stone model we prefer to start studying the deterministic dynamical problem. This case shades some light on what may happen in the stochastic case for small noises and just gives the reader additional intuition and deeper insight.

2.2 One-Period Problem

Consider first a **one-period problem** fixing the generation of time period t. The state of the generation is described by its wealth $w(t)$. The generation chooses consumption $c(t)$ and human capital investment $k(t)$ so as to maximize the expected utility $U(c(t), w(t+1))$ of its own consumption and child's wealth, maybe random, subject to a budget constraint. Wealth of the child $w(t+1)$ is an outcome of the parent's investment $k(t)$ which is described by the transition function: $w(t+1) = f(k(t), s(t+1))$, where $s(t+1)$ is a random variable realized in the next period of time. We assume that generation t has correct beliefs about the distribution μ of $s(t+1)$.

Once we can write $f(k(t), s(t+1))$ instead of $w(t+1)$, there will be no confusion between $w(t)$ and $w(t+1)$ anymore and we can omit the period of time for the rest of the subsection. Any fixed generation faces the following problem:

$$V(c,k) \equiv E_\mu U(c, f(k,s)) \to max_{c,k} \tag{1}$$

$$s.t.\ c + k \le w,\ \ c \ge 0,\ \ k \ge 0, \tag{2}$$

where w stands for $w(t)$.

On the utility function $U(x, y)$, defined on $\mathbb{R}^2_+$, and the transition function $f(k, s)$, defined on $\mathbb{R}_+ \times \mathbb{R}$, we put the following assumptions:

---

[4] Bowles and Gintis (2002), Björklund, Jäntti, and Solon (2005) and more recent Black et al. (2020).

[5] See, for example, Youngson and Whitelaw (2008).

- A1. $U(x, y)$ is strictly increasing in x and y.
- A2. $f(k, s)$ is non-decreasing in k for each s.
- A3. $U(x, y)$ is upper-semicontinuous in both x and y, $f(k, s)$ is upper-semicontinuous in k for any s.
- A4. $U(x, f(k, s))$ is Lebesgue integrable over the distribution $\mu(s)$ for any x and k.
- A5. $U(x, y)$ is supermodular ($U''_{xy} \geq 0$ if it is differentiable).
- A6. $U(x, y)$ is strictly concave in x.
- A7. $f(k, s)$ is non-negative.

Assumptions A1, A2, A6, A7 are standard and seem to be very natural. Note that A6 yields that $U(x, y)$ is continuous in x. A3 and A4 are more technical, however they are crucial already for this subsection. Supermodularity used in A5 is equivalent to the property of increasing differences. And since parental own consumption and child's wealth seem to be complementary goods, A5 is also a realistic assumption.

The first observation one has to make is that due to assumption A1 the budget constraint 2 is tight. Thus, we are looking for a maximum of upper-semicontinuous function (guaranteed by A3) on the compact set, and exactly this allows us to write maximum instead of supremum in 1. More formally we have

Theorem 1. *Assume A1-A4 are fulfilled. Then for all values of wealth* $w \geq 0$ *there exists at least one pair of consumption value* $c \geq 0$ *and investment value* $k \geq 0$ *that solves the optimization problem 1-2.*

The detailed proof can be found in Appendix.

Note that Theorem 1 does not guarantee uniqueness of the problem 1-2 solution. Having this theorem, we can introduce the notion of an optimal policy.

Definition 1. *The optimal policy is the correspondence* $\pi : \mathbb{R}_+ \to \mathbb{R}_+$ *given by*

$$\pi(w) = \{k \geq 0 : \textit{there is } c \geq 0 \textit{ such that } c + k \leq w \textit{ and } (c, k) \in argmax_{c,k} V(c, k)\}.$$

Let us underline this one more time. It might happen that the solution to the problem 1-2 is not unique (the reader will see that it will be true for our stepping-stone model), and thus we cannot talk about the policy function, the best we can do is to look at the correspondence. The correspondence π shows what investment level (or levels) is (are) optimal for a given amount of parental wealth w.

In terms of the optimal policy the statement of Theorem 1 can be reformulated in the following form:

$$\text{for all } w \geq 0 \text{ the set } \pi(w) \text{ is nonempty.}$$

Using the Berge maximum theorem, we can say something about properties of the correspondence $\pi(w)$. Precisely,

Theorem 2. *Given assumptions A1-A4, the correspondence π, defined above, is upper-hemicontinuous at every continuity point of the function $V(c,k)$.*

In the Becker-Tomes paper capital investment in a child is evidently a normal good. Under assumptions A1-A5 it turns out to be true in our case as well. Let us check that the optimal choice of parental investment k increases with disposable wealth w.

Theorem 3. *Assume A1-A6 are fulfilled. If $w' > w''$ and if k' is optimal for w', k'' is optimal for w'', then $k' \geq k''$.*

Summarizing this section, we can state that under assumptions A1-A5 any generation t, having arbitrary wealth $w(t) \geq 0$, can choose an optimal parental investment $k(t) \geq 0$ that leads the next generation to wealth $w(t+1)$ once $s(t+1)$ is realized. We also need A7 to guarantee that the latter is non-negative. Moreover, we know that having more wealth generations will invest in their children at least not less. The latter reflects intuition one can have assuming that individuals are rational and take care of the next generation's wealth. Therefore, we can study the dynamic process observing wealth of some dynasty that has initial wealth $w(0)$.

Set for the rest of the paper that all the assumptions A1-A7 are fulfilled.

2.3 The Deterministic Stepping-Stone Model

As it was described above, the generation t parent allocates her wealth $w(t)$ between consumption $c(t)$ and capital investment $k(t)$ so as to maximize the utility $U(c(t), w(t+1))$ subject to a budget constraint. Wealth of children $w(t+1)$ is an outcome of the parental investment $k(t)$. It is described by the transition function: $w(t+1) = f(k(t))$ (the random part $s(t+1)$ of the transition function is omitted until the stochastic case consideration), where

$$f(k) = \begin{cases} w^0 & \text{if } k \ < k^1, \\ w^1 & \text{if } k^1 \leq k < k^2, \\ \ldots & \\ w^N & \text{if } k \geq k^N, \end{cases} \tag{3}$$

Here all the values $k^i > k^0 = 0$ and $w^j \geq 0$ are exogenous. $k^i > k^{i-1}$ for any $i = 1, ..., N$, $w^j > w^{j-1}$ for any $j = 1, ..., N$. The value k^i is the lower bound for the amount of investment which allows the child to get a certain level of training bringing him wealth w^i. (Going forward, in the stochastic stepping-stone model we will consider random values $\widetilde{w}^i$ instead of given values w^i, in particular $w^i + s$, where s is some "noise".) The child's wealth will be w^{i+1} in case if investment exceeds the next threshold and so on. Thus, in the deterministic case it is assumed that going through certain kind of training assures getting predetermined amount of wealth. This means at least that all agents are homogeneous in terms of their abilities, luck, etc. which is certainly far from reality. Therefore, this model has to be perceived as a starting point in moving towards the stochastic case.

There is some example of stepping-stone $f(k)$ in the Figure 1. The value $w^0 > 0$ may be interpreted as some living cost benefit that every person gets no matter how small the parental investment was. However, we do not exclude the case $w^0 = 0$, and our results are entirely general.

What one can expect to see in this framework? There are a finite number of distinct investment levels. Some of these levels might be self-sustaining, which means that if a dynasty member makes an investment at that level, the return is such that the child will want to and be able to invest the same amount in the next generation. Other investment levels may be transitory. Some dynasties can "step up" through the generations to the higher levels, other families, on the contrary,

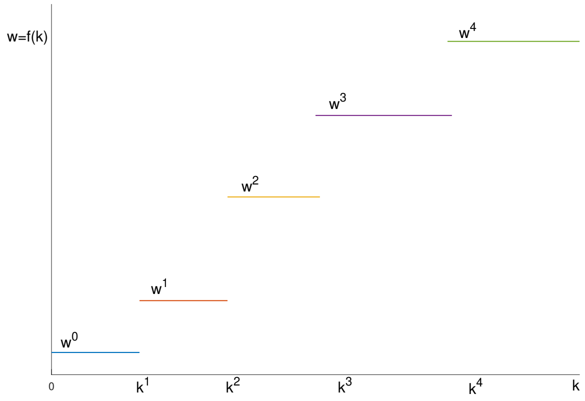

Figure 1: Stepping stones, example

may "step down" to the lower levels than the initial one. The situation evidently depends on exogenous values k^i, w^i and on the form of parental utility function, more precisely, on elasticity of substitution between parents' consumption and childrens' wealth (strictly speaking, additional assumptions on smoothness are needed to talk about the latter). For example, if sufficiently small levels of investments can lead the next generation to sufficiently large amount of wealth, then dynasties will typically move to higher wealth levels in case parents are not total egoists.

As for studying poverty, the model may exhibit "scarcity-driven" poverty traps (Ghatak (2015)). Under extreme scarcity the choice is done in favor of consumption. This in turn yields small outcome for the child who in turn will not be able to invest sufficient amount in her offspring either, and so on. The dynasty, therefore, stays in the poverty trap forever.

From the technical point of view the first observation one can make is that since utility strictly increases with consumption (A1), the optimal value of investment will be always at the thresholds.

Proposition 1. *If the transition function is defined by 3, then the following is true for the optimal policy:*

$$\pi(w) \subset \{0, k^1, ..., k^N\} \textit{ for any } w \geq 0. \tag{4}$$

Proof. Suppose for some i there exists $k^i < k < k^{i+1}$ such that $k \in \pi(w)$ for some wealth $w \geq 0$. Then $\max_{c,k} V(c,k) = U(w-k, w^i)$. At the same time $V(c,k^i) =$

$U(w-k^i, w^i) > U(w-k, w^i)$ because $U(x,y)$ strictly increases in x. This is a contradiction. Thus, $k \notin \pi(w)$. □

The intuition standing behind this statement is extremely simple. Since it is known to a parent that investment k such that $k^i < k < k^{i+1}$ does not give her child any more than investment k^i does, the parent prefers to use the difference $k - k^i$ for her own consumption. Thus, there is no point in investing anything between k^i and k^{i+1}, and the set of all optimal capital investments is finite.

One more simple observation that we can make using the Theorem 3 is given below.

Proposition 2. *If* $w'' > w'$, $k \in \pi(w')$ *and* $k \in \pi(w'')$, *then* $k \in \pi(w)$ *for any* w *such that* $w' < w < w''$.

Proof. Suppose the opposite. Then there exists $\overline{w}$ such that $w' < \overline{w} < w''$ and $k \notin \pi(\overline{w})$. According to Theorem 1 there exists $\overline{k} \neq k$ such that $\overline{k} \in \pi(\overline{w})$. From Theorem 3 we have $\overline{k} \geq k$ because $\overline{w} > w'$ and $k \in \pi(w')$, at the same time $k \geq \overline{k}$ because $w'' > \overline{w}$ and $k \in \pi(w'')$. This is a contradiction of $\overline{k} \neq k$. □

Note that it might happen that for some $1 \leq i \leq N$ there is no such amount of wealth $w \geq 0$ such that $k^i \in \pi(w)$. It actually means that some threshold levels of investment may be never optimal. This situation may confuse the following notations, thus, let us assume without any loss of generality that from the very beginning we deal with a stepping function of the form 3 such that for any k^i, $1 \leq i \leq N$ there exists more than one value of wealth $w \geq 0$ meeting the following condition: $k^i \in \pi(w)$. It will become clear why do we assume that from the proof of the next proposition that is a little less evident on the one hand and is a key moment for the whole paper on the other hand.

Proposition 3. *The non-negative semi-axis* $\mathbb{R}_+$ *can be represented as the union of intervals* $\bigcup_{i=0}^{N} W(i)$, *where* $W(0) = [0, w_1]$, $W(1) = [w_1, w_2]$, ..., $W(i) = [w_i, w_{i+1}]$, ..., $W(N-1) = [w_{N-1}, w_N]$, $W(N) = [w_N, +\infty)$, *the values* $0 = w_0 < w_1 < ... < w_N$ *depend on the values* w^0,...,w^N *and* k^1, ..., k^N, *such that*

1. k^i *is optimal inside of the interval* $W(i)$, $i = 0, ..., N$, *i. e.* $k^i \in \pi(w)$ *for all* w *such that* $w_i < w < w_{i+1}$, $i = 0, ..., N-1$, $k^N \in \pi(w)$ *for all* $w > w_N$;

2. *k^i is always optimal for w_i ($k^i \in \pi(w_i)$), i. e. at the left endpoint of the interval $W(i)$, $i = 0, ..., N$;*
3. *k^i may be or may be not optimal for w_{i+1}, i. e. at the right endpoint of the interval $W(i)$, $i = 0, ..., N-1$.*

Note that this case may degenerate to $\mathbb{R}_+ = [0, +\infty)$ when $k = 0$ is optimal for any values of wealth $w \geq 0$. Such situation is certainly not interesting for our further analysis.

Proof. First, let us prove that the set of wealth w such that there are at least two capital investments k' and k'' which are optimal for w (i. e. w such that $\#\pi(w) > 1$, where $\#$ stands for cardinality of a set) is finite.

Suppose $w' > w$ and $\#\pi(w') > 1$, $\#\pi(w) > 1$. According to Theorem 3 $\min\limits_{k^i \in \pi(w')} k^i \geq \max\limits_{k^j \in \pi(w)} k^j$. The set of all possible optimal investments contains $N+1$ elements (including $k^0 = 0$). Thus, the cardinality of the set of wealth for which π is multivalued is less or equal to N. It can equal N in the only one case, when the correspondence π is two-valued for every wealth w_i from this set and $\pi(w_i) = \{k^i, k^{i-1}\}$, $i = 1, ..., N$. According to the Proposition 2 in this case k^{i-1} is optimal in $[w_{i-1}, w_i]$ and k^i is optimal in $[w_i, w_{i+1}]$.

If π is more than two-valued at some point w', then $\min\limits_{k^i \in \pi(w')} k^i$ is optimal to the left from w' and $\max\limits_{k^i \in \pi(w')} k^i$ is optimal to the right. All the other $k^i \in \pi(w')$ are optimal only for the one value of wealth w'. To get $N+1$ intervals $W(i)$ such that k^i is optimal in the interior of $W(i)$ we assume that there is no k^i such that k^i is optimal for only one value of w.

"Switches" from k^i to k^{i+1} may happen abruptly meaning that k^i may be optimal in $[w_i, w_{i+1})$, and $\pi(w_{i+1}) = k^{i+1}$ is one-valued. It is easy to construct such examples. However, we can guarantee that k^i is always in $\pi(w_i)$. Suppose it is not, then some other $k \in \pi(w_i)$, it can be only k^{i-1} since k^{i-1} is optimal to the left from w_i. $k^{i-1} \in \pi(w_i)$ and $k^i \notin \pi(w_i)$ yields that $U(w_i - k^{i-1}, w^{i-1}) > U(w_i - k^i, w^i)$. Since the inequality is strict there exists $\varepsilon > 0$ such that $U(w_i - k^{i-1}, w^{i-1}) > U(w_i - k^i, w^i) + \varepsilon$. Using upper-semicontinuity of U in the first argument, for this $\varepsilon > 0$ we can find $\delta > 0$ such that $U(w_i + \delta - k^i, w^i) < U(w_i - k^i, w^i) + \varepsilon$. Taking into account the fact that k^i is optimal to the right from w_i and combining the inequalities from above, we get the following chain of inequalities: $U(w_i + \delta - k^{i-1}, w^{i-1}) < U(w_i + \delta - k^i, w^i) < U(w_i - k^i, w^i) + \varepsilon < U(w_i - k^{i-1}, w^{i-1})$. But this is a contradiction of the fact that U is strictly increasing in the first argument.

□

This statement tells us that we can partition the non-negative real numbers into the intervals $W(i)$ such that investment k^i is optimal in the interior of $W(i)$. k^i is optimal at the left endpoint of $W(i)$, and it may be or may be not optimal at the right endpoint. The thresholds of the intervals are denoted by w_i with the lower subscripts (recall that w^i with upper subscripts denote levels of childrens' wealth in 3). Thus, we know what amounts of wealth allow a parent to provide her child certain quality of education.

Let us try to give a reader some geometrical intuition of how the values w_i can be found, i. e. at what values of wealth we switch (in terms of optimality) from one capital investment level to another one. Recall the definition of function $V(c, k) = U(c, f(k))$ from 1. That means

$$V(c,k) = \begin{cases} U(c, w^0) & \text{if } k \ < k^1, \\ U(c, w^1) & \text{if } k^1 \leq k < k^2, \\ \dots & \\ U(c, w^N) & \text{if } k \geq k^N. \end{cases}$$

Since any intermediate investments (i. e. k such that $k^i < k < k^{i+1}$) do not allow parents to switch from $U(c, w^i)$ to $U(c, w^{i+1})$ that corresponds to higher level of child's wealth, the amount of capital $k - k^i$ should be consumed. Thus, only several capital investments can be optimal in case of the stepping-stone technology (see Proposition 1), and we have to compare with each other $V(w - k^i, k^i) = U(w - k^i, w^i)$ for different values of k^i. This comparison can be done for any amount of wealth w (note that capital investment k^i becomes feasible for a parent only for amount of wealth w such that $w \geq k^i$).

Figure 2 illustrates everything said above for the following example: $U(c, w) = log(c + 1) + log(w + 1)$, $k^0 = 0$, $w^0 = 0$, $k^1 = 1$, $w^1 = 2$, $k^2 = 2$, $w^2 = 4$, $k^3 = 3$, $w^3 = 12$. As the result, w_1 is the same as k^1, this means that if a parent has the amount of wealth k^1 she invests everything she has in the child's training. Thus, switching from zero investment to the investment level k^1 passes through zero consumption (theoretically such situation may emerge in this model). The next switching, from k^1 to k^2, takes place at the point w_2 that is somewhere in between k^2 and k^3 and that corresponds to nonzero consumption $w_2 - k_2$. The exact value of w_2 may be found as the solution of $V(w - k^1, k^1) = V(w - k^2, k^2)$ with respect to w. Note that both k^1 and k^2 are optimal for amount of wealth w_2, it means that $\#\pi(w_2) = 2$. And the last switching from k^2 to k^3 goes through zero consumption again and happens at $w = k^3$, the jump is observed.

This example also gives an illustration to the point 3 of Proposition 3, k^1 is optimal at the right endpoint of $W(1)$, at the same time k^0 is not optimal at the right endpoint of $W(0)$ and k^2 is not optimal at the right endpoint of $W(2)$.

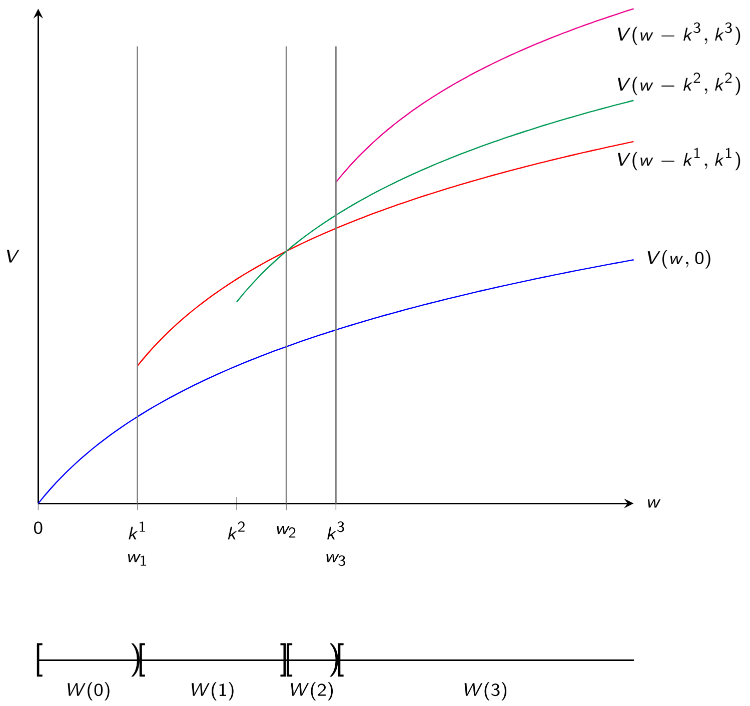

Figure 2: Switching from one level of investment to another

Dynasty's wealth dynamics is determined by an initial wealth $w(0)$ and by the relative positions of w^i and w_i. In the example described above

- a parent with initial wealth $w(0) < k^1 = 1$ invests zero in her child, thus, the child gets zero in the next period, and all the subsequent generations have zero wealth too;

- a parent with initial wealth $w(0)$ such that $w_1 = k^1 \leq w(0) < w_2$, i. e. $w(0) \in W(1)$, invests in the child k^1, the child in turn gets $w^1 \in W(1)$, consequently, the same level of parental investment k^1 is optimal again, and so on, the whole dynasty is stuck in $W(1)$, having the amount of wealth w^1;
- a parent with initial wealth $w(0) \in W(2)$ invests in the child k^2, the child gets $w^2 \in W(3)$, and, thus, invests in her child k^3, thus, the grandchild of the first person gets $w^3 \in W(3)$ for which k^3 is optimal again, therefore, such dynasty is "stepping up" from $W(2)$ to $W(3)$, having the amount of wealth w^3 starting from the grandchild of the progenitor;
- the level of investment k^3 is self-sustaining, thus, the dynasty starting with initial wealth $w(0) \in W(3)$ stays in $W(3)$ forever having w^3 for all the offsprings.

Theoretically it may happen that either w^i for some i or $w(0)$ is at the boundary of $W(j)$ for some j, in this case it's not guaranteed that an optimal investment is unique for this amount of wealth (recall point 3 of Proposition 3). Thus, there is no rule that tells us what of the two values of investment a parent has to choose. Since we talk here only about several possible values of wealth (the set of these values is finite) and since this is not of much interest, assume for the rest of the paper that neither $w(0)$ nor any of w^i are at the intervals boundaries. Then dynamics is strictly determined.

We can think of $W(i)$ as of wealth classes. Our model then predicts how a dynasty will be moving across the classes if the progenitor is from class $W(j)$ for some j (this means $w(0) \in W(j)$). Depending on the parameters of the model we can get different relative positions of w^i and w_i and, thus, different dynamic patterns. So dynamics can be arbitrary, some dynasties stay forever in the class they started from, some move to upper classes, some to lower. There are no cycles and there are no movements that cross each other. The latter means that for $w(0) > \widehat{w}(0)$ it is guaranteed that $w(T) \geq \widehat{w}(T)$ for any moment of time $T > 0$, where w and $\widehat{w}$ denote wealth of two different dynasties. Thus, a dynasty richer at the outset than another one never becomes poorer in the future. This in turn yields that there always exists some attractor because the number of classes is finite. The last observation is that if $w^0 = 0$ there is always a stable state $W(0)$.

Dynamics can be described by a graph. For example, Figure 3 describes the situation $w^0 \in W(0)$, $w^1, w^2, w^3 \in W(2)$. For any initial wealth $w(0)$, exceeding the left boundary of $W(1)$, i. e. w_1, dynasty's wealth converges to the values w^2.

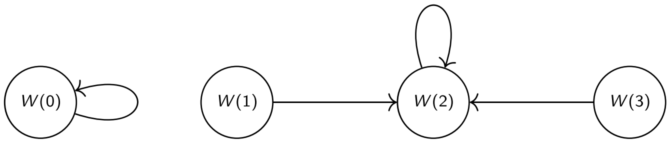

Figure 3: Dynamics described by a graph

Figure 4 gives one more example where we draw not only a graph, but relative positions of w^i and w_i as well. Note that as it was mentioned above dynamics also depends on the values of k^i, because w_i depend on both k^i and w^i. In Figure 4 $W(0)$ is an example of "poverty trap"[6], a low-wealth steady state. $W(3)$ is an example of "affluence trap", a high-wealth attractor. $W(1)$ is another self-sustaining class somewhere in-between. $W(2)$ is a transition state.

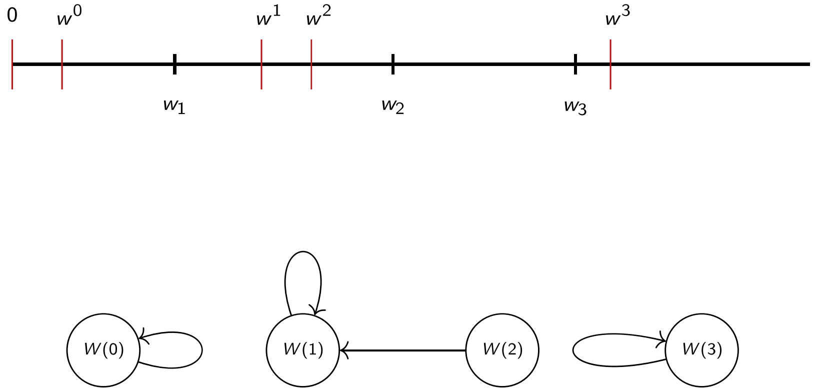

Figure 4: Dynamics described by a graph and relative positions of w^i and w_i

---

[6] The most commonly used in economic literature definition of a poverty trap is the following. "A poverty trap is any self-reinforcing mechanism which causes poverty to persist". (Azariadis and Stachurski (2005))

Summing up, the deterministic stepping-stone model describes the case in which certain levels of parental investment made in the period t guarantee that children will get certain amounts of wealth in the period $t+1$. In this framework we can identify wealth classes such that parents from different classes invest in their children different amounts of capital. Proceeding from this, we can analyze dynamics of wealth of the dynasty whose progenitor has given initial amount of wealth. The model may exhibit multiple inescapable steady states. Inescapable steady states that correspond to low wealth levels may be understood as poverty traps, in the opposite case we deal with affluence traps.

Thus, the deterministic environment provides theoretical foundations for understanding poverty (or affluence) traps. "Self-reinforcing mechanism" that underlies poverty persistence is impossibility for a parent to invest in a child enough to provide an access to good quality of education or training. Persistent poverty in this case is not a transitory phenomenon – the same families stay poor period after period. It is also evident that in the framework described in this section there is no mobility observed after a certain finite amount of periods so that eventually any movements across wealth classes just vanish. In the next section we allow mobility to exist in any given time period considering the stochastic environment.

2.4 The Stochastic Stepping-Stone Model

The Evolution of the Distribution of Wealth

Now consider the case when the child's outcome is random – the return to parental investment depends on unobservables. A definite level of parental investment allows the child to get a definite education level (or level of training). At the same time as the result the child gets a random amount of wealth that depends on his abilities etc. We assume that though this uncertainty is uncovered in the period, next to the period when parental investment is made, a parent has correct beliefs about these random factors that will affect the child's wealth at the moment she makes the decision how much to invest in her child. Technically it means that the parent knows the distribution of the child's wealth. It presumes that a parent properly estimates the child's abilities, the level of unpredictability of how the real child's wealth outcome will depend on the quality of education and training she got etc. We realize that this is a pretty strong assumption but it seems difficult to imagine some kind of theoretical modelling without similarly vigorous assumptions.

So here we deal with the transition function that was used in section 2.2, i. e. with $f(k,s)$, where s is a random variable that embraces the random factors described above. $f(k,s)$ is defined as follows:

$$f(k,s) = \begin{cases} \max\{0, w^0 + s\} & \text{if } k \ < k^1, \\ \max\{0, w^1 + s\} & \text{if } k^1 \le k < k^2, \\ \dots \\ \max\{0, w^N + s\} & \text{if } k \ge k^N, \end{cases} \tag{5}$$

where all the values $k^i > k^0 = 0$ and $k^i > k^{i-1}$ for any $i = 1, ..., N$, $w^j > w^{j-1}$ for any $j = 1, ..., N$, s is a continuously distributed[7] random variable. In this paper we focus on the case $\widetilde{w}^j = max\{0, w^j + s\}$, but it is not difficult to track that some of our results are true for the general case of random variables $\widetilde{w}^j(s)$ once we have the following assumptions: $\widetilde{w}^j(s)$ are continuously distributed non-negative random variables that strictly increase with j in the sense of stochastic dominance (note that both these conditions are fulfilled for $\widetilde{w}^j = max\{0, w^j + s\}$). In this case increasing in the sense of stochastic dominance is modification of the assumption A2.

Let $g^j(s)$ denote the density of $\widetilde{w}^j(s)$. $V(c,k) \equiv E_\mu U(c, f(k,s))$, defined in 1, looks as follows:

$$V(c,k) = \begin{cases} U_0(c) & \text{if } k \ < k^1, \\ U_1(c) & \text{if } k^1 \le k < k^2, \\ \dots \\ U_N(c) & \text{if } k \ge k^N, \end{cases}$$

where the sequence of functions $U_j(c) = \displaystyle\int_0^{+\infty} U(c, w^j(s)) g^j(s) ds$, $j = 0, 1, ..., N$, is point-wise strictly increasing due to the assumptions that $\widetilde{w}^j$ strictly increase with j in the sense of stochastic dominance and A1. Here we evidently assume that $U(c, w^j(s)) g^j(s) ds$ is Lebesgue integrable on the positive half-axis, i. e.

$$U(c, w^j(s)) g^j(s) \in L^1([0, +\infty))$$

for all $0 \le j \le N$ (this is actually reformulation of A4).

---

[7] Assumption about continuous distribution is crucial in the following sense. In the previous section we assumed that none of w^i is not at a boundary of any $W(j)$. When we transit from the deterministic case to stochastic we have to reassure ourselves that none of $\widetilde{w}^i$ cannot "drop" on a boundary or that this event has a zero probability.

Then it is not difficult to understand that propositions of section 2.3 are true here as well. For example, any intermediate investments (i. e. k such that $k^i < k < k^{i+1}$) do not allow parents to switch from $U_i(c)$ to the next regime of consumption $U_{i+1}(c)$ that corresponds to higher level of child's wealth and, thus, the amount of capital $k - k^i$ should be consumed. Propositions 1 and 2 are true without any changes. Proposition 3 has to be modified in an evident way.

Proposition 4. *The non-negative semi-axis* $\mathbb{R}_+$ *can be represented as the union of the intervals* $\bigcup_{i=0}^{N} W(i)$, *where* $W(0) = [0, w_1]$, $W(1) = [w_1, w_2]$, ..., $W(i) = [w_i, w_{i+1}]$, ..., $W(N-1) = [w_{N-1}, w_N]$, $W(N) = [w_N, +\infty)$, *the values* $0 = w_0 < w_1 < ... < w_N$ *depend on the values* k^1, ..., k^N *and on the distributions of* $\widetilde{w}^0$, $\widetilde{w}^1$, ..., $\widetilde{w}^N$, *such that*

1. k^i *is optimal inside of the interval* $W(i)$, $i = 0, ..., N$, *i. e.* $k^i \in \pi(w)$ *for all* w *such that* $w_i < w < w_{i+1}$, $i = 0, ..., N-1$, $k^N \in \pi(w)$ *for all* $w > w_N$;
2. k^i *is always optimal for* w_i ($k^i \in \pi(w_i)$), *i. e. at the left endpoint of the interval* $W(i)$, $i = 0, ..., N$;
3. k^i *may be or may be not optimal for* w_{i+1}, *i. e. at the right endpoint of the interval* $W(i)$, $i = 0, ..., N-1$.

We repeat this proposition intentionally in order to draw reader's attention to the fact that the endogenously determined boundaries of the wealth classes $W(i)$ are certainly not random in the stochastic case, but they do depend on the distributions of random values $\widetilde{w}^j(s)$, i. e. on all the values w^j and on the distribution of s in the case $\widetilde{w}^j = max\{0, w^j + s\}$.

We turn now to analysis of the evolution of wealth, i. e. to study of dynasties movements across different wealth classes. This process was determined in section 2.3, and we could say what exactly would be the wealth of a dynasty's member depending on the value of the progenitor's wealth. We knew how the dynasty would be moving across wealth classes and we used graphs to visualize this process. Since child's wealth levels are random now, in this section we deal with the stochastic process, and the only thing we can talk about is the probabilistic distribution of wealth. However, we still can analyze the evolution of wealth thinking of the evolution of the distribution of wealth. This process is Markovian because the wealth distribution of any fixed generation depends on parental investment which is determined by parental wealth and, thus, does not depend on the previous generations wealth.

Recall that the evolution of wealth in the deterministic case is defined once we assumed that no w^i is on the boundary of any $W(j)$. We had to assume this since $\pi(w)$ (see definition 1) is multivalued for some amounts of wealth w. For the same reason here we start with some selection $\bar{\pi}$ from correspondence π. This essentially means that we just forcedly choose one concrete value of optimal investment for such values of wealth that give us several optimal k according to the optimization problem 1-2. All selections differ from each other only at the wealths w_i, where both k^{i-1} and k^i are optimal. Let us fix some selection $\bar{\pi}$ from π. The selection $\bar{\pi}$ describes Markov process. For measurable $A \subseteq \mathbb{R}_+$ the probability that the dynasty's generation $t+1$ has the wealth $w_{t+1} \in A$ depends only on the wealth of the previous generation t and is described as follows:

$$P_{\bar{\pi}}(w_{t+1} \in A | w_0, \ldots, w_t) = P_{\bar{\pi}}(w_{t+1} \in A | w_t) = P_{\bar{\pi}}(w_t, A) = \int_A g^{\bar{i}}(s) ds, \tag{6}$$

where $\bar{i}$ equals i such that $k^i \in \bar{\pi}(w_t)$. This process is not a “textbook” example of Markov process because of the jumps at w_i, more precisely, $P_{\bar{\pi}}(w_t, A)$ is not continuous in w_t. Nonetheless, as we show below, under some standard assumptions the process has a unique invariant distribution. Furthermore, the generation t distribution of wealth converges weakly to this invariant distribution from any initial wealth distribution.

A parent with wealth w such that $w_i < w < w_{i+1}$ (w is in interior of $W(i)$) chooses capital investment k^i. The child's wealth will be $\widetilde{w}^i(s)$, drawn from density $g^i(s)$. Let p_{ij} denote the probability that $\widetilde{w}^i \in W(j)$, $i = 0, ..., N$, $j = 0, ..., N$:

$$p_{ij} = P(w_{t+1} \in W(j) | w_t \in W(i)) = \int_{W(j)} g^i(s) ds. \tag{7}$$

Thus, p_{ij} is the probability that a child will be in the class $W(j)$ if her parent has wealth w such that $w_i < w < w_{i+1}$. This value does not depend on the selection[8] and on the time period we are looking at. Row vector $p_{i*} = [p_{i0}, p_{i1}, ..., p_{iN}]$ may be interpreted as the lottery faced by a child whose parent belongs to class $W(i)$.

---

[8] Note that the probability of finding a child in a class $W(j)$ if her parent has wealth w_i or w_{i+1} indeed depends on the selection $\bar{\pi}$, but it does not affect the evolution of the distribution of wealth as a reader will see in a moment.

Denote the distribution of wealth over $\mathbb{R}_+$ at time period t by ν_t. Then in the next period of time $t+1$ for any measurable $A \subseteq \mathbb{R}_+$ we will have

$$\nu_{t+1}(A) = \int_{\mathbb{R}_+} P_{\bar{\pi}}(w_t, A) d\nu_t = \int_{\bigcup_{i=0}^{N} W(i)} P_{\bar{\pi}}(w_t, A) d\nu_t =$$

$$= \sum_{i=0}^{N} \int_{W(i)} P_{\bar{\pi}}(w_t, A) d\nu_t = \sum_{i=0}^{N} \nu_t(W(i)) \int_A g^i(s) ds. \quad (8)$$

This equation entirely describes the evolution of the distribution of wealth ν_t once the initial wealth distribution ν_0 is given. Note that dependence on the selection $\bar{\pi}$ is lost in the last equality. Two selections $\bar{\pi}$ and $\bar{\bar{\pi}}$ from correspondence π can differ from each other only at the points w_i, but $\nu_t(w_i) = 0$ which essentially means that the probability of drawing any concrete value of wealth, including w_i, is zero. Therefore, a concrete choice of the selection does not affect the distribution of wealth.

First, from equation 8 we see that the distribution ν_{t+1} over $\mathbb{R}_+$ is actually defined through the numbers $\{\nu_t(W(j))\}_{j=0}^{N}$, i. e. the measure ν_t matters for ν_{t+1} only through these numbers. Second, using 8, we immediately can describe the evolution of these numbers in time as follows:

$$\nu_{t+1}(W(j)) = \sum_{i=0}^{N} \nu_t(W(i)) \int_{W(j)} g^i(s) ds = \sum_{i=0}^{N} \nu_t(W(i)) \cdot p_{ij}, \ j = 0, ..., N. \quad (9)$$

The matrix $P = \{p_{ij}\}_{i,j=0}^{N}$ is $(N+1) \times (N+1)$ Markov matrix, and equation 9 can be rewritten in the vector-matrix notations as follows:

$$(\nu_{t+1}(W(0), \ldots, \nu_{t+1}(W(N)) = (\nu_t(W(0), \ldots, \nu_t(W(N)) \cdot P. \quad (10)$$

The Invariant Distribution

It is a well known fact that if the matrix P is irreducible then there is an invariant probability distribution for the process described by 10. Let us denote this invariant probability vector by $q^* = (q_0^*, ..., q_N^*)$. If the matrix P is in addition primitive, then the invariant vector is unique and the probability vector $(\nu_t(W(0), \ldots, \nu_t(W(N))$ converges to q^* as $t \to \infty$.

Then, assuming that the matrix P is primitive and going back to the process of the evolution of wealth given by 8, we can state that it has a unique invariant distribution ν^* defined as follows:

$$\nu^*(A) = \sum_{i=0}^{N} q_i^* \int_A g^i(s)ds \tag{11}$$

for any measurable $A \subseteq \mathbb{R}_+$. The distribution ν_t converges to ν^*.

Let us assume for the rest of the section that the matrix P is primitive (and, subsequently, irreducible). Irreducibility means that for any pair (i,j) there is k such that $p_{ij}^k > 0$, where p_{ij}^k is the (i,j) element of the matrix P^k. The latter suggests that in a graph associated with P (this graph has $N+1$ vertices, and an edge connects the vertex i to the vertex j if and only if $p_{ij} > 0$) there is a path of length k from i to j, so the graph is strongly connected. In our case one can interpret this as that any dynasty has a non-zero probability of moving from any class $W(i)$ to any class $W(j)$ in k steps, where k depends on i and j. Primitivity is a stronger assumption indicating that there is k such that for any i and j the probability of moving from any $W(i)$ to any $W(j)$ in k periods of time is non-zero. If the matrix is irreducible, then primitivity is guaranteed if P has at least one non-zero diagonal element, in our case this means that there exists i such that g^i is strictly positive on some open interval in $W(i)$. The latter can be interpreted as that there is some class $W(i)$ such that if a dynasty is in $W(i)$ at time period t then it will stay in $W(i)$ at time period $t+1$ with non-zero probability.

Summing up, if all our assumptions are fulfilled, we have

Theorem 4. *Assume that the* $(N+1)\times(N+1)$ *matrix* $P = \{p_{ij}\}_{i,j=0}^N$, *defined by (7), is irreducible and that there exists the density function* g^i *which is strictly positive on some open interval in* $W(i)$, $i = 0, 1, ..., N$. *Then the Markov process with any initial distribution* ν_0 *and transition probability* $P_{\bar{\pi}}$, *defined in (6), is ergodic. The invariant distribution* ν^* *has density* $\sum_{i=0}^{N} q_i^* g^i(s)$ *where* q^* *is the invariant vector of the matrix* P. *Finally,* ν^* *is independent of a selection* $\bar{\pi}$.

As an attentive reader could notice, until this moment we never used the special form of the random values $\widetilde{w}^j = max\{0, w^j + s\}$. The next question we are going to study is whether the picture we observed in the deterministic case will give us the right impression of what is going on in the stochastic case with $\widetilde{w}^j = max\{0, w^j + s\}$ for low levels of "noise" s.

Local Comparative Statics

Suppose the utility function is additively separable, that is $U(x,y) = Q(x) + B \cdot R(y)$, where $B > 0$ is a parameter characterizing the degree of intergenerational altruism (the greater B, the more bountiful each parent is). To fulfill assumptions of section 2.2 we need to suppose that $Q(x)$ and $R(y)$ are increasing non-negative functions of their arguments. Also, let us assume that $Q(\cdot) \in C^1$, $Q''(\cdot)$ exists and $Q''(\cdot) < 0$ ($U(x,y)$ must be strictly concave in x), $R(\cdot)$ is upper-semicontinuous.

Then in the deterministic case $V(c,k) = U(c, f(k)) = Q(c) + B \cdot R(f(k))$ and

$$V(c,k) = \begin{cases} Q(c) + B \cdot R(w^0) & \text{if } k \ < k^1, \\ Q(c) + B \cdot R(w^1) & \text{if } k^1 \le k < k^2, \\ \dots & \\ Q(c) + B \cdot R(w^N) & \text{if } k \ge k^N. \end{cases}$$

As before, to find w_i we need to compare $V(w - k^i, k^i)$ for different values of k^i. Assuming that there are no switches from one investment to another through the zero consumption, w_i, $1 \le i \le N$, may be found as solutions of the following equations with respect to w:

$$Q(w - k^{i-1}) + B \cdot R(w^{i-1}) = Q(w - k^i) + B \cdot R(w^i), \tag{12}$$

where $k^0 = 0$.

The latter may be rewritten as

$$Q(w - k^{i-1}) - Q(w - k^i) = B \cdot \left(R(w^i) - R(w^{i-1}) \right), \tag{13}$$

where the right-hand side is greater than zero. The left-hand side is a function of w. Since the marginal utility of consumption $Q'(\cdot)$ strictly decreases, $Q'(w - k^{i-1}) < Q'(w - k^i)$. Thus, $Q(w - k^{i-1}) - Q(w - k^i)$ strictly decreases with respect to w. Consequently, for larger B we will have smaller w_i as the solution of (12). Intuitively, that means a more altruistic parent needs to have smaller wealth w to switch to a larger investment than a less altruistic one.

Finally, we just concluded that values $\widehat{w}_i$ corresponding to $\widehat{B} > B$ are smaller than values w_i that correspond to B. If originally we have a situation where all capital investments appear as optimal ones for some w and where a parent does not switch from k^i to k^{i+1} through the zero consumption, we may destroy this ideal picture perturbing B much. Thus, we talk only about small perturbations of B and study the local comparative statics. To be more precise, under our assumptions w_i are continuously differentiable (the implicit function theorem applied to (12)) in B and $\dfrac{\partial w_i}{\partial B} < 0$.

Let us consider a stochastic case now. Suppose that $\widetilde{w}^i(s)$ are continuously distributed (with $g^i(s)$ densities) non-negative random variables that strictly increase with i in the sense of stochastic dominance. In this case

$$V(c,k) = \begin{cases} Q(c) + B \cdot \left(\displaystyle\int_0^{+\infty} R(\widetilde{w}^0(s)) g^0(s) ds \right) & \text{if } k \ < k^1, \\ Q(c) + B \cdot \left(\displaystyle\int_0^{+\infty} R(\widetilde{w}^1(s)) g^1(s) ds \right) & \text{if } k^1 \le k < k^2, \\ \dots & \\ Q(c) + B \cdot \left(\displaystyle\int_0^{+\infty} R(\widetilde{w}^N(s)) g^N(s) ds \right) & \text{if } k \ge k^N. \end{cases}$$

It is clear that all argumentation made above may be repeated so that values $\widehat{w}_i$ corresponding to $\widehat{B} > B$ are smaller than values w_i that correspond to B. Analogue of equation 13 is the following:

$$Q(w - k^{i-1}) - Q(w - k^i) = B \cdot \left(\int_0^{+\infty} R(\widetilde{w}^i(s)) g^i(s) ds - \int_0^{+\infty} R(\widetilde{w}^{i-1}(s)) g^{i-1}(s) ds \right),$$

thus, the right-hand side is positive in this case too ($ER(\widetilde{w}^i) - ER(\widetilde{w}^{i-1}) > 0$ because $\widetilde{w}^i$ increase with i in the sense of stochastic dominance).

Next, let us analyze how the matrix P changes if we increase B to $\widehat{B}$ that is enough close to B. Denote corresponding to B probabilities of moving from class i to j by p_{ij}, and corresponding to $\widehat{B}$ – by $\widehat{p}_{ij}$.

First, we know that boundaries w_i of all classes move to the left. That means that the lowest class $W(0)$ becomes narrower and the highest class $W(N)$, on the contrary, becomes wider. $p_{i0} = \displaystyle\int_0^{w_1} g^i(s) ds$ and $p_{iN} = \displaystyle\int_{w_N}^{+\infty} g^i(s) ds$. Thus, $\widehat{p}_{i0} < p_{i0}$ and $\widehat{p}_{iN} > p_{iN}$ for any $0 \le i \le N$ (intuitively, it becomes easier to get to the highest class and more difficult to drop to the lowest class from any class). All other classes may shrink or stretch depending on whether the left boundary moves more than the right one or vice versa. However, we can guarantee that $\widehat{p}_{i0} + \widehat{p}_{i1} \leq p_{i0} + p_{i1}$ because w_2 moves to the left; $\widehat{p}_{i0} + \widehat{p}_{i1} + \widehat{p}_{i2} \leq p_{i0} + p_{i1} + p_{i2}$ because w_3 moves to the left; and so on up to the point $\widehat{p}_{i0} + \widehat{p}_{i1} + \widehat{p}_{i2} + ... + \widehat{p}_{i,N-1} \leq p_{i0} + p_{i1} + p_{i2} + ... + p_{i,N-1}$, for any $0 \leq i \leq N$. Here we used that for $k \leq N-1$

$$p_{i0} + p_{i1} + p_{i2} + ... + p_{ik} = \int_0^{w_{k+1}} g^i(s)ds.$$

The inequalities listed above imply that the matrix $\widehat{P}$ stochastically dominates the matrix P [9].

$\widehat{P} \succcurlyeq P$ is not enough to guarantee the corresponding stochastic dominance relation for invariant distributions (see Dardanoni (1995)), but if we add monotonicity[10] then it becomes sufficient. In our framework each row stochastically dominates the row above it (since $\widetilde{w}^i$ increase with i, a child of the parent from class j faces a better lottery than a child of the parent from class i, where $j > i$). Thus, we have both monotonicity and $\widehat{P} \succcurlyeq P$, so $\widehat{q}^* \succcurlyeq q^*$, where $\widehat{q}^*$ and q^* are the invariant distributions of $\widehat{P}$ and P, correspondingly. This means that if parents are more altruistic, then a member of any dynasty (no matter what class a progenitor belongs to) faces a better lottery in the long run.

If instead of increasing altruism we switch from one production function to another, where the second one "pointwise" stochastically dominates the other, meaning that $\widehat{w}^i$ stochastically dominates $\widetilde{w}^i$ for every i, things will get more complicated. We still need to assume that perturbations are small, so that we do not destroy the picture completely (that means we need $E\widehat{w}^i$ to be enough close to $E\widetilde{w}^i$ for each i). Even in the case of an additively separable utility function (and even for a linear R), we need to assume that for every i

$$ER(\widehat{w}^i) - ER(\widehat{w}^{i-1}) > ER(\widetilde{w}^i) - ER(\widetilde{w}^{i-1}).$$

And this has something to do with both function R and how perturbations of $\widetilde{w}^i$ and $\widetilde{w}^{i-1}$ are compared to each other. Intuitively these two aspects (what are expected

---

[9] According to the definitions vector $\widehat{\pi}$ stochastically dominates vector π, $\widehat{\pi} \succcurlyeq \pi$, if $\widehat{\pi}_1 + \widehat{\pi}_2 + ... + \widehat{\pi}_k \leq \pi_1 + \pi_2 + ... + \pi_k$ for any $k = 1, 2, ..., n-1$, where n is the size of vectors. Matrix $\widehat{P}$ stochastically dominates matrix P, $\widehat{P} \succcurlyeq P$, if $\pi\widehat{P} \succcurlyeq \pi P$ for any vector π. $\widehat{P} \succcurlyeq P$ is equivalent to $\widehat{P}T \leq PT$, where T is the upper-triangular matrix with zeros below the main diagonal and ones elsewhere (postmultiplying P by T transforms each row to a cumulative density, and that gives us that $\pi\widetilde{P} \succcurlyeq \pi P$ for any distribution π). For example, see Dardanoni (1995), and references there.

[10] A matrix is called monotone if each row stochastically dominates the row above it.

changes of child's wealth and how these changes affect parental utility) indeed must matter.

The answer will be straightforward if we consider a linear[11] function R and $\widetilde{w}^j = max\{0, w^j + s\}$, where noise s is continuously and symmetrically distributed around zero, slightly perturbing variation of s. In that case $ER(\widehat{w}^i) - ER(\widehat{w}^{i-1}) > ER(\widetilde{w}^i) - ER(\widetilde{w}^{i-1})$ for each i, where $\widehat{w}^i$ corresponds to the larger variance. Thus, in this special case a parent will switch to a larger investment, having less wealth, if the variance of the noise increases[12]. As before, that would lead a member of any dynasty to a better lottery in the long run.

Also, the stochastic dominance result will hold if we perturb only $\widetilde{w}^0$ or only $\widetilde{w}^N$. Suppose $\widehat{w}^0$ stochastically dominates $\widetilde{w}^0$ and all other values remain unchanged. Then w_1 moves to the left, and all other w_i remain the same. All argumentation made previously holds, and the invariant distribution, corresponding to the perturbed production function, stochastically dominates the invariant distribution of the unperturbed. If $\widehat{w}^N$ stochastically dominates $\widetilde{w}^N$, w_N moves to the left and, consequently, the same result is true. However, if we perturb $\widetilde{w}^i$, where $i \neq 0, N$, w_i moves to the left, but w_{i+1} moves to the right and, thus, stochastic dominance results are no longer necessarily true.

Note that when we consider perturbations of parameters of the parental utility function, only the invariant distribution changes. If we disturb some parameters of the production function $f(k, s)$, that affects both the invariant distribution and wealth levels if we disturb w^i (or optimal investment levels if we disturb k^i values).

3 Stochastic Stability of the Deterministic Steady States

Randomness implies that a given investment can "land" a child anywhere. In this section we want to understand how the stochastic process of the evolution of wealth tracks the deterministic process. To do this we will examine what happens as the variance in the noise term collapses to zero. We parametrize the noise and analyze how the stationary distribution of wealth behaves as the "noise" disappears. One

---

[11] If R is not linear, the result will depend on convexity of R at different w^i.

[12] This depends on the fact that we cut random wealth at zero.

might expect that the deterministic stable states (including maybe poverty and affluence traps) would be seen in the invariant distribution as modes, where mass piles up in the long run (the corresponding steady states are called stochastically stable). A dynasty in any year would have a high probability of being at one or another of the deterministic-self-sustaining states, and low probability of being in a transition state. This view turns out to be false. For high levels of stochasticity, the modes are barely visible. For low levels, there will be multiple modes, but typically one will dominate the others, and that domination increases as noise is reduced. In the low-noise limit, invariant distributions converge to point-mass on a single state.

Stochastically stable states are identified by introducing a family of random perturbations into a deterministic system in a reasonable way, and finding the limit of invariant distributions as the perturbations shrink to 0. Let us start studying this problem with considering some example that also must give a good illustration for the previous sections.

3.1 Example

First we turn to the deterministic case. Consider the utility function $U(c, w) = log(c + 0.1) + 1.5w$ and the stepping-stone transition fuction $f(k)$ of the following form

$$f(k) = \begin{cases} 0.3 & \text{if } k \ < 1, \\ 1 & \text{if } 1 \leq k < 2 \\ 2.4 & \text{if } 2 \leq k < 3 \\ 3.5 & \text{if } 3 \leq k < 4.5 \\ 4.5 & \text{if } k \ \geq 4.5 \end{cases} \tag{14}$$

Note that the functions U and f satisfy all the assumptions we made above.

First, find the values of w_i, i. e. the boundaries of the wealth classes $W(i)$. Figure 5 is analogue of the Figure 2 presented in section 2.3. The example is constructed in the way that there are no jumps corresponding to parental zero consumption. Analytic expressions for the values w_i are

$$w_i = \frac{(k^i - 0.1)exp(1.5(w^i - w^{i-1})) - (k^{i-1} - 0.1)}{exp(1.5(w^i - w^{i-1})) - 1}, \quad 1 \leq i \leq 4. \tag{15}$$

Computing the wealth classes $W(i)$, we get $W(0) \approx [0; 1.44]$, $W(1) \approx [1.44; 2.04]$, $W(2) \approx [2.04; 3.14]$, $W(3) \approx [3.14; 4.83]$, $W(4) \approx [4.83; +\infty)$.

Relative positions of w_i and w^i and the graph describing movements across the wealth classes are presented in the Figure 6.

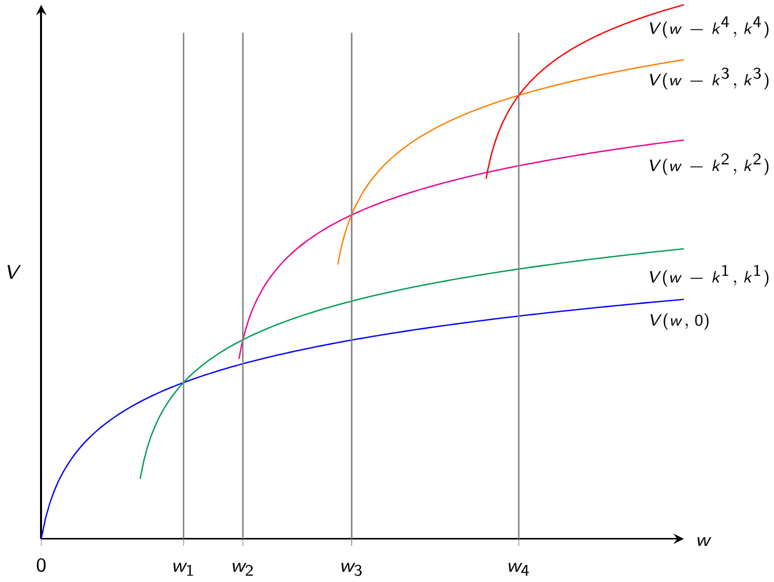

Figure 5: Switching from one level of investment to another, $U(c, w) = log(c+0.1)+1.5w$, $f(k)$ of the form 14.

Now let us perturb the values w^j from 14 and consider the stochastic stepping-stone transition function with random wealths $\widetilde{w}^j = max\{0, w^j + s\}$, where the random variable s has Laplace distribution, i. e. the density of s is $h_\lambda(s) = \dfrac{\lambda}{2} exp(-\lambda|s|)$. λ is a parameter that allows us to control the strength of "noise". $\widetilde{w}^j = max\{0, w^j + s\}$ shrinks around the value w^j as λ becomes large. In the Figure 7 one can see how the distribution of s shrinks around zero as λ grows.

As we know, in this case the values w_i depend on the distribution of s, i. e. they depend on the parameter λ. In our example we can find them analytically

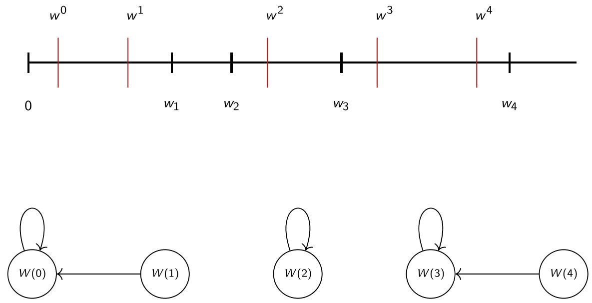

Figure 6: w_i relatively to w^i and the dynamics graph, $U(c, w) = log(c + 0.1) + 1.5w$, $f(k)$ of the form 14.

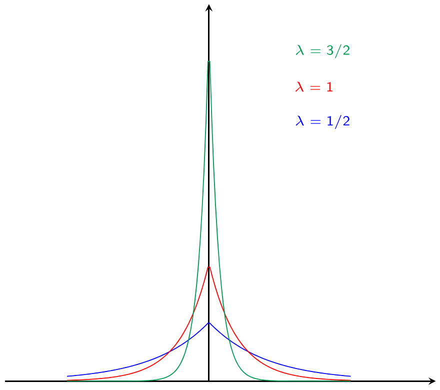

Figure 7: Laplace distribution

since it is possible to find an expected value $E_\lambda U(c, f(k, s))$ directly. Evidently $E_\lambda U(c, f(k, s)) = log(c + 0.1) + 1.5E_\lambda \widetilde{w}^i$ for $k^i \leq k < k^{i+1}$, where

$$1.5E_\lambda \widetilde{w}^i = 1.5 \int_{-w^i}^{+\infty} (w^i + s) h_\lambda(s) ds = 1.5 \left(w^i + \frac{1}{2\lambda} exp(-\lambda w^i) \right).$$

For brevity let us introduce the notation $b(\lambda, w^i) = \dfrac{1}{2\lambda} exp(-\lambda w^i)$. Note that as λ goes to infinity $b(\lambda, w^i)$ goes to zero, thus, $1.5E_\lambda \widetilde{w}^i \to 1.5w^i$ as $\lambda \to +\infty$ so that $E_\lambda U(c, f(k, s)) \to U(c, f(k))$.

Actually for fixed λ the problem of finding $w_i(\lambda)$ is not different from the problem of finding w_i in the deterministic case. Analytic expressions for $w_i(\lambda)$ look as follows:

$$w_i(\lambda) = \frac{(k^i - 0.1)exp\left(1.5(w^i - w^{i-1} + b(\lambda, w^i) - b(\lambda, w^{i-1}))\right) - (k^{i-1} - 0.1)}{exp\left(1.5(w^i - w^{i-1} + b(\lambda, w^i) - b(\lambda, w^{i-1}))\right) - 1}.$$

$w_i(\lambda) \to w_i$ as $\lambda \to +\infty$. An interested reader can also check that $w_i < w_i(\lambda)$ for any i and for any λ.

Knowing the exact values of $w_i(\lambda)$, compute the class transition probabilities.

$$p_{ij}(\lambda) = \begin{cases} \dfrac{\lambda}{2}\displaystyle\int_{-\infty}^{w_1(\lambda)} exp(-\lambda\left|s - w^i\right|)ds & \text{if } j = 0, \\ \dfrac{\lambda}{2}\displaystyle\int_{w_j(\lambda)}^{w_{j+1}(\lambda)} exp(-\lambda\left|s - w^i\right|)ds & \text{if } j \neq 0, N, \\ \dfrac{\lambda}{2}\displaystyle\int_{w_N(\lambda)}^{+\infty} exp(-\lambda\left|s - w^i\right|)ds & \text{if } j = N. \end{cases}$$

Final expressions for $p_{ij}(\lambda)$ are the following. For $j \neq 0, N$ and $0 \leq i \leq N$

$$p_{ij} = \begin{cases} \frac{1}{2}\left(exp\left[-\lambda(w_j - w^i)\right] - exp\left[-\lambda(w_{j+1} - w^i)\right]\right) & \text{if } w^i < w_j, \\ \frac{1}{2}\left(2 - exp\left[-\lambda(w^i - w_j)\right] - exp\left[-\lambda(w_{j+1} - w^i)\right]\right) & \text{if } w_j < w^i < w_{j+1}, \\ \frac{1}{2}\left(exp\left[-\lambda(w^i - w_{j+1})\right] - exp\left[-\lambda(w^i - w_j)\right]\right) & \text{if } w^i > w_{j+1}. \end{cases}$$

The last cases to consider are $j = 0$, $j = N$ with $0 \leq i \leq N$:

$$p_{i0} = \begin{cases} \frac{1}{2}exp(-\lambda(w^i - w_1)) & \text{if } w^i > w_1, \\ \frac{1}{2}(2 - exp(-\lambda(w_1 - w^i))) & \text{if } w^i < w_1. \end{cases}$$

$$p_{iN} = \begin{cases} \frac{1}{2}exp(-\lambda(w_N - w^i)) & \text{if } w^i < w_N, \\ \frac{1}{2}(2 - exp(-\lambda(w^i - w_N))) & \text{if } w^i > w_N. \end{cases}$$

Knowing the matrix $P(\lambda)$, we can find invariant distributions for different values of λ directly. Figure 8 depicts invariant distribution over the classes $W(i)$ as λ grows. The lines a reader can see in the graphs are $q_0^*(\lambda) = \nu_\lambda^*(W(0))$,..., $q_4^*(\lambda) = \nu_\lambda^*(W(4))$. This numerical study shows that only the class $W(0)$ which can be regarded as the poverty trap is stochastically stable. The latter means that

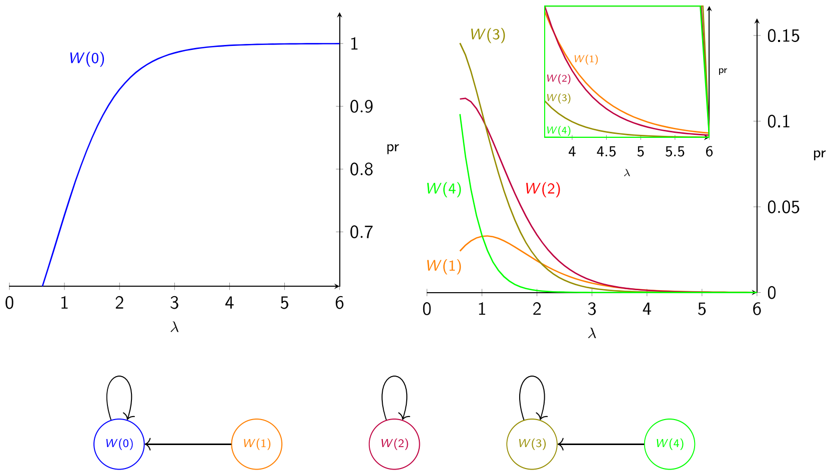

Figure 8: Stochastic stability, Laplace distribution numerical example.

$\lim_{\lambda \to +\infty} q_0^*(\lambda) = 1$ and $\lim_{\lambda \to +\infty} q_i^*(\lambda) = 0$ for $1 \le i \le 4$, i. e. for small noises (for large λ) invariant distribution mass piles up in the long run over the class $W(0)$ exclusively.

Another interesting observation is the following. One might expect that for large values of λ deterministic attractors would "compete" with each other, more precisely, for λ larger than some $\bar{\lambda}$ it would be true that $q_0^*(\lambda)$, $q_2^*(\lambda)$ and $q_3^*(\lambda)$ are larger than $q_1^*(\lambda)$ and $q_4^*(\lambda)$. It turns out that this expectation is false as zoom in on the part of the graph on the right shows (see Figure 8).

Next, we turn to studying a more general case. We will get back to this example at the end of the next section to illustrate some results holding in this more general case.

3.2 More General Case

Now, turn to a more general class of distributions of s that is possible to analyze theoretically. Suppose that s has a density $h_i(s;\lambda)$ on $\mathbb{R}$

$$h_\lambda(s) = \exp(-\lambda h(s))/Z(\lambda), \tag{16}$$

where $Z(\lambda)$ is a normalizing constant such that $\int_{\mathbb{R}} h_\lambda(s)ds = 1$. Note that Laplace distribution considered above submerges into this class. However, in this section we shall assume that the non-negative function $h \in C^2$ (that is not true for Laplace distribution) has a minimum of 0 at 0, and $h''(0) > 0$.

- AI. $h(s)$ is non-negative, C^2-smooth;
- AII. $h(s)$ has a global minimum at 0, and $h(0) = 0$;
- AIII. $h^{''}(0) > 0$.

The class transition probabilities can be computed according to the following formulas:

$$p_{ij}(\lambda) = \begin{cases} \dfrac{1}{Z(\lambda)} \displaystyle\int_{-\infty}^{w_1(\lambda)} exp(-\lambda h(s - w^i))ds & \text{if } j = 0; \\ \dfrac{1}{Z(\lambda)} \displaystyle\int_{w_j(\lambda)}^{w_{j+1}(\lambda)} exp(-\lambda h(s - w^i))ds & \text{if } j \neq 0, N; \\ \dfrac{1}{Z(\lambda)} \displaystyle\int_{w_N(\lambda)}^{+\infty} exp(-\lambda h(s - w^i))ds & \text{if } j = N, \end{cases}$$

where $0 \le i, j \le N$.

To study the limit distribution over the wealth classes $W(i)$ (i. e. the behaviour of $\nu^*_\lambda(W(i))$ as λ goes to infinity) we use technique developed in Friedlin and Wentzell (1984), Kifer (1988).

Let Γ_n, $0 \le n \le N$, be the set of trees with vertices labelled $W(0), \ldots, W(N)$ and edges directed to the root $W(n)$. Define two numbers $\theta_\gamma(\lambda)$ and $r_n(\lambda)$:

$$\theta_\gamma(\lambda) = \prod_{i \to j \in \gamma} p_{ij}(\lambda), \quad r_n(\lambda) = \sum_{\gamma \in \Gamma_n} \theta_\gamma(\lambda). \tag{17}$$

To find the number $\theta_\gamma(\lambda)$ for a fixed tree $\gamma \in \Gamma_n$ we have to multiply $p_{ij}(\lambda)$ for all pairs (i,j) such that the directed edge $i \to j$ is in γ. Note that $\theta_\gamma(\lambda) \to 0$ in all the cases except the situation when the root $W(n)$ is an attractor and all the edges contained in γ are also contained in the deterministic graph. In such case $\theta_\gamma(\lambda) = 1$. To compute $r_n(\lambda)$ we have to construct all possible trees with vertices $W(0), \ldots, W(N)$ and edges directed to the root $W(n)$, to find the value $\theta_\gamma(\lambda)$ for every such tree and finally to sum up all these values.

Knowing the numbers $r_0(\lambda)$, ..., $r_N(\lambda)$, we can characterize the limit behaviour of ν^*_λ as follows:

Lemma 1. *(Kifer (1988), ch. 1, lemma 5.5)*
For any n *and* m *such that* $0 \le n, m \le N$

$$\lim_{\lambda\to+\infty} \frac{r_n(\lambda)}{r_m(\lambda)} = \lim_{\lambda\to+\infty} \frac{q^*_n(\lambda)}{q^*_m(\lambda)} = \lim_{\lambda\to+\infty} \frac{\nu^*_\lambda\{W(n)\}}{\nu^*_\lambda\{W(m)\}}.$$

Thus, to understand what wealth class or classes dominate the others in the limit invariant distribution we should analyze the limit behaviour of $r_n(\lambda)$. The latter in turn leads to study the asymptotic behaviour of $p_{ij}(\lambda)$. To do that we use Laplace approximations. More accurately, we replace the functions $r_n(\lambda)$ with equivalent functions of an easier form to understand what the limits of these ratios are. Assumptions made at the beginning of this subsection are exactly needed for carrying out these approximations.

First, approximate the normalizing constant $Z(\lambda)$. By definition $Z(\lambda) = \int_{-\infty}^{+\infty} exp(-\lambda h(s))ds$. Under assumptions AI-AIII Laplace's approximation method (Olver (1997), chapter 3, §7, or Zorich (2004), §19.2) can be applied for this integral. As the result, we have

$$Z(\lambda) \sim \sqrt{2\pi/\lambda h''(0)},\ \lambda \to +\infty, \tag{18}$$

where equivalence of two functions of λ is understood in the standard sense, i. e. it means that $\lim\limits_{\lambda\to+\infty} \dfrac{Z(\lambda)}{\sqrt{2\pi/\lambda h''(0)}} = 1$.

Now turn to the integrals $\displaystyle\int_{w_j(\lambda)}^{w_{j+1}(\lambda)} exp(-\lambda h(s - w^i))ds$, where $w_j = -\infty$ if $j = 0$ and $w_j = +\infty$ if $j = N$. The approximations we use are not valid if the limits of integration vary with λ. We assume that $w_j(\lambda) \to w_j$ as $\lambda \to +\infty$ for $j \neq 0, N$, where w_j are the boundaries of the classes in the deterministic case (zero noise). The sufficient condition for this is that $U(x, y)$ is continuous in y.[13]

$w_j(\lambda) \to w_j$ as $\lambda \to +\infty$ means that for any $\varepsilon > 0$ there exists $\bar{\lambda}$ such that $|w_j(\lambda) - w_j| < \varepsilon$ for any $\lambda \geq \bar{\lambda}$. Thus, for enough large λ we have $w_j - \varepsilon < w_j(\lambda) < w_j + \varepsilon$, and we can underestimate and overestimate the true transition probabilities by integrating over the regions $[w_j + \varepsilon, w_{j+1} - \varepsilon]$ and $[w_j - \varepsilon, w_{j+1} + \varepsilon]$ correspondingly.

Using technique described above, we get the lower and upper estimates for $\theta_\gamma(\lambda)$. Finally, we get the following result

Lemma 2. *For any two trees* γ *and* γ', *if* $c(\gamma) < c(\gamma')$, *then* $\lim\limits_{\lambda \to +\infty} \dfrac{\theta_\gamma(\lambda)}{\theta_{\gamma'}(\lambda)} = +\infty$, *where* $c(\gamma) = \sum\limits_{\substack{i \to j \\ w^i < w_j}} h(w_j - w^i) + \sum\limits_{\substack{i \to j \\ w^i > w_{j+1}}} h(w_{j+1} - w^i)$ *is the cost of the tree* γ. *Here we take the sum over the all edges* $i \to j$ $(0 \leq i, j \leq N)$ *that are contained in the tree* γ. *We assign different "weights" for certain edges* $i \to j$: *if* $w^i < w_j$ *then the weight is* $h(w_j - w^i)$, *if* $w_j < w^i < w_{j+1}$ *then the weight is zero, if* $w^i > w_{j+1}$ *then the weight is* $h(w_{j+1} - w^i)$.

The lemma actually means that θ_γ goes to zero much slower than $\theta_{\gamma'}$ does as λ goes to infinity.

Proof. We use the following estimates for the true transition probabilities:

$$p^l_{ij}(\lambda) = \int_{w_j+\varepsilon}^{w_{j+1}-\varepsilon} h_\lambda(s - w^i)ds < p_{ij}(\lambda) < \int_{w_j-\varepsilon}^{w_{j+1}+\varepsilon} h_\lambda(s - w^i)ds = p^h_{ij}(\lambda), \qquad (19)$$

assuming that λ here is large enough so that $w_j - \varepsilon < w_j(\lambda) < w_j + \varepsilon$. Here $0 \leq i, j \leq N$, $w_0 \pm \varepsilon = -\infty$, $w_N \pm \varepsilon = +\infty$.

The limits of integration in both $p^l_{ij}(\lambda)$ and $p^h_{ij}(\lambda)$ do not depend on λ so

---

[13] A5 actually guarantees that it is also continuous in x.

Laplace approximations can be applied, and we get

$$p^l_{ij}(\lambda) \sim \begin{cases} \dfrac{\exp(-\lambda h(w_j + \varepsilon - w^i))}{\lambda Z(\lambda) h'(w_j + \varepsilon - w^i)} & \text{if } w^i < w_j, \\ 1 & \text{if } w_j < w^i < w_{j+1}, \\ \dfrac{-\exp(-\lambda h(w_{j+1} - \varepsilon - w^i))}{\lambda Z(\lambda) h'(w_{j+1} - \varepsilon - w^i)} & \text{if } w^i > w_{j+1}, \end{cases} \tag{20}$$

$$p^h_{ij}(\lambda) \sim \begin{cases} \dfrac{\exp(-\lambda h(w_j - \varepsilon - w^i))}{\lambda Z(\lambda) h'(w_j - \varepsilon - w^i)} & \text{if } w^i < w_j, \\ 1 & \text{if } w_j < w^i < w_{j+1}, \\ \dfrac{-\exp(-\lambda h(w_{j+1} + \varepsilon - w^i))}{\lambda Z(\lambda) h'(w_{j+1} + \varepsilon - w^i)} & \text{if } w^i > w_{j+1}. \end{cases} \tag{21}$$

Note that all expressions on the right in both 20 and 21 are positive because $h'(s) > 0$ for $s > 0$ and $h'(s) < 0$ for $s < 0$ according to AII).

Fix some tree γ. Define $\theta^l_\gamma(\lambda, \varepsilon) = \prod\limits_{i \to j \in \gamma} p^l_{ij}(\lambda)$ and $\theta^h_\gamma(\lambda, \varepsilon) = \prod\limits_{i \to j \in \gamma} p^h_{ij}(\lambda)$. Because of 19 $\theta^l_\gamma(\lambda) < \theta_\gamma(\lambda) < \theta^h_\gamma(\lambda)$.

For the trees γ and γ' such that $c(\gamma) < c(\gamma')$ we have

$$\lim_{\lambda \to \infty} \frac{\theta_\gamma(\lambda)}{\theta_{\gamma'}(\lambda)} \geq \lim_{\lambda \to \infty} \frac{\theta^l_\gamma(\lambda, \varepsilon)}{\theta^h_{\gamma'}(\lambda, \varepsilon)} =$$

$$= \lim_{\lambda \to \infty} \frac{\prod\limits_{\substack{i \to j \in \gamma: \\ w^i < w_j}} \dfrac{\exp(-\lambda h(w_j + \varepsilon - w^i))}{\lambda Z(\lambda) h'(w_j + \varepsilon - w^i)} \prod\limits_{\substack{i \to j \in \gamma: \\ w^i > w_{j+1}}} \dfrac{-\exp(-\lambda h(w_{j+1} - \varepsilon - w^i))}{\lambda Z(\lambda) h'(w_{j+1} - \varepsilon - w^i)}}{\prod\limits_{\substack{i \to j \in \gamma': \\ w^i < w_j}} \dfrac{\exp(-\lambda h(w_j - \varepsilon - w^i))}{\lambda Z(\lambda) h'(w_j - \varepsilon - w^i)} \prod\limits_{\substack{i \to j \in \gamma': \\ w^i > w_{j+1}}} \dfrac{-\exp(-\lambda h(w_{j+1} + \varepsilon - w^i))}{\lambda Z(\lambda) h'(w_{j+1} + \varepsilon - w^i)}}.$$

Here $\varepsilon < \varepsilon_1$ is small enough so that inequalities 19 hold.

Looking at the denominators of the fractions which appear inside the products we see that they contain expressions $h'(\cdot)$ that do not depend on λ, but only adjust the sign of the limit. The product of the terms $\lambda Z(\lambda)$ has the polynomial growth rate not depending on the concrete amount of these terms. Thus, asymptotic behaviour of the entire expression is defined by the exponents.

For any tree γ and for any fixed $\varepsilon > 0$ denote

$$c^l(\gamma, \varepsilon) = \sum_{\substack{i \to j \\ w^i < w_j}} h(w_j + \varepsilon - w^i) \ + \sum_{\substack{i \to j \\ w^i > w_{j+1}}} h(w_{j+1} - \varepsilon - w^i),$$

$$c^h(\gamma, \varepsilon) = \sum_{\substack{i \to j \\ w^i < w_j}} h(w_j - \varepsilon - w^i) \ + \sum_{\substack{i \to j \\ w^i > w_{j+1}}} h(w_{j+1} + \varepsilon - w^i).$$

In the introduced notations we get

$$\lim_{\lambda \to \infty} \frac{\theta_\gamma(\lambda)}{\theta_{\gamma'}(\lambda)} \geq \lim_{\lambda \to \infty} \frac{\exp(-\lambda c^l(\gamma, \varepsilon))}{\exp(-\lambda c^h(\gamma', \varepsilon))} = \lim_{\lambda \to \infty} \exp\left(-\lambda (c^l(\gamma, \varepsilon) - c^h(\gamma', \varepsilon))\right) \tag{22}$$

Because of continuity of h we get $c^l(\gamma, \varepsilon) \to c(\gamma)$ and $c^h(\gamma, \varepsilon) \to c(\gamma)$ as $\varepsilon \to 0$. If $c(\gamma) < c(\gamma')$, then for ε small enough $c^l(\gamma, \varepsilon)$, $c^h(\gamma, \varepsilon)$ and $c^l(\gamma', \varepsilon)$, $c^h(\gamma', \varepsilon)$ are in the two ε-neighborhoods of $c(\gamma)$ and (γ') correspondingly, and these neighborhoods do not intersect. This means that there exists $\varepsilon_2 > 0$ such that for $\varepsilon < \varepsilon_2$ necessarily $c^l(\gamma, \varepsilon) < c^h(\gamma', \varepsilon)$. Finally, for $\varepsilon < \min\{\varepsilon_1, \varepsilon_2\}$ from 22 we conclude that $\lim\limits_{\lambda \to +\infty} \dfrac{\theta_\gamma(\lambda)}{\theta_{\gamma'}(\lambda)} = +\infty$.

□

The lemma yields the key result.

Theorem 5. *If for some state m there is a tree $\overline{\gamma} \in \Gamma_m$ such that $c(\overline{\gamma}) < c(\gamma')$ holds for all other trees $\gamma' \in \bigcup\limits_{i \in \{0,1,\ldots,N\}} \Gamma_i$, then $q^*_m(\lambda) \to 1$.*

Proof. Let us take any arbitrary $n \neq m$.

$$\lim_{\lambda\to\infty}\frac{q_n^*(\lambda)}{q_m^*(\lambda)}=\lim_{\lambda\to\infty}\frac{\sum\limits_{\gamma'\in\Gamma_n}\theta_{\gamma'}(\lambda)}{\sum\limits_{\gamma\in\Gamma_m}\theta_\gamma(\lambda)}=\lim_{\lambda\to\infty}\frac{\sum\limits_{\gamma'\in\Gamma_n}\theta_{\gamma'}(\lambda)}{\theta_{\overline{\gamma}}(\lambda)\left(1+\sum\limits_{\substack{\gamma\in\Gamma_m\\ \gamma\neq\overline{\gamma}}}\dfrac{\theta_\gamma(\lambda)}{\theta_{\overline{\gamma}}(\lambda)}\right)}=$$

$$=\lim_{\lambda\to\infty}\sum_{\gamma'\in\Gamma_n}\frac{\theta_{\gamma'}(\lambda)}{\theta_{\overline{\gamma}}(\lambda)\left(1+\sum\limits_{\substack{\gamma\in\Gamma_m\\ \gamma\neq\overline{\gamma}}}\dfrac{\theta_\gamma(\lambda)}{\theta_{\overline{\gamma}}(\lambda)}\right)}.$$

According to the lemma 2 $\lim\limits_{\lambda\to+\infty}\dfrac{\theta'_\gamma(\lambda)}{\theta_{\overline{\gamma}}(\lambda)}=0$ for any tree $\gamma'\in\bigcup\limits_{i\in\{0,1,\dots,N\}}\Gamma_i$, $\gamma'\neq\overline{\gamma}$. Thus, in the last line of the previous formula the bracket in the denominators of all terms of the sum goes to the unit, and the fraction goes to zero. Then the whole sum also goes to zero since we have the finite number of terms.

Therefore,

$$\lim_{\lambda\to\infty}\frac{q_n^*(\lambda)}{q_m^*(\lambda)}=0$$

for any $n \neq m$. At the same time $0\leq q_n^*(\lambda)\leq 1$ and $\sum\limits_{n=1}^{N}q_n^*(\lambda)=1$ for any λ. Consequently, $q_n^*(\lambda)\to 0$ for any $n\neq m$ and $q_m^*(\lambda)\to 1$ as $\lambda\to\infty$. □

This theorem allows us to find the state (the class) that is stochastically stable, i. e. that one that dominates all the other classes in the low-noise limit in the long run. The result guarantees that if there exists the state m such that some tree with the root m is the least-cost among all the trees that we can construct with all the other roots, then the class $W(m)$ is stochastically stable.

Since $\bigcup_{\substack{i\in\{0,1,\dots,N\}\\ i\neq m}}\Gamma_i$ is a finite set, typically there will be a unique stochastically stable state. To construct an example where there will be several stochastically stable states is another complicated task. This situation takes place no matter how many attractors there were in the deterministic case, however, we can state that if the class $W(m)$ is stochastically stable, then this class is self-sustaining in the deterministic case.

Theorem 6. *If $W(m)$ is stochastically stable, then it is an attractor in the deterministic dynamics.*

Proof. Suppose that $W(m)$ is not an attractor in the deterministic graph. Then it is in the basin of attraction $B(n)$ of some attractor $n \neq m$. Let us show that for any tree $\gamma \in \Gamma_m$ there exists a cheaper tree $\gamma' \in \Gamma_n$ meaning that $W(m)$ is not stochastically stable.

Proceed by induction on the path length d from $W(m)$ to $W(n)$ in the deterministic graph. Suppose $d = 1$. Fix some tree $\gamma \in \Gamma_m$, remove from γ the one edge that goes out of state n and add the edge from m to n. Regardless of distance between n and m in the tree γ once these procedures are accomplished we get as a result some tree $\gamma' \in \Gamma_n$. $c(\gamma') < c(\gamma)$ since we replaced a nonzero cost edge with the edge that has the zero cost.

Suppose that the claim is true for all states in $B(n)$ within distance d, and suppose that m has distance $d+1$ from n. There is a link from m to some p within distance d of n in the deterministic graph. Within the tree γ remove the edge out of p and add an edge from m to p to create a tree $\gamma' \in \Gamma_p$. Obviously $c(\gamma') < c(\gamma)$ by the same argument we used above. Now p is in a basin of attraction $B(n)$ and distance from p to n is d. By induction assumption using $\gamma' \in \Gamma_p$ it is possible to construct a tree $\gamma'' \in \Gamma_n$ such that $c(\gamma'') < c(\gamma') < c(\gamma)$. So the cheaper tree with a vertex n is constructed.

□

Note that the stochastically stable attractor is not determined by the graph. That is, there are graphs for which the identity of the unique stochastically stable state depends on the shape of the function h and/or magnitudes of $|w^i - w_j|$ for all possible i and j.

Having some additional constraints on the function h, we can simplify the procedure of looking for the least-cost tree. However, it is still not so straightforward as we would wish.

4 Conclusions

Based on the logic of classic papers of intergenerational human capital development, this paper offers a stepping-stone model of the evolution of wealth. Assuming that in each period a parent allocates her wealth between consumption and investment in child's human capital, we consider both deterministic and stochastic environments. In the former case child's wealth is unambiguously defined by parental investment chosen from a finite number of possible investment levels; in the latter we allow wealth also to be affected by a stochastic term. This stochastic shock does not have an explicit interpretation. We could use it to discuss, for instance, macroeconomic shocks or the distribution of unobservables, such as children's genes.

In the deterministic framework there is no mobility observed after a certain finite amount of periods so that eventually any movements across wealth classes just vanish. The deterministic model may exhibit multiple inescapable steady states. Thus, inescapable steady states that correspond to low wealth levels may be understood as poverty traps. Similarly we can define affluence traps.

In the stochastic environment mobility exists in any given time period, so there are no inescapable states. The stochastic framework brings us to the Markov model of the evolution of wealth that has a unique invariant distribution. First, we study how small perturbations of parameters of the model change the stationary wealth distribution. Second, we analyze the difference between the deterministic model and the stochastic model with a small amount of noise.

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. The multiple deterministic attractors play a special role in determining the shape of the invariant distribution which can be seen when the noise is small. As the noise becomes small, the step-by-step transitions of the stochastic process look increasingly like the deterministic evolution. Beyond this, some attractors are favored over others. Typically there will be a unique attractor around which the invariant distribution will concentrate. Therefore, poverty traps can arise naturally in stochastic models as they do in deterministic models, but their consequences are more complex than a multi-peaked wealth distribution. The latter suggests that in cross-section or in short panels it will be hard to find poverty traps by looking for multiple modes.

Appendix

Theorems 1 and 2.

Proof. Having A4, we guarantee that the function $V(c,k)$, defined in 1, exists.

The next step is to show that $V(c,k)$ is upper-semicontinuous (usc). First, $U(c,f(k,s))$ is usc in c just by the assumption and in k as the composition of two usc functions. This is true for any s. Second, we need to show that $\int_{-\infty}^{+\infty} U(c,f(k,s))d\mu(s)$ is usc in c and k. For instance, let us prove this for c, the same argumentation works for k. Let c_n be a sequence converging to arbitrarily chosen point c_0. Introduce the sequence of functions $F_n(s) = U(c_n, f(k,s))$ and the function $F_0(s) = U(c_0, f(k,s))$, k and s are perceived as parameters. Since $U(c,f(k,s))$ is usc in c for any k and s we have $F_n(s) \leq F_0(s)$ for any k and s. Reverse Fatou's lemma gives us that $\limsup_n \int_{-\infty}^{+\infty} F_n(s)d\mu(s) \leq \int_{-\infty}^{+\infty} \limsup_n F_n(s)d\mu(s)$. Using $F_n(s) \leq F_0(s)$, we finally get $\limsup_n \int_{-\infty}^{+\infty} F_n(s)d\mu(s) \leq \int_{-\infty}^{+\infty} F_0(s)d\mu(s)$, and this means that $U(c,f(k,s))$ is usc at the point c_0. To use reverse Fatou's lemma we need to be sure that there exists a function $G(s) \in L^1(-\infty,+\infty)$ such that $U(c_n, f(k,s)) \leq G(s)$. This condition is fulfilled, because c_n as a converging sequence is bounded, i. e. $c_n \leq \bar{c}$ for some $\bar{c} \geq 0$, and, thus, $U(c_n, f(k,s)) \leq U(\bar{c}, f(k,s))$ and the latter is an integrable function in s for all k according to the assumption A4. Therefore, $V(c,k)$ is usc at c_0 that was chosen arbitrarily so $V(c,k)$ is usc in c.

Because of A1 the budget constraint 2 is tight, therefore we maximize usc $V(c,k)$ on the compact set $\{(c,k) \in \mathbb{R}^2 : c \geq 0, k \geq 0, c+k \leq w\}$. Then the optimal value function is actually a maximum and according to the Berge theorem the "argmax" correspondence $\pi(w)$ is upper-hemicontinuous at every continuity point of $V(c,k)$. $\square$

Theorem 3.

Proof. First, it is easy to see that if $U(x,y)$ is supermodular (equivalently, has increasing differences) and $f(k,s)$ is non-decreasing in k for any s, then $U(c,f(k,s))$ is supermodular for any s. Second, $V(c,k) = \int_{-\infty}^{+\infty} U(c, f(k,s))d\mu(s)$ is supermodular just because integral of the sum is the sum of integrals. Thus, $V(c,k)$ is supermodular.

Let $w' > w''$, $k' \in \pi(w')$, $k'' \in \pi(w'')$. Suppose that the statement of the theorem is false, i. e. $k' < k''$. Since $k'' \in \pi(w'')$, $w'' - k'' \geq 0$, consequently, $w' - k'' > 0$ which means that the investment level k'' is feasible for w'. k' is optimal for w' and, thus, $V(w'-k',k') \geq V(w'-k'',k'')$. On the other hand, k' is feasible for w'', and $V(w''-k'',k'') \geq V(w''-k',k')$ because k'' is optimal for w''. Combining these two inequalities, we get

$$V(w'-k',k') - V(w''-k',k') \geq V(w'-k'',k'') - V(w''-k'',k''). \tag{23}$$

Then, $w'-k'' > w''-k''$, $k'' > k'$ and V is supermodular, therefore,

$$V(w'-k'',k'') - V(w''-k'',k'') \geq V(w'-k'',k') - V(w''-k'',k'). \tag{24}$$

The right-hand side of 23 and the left-hand side of 24 coincide, and we get that

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

Equivalently,

$$V(w'-k',k') - V(w'-k'',k') \geq V(w''-k',k') - V(w''-k'',k'). \tag{25}$$

For simplicity let us introduce the following notations: $x = w'-k''$, $y = w''-k''$, $h = k''-k'$. Note that $x > y$ and $h > 0$. Inequality 25 in these notations has the following form:

$$V(x+h,\cdot) - V(x,\cdot) \geq V(y+h,\cdot) - V(y,\cdot)$$

It is easy to see that $V(c,k)$ is strictly concave in c because of A6, and the last inequality contradicts this fact. Therefore, our initial supposition was false, and $k' \geq k''$. □

References

Azariadis, Costas and John Stachurski. 2005. “Poverty traps.” In *Handbook of Economic Growth*, 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.

Björklund, Anders, Markus Jäntti, and Gary Solon. 2005. “Influences of nature and nurture on earnings variation.” *Unequal chances: Family background and economic success* :145–164.

Black, Sandra E, Paul J Devereux, Petter Lundborg, and Kaveh Majlesi. 2020. “Poor little rich kids? The role of nature versus nurture in wealth and other economic outcomes and behaviours.” *The Review of Economic Studies* 87 (4):1683–1725.

Bowles, Samuel and Herbert Gintis. 2002. “The inheritance of inequality.” *Journal of economic Perspectives* 16 (3):3–30.

Carter, Michael R and Christopher B Barrett. 2006. “The economics of poverty traps and persistent poverty: An asset-based approach.” *The Journal of Development Studies* 42 (2):178–199.

Dardanoni, Valentino. 1995. “Income distribution dynamics: monotone Markov chains make light work.” *Social Choice and Welfare* 12 (2):181–192.

Friedlin, Mark I and Alexander D Wentzell. 1984. *Random perturbations of dynamical systems*. Berlin Heidelberg New York: Springer.

Ghatak, Maitreesh. 2015. “Theories of poverty traps and anti-poverty policies.” *The World Bank Economic Review* 29 (suppl_1):S77–S105.

Kifer, Yuri. 1988. “Random perturbations of dynamical systems.” In *Progress in probability and statistics*, vol. 16. Birkhäuser Boston, 294.

Loury, Glenn C. 1981. “Intergenerational transfers and the distribution of earnings.” *Econometrica: Journal of the Econometric Society* :843–867.

Olver, Frank. 1997. *Asymptotics and special functions*. A K Peters.

Youngson, Neil A and Emma Whitelaw. 2008. “Transgenerational epigenetic effects.” *Annual Review of Genomics and Human Genetics* 9:233–257.

Zorich, Vladimir A. 2004. *Mathematical analysis II*. Springer.