Like Father, Like Son: Social Network Externalities and Parent-Child Correlation in Behavior

Antoni Calvó-Armengol* and Matthew O. Jackson†

This version: March 21, 2008
forthcoming in the *American Economic Journal: Microeconomics*

Abstract

We build an overlapping generations model where an individual sees higher returns to adopting a behavior as a larger number of their neighbors adopt the behavior. We show that overlap in the state of a parent and child's neighborhood can lead to correlation in parent-child behavior independent of any parent-child interaction. Increasing the sensitivity of individual decisions to the state of their social community leads to increased parent-child correlation, as well as less efficient (more costly) behavior on average in the society. We also show that this model is distinguished from a direct parental influence model, in that it predicts increased generational effects, implying residual correlation between children and grandparents after including parental information.

Keywords: Networks, Social Interaction, Parent-Child Correlation, Education, Human Capital.

JEL Classification Numbers: A14, J64, J31, J70.

---

*Toni passed away in November of 2007, after the paper was written. His friendship, energy, and talents are sorely missed.

†Department of Economics, Stanford University, Stanford CA 94305, USA. http://www.stanford.edu/∼jacksonm Email: jacksonm@stanford.edu Financial support from the Guggenheim Foundation, the Lee Center for Advanced Networking, the Center for Advanced Studies in the Behavioral Sciences, and the NSF under grants SES-0316493 and SES-0647867 is gratefully acknowledged. We are very grateful to Simona Comi for making her data available to us. We also thank Rachel Kranton and Alison Watts for helpful conversations and discussions, and Andy Postlewaite and several anonymous referees for helpful comments and suggestions.

1 Introduction

Correlations in the behavior and welfare of parents and their children have been widely documented. For example, parents and children exhibit strong correlations in their political affiliation, religion, education levels, substance use, language proficiency, income and wealth.[1] There are many explanations for these, generally involving some sort of direct influence of the parent on the child through channels ranging from genetics to household environment to bequests. While we do not wish to dispute such channels here, we develop a model that illustrates that one can derive quite strong parent-child correlations without any direct relationship between them, but simply through the sensitivity of both a parent and child decisions to an overlapping social environment.

To fix ideas, suppose that we find that conditional on a parent having pursued education beyond high school there is an X percent chance that the child will have pursued education beyond high school, while if the parent did not pursue education beyond high school then there is a Y percent chance that the child will have pursued education beyond high school, where Y<X. We might be tempted to conclude that this inequality in education outcomes reflects parental or household influence, and the difference between X and Y could be attributed to how the decisions of the parent or familial circumstances affect the decision of the child. We examine a different driver of the same phenomenon and show that this difference in outcomes might alternatively be attributed, in part or in whole, to social circumstances. That is, it might be that the child's decision is based on how many other people in the community are pursuing higher education. Given that the parent and child live in the same community, there is a substantial overlap in the surroundings that have influenced their decision, and what appears to be a parental influence, could in fact be an indirect social influence. In fact, we show that these two different potential influence models lead to exactly the same set of possible distributions over parent and child behavior.

Our purpose in this paper is to take the potential social influence seriously, to provide a model to analyze it, and to see how changes in sensitivity to social surroundings translate into changes in the observed parent-child behaviors.

The model that we propose is an overlapping-generations model. There are a number of single-agent families or "dynasties" that form a community. At any point in time, the agents representing these dynasties are taking actions. These are binary actions, which can be seen as things like investing in education or not, smoking or not, etc. So, a description of the state of the community is just a list of which dynasties are taking action 1 and which

[1] The hereditary transmission of political affiliation is discussed, for instance, in Thomson (1971) and Abramanson (1973); and Piketty (1995) surveys this literature. Bisin and Verdier (2000) analyze the family transmission of religious traits, and discuss numerous empirical works relating the religious affiliation of children to their social environment, e.g., Hoge *et al.* (1982). The social determinants of substance use (including alcohol and smoking, among others) are discussed thoroughly in the social epidemiology literature, e.g., Berkman and Kawachi (2000) and Galea *et al.* (2004). See also Johnson (1996) for a survey of family influence on the traits of children and social behavior.

are taking action 0. In each period one person ("the parent") is randomly selected out of the community and replaced by another ("the child"). The child then makes a decision of which action to take for their lifetime and their decision is dependent on the state of the community. The decision exhibits complementarities, so that the higher the number of agents in the community who have adopted action 1, the greater the propensity of the child to choose action 1. This results in a Markov chain, and we can analyze its properties. We show that the complementarities in decisions relative to the social state lead to positive correlations between parent and child decisions. In particular, we show that the parent-child correlation in behavior is exactly proportional to the variance in the child's decision across the social state. Thus, we can tie intergenerational correlation directly to the sensitivity of the child's decision to the social neighborhood. As we show, very different patterns of social sensitivity can result in similar intergenerational correlations in behavior. By backing out the social sensitivity from observed correlations, we see that seemingly similar correlations come from very different social patterns.

Beyond comparative statics in social sensitivity, we also examine the efficiency of decision making in the context of social influence. We outline an inefficiency that results in the following sense. An individual with a relatively low cost for taking action 1 might not take it if his or her neighbors are not taking the action, while at the same time another individual with a higher cost for taking action 1 might take it because his or her neighbors are taking the action. We show that as we increase the sensitivity to social circumstances, this effect increases and the overall investment costs in the society increase. In a sense, the "wrong" people are taking actions in that if we could rearrange who had which cost of taking action 1 in the society, the members of the society would be better off. We call this a "cost-inefficiency," and show that it is proportional to the social sensitivity of decisions in some particular cases.

We also discuss how our Markov model can be applied more generally, and that it can be seen as generalizing imitation investment models (e.g., Kirman (1993)).

As part of the analysis, we note that if one only looks at parent-child data, our social indirect influence model is observationally equivalent to a direct parental influence model without any role for the social setting. In other words, a direct parent-child interaction cannot be distinguished from the social channel as a driver for intergenerational correlation in behavior (and vice-versa), and so a system with both sources of interaction would not be identified. In order to distinguish between the different influences, one needs to look beyond parental and child outcomes. An obvious possibility is to use time series data about the social surroundings to identify the social influence and distinguish it from direct parental influence. We provide an alternative identification strategy that does not use information about the community status, but instead uses time series data about dynasties' history across two (or more) generations.

1.1 Related Literature

There are other papers that have tied intergenerational relations to social situation, e.g., Borjas (1992) and Bénabou (1996). However, in such models the social setting is something that amplifies the parent-child interaction (e.g., in Borjas affecting the parent's decision for bequests); while in our model there is no parent-child interaction at all and correlations derive entirely from the social channel. Also, our model is meant to be portable across a variety of applications, and is not directly tailored with setting-specific assumptions.

In terms of the identification issue, the difficulty of distinguishing between direct parental and social influences is certainly not new.[2] Our model, nonetheless, makes this issue clear. It is worth contrasting the identification that we note with the reflection problem described by Manski (1993). In Manski's formulation, the researcher posits that there are social effects, but does not observe them directly and instead infers them based on the attributes of the individual in question. Since the missing social effects are inferred from the individual's attributes, they end up simply "reflecting" back the individual's attributes and not being identified. Here we show that two very different models lead to exactly the same observed patterns in parent-child relations. In particular, any specification of our social interaction model, regardless of the specific formulation of how the decision probabilities vary with the community is equivalent to some direct parent-child specification, and vice versa. This is a different point from the reflection problem, which derives from either a linear in means formulation or another form of separability.[3] Regardless of specific differences in how the identification problem noted here differs from the reflection problem (which has become a canonical example of identification issues), the point that there can be multiple explanations for the same data is obviously not new. Methods of overcoming this issue are standard: for instance, finding instruments for, or direct observations of, the social drivers of individual behavior. Beyond this our model provides some time series predictions that allow one to distinguish it from direct effects without having an instrument or observation of the social surroundings.

Section 2 presents the model of individual decisions with random over-lapping generations and community influence. Section 3 contains the results on identification, and the comparative statics of intergenerational correlation (and related investment costs) with respect to the social sensitivity of behavior. Section 4 provides results on a special case of the model that we call the threshold model.[4] All proofs are relegated to an appendix.

[2] See, for instance, Corcoran *et al.* (1992).

[3] In Section 3 of Manski (1993), there is a more general formulation than the linear one, which makes clear that the reflection problem arises from a separability of expectations based on social factors and other factors.

[4] Supplementary material provides an exercise of fitting the threshold model to data from European countries.

2 The Model

2.1 Dynasties and Random Over-Lapping Generations

We analyze a given community or neighborhood of individuals. There are n dynasties in the community indexed by $i = 1, ..., n$. Time evolves in discrete periods indexed by $t = 1, 2, \ldots$. Each generation of a given dynasty consists of one player. When there is no confusion, we identify current generation players by their dynasty index i.

At the beginning of each period, one dynasty is randomly chosen and its member replaced by a new player. This happens with equal probability across dynasties.[5] We refer to the new player as the child, and to the old player as the parent.

2.2 Individual Decisions and the Social Externality

New players choose between two actions, high (denoted 1) and low (denoted 0). Let $k_{-i}^t \in \{0, 1, \ldots, n-1\}$ denote the number of high-action players other than i at time t, and k^t be the total number of high-action players in the community at time t.

$\pi(0, k_{-i})$ and $\pi(1, k_{-i})$ denote player i's expected payoffs conditional on the current actions by the other players being k_{-i} and conditional on i choosing 0 or 1, respectively. Given the Markov properties of the model, this is the same function in any period. We normalize $\pi(0, k_{-i}) = 0$ for all k_{-i}, and assume that $\pi(1, k_{-i})$ is non-decreasing in k_{-i}, which corresponds to a complementarity in actions.[6]

Taking the high action is costly, and that new players are assigned a randomly drawn cost c of taking this action, which is described by a cumulative distribution function $F(c)$. Because of this cost, we say that a player "invests" in the high action.

Thus, a player i's decision is characterized by a vector $p = (p_0, \ldots, p_{n-1})$, where $p_k = F(\pi(1, k))$ is the probability that he or she invests in the high action when k other players are high-action. Player i plays the low action 0 with complementary probability $1 - p_k$. The non-decreasing nature of $\pi(1, \cdot)$ implies that p_k is also non-decreasing in k.

This formulation assumes that the only relevant state variable is the state of other agents actions, although one could enrich the model with state variables. It is worth noting that this formulation avoids multiple equilibrium issues that plague games with complementarities. It is avoided via two key assumptions. First, that only one agent moves at a time. Second, an agent makes a choice based only on the current actions of the other agents, and does not base the decision on beliefs about how future agents will act. With forward looking agents,

[5] Therefore, the expected life span of any generation in dynasty i is $\sum_{t=1}^{\infty} t \frac{1}{n}(1 - \frac{1}{n})^{t-1} = n$. Thus, the length of a time period could be taken to be proportional to the expected lifetime of an individual divided by n.

[6] We discuss this complementarity in more detail with regards to the application to education and human capital investment. Complementarities arise from a very wide variety of sources including direct social pressures, informational externalities, production complementarities, network externalities, improved job contacts, etc. At this point we can be completely agnostic about the source of the complementarity.

one could derive multiple equilibrium reactions to the same current state, which is precluded in our formulation.

For most of the analysis that follows, one can work directly with the specification of p as the primitive of the model. However, when we return to a discussion of the efficiency of the model it will be useful to keep track of the costs and benefits to taking an action.

2.3 A Markov Process

The random overlapping generations model, together with the individual investment decisions with idiosyncratic costs generates a Markov process. The state is the number of players k^t at the high action at the end of a period t, and transition probabilities can be derived from the vector $p = (p_0, ..., p_{n-1})$. These Markov transition probabilities are:[7]

$$\begin{aligned}
\Pr\{k^{t+1} = m+1 \mid k^t = m\} &= \frac{n-m}{n} p_m, \text{ for } 0 \leq m \leq n-1 \\
\Pr\{k^{t+1} = m-1 \mid k^t = m\} &= \frac{m}{n}\left(1 - p_{m-1}\right), \text{ for } 1 \leq m \leq n \\
\Pr\{k^{t+1} = m \mid k^t = m\} &= \frac{n-m}{n}\left(1 - p_m\right) + \frac{m}{n} p_{m-1}, \text{ for } 0 \leq m \leq n \\
\Pr\{k^{t+1} = m' \mid k^t = m\} &= 0, \text{ otherwise.}
\end{aligned}$$

This is a finite-state irreducible and aperiodic Markov process. We characterize the unique long-run steady-state distribution of this process.[8]

The steady-state of the Markov process can thus be described by $\mu = (\mu_0, \ldots, \mu_n)$, where μ_k is the probability that k players take the high action. As the steady-state distribution depends on the vector of investment probabilities, we write $\mu(p)$.

Given p, the average probability $\overline{p}$ that any given player takes the high action under the steady-state distribution $\mu(p)$ is:

$$\overline{p} = \sum_{k=0}^{n} \mu_k \frac{k}{n}. \tag{1}$$

The p and the induced $\mu(p)$, together with the dynastic birth and death process, also induces a long-run distribution over the four possible parent-child action combinations,

[7] Note that the economic derivation of these probabilities gives one justification to them. One can also provide alternative justifications for these probabilities simply directly appealing social pressure effects, social-psychological effects, roles models, conformity, and so forth. The important aspect of our approach is to examine how behavior varies over generations as these vary.

[8] For such a Markov process, it is well-known that it has a unique steady-state distribution that has several nice features. First, the steady-state distribution represents the relative frequencies spent in each state over long time horizons. Second, starting with a random draw from the steady-state distribution, the distribution over the next period's states is governed by the same distribution. Third, starting from any state, given a long enough horizon, the probability that one will end up in any give state is approximately given by the steady-state distribution.

00, 01, 10 and 11:

		Child Action 0	Child Action 1
Parent Action	0	$\sum_{k=0}^{n}\mu_k\frac{n-k}{n}(1-p_k)$	$\sum_{k=0}^{n}\mu_k\frac{n-k}{n}p_k$
	1	$\sum_{k=1}^{n}\mu_k\frac{k}{n}(1-p_{k-1})$	$\sum_{k=1}^{n}\mu_k\frac{k}{n}p_{k-1}$

(2)

Given these probabilities, it is straightforward to compute the parent-child intergenerational correlation in actions under the steady-state distribution, denoted $Cor\,(p)$, which is one of our objects of interest.

Note also that, although it is not obvious from (2), the long-run distribution over parent-child actions is symmetric; that is, the probabilities assigned to the action combinations 01 and 10 are the same. Formally, it is easily checked that:

$$\sum_{k=1}^{n}\mu_k\frac{k}{n}(1-p_{k-1})=\sum_{k=0}^{n}\mu_k\frac{n-k}{n}p_k,\text{ for all }p.$$

Indeed, the marginal probability to take the high action at steady state is the same for both the parent and the child, and equal to $\overline{p}$. Given that this marginal probability for the parent (resp. the child) is the sum of the joint probabilities for the action combinations 10 and 11 (resp, 01 and 11), the symmetry follows.

3 Parent-child Correlation in Behavior

3.1 Monotonicity of the Steady-State Distribution of Actions

Consider a vector of investment probabilities $p=(p_0,...,p_{n-1})$. We begin with an obvious result that establishes monotonicity of the steady-state distribution of individual actions (in the sense of first-order stochastic dominance) with respect to the investment vector p.

We write $p'\geq p$ if $p'_k\geq p_k$, for all k, and $p'>p$ if $p'_k>p_k$, for all k.

Proposition 1 *Let μ and μ' denote the steady-state distribution associated with p and p', respectively. If $p'\geq p$, then μ' first-order stochastically dominates μ. When $p'>p$, the first-order stochastic dominance is strict.*

A greater propensity to invest in the high action shifts the mass to higher realizations at steady-state. This simple monotonicity result on $\mu\,(p)$, though, does not translate into an increase in parent-child correlation. This is because a increase in p does not necessarily increase the *sensitivity* of the investment decision to the community choices. We illustrate this with a simple example.

Example 1 (A Dyad: $n = 2$) Investment probabilities are $p = (p_0, p_1)$. Figure 1 summarizes the transition probabilities. A darkened (resp. empty) node represents a high-action (resp. low-action) player. Arrows represent state transitions with their corresponding probabilities. The probability with which the state does not change is indicated by the expression near the dyad.

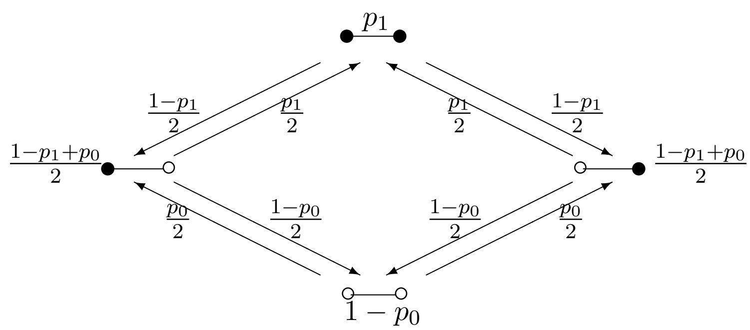


Figure 1: A Two Dynasty Markov Process

The spread $p_1 - p_0$ reflects the sensitivity of individual decisions to community choices. If p_0 is close to p_1, then one player's choice is largely independent of the other player's choice. If the gap between these probabilities is large, then the decisions are very sensitive to the choice of the other player.

Direct calculations give the steady-state distribution $\mu(p_0, p_1)$ over actions, from which we obtain the distribution over parent-child behavior. The corresponding parent-child correlation in actions is:[9]

$$Cor\,(p) = (p_1 - p_0)^2.$$

Thus, the serial correlation is dependent directly on the sensitivity of behavior to the state of the neighbor, as captured by $p_1 - p_0$.

We see here that an increase in the vector of investment probabilities, $p' \geq p$, will not necessarily lead to an increase in parent-child correlation, unless the spread in p' is larger than the spread in p.

3.2 The Identification Issue: Equivalence to Direct Parent-Child Influence

Our model shows that the social setting can be a driver for arbitrary parent-child correlations, even without direct relationships between them. However, parent-child correlations could also result from a direct parental influence on child's behavior (including genetics, etc.). The

[9] The marginal long-run probability that an agent takes a high action is $\overline{p} = \mu_2 + \frac{1}{2}\mu_1$. The joint probability with which two consecutive generation members of the same dynasty take a high action is $\eta = p_1\mu_2 + \frac{1}{2}p_0\mu_1$. The intergenerational correlation can then be computed as $(\eta - \overline{p}^2)/\overline{p}(1 - \overline{p})$. The expression follows from simple algebra. Note that the *intergenerational correlation* involves the joint probability of a high action for *consecutive generations* of a same dynasty, η. Instead, the *cross-sectional correlation* involves the joint probability of a high action for *contemporaneous generations*, μ_2.

next result shows that a model that has child decisions *directly* dependent upon parental influence is observationally equivalent to an *indirect* social influence model. This implies that, given data on the parent-child distribution of actions (2), the indirect social channel and the direct parental channel cannot be identified against each other.

The direct parent-child influence model is defined by a pair of probabilities (d_0, d_1). The child's investment probability is d_1 if the parent's state is 1 and d_0 if the parent's state is 0. Presuming that $d_1 \geq d_0$, this generates all possible symmetric distributions over joint parent-child outcomes such that intergenerational correlation is nonnegative.

PROPOSITION **2** *The direct parental influence model and the social influence model (even with $n = 2$) generate exactly the same set of distributions over joint parent-child outcomes. These are the distributions such that (a) correlation is nonnegative,*[10] *and (b) the probability that the parent state=1, child state=0 is the same as the probability that the parent state=0, child state=1.*

Proposition 2 implies that observing serial correlations in parent and child behaviors cannot be attributed unambiguously to a direct parental influence, regardless of how strong or weak that correlation is. It can also be explained by an indirect social model. That is, a correlation in parent-child behavior is as consistent with social effects as it is with direct parent-child influence. While reality is somewhere in between, Proposition 2 implies that one needs to look beyond parental and child outcomes in order to distinguish between the different influences.

While the proof is offered for the case of $n = 2$, it is not unique to that case. For instance, with larger n one can still get a full spectrum of correlations, by increasing the sensitivity of an individual's choice to the social state. For example, a version of the model where an individual's decision depends on that of the majority of the neighbors can replicate the behavior in the dyad case.

Because of the identification problem pointed at here, which is not about handling a misspecified model but about observational equivalence, one can think about at least two different ways to improve the prospects of identification.

The first and obvious possibility to distinguishing between the two drivers is to gather information about the community status over time. However, rich enough data sets need not always be available.[11] Hence we mention another approach to distinguishing between the models that builds on a deeper theoretical understanding of the workings of the direct and the indirect channels in generating time series predictions.

In the direct parental influence model, conditional on the parent's state, the child's state is independent of the previous history. That is, the child's investment decision is

[10] The proposition extends to the case of negative correlations if we extend the direct parental influence model to allow $d_1 < d_0$ and similarly we extend the social influence model to allow $p_1 < p_0$.

[11] An exception is a quantitative study of political dynasties in the United States Congress since 1789 in Dal Bó et al. (2007).

described by a Bernoulli process whose success probability d_0 or d_1 depends on the parent's state. The parent's state specifies all the historical information relevant to description of the future investment decision of the offspring. In contrast, in the social influence model, the distribution over the child's state is not independent of the history beyond the parent's state. That is, the investment decisions of two children with identical parents' states are generally described by two different random processes. With social influences, the past history of the process beyond that of their parents has an impact on investment decisions by children. In particular, longer strings of investment in a dynasty's history are indirect indicators that the neighbors' have invested and so a history of investment makes it more likely that the child will invest.

PROPOSITION **3** *In the direct parental influence model the probability that the Child's state is 1 conditional upon the Parent's state being 1 and the Grandparent's state being 1 is equal to the probability that the Child's state is 1 conditional upon the Parent's state being 1. In contrast, in the dyadic social influence model, if* $1 > p_1 > p_0 > 0$*, then the probability that the Child's state is 1 conditional upon the Parent's state being 1 and the Grandparent's state being 1 is greater than the probability that the Child's state is 1 conditional upon the Parent's state being 1.*

This result shows two different effects. First, the fact that a child's grandparent is invested makes it more likely that other people in the community will have invested and this makes it more likely that the child will invest. Second, the fact that a child's grandparent is invested is raises the posterior probability of higher investment levels in the community at the time the grandparent invested, and these then lead to higher probabilities of neighbors being invested over time and a higher probability that the child will invest.

We note that this does not necessarily provide a sure technique for identification, as one could also imagine direct effects from multiple generations (that is, grandparents directly influencing children beyond parental influence). More generally, one could claim that the span of direct influence across consecutive generations is bounded from above, but intergenerational effects never completely disappear from the social influence model. The ability to distinguish this, or to know the upper bound on direct influence, might be limited, and so the inclusion of variables that help identify social effects more directly is always useful.

3.3 Parent-Child Correlation and the Social Sensitivity of Behavior

We now explore the relationship between the social sensitivity, as captured by p, and the parent-child correlation in behavior, $Cor\,(p)$. The former subsumes cross-sectional (contemporaneous) influences within members of the same group, while the latter reflects intergenerational (serial) dependence in behavior across consecutive members of a same dynasty. The analysis to follow shows how intergenerational correlation can arise purely as the result of (repeated) cross-sectional influences, and establishes a relationship between the size and the sign of these two a priori orthogonal forms of covariation.

Recall that a player i's decision is characterized by a vector $p = (p_0, \ldots, p_{n-1})$, where p_k is the probability that he or she invests in the high action when k other players have selected the high-action. Intuitively, increasing the spread in p_k's across k makes the investment decision more sensitive to the community choices. That is, increasing p_k for high levels k and/or decreasing it for lower k leads investment to be more dependent on the state of the community. In the example of the dyad, for instance, a measure of an increasing sensivity is given by the spread $p_1 - p_0$. More generally, we capture this notion of increasing the spread in p_k's through the variance in the probability that a given player adopts the high action. We compute this variance as follows.

Given p and the associated steady-state distribution $\mu(p)$ over population choices, let $\mu^{-i}(p) = \left(\mu_0^{-i}, ..., \mu_{n-1}^{-i}\right)$ be the long-run distribution for the action choices in the population when some arbitrary player i is excluded from the group. So μ_k^{-i} is the steady-state probability that k players other than some player i take the high action. Straightforward calculations show that $\mu^{-i}(p)$ can be written directly in terms of the steady state distribution $\mu(p)$ as follows:[12]

$$\mu_k^{-i} = \frac{n-k}{n}\mu_k + \frac{k+1}{n}\mu_{k+1}, \text{ for } k = 0, ..., n-1.$$

Let $Var(p)$ denote the variance in investment probabilities. Given that player i makes his or her investment decision with probability p_k when there are exactly k other high actions in the community, which is an event with long-run probability μ_k^{-i}, the variance of investment probabilities is then

$$Var(p) = \sum_{k=0}^{n-1} \mu_k^{-i} p_k^2 - (\sum_{k=0}^{n-1} \mu_k^{-i} p_k)^2. \tag{3}$$

Differences in the coordinates of p reflect differences in investment choices accruing solely from variations in social circumstances. The vector $\mu^{-i}(p)$, precisely, quantifies the ocurrence with which, in the long-run, the composition of the social milieu varies for different agents at the time their investment decision is made. Then, $Var(p)$ accounts both for the variability in social circumstances and for the dependence of investment choices to this fluctuating social setting.

The next result states that increased social sensitivity or influences exerted within periods and across dynasties induces higher correlation in behavior across periods and for members of the same dynasty. More precisely, we provide a complete characterization of the relative ordering of intergenerational correlations as dependent on properties of the vector p of contemporaneous social influences.

[12] The number of states with k high types, not counting i, is $\binom{n-1}{k}$. There are exactly $\binom{n}{k+1}$ states with $k+1$ high types. The fraction of such states for which i is of high type is $\binom{n-1}{k}/\binom{n}{k+1} = (k+1)/n$. Also, there are $\binom{n}{k}$ states with k high types, and the fraction of such states for which i is of low type is $\binom{n-1}{k}/\binom{n}{k} = (n-k)/n$.

THEOREM 1 *The parent-child correlation in behavior starting from the steady state is higher under p' than under p if and only if*

$$\frac{Var(p')}{\overline{p}'(1-\overline{p}')} > \frac{Var(p)}{\overline{p}\,(1-\overline{p})}. \tag{4}$$

A corollary of Theorem 1 is the following.

COROLLARY 1 *The parent-child correlation starting from the steady state is higher under p' than under p if $Var(p') > Var(p)$ and $\overline{p} = \overline{p}'$.*

Keeping track of the variation in investment ($Var(p)$) alone is not enough to tie down the parent-child correlation, as it is this variance relative to the overall variation in a player's state that is important. This normalization is necessary since it is only relative variations that matter in correlation comparisons, rather than absolute levels. The relevant normalizing variation is the variance of any given player i's action choice, which is simply $\overline{p}\,(1-\overline{p})$, where $\overline{p}$ is the average investment probability defined in (1). If all players were facing identical social circumstances, they would all invest in the high action with average probability $\overline{p}$, leading to a variance for investment decisions equal to $\overline{p}\,(1-\overline{p})$. Instead, the long-run variability in social circumstances is a source of variability in investment probabilities subsumed by $Var\,(p)$.

As outlined in the case of a dyad (Example 1), the intuition behind the theorem and corollary is that increased direct (cross-sectional) correlation of both parent and child choices with the state of the community at the time they each make their own investment decision leads to higher indirect (serial) correlation between parent and child's state. While this intuition suggests that intergenerational correlation will vary with sensitivity to social circumstances, the fact that the intergenerational correlation is always exactly directly proportional to the variance of the investment probabilities is a bit unexpected. The proof takes some work, which appears in the appendix.

3.4 Cost efficiency of behavior and social sensitivity

One of the reasons that we are interested in correlation in behavior across generations is that it suggests that there are influences which distort behavior away from efficient levels. For example, suppose that one player invests in the high action even though that player has a high cost due to the fact that a large fraction of his or her community is taking the high action, while another player does not invest despite a much lower cost due to the fact that a low fraction of his or her community is taking the high action. This leads to more costly investment than would be present in a world where the players were switched. In a perfect world, where we could exchange these two players (and make some transfers) we would have a Pareto improvement.

We use the term cost efficiency to emphasize that this analysis examines the costs only. It does not represent a measure of Pareto efficiency unless the planner has the ability to make transfers and to move agents around based on their inherent costs. The idea is that a society could generate more total welfare if individuals with lower costs were placed in settings with higher investment by neighbors, and vice versa. Thus, a failure of cost efficiency is a weak form of inefficiency, in that it is only inefficient when measured relative to strong assumptions on transfers and ability to move agents. Nonetheless, it does clarify some specific aspects of investment behavior that are present in socially sensitive societies: relatively untalented individuals can end up investing while relatively talented individuals do not.

A rough measure that keeps track of this distortion is the total investment expenditure to get a given average level of investment. That is, consider two different communities whose investment decisions are described by p and p'. To keep a level playing field, compare communities with the same average steady-state investment rates so that $\overline{p} = \overline{p}'$. We can then compare their respective average costs of investment to see how their comparison depends on the relationship between p and p'.

To do this, we derive the expected costs of investment in steady state. To get a closed-form solution, we examine a case where any given player's costs of investment are uniformly distributed on $[0, C]$. Recall, that the model is one where the benefits depend on the community choices (k_{-i}) and then the costs are simply individual-specific. In particular, recall that $p_k = F(\pi(1, k))$. That is, the probability that i chooses the high action when k players are high-action is the probability that i's cost of investment is below $\pi(1, k)$, which is i's expected payoff from the high action.

Given that costs are uniformly distributed on $[0, C]$, $p_k = \pi(1, k)/C$. Conditional on investing when $k_{-i} = k$ we then can conclude that the expected costs are $\pi(1, k)/2$. Thus, these are $Cp_k/2$. Then, the average cost per capita of choosing the high action can be written as:

$$Cost(p) = \sum_{k=0}^{n-1} \frac{\mu_k^{-i} p_k}{\overline{p}} \frac{C}{2} p_k.$$

Here, $\mu_k^{-i} p_k / \overline{p}$ is the conditional probability that there were k other high-action players at the time that i made his or her choice, conditional on i investing. We can then conclude the following.

PROPOSITION **4** *If the steady-state average investment rates are the same, $\overline{p}' = \overline{p}$, then under the above formulation $Cost(p') > Cost(p)$ if and only if $Var(p') > Var(p)$.*

For two processes p and p' that have the same overall percentage of high-action players $\overline{p}$, the more sensitive one has a higher overall cost associated with it. It is thus less efficient in terms of the costs of investment for the same average level of investment.

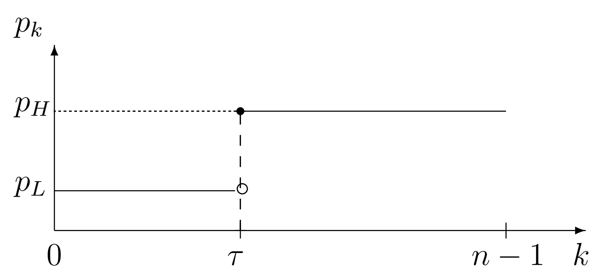
4 A Threshold Model

We now describe a specific case of our model that we call the "threshold" model. This is useful as it is a very sparse model, and yet still gives rich dynamics and has a very natural interpretation.

More precisely, we say that there are threshold investment decisions if there exists $1 \geq p_H \geq p_L \geq 0$ and $\tau \in \{1, ..., n-1\}$ such that $p_k = p_H$ if $k \geq \tau$ and $p_k = p_L$ is $k < \tau$. The difference between the high and the low investment probability is $p_H - p_L$. Holding τ fixed, when p_H increases and/or p_L decreases, individual investment decisions become more sensitive to the state of the community. This is a generalization of the dyad example to an arbitrary population size n.

The idea behind this is that there is some threshold number of neighbors such that if at least that many neighbors have invested then investment becomes attractive for the individual and below which it does not. Allowing the probabilities to differ from 0 and 1 means that decisions are not completely deterministic but also noisy and possibly influenced by some idiosyncratic and exogenous factors.

The transition probabilities are depicted below.



For the threshold model, the steady-state distribution over action choices can be readily computed and is equal to:

$$\mu_k = \begin{cases} \mu_0 \binom{n}{k} \left(\frac{p_L}{1-p_L}\right)^k, \text{ for } k < \tau, \text{ and} \\ \mu_0 \left[\frac{p_L(1-p_H)}{p_H(1-p_L)}\right]^\tau \binom{n}{k} \left(\frac{p_H}{1-p_H}\right)^k, \text{ for } k \geq \tau, \end{cases}$$

where μ_0 normalizes the sum to one.[13]

Despite this closed-form expression for the steady-state distribution μ for the aggregate actions within the group, the resulting bivariate process (2) for the parent-child outcomes is still harder to analyze than in the simple dyad case. In particular, simply increasing the spread between p_H and p_L need not increase the steady-state intergenerational correlation when $n > 2$, without an additional condition. Holding τ fixed, an increase in p_H and a decrease in p_L that keep the average investment level $\overline{p}$ constant lead to an increase in intergenerational correlation. Otherwise, it need not.

If (p_H, p_L) and (p'_H, p'_L) are such that $p'_H > p_H$, $p'_L < p_L$ and $\overline{p}$ is the same under (n, τ, p_H, p_L) than under (n, τ, p'_H, p'_L), we say that (p'_H, p'_L) is a $\overline{p}$-preserving spread of (p_H, p_L).

[13]The expressions for the steady-state probabilities follow from iterated application of (5) in the appendix.

PROPOSITION 5 *In the threshold model, an $\overline{p}$-preserving spread increases the (steady state) intergenerational correlation in human capital investments.*[14]

We prove Proposition 5 by applying Theorem 1 (in fact, Corollary 1) and showing that an $\overline{p}$-preserving spread leads to an increase in the variance.

4.1 Education and Social Mobility

Consider an application where the low and high actions correspond, respectively, to investing in low and high human capital levels. Let $\pi(0, k_{-i})$ and $\pi(1, k_{-i})$ be the expected discounted stream of wages conditional on i being of human capital level 0 or 1, respectively. Idiosyncratic investment costs encompass innate skills and abilities, as well as direct costs of education. Note that this assumption completely abstracts from any correlation in costs of education across generations. *We do not do this because we believe that costs are independent across generations, but rather because we wish to isolate the social setting as a driver of social immobility.*[15,16]

The critical property in applying our results from above is that the vector of investment decisions p_k is non-decreasing in k. Economists and sociologists have proposed and documented a variety of channels for this positive social externality, including role-model theory, peer-pressure for conformity, heredity of cultural traits or skills.[17] Given the direct modeling

[14] This refers to the correlation between parent and child investment states of any given dynasty at any time, $Cor\,(p)$, when the initial distribution is the steady state distribution μ.

[15] Intergenerational correlation in costs of education has been studied as a driver of social immobility, both theoretically and empirically. See Bowles and Gintis (2002) and references therein. We deliberately isolate social effects as a driver for observed social mobility patterns to work out their possible functioning and sizeable impact. In particular, we abstract from any direct source of parent-child correlation, be it heredity of cultural or genetic traits, or any other direct economic linkage mechanism across generations. Therefore, our approach is complementary to the already existing theories of social mobility, such as: wealth transfers in the form of bequests by altruistic parents or genetic inheritance of productive abilities (Becker and Tomes 1979), imperfect credit markets imposing borrowing constraints at the bottom of the earnings distribution (Loury 1981), and segregation of individuals into homogeneous communities that spur positive spillover effect (Bénabou 1993, 1996 and Durlauf 1996).

[16] Economists' measure of social mobility is mostly based on parent-child correlation of (log) earnings and income. Recent estimates of the intergenerational correlation of long-run log earnings lie in the range $[.4, .6]$ for the U.S. and the U.K., and $[.2, .4]$ for Germany and Sweden (see, e.g. Björklund and Jännti 1999, Solon 2002, Piketty 2000 and references therein, in particular, Solon 1992, Zimmerman 1992, Björklund and Jännti 1997, Dearden *et al.* 1997, and Mulligan 1997. Sociologists' analyses of intergenerational correlation focus on mobility tables. Given a hierarchy of occupational classes (or social classes), mobility tables relate the children-class destination to the parents-class origin. Odds ratios compute the relative likelihood of identical versus different parent-child class. Estimated odds ratios vary from 1 and 15 depending on the occupational class and for a large set of countries (see, e.g., Ganzenboom and Treiman (1996) and Erikson and Goldthorpe 1992, 2002).

[17] Role-model theories assert that children learn how to behave by observing the adults in their social network (Jencks and Mayer 1990). Exogenous norm-enforcing mechanisms also induce conformism among peers (Coleman 1990). A setting where preferences for conformity are endogenized appears in Calvó-Armengol

of the social interaction through the variation of p_k as a function of k, the application to social mobility of our model encompasses all such channels. Social networks, which are pervasively used in labor markets to disseminate job information, can also lead to substantial complementarities in human capital investment decisions (see Calvó-Armengol and Jackson 2004,2007 for details).

All of our previous results apply. In particular, Theorem 1 relates the intergenerational correlation in human capital to the strength of the social externality in education choices, and vice-versa. Proposition 4 provides an efficiency assessment of observed levels of intergenerational correlation.

For instance, suppose that a child pays attention to the social community with some probability. More precisely, with probability q the child's investment decision is governed by p_k, and with probability $(1-q)$ it is given by $\overline{p}$. Note that changes in q do not modify the average investment probability $\overline{p}$, but do affect its sensitivity to the social situation. Applying Theorem 1 (or its Corollary 1), we conclude the following.

Corollary **2** *The parent-child correlation in human capital investments starting from the steady state distribution, and the cost inefficiency, both increase with the probability that the child invests as a function of the community, q.*

To the extent that the overlap of the social universe across generations is higher for the two-tails of the income distribution, this finding is consistent with the $U-$shaped intergenerational income correlation curve as a function of income levels documented by Cooper *et al.* (1994).[18] Beyond issues of equity and equality of opportunity, the previous result also relates social mobility to economic efficiency in a way that echoes conclusions found by modeling immobility due to other sources. For instance, inefficiencies are also present in incomplete market models of social mobility (Loury 1981, and Banerjee and Newman 1991) and assortative matching models (Becker 1973, Cole, Mailath, and Postlewaite 1994).

In *Supplementary Material* to this manuscript, we fit the threshold model to data on parent-child education levels in European countries. In light of the identification issue addressed in Proposition 2, this exercise is meant to illustrate how the model can generate behavior consistent with data, and as an alternative explanation to that of direct parent-child influence. This is not meant as a test of the model.

and Jackson (2004,2007) for the case of "drop-out" decisions based on the state of one's network contacts. Heredity of preferences theory presumes that low-class children inherit preferences that discount future payoffs more than children from wealthier parents, and hence they invest less in human capital (Boudon 1973). Heredity of cultural traits theories claims that high-class children inherit better suited attitudes and aptitudes for culture and education, and then correlation follows from a parental earnings bias in social capital endowment for the child (Bourdieu and Passeron 1964).

[18] See Wright Mills (1945) for a seminal analysis of the social endogamy of the business elite in the U.S. and Wilson (1987) and Jencks and Mayer (1990) for an analysis of the social and economic consequences of living in the inner city in the U.S.

5 Conclusion

We have provided a parsimonious model that shows how individual decisions depend on the choices of the group, and how this leads to parent-child correlation. Increased sensitivity of an individual's decision to the choices in his or her social neighborhood corresponds to increased parent-child correlation. We show how the social driver and the parental channel are observationally equivalent, and describe a possible identification strategy. We provide an application to educational decisions where social networks, which are pervasively used in labor markets to disseminate job information, generate the sort of complementarities in human capital investment decisions that are considered in our model. We also analyze a special case of the model, the threshold investment model, for which the long-run steady-state distribution is easily characterized.

5.1 Other Applications

Beyond the analysis of parent-child correlations, our model provides a general set up to generate time series predictions from cross-sectional peer effects and, more generally, from imitative behavior. Consider a finite number of agents that take a binary 0-1 decision. Time is discrete and, each period, one randomly selected agent can revise his or her previous choice. This agent then samples some individuals around and imitates the most popular decision in this sample, possibly with some noise. Popularity can correspond to some measure of relative payoff, or simply reflect the relative frequencies of both actions in the sample.[19] Clearly, in this set up where imitation spurs a within-periods and across-agents community positive externality in individual decisions. Our analysis suggest that cross-sectional imitation induces across-periods persistence in individual and aggregate behavior in the long-run.

A particular (simple) example of this imitative process is the following. With some small probability ε, the revised choice follows a coin toss. With the complementary probability $1-\varepsilon$, instead, the agent who's choice is to be revised is matched randomly with one other agent, all matches being equally likely, and adopts the action of the matched partner. The resulting probability of choosing 1 conditional on k of the other $n-1$ people in the society having chosen 1 is:

$$p_k = \frac{1}{2}\varepsilon + (1-\varepsilon)\frac{k}{n-1}.$$

Clearly, this process induces a vector of investment probabilities which are non-decreasing in k, and fits our general model. This model is analyzed in Kirman (1993), in the context of ant behavior, as a proxy for some sorts of investment behavior.[20] He shows that the continuous population limit of the steady-state distribution for this process is a Beta distribution, which corresponds to overcrowding in choices. Here, changes in ε do not change

[19]Sobel (2000) discusses the connections between these models of imitation and the literature on social learning.

[20]For a related model, see also Ellison and Fudenberg (1995).

the average probability to choose action 1, which is (roughly) one-half, but does affect the sensitivity of the choice to the social state. From our results we can conclude that the time series persistence of the overcrowded action starting from the steady state decreases with ε, a result that corroborates a similar numerical observation in Kirman (1993).

References

[1] Abramson, Paul R. 1973. "Intergenerational Social Mobility and Partisan Preferences in Britain and Italy." *Comparative Political Studies*, 6: 224-234.

[2] Banerjee, Abhijit V. and Andrew F. Newman. 1991. "Risk-bearing and the Theory of Income Distribution." *Review of Economic Studies*, 58(2): 211-235.

[3] Becker, Gary S. 1973. "A Theory of Marriage: Part I." *Journal of Political Economy*. 81(4): 813-846.

[4] 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.

[5] Bénabou, Roland. 1993. "Workings of a City: Location, Education, Production." *Quarterly Journal of Economics*, 108(3): 619–653.

[6] Bénabou, Roland. 1996. "Equity and Efficiency in Human Capital Investment: The Local Connection." *Review of Economic Studies*, 63(2): 237-264.

[7] Berkman, Lisa F. and Ichiro Kawachi. 2000. *Social Epidemiology*, New York: Oxford University Press.

[8] Bisin, Alberto and Thierry Verdier. 2000. "Beyond the Melting Pot: Cultural Transmission, Marriage, and the Evolution of Ethnic and Religious Traits." *Quarterly Journal of Economics*, 115(3): 955-988.

[9] Björklund, Anders and Markus Jäntti. 1997. "Intergenerational Mobility in Sweden Compared to the United States." *American Economic Review*, 87(5): 1009-1018.

[10] Björklund, Anders and Markus Jäntti. 1999. "Intergenerational Mobility of Socioeconomic Status in Comparative Perspective." manuscript, University of Stockholm.

[11] Borjas, George J. 1992. "Ethnic Capital and Intergenerational Mobility." *Quarterly Journal of Economics*, 107(1): 123-150.

[12] Boudon, Raymond. 1973. *L'Inégalité des Chances*, Paris: Armand Colin.

[13] Bourdieu, Pierre and Jean Claude Passeron. 1964. *Les Héritiers*, Paris: Editions de Minuit.

[14] Bowles, Samuel and Herbert Gintis. 2002. "The Inheritance of Inequality." *Journal of Economic Perspectives*, 16(3): 3-30.

[15] Calvó-Armengol, Antoni and Matthew O. Jackson. 2004. "The Effects of Social Networks on Employment and Inequality." *American Economic Review,* 94(3): 426-454.

[16] Calvó-Armengol, Antoni and Matthew O. Jackson. 2007. "Networks in Labor Markets: Wage and Employment Dynamics and Inequality." *Journal of Economic Theory,* 132(1): 27-46.

[17] Card, David and Alan B. Krueger. 1992. "School Quality and Black-White Relative Earnings: A Direct Assessment." *Quarterly Journal of Economics*, 107(1): 151-200.

[18] Carneiro, Pedro and James J. Heckman. 2002. "The Evidence on Credit Constraints in Post-Secondary Schooling." manuscript, Chicago University.

[19] Cole, Harold, George J. Mailath, and Andrew Postlewaite. 1994. "Class Systems and the Enforcement of Social Norms." *Journal of Public Economics,* 70(1): 157-183.

[20] Coleman, James. 1990. *Foundations of Social Theory*, Cambridge: Harvard University Press.

[21] Comi, Simona. 2003. "Intergenerational Mobility in Europe: Evidence from ECHP." manuscript, Università degli Studi di Milano.

[22] Cooper, Suzanne, Steven Durlauf, and Paul Johnson. (1994) "On the Evolution of Economic Status Across Generations." *American Statistical Association,* Papers and Proceedings 50-58.

[23] Corcoran, Mary, Roger Gordon, Deborah Laren, and Gary Solon. 1992. "The Association Between Men's Economic Status and their Family and Community Origins." *Journal of Human Resources*, 27(4): 575-601.

[24] Dal Bó, Ernesto, Dal Bó, Pedro and Jason Snyder. 2007. "Political Dynasties." manuscript, U.C. Berkeley.

[25] Dearden, Lorraine, Steve Machin, and Howard Reed. 1997. "Intergenerational Mobility in Britain." *Economic Journal,* 107: 47-66.

[26] Durlauf, Steven. 1996. "A Theory of Persistent Income Inequality." *Journal of Economic Growth*, 1(1): 75-93.

[27] Ellison, Glenn and Drew Fudenberg. 1995. "Word-of-Mouth Communication and Social Learning." *Quarterly Journal of Economics*, 110(1): 93-125.

[28] Erikson, Robert and John G. Goldthorpe. 1992. *The Constant Flux: A Study of Class Mobility in Industrial Societies*, Oxford: Clarendon.

[29] Erikson, Robert and John G. Goldthorpe. 2002. "Intergenerational Inequality: A Sociological Perspective." *Journal of Economic Perspectives* 16(3): 31-44.

[30] Galea, Sandro, Arijit Nandi, and David Vlahov. 2004. "The Social Epidemiology of Substance Use." *Epidemiologic Reviews,* 26: 36-52.

[31] Ganzeboom, Harry and Donald Treiman. 1996. "Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations." *Social Science Research,* 25(3): 201-239.

[32] Goux, Dominique and Eric Maurin. 1997. "Meritocracy and Social Heredity in France: Some Aspects and Trends." *European Sociological Review,* 13(2)159-178.

[33] Heckman, James J. and Alan B. Krueger. 2004. *Inequality in America: What Role for Human Capital Policies?*, Cambridge: MIT Press.

[34] Hoge, Dean, Gregory Petrillo, and Ella Smith. 1982. "Transmission of Religious and Social Values from Parents to Teenage Children." *Journal of Marriage and the Family,* 44(3): 569-580.

[35] Ioannides, Yannis and Linda Datcher Loury. 2004. "Job Information Networks, Neighborhood Effects and Inequality." *Journal of Economic Literature,* 42(4): 1056–1093.

[36] Jencks, C. and S. Mayer. 1990. "The Social Consequences of Growing Up in a Poor Neighborhood: A Review." in *Concentrated Urban Poverty in America*, Washington: National Academy.

[37] Johnson, D. 1996. "Fathers Presence Matters: A Review of the Literature." *National Center on Fathers and Families*, working paper.

[38] Kane, Thomas. 1994. "College Entry by Blacks Since 1970: The Role of College Costs, Family Background, and the Returns to Education." *Journal of Political Economy,* 102(5): 878-911.

[39] Kirman, Alan. 1993. "Ants, Rationality, and Recruitment." *Quarterly Journal of Economics,* 108(1): 137-156.

[40] Kling, Jeffery, Liebman, Jeffrey and Lawrence Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica,* 75(1): 83-119.

[41] Loury, Glenn C. 1981. "Intergenerational Transfers and the Distribution of Earnings." *Econometrica* 49(4): 843-867.

[42] Mulligan, Casey. 1997. *Parental Priorities and Economic Inequality,* Chicago: University of Chicago Press.

[43] Piketty, Thomas. 1995. "Social Mobility and Redistributive Politics." *Quarterly Journal of Economics*, 110(3), 551-584.

[44] Piketty, Thomas. 2000. "Theories of Persistent Inequality and Intergenerational Mobility." in *Handbook of Income Distribution, Vol. 1*, Anthony B. Atkinson and François Bourguignon, eds., Elsevier: Amsterdam.

[45] Sobel, Joel. 2000. "Economists' Models of Learning." *Journal of Economic Theory*, 94(2): 241-261.

[46] Solon, Gary. 1992. "Intergenerational Income Mobility in the United States." *American Economic Review*, 82(3): 393-408.

[47] Solon, Gary. (2002) "Cross-Country Differences in Intergenerational Earnings Mobility." *Journal of Economic Perspectives*, 16(3): 59-66.

[48] Thomson, Kenneth. 1971 "A Cross-National Analysis of Intergenerational Social Mobility and Political Orientation." *Comparative Political Studies* 4(3): 3-20.

[49] Wilson, William J. 1987. *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, Chicago: University of Chicago Press.

[50] Wright Mills, C. 1945. "The American Business Elite: A Collective Portrait." *Journal of Economic History* 5: 20-44.

[51] Zimmerman, David J. 1992. "Regression Toward Mediocrity in Economic Stature." *American Economic Review* 82(3): 409-429.

Appendix

Proof of Proposition 1: We first show that $\mu_{k+1} = a_k \mu_k$ for all $0 \leq k \leq n-1$, where

$$a_k = \frac{n-k}{k+1}\left(\frac{p_k}{1-p_k}\right). \tag{5}$$

Indeed, consider a state where exactly k players take the high action. At steady-state, the inflow to and the outflow from this state exactly balance each other. This is written as

$$\begin{aligned} \mu_0 p_0 &= \mu_1 \frac{1}{n}(1-p_0), \text{ for } k=0 \\ \mu_{k-1}\frac{n-(k-1)}{n}p_{k-1} + \mu_{k+1}\frac{k+1}{n}(1-p_k) &= \mu_k \frac{k}{n}(1-p_{k-1}) + \mu_k \frac{n-k}{n}p_k, \text{ for } 1 \leq k \leq n-1 \\ \mu_{n-1}\frac{1}{n}p_{n-1} &= \mu_n(1-p_{n-1}), \text{ for } k=n, \end{aligned}$$

and the identity (5) follows immediately.

The fact that $\mu_{k+1} = a_k \mu_k$, for all $0 \leq k \leq n-1$ implies that $\mu_i/\mu_j = a_j a_{j+1} \cdots a_{i-1}$, for all $i > j$. Noting that $p'_k \geq p_k$ implies $a'_k \geq a_k$ (with corresponding strict inequalities), it follows that

$$\mu'_i/\mu'_j \geq \mu_i/\mu_j$$

for all $i > j$, with strict inequality for some pairs when $p' \neq p$. Given that $\sum_{k=0}^{n} \mu'_k = 1 = \sum_{k=0}^{n} \mu_k$, the result follows directly. ■

Proof of Proposition 2: Let $n = 2$ and consider the direct parental influence model where investment probabilities are $d = (d_0, d_1)$, where d_i is the investment probability when parental status is $i \in \{0, 1\}$. Let $1 \geq d_1 \geq d_0 \geq 0$.

Let μ_0 be the probability of any agent being a 0. The probability that a child is a 0 is then $\mu_0 = (1-d_0)\mu_0 + (1-d_1)(1-\mu_0)$, or:

$$\mu_0 = \frac{1-d_1}{1-d_1+d_0}.$$

Then we directly find the distribution $\nu'(d_0, d_1)$ over parent-child behavior:

	0	1
0	$\frac{(1-d_1)(1-d_0)}{1+d_0-d_1}$	$\frac{d_0(1-d_1)}{1+d_0-d_1}$
1	$\frac{d_0(1-d_1)}{1+d_0-d_1}$	$\frac{d_0 d_1}{1+d_0-d_1}$

Consider now the dyad model, defined by $p = (p_0, p_1)$, with $1 \geq p_1 \geq p_0 \geq 0$. The steady-state distribution $\nu(p_0, p_1)$ over parent-child behavior is now:

	0	1
0	$\frac{(1-p_0)^2(1-p_1)+(1-p_1)^2 p_0}{1+p_0-p_1}$	$\frac{p_0(1-p_1)(1-p_0+p_1)}{1+p_0-p_1}$
1	$\frac{p_0(1-p_1)(1-p_0+p_1)}{1+p_0-p_1}$	$\frac{p_0 p_1^2+p_0^2(1-p_1)}{1+p_0-p_1}$

We proceed in two steps.

We first show that, given any bivariate distribution $\nu'(d_0, d_1)$ for the direct influence model, there exists some $p = (p_0, p_1)$, with $1 \geq p_1 \geq p_0 \geq 0$ that replicates it, that is, such that $\nu(p_0, p_1) = \nu'(d_0, d_1)$. Notice that:

$$\frac{\nu'_{11}}{\nu'_{10}} = \frac{d_1}{1-d_1} \text{ and } \frac{\nu'_{00}}{\nu'_{10}} = \frac{1-d_0}{d_0}$$

Also,

$$\frac{\nu_{11}}{\nu_{10}} = \frac{p_1^2 + p_0(1-p_1)}{1-(p_1^2+p_0(1-p_1))}.$$

So, a first equation expressing (p_0, p_1) as a function of (d_0, d_1) is:

$$d_1 = p_1^2 + p_0(1-p_1) = p_1(p_1 - p_0) + p_0. \tag{6}$$

From the expression of $\nu_{11}(p_0, p_1)$ and (6) we get:

$$\nu_{11} = \frac{p_0 d_1}{1 + p_0 - p_1}.$$

To simplify notations, let $\lambda = d_1/\nu'_{11} \geq 1$. A second equation expressing (p_0, p_1) as a function of (d_0, d_1) (and $\nu'(d_0, d_1)$) is then:

$$p_1 - p_0 + \lambda p_0 = 1. \tag{7}$$

We solve (6) and (7) with unknowns (p_0, p_1) such that $0 \leq p_0 \leq p_1 \leq 1$. Using the fact that $p_1 - p_0 = 1 - \lambda p_0$ from (7) we rewrite (6) as:

$$p_1 + p_0 - \lambda p_0 p_1 = d_1.$$

So the system of two equations with two unknowns is:

$$\begin{cases} p_1 - p_0 + \lambda p_0 = 1 \\ p_1 + p_0 - \lambda p_0 p_1 = d_1 \end{cases}$$

Solving for p_0 in the first equation and plugging in the second one gives:

$$\begin{cases} p_1 - p_0 + \lambda p_0 = 1 \\ \lambda p_1^2 - 2p_1 + 1 - (\lambda - 1)\, d_1 = 0 \end{cases} \tag{8}$$

We solve the second order equation in p_1. The highest root of this second-order polynomial is:

$$p_1^* = \frac{1}{\lambda}\left[1 + \sqrt{(\lambda - 1)(\lambda d_1 - 1)}\right].$$

p_0^* is then deduced from the first equation of (8):

$$p_0 = \frac{1}{\lambda}\left[1 - \sqrt{\frac{\lambda d_1 - 1}{\lambda - 1}}\right].$$

Clearly, $v(p_0^*, p_1^*) = \nu'(d_0, d_1)$. We check that $1 \geq p_1^* \geq p_0^* \geq 0$. First, $1 \geq p_1^*$ is equivalent to $\lambda \geq \lambda d_1$, which is true because $d_1 \leq 1$. Second, $p_0^* \geq 0$ is equivalent to $\lambda \geq \lambda d_1$, which again is true. Finally, simple algebra shows that $p_1^* \geq p_0^*$ also holds.

Second, we show that the set of bivariate distributions generated by the parental direct influence model corresponds to all symmetric bivariate distributions with non-negative correlation. We want to generate the following symmetric bivariate:

	0	1
0	a	$\frac{1-(a+b)}{2}$
1	$\frac{1-(a+b)}{2}$	b

(9)

with the following transition probabilities (child in columns; parent in rows):

$$\begin{bmatrix} 1-d_0 & d_0 \\ 1-d_1 & d_1 \end{bmatrix}.$$

Without any loss of generality, we can write $d_1 = 1 - xd_0$. Let $a, b \geq 0$, $a + b \leq 1$. We solve for the bivariate distribution $\nu(d_0, d_1)$, and equate it with the bivariate distribution (9), that is:

$$\begin{cases} a = \frac{x(1-d_0)}{1+x} \\ b = \frac{1-xd_0}{1+x} \end{cases}.$$

We have:

$$d_0^* = 1 - a\frac{1+x}{x},$$

and plugging into the second equation gives, after some algebra:

$$x^* = \frac{1+a-b}{1+b-a}$$

We need to check that $1 - x^* d_0^* \geq 0$, that is, $x^* \leq (1+a)/(1-a)$, which is true. We also impose $1 - x^* d_0^* = d_1^* \geq d_0^*$, that is, $1 \geq (1 + x^*) d_0^*$. Using the expressions for x^* and d_0^*, this inequality becomes, after some algebra:

$$1 + 2ab \geq (1-a)^2 + (1-b)^2$$

One can easily check that this is equivalent to requiring that the correlation corresponding to (9) be non-negative. ■

Proof of Proposition 3: The proof is related to that of Proposition 3 in Calvó-Armengol and Jackson (2004).

Consider a dyad and $s \in \{0,1\}^2$ a state representing the actions taken by the dyad members. We denote random variables by capital letters, S, and realizations by small letters, s. Let P be the two dynasty Markov process with transitions in Figure 1. So $P_{ss'}$ is the probability that $S_t = s'$ conditional on $S_{t-1} = s$.

We introduce some useful definitions and notations.

Let μ and ν be two probability distributions on S. We say that μ *dominates* ν if $\mathbb{E}_\mu(f) \geq \mathbb{E}_\nu(f)$ for all nondecreasing function $f : \mathbb{R} \to \{0,1\}^2$. We say that the domination is strict when the inequality is strict for some f. Domination extends first-order stochastic dominance to random vectors.

Let $\mathcal{E} = \{E \subset \{0,1\}^2 : s \in E, s' \geq s \Rightarrow s' \in E\}$. The elements of $\mathcal{E}$ are called increasing sets. Increasing sets are useful to handle coordinate-wise partial orderings, and to extend dominance relationships on random variables to random vectors.

Lemma 1 *Let μ and ν be two probability distributions on $\{0,1\}^2$. Then, μ dominates ν if and only if $\mu(E) \geq \nu(E)$, for all $E, E' \in \mathcal{E}$. The domination is strict if the inequality is strict for at least one E.*

Proof of Lemma 1: The result follows from Lemma 6, p. 444 in Calvó-Armengol and Jackson (2004).[21] ■

Given a measure ξ on $\{0,1\}^2$, let ξP denote the measure induced by multiplying the (1×4) vector ξ by the (4×4) transition matrix P. This is the distribution over states induced by a starting distribution ξ multiplied by the transition probabilities P.

In what follows we assume that $1 > p_1 > p_0 > 0$.

Lemma 2 *Let μ and ν be two probability distributions on $\{0,1\}^2$. If μ dominates (resp. strictly dominates) ν, then μP dominates (resp. strictly dominates) νP.*

Proof of Lemma 2: This result corresponds to Lemma 9, p. 445 in Calvó-Armengol and Jackson (2004). A crucial step in the proof uses the fact that $P_{s'E} \geq P_{sE}$, for all $s' \geq s$ and $E \in \mathcal{E}$, with a strict inequality for at least one E when $s' \neq s$.

We prove this. We establish the result for pairs of adjacent states, that is, $s' = (1,1)$ and $s \in \{(1,0), (0,1)\}$, or $s' \in \{(1,0), (0,1)\}$ and $s = (0,0)$. The more general statement then follows from a chain of comparisons across such adjacents states.

Let $s' \geq s$ adjacent, and suppose without loss of generality that $s'_1 > s_1$ and $s'_2 = s_2$. Let $\Pr_s(S_{2,t})$ be the probability distribution on $S_{2,t}$ conditional on $S_{t-1} = s$. If $s = (0,0)$ and $s' = (1,0)$, then $\Pr_s(S_{2,t})$ (resp. $\Pr_{s'}(S_{2,t})$) is a Bernoulli distribution with success probability $p_0/2$ (resp. $p_1/2$). If $s = (0,1)$ and $s' = (1,1)$, then $\Pr_s(S_{2,t})$ (resp. $\Pr_{s'}(S_{2,t})$) is a Bernoulli distribution with success probability $(1+p_0)/2$ (resp. $(1+p_1)/2$). Given that $p_1 > p_0$, it follows that $\Pr_{s'}(S_{2,t})$ strictly first-order stochastically dominates $\Pr_s(S_{2,t})$. Then, given that $s'_1 > s_1$, it follows that $\Pr_{s'}(S_t)$ strictly dominates $\Pr_s(S_t)$. Lemma 1 then implies that $P_{s'E} \geq P_{sE}$, for all $s' \geq s$ and $E \in \mathcal{E}$, with a strict inequality for at least one E when $s' \neq s$. ■

For any t, let E^t_{i0} (resp. E^t_{i1}) be the set of s_t such that $s_{i,t} = 0$ (resp. $s_{i,t} = 1$). For instance, $E^t_{20} = \{(0,0), (1,0)\}$. Let μ^* be the steady-state distribution for the two dynasty Markov process:

$$\begin{bmatrix} \mu^*_{11} \\ \mu^*_{01} \\ \mu^*_{10} \\ \mu^*_{00} \end{bmatrix} = \frac{1}{1+p_0-p_1} \begin{bmatrix} p_0 p_1 \\ p_0(1-p_1) \\ p_0(1-p_1) \\ (1-p_0)(1-p_1) \end{bmatrix}.$$

We now state another useful result.

Lemma 3 *$\mu^*(. \mid E^t_{i1})$ strictly dominates $\mu^*(. \mid E^t_{i0})$, for all i and t.*

Proof of Lemma 3: By Lemma 1, it is enough to check that $\mu^*(E \mid E^t_{i1}) \geq \mu^*(E \mid E^t_{i0})$, for all $E, E' \in \mathcal{E}$, with at least one strict inequality. In the next table, the rows correspond

[21] Variations of a more general version of this lemma appear in the statistics literature.

to the elements of $\mathcal{E}$, while the columns give the corresponding conditional probabilities for $i=1$;[22] the case $i=2$ is symmetric.

$\mathcal{E}$	$\mu^*(. \mid E_{10}^t)$	$\mu^*(. \mid E_{11}^t)$
$\{(1,1)\}$	0	$\mu_{11}^*/(\mu_{10}^*+\mu_{11}^*)$
$\{(1,0),(1,1)\}$	0	1
$\{(0,1),(1,1)\}$	$\mu_{01}^*/(\mu_{00}^*+\mu_{01}^*)$	$\mu_{11}^*/(\mu_{10}^*+\mu_{11}^*)$
$\{(0,0),(1,1),(1,0),(1,1)\}$	1	1

To establish strict domination, we only need to check that $\mu_{11}^*/(\mu_{10}^*+\mu_{11}^*) \geq \mu_{01}^*/(\mu_{00}^*+\mu_{01}^*)$, equivalent to $p_1 \geq p_0$, which is true under our assumption that $p_1 > p_0$. ■

We are now ready to prove Proposition 3.

For any $t > t' \geq 0$ and any $i = 1,2$, let $h_{i1}^{t',t}$ be the event that $S_{it'} = S_{it'+1} \cdots = S_{it-1} = S_{it} = 1$. Let $h_{i0}^{t',t}$ be the event that $S_{it'} = 0$ and $S_{it'+1} \cdots = S_{it-1} = S_{it} = 1$. So, $h_{i0}^{t',t}$ and $h_{i1}^{t',t}$ differ only in i's status at date t'.

Let $t > t'+1$. We want to show that:

$$P\left(S_{i,t+1} = 1 | h_{i1}^{t't}\right) > P\left(S_{i,t+1} = 1 | h_{i1}^{t'+1,t}\right). \tag{10}$$

Since $P\left(S_{i,t+1} = 1 | h_{i1}^{t'+1,t}\right)$ is a weighted average of $P\left(S_{i,t+1} = 1 | h_{i0}^{t't}\right)$ and $P\left(S_{i,t+1} = 1 | h_{i1}^{t't}\right)$, (10) is equivalent to showing that

$$P\left(S_{i,t+1} = 1 | h_{i1}^{t't}\right) > P\left(S_{i,t+1} = 1 | h_{i0}^{t't}\right). \tag{11}$$

By Bayes' rule,[23]

$$P\left(S_{i,t+1} = 1 | h_{i1}^{t't}\right) = \frac{P\left(S_{i,t+1} = 1, h_{i1}^{t't}\right)}{P\left(S_{i,t+1} = 1, h_{i1}^{t't}\right) + P\left(S_{i,t+1} = 0, h_{i1}^{t't}\right)}$$

and

$$P\left(S_{i,t+1} = 1 | h_{i0}^{t't}\right) = \frac{P\left(S_{i,t+1} = 1, h_{i0}^{t't}\right)}{P\left(S_{i,t+1} = 1, h_{i0}^{t't}\right) + P\left(S_{i,t+1} = 0, h_{i0}^{t't}\right)}$$

Using the two above equations to rewrite (11), and rearranging terms leads to:

$$P\left(S_{i,t+1} = 1, h_{i1}^{t't}\right) P\left(S_{i,t+1} = 0, h_{i0}^{t't}\right) > P\left(S_{i,t+1} = 1, h_{i0}^{t't}\right) P\left(S_{i,t+1} = 0, h_{i1}^{t't}\right). \tag{12}$$

Dividing each side of the above inequality by $\mu^*(E_{i0}^{t'})\mu^*(E_{i1}^{t'})$, we deduce that to establish (12) it is enough to show that:

$$\frac{P\left(S_{i,t+1} = 1, h_{i1}^{t't}\right)}{\mu^*(E_{i1}^{t'})} > \frac{P\left(S_{i,t+1} = 1, h_{i0}^{t't}\right)}{\mu^*(E_{i0}^{t'})} \tag{13}$$

[22] We have $E_{10}^t = \{(0,0),(0,1)\}$ and $E_{11}^t = \{(1,0),(1,1)\}$.

[23] Note that $\left(S_{i,t+1} = 1, h_{i1}^{t',t}\right) = h_{i1}^{t',t+1}$.

and

$$\frac{P\left(S_{i,t+1}=0,h_{i0}^{t't}\right)}{\mu^*(E_{i0}^{t'})} > \frac{P\left(S_{i,t+1}=0,h_{i1}^{t't}\right)}{\mu^*(E_{i1}^{t'})}. \tag{14}$$

Let us show (13), as the argument for (14) is analogous.

Then,

$$\frac{P\left(S_{i,t+1}=1,h_{i1}^{t't}\right)}{\mu^*(E_{i1}^{t'})} = \sum_{s_{t'}\in\{0,1\}^2}\sum_{s_{t'+1}\in E_{i1}^{t'+1}}\cdots\sum_{s_{t+1}\in E_{i1}^{t+1}}\mu^*(s_{t'}|E_{i1}^{t'})P_{s_{t'}s_{t'+1}}\cdots P_{s_t s_{t+1}}.$$

Similarly

$$\frac{P\left(S_{i,t+1}=1,h_{i0}^{t't}\right)}{\mu^*(E_{i0}^{t'})} = \sum_{s_{t'}\in\{0,1\}^2}\sum_{s_{t'+1}\in E_{i1}^{t'+1}}\cdots\sum_{s_{t+1}\in E_{i1}^{t+1}}\mu^*(s_{t'}|E_{i0}^{t'})P_{s_{t'}s_{t'+1}}\cdots P_{s_t s_{t+1}}.$$

By Lemma 3, $\mu^*\left(\cdot \mid E_{i1}^t\right)$ strictly dominates $\mu^*\left(\cdot \mid E_{i0}^t\right)$. Then, by the above equations, and applying Lemma 2 iteratively, we derive the desired conclusion.[24] ■

Proof of Theorem 1: Let us consider any dynasty i. Consider any point in time (having started at time 0 from the steady state distribution) where a given newborn in role i is faced with the choice to invest. Let X be the number of other players who have chosen the high action 1 at that point in time. We know that

$$\mu_k^{-i} = \Pr[X=k] = \frac{n-k}{n}\mu_k + \frac{k+1}{n}\mu_{k+1}. \tag{15}$$

Let Z be i's parent's action and Y be i's choice. We can write the covariance of the parent and child's choices as

$$Cov = \Pr[Z=Y=1] - \Pr[Z=1]\Pr[Y=1].$$

We write this as

$$Cov = \left(\sum_{k=0}^{n-1}\Pr[Z=1|X=k]\Pr[X=k|Y=1]\Pr[Y=1]\right) - \Pr[Z=1]\Pr[Y=1].$$

[24] The fact that $\mu^*\left(\cdot \mid E_{i1}^t\right)$ strictly dominates $\mu^*\left(\cdot \mid E_{i0}^t\right)$ only requires that $p_1 \geq p_0$, as we already have a strict inequality for the probabilities of the increasing set $\{(1,0),(11)\}$:

$$1 = \mu^*\left(\{(1,0),(11)\} \mid E_{i1}^t\right) > \mu^*\left(\{(1,0),(11)\} \mid E_{i0}^t\right) = 0$$

Lemma 2, instead, requires $p_1 > p_0$ to establish the strict dominance conclusion. To establish strict inequality above, we need to check that a strict inequality holds for at least one increasing set contained in E_{i1}^τ, for all $t' < \tau \leq t+1$ and for all $i=1,2$. Take for instance $E_{21}^\tau = \{(0,1),(1,1)\}$ where agent 2 takes the high action. Let $s=(0,1)$ and $s\prime=(1,1)$. Then, the argument used in the proof of Lemma 2 for these two adjacent states $s' \geq s$ shows, precisely, that strict domination holds across the two probability distributions conditional on them.

By definition, $\Pr[Y=1]=\overline{p}>0$. It then follows that

$$Cov=\overline{p}\left(\sum_{k=0}^{n-1}\Pr[Z=1|X=k]\Pr[X=k|Y=1]\right)-\overline{p}\Pr[Z=1]$$

or

$$Cov=\overline{p}\sum_{k=0}^{n-1}\Pr[Z=1|X=k]\left(\Pr[X=k|Y=1]-\Pr[X=k]\right). \tag{16}$$

Note that $\Pr[Z=1,X=k]$ (the probability that the parent is high-action and there are k others high-actions) is equal to $\mu_{k+1}(k+1)/n$. Note also that

$$\Pr[X=k|Y=1]=\Pr[Y=1|X=k]\frac{\Pr[X=k]}{\Pr[Y=1]}=p_k\frac{\Pr[X=k]}{\overline{p}}.$$

Then, using (15), we rewrite (16) as

$$Cov=\overline{p}\sum_{k=0}^{n-1}\frac{\mu_{k+1}\frac{k+1}{n}}{\mu_k\frac{n-k}{n}+\mu_{k+1}\frac{k+1}{n}}\left(\Pr[X=k]\frac{p_k}{\overline{p}}-\Pr[X=k]\right). \tag{17}$$

Using the fact that $\mu_{k+1}=a_k\mu_k$ as established in the proof Proposition 1, we rewrite (16) as

$$Cov=\overline{p}\sum_{k=0}^{n-1}\frac{a_k\frac{k+1}{n}}{\frac{n-k}{n}+a_k\frac{k+1}{n}}\left(\Pr[X=k]\frac{p_k}{\overline{p}}-\Pr[X=k]\right),$$

which, using the expression for a_k in (5) gives:

$$Cov=\sum_{k=0}^{n-1}p_k\Pr[X=k]\left(p_k-\overline{p}\right). \tag{18}$$

Normalizing, this implies that

$$Cor(p)=\sum_{k=0}^{n-1}\frac{p_k}{\overline{p}}\Pr[X=k]\left(\frac{p_k-\overline{p}}{1-\overline{p}}\right). \tag{19}$$

We rewrite this as

$$Cor(p)=\frac{\sum_{k=0}^{n-1}p_k\mu_k^{-i}\left(p_k-\overline{p}\right)}{\overline{p}(1-\overline{p})}.$$

Thus,

$$Cor(p)=\frac{Var(p)}{\overline{p}(1-\overline{p})},$$

and the Theorem follows directly. ■

Proof of Proposition 4: We write

$$Cost(p)=\frac{C}{2\overline{p}}\sum_{k=0}^{n-1}\mu_k^{-i}p_kp_k, \tag{20}$$

which, by (19) we rewrite as

$$Cost(p) = \frac{C}{2}\left((1-\overline{p})Cor(p) + \overline{p}\right).$$

The result then follows from Theorem 1 and the fact that $\overline{p}' = \overline{p}$. ■

Proof of Proposition 5: From the proof of Theorem 1, we can write

$$Var(p) = \overline{p}(1-\overline{p})Cor(p) = \sum_{k=0}^{n-1} \Pr[X=k] p_k \left(p_k - \overline{p}\right).$$

Given the threshold model, we can rewrite this as

$$Var(p) = p_L \left(p_L - \overline{p}\right) \sum_{k=0}^{\tau-1} \Pr[X=k] + p_H \left(p_H - \overline{p}\right) \sum_{k=\tau}^{n-1} \Pr[X=k]. \tag{21}$$

Note that

$$\overline{p} = p_L \sum_{k=0}^{\tau-1} \Pr[X=k] + p_H \sum_{k=\tau}^{n-1} \Pr[X=k]. \tag{22}$$

Then, using (22) we rewrite (21) as

$$Var(p) = p_L^2 \sum_{k=0}^{\tau-1} \Pr[X=k] + p_H^2 \sum_{k=\tau}^{n-1} \Pr[X=k] - \overline{p}^2 \tag{23}$$

Consider a differential change (dp_H, dp_L) in (p_H, p_L) such that $dp_H > 0$, $dp_L < 0$, and $(p_H + dp_H, p_L + dp_L)$ is a $\overline{p}$-preserving spread of (p_H, p_L), that is, $d\overline{p} = 0$. In what follows, we use the following notation:

$$\sigma_\alpha^\beta = \sum_{k=\alpha}^{\beta} \Pr[X=k], \text{ for } \alpha \leq \beta.$$

Taking a differential of (23) under the constraint $d\overline{p} = 0$ gives

$$dVar(p) = 2p_L \sigma_0^{\tau-1} dp_L + 2p_H \sigma_\tau^{n-1} dp_H + p_L^2 d\sigma_0^{\tau-1} + p_H^2 d\sigma_\tau^{n-1}. \tag{24}$$

Using (22), the condition $d\overline{p} = 0$ becomes

$$\sigma_0^{\tau-1} dp_L + \sigma_\tau^{n-1} dp_H + p_L d\sigma_0^{\tau-1} + p_H d\sigma_\tau^{n-1} = 0. \tag{25}$$

Multiplying (25) by $p_H + p_L$ and substracting it from (24) gives

$$dVar(p) = (p_H - p_L)\left[\sigma_\tau^{n-1} dp_H - \sigma_0^{\tau-1} dp_L\right] - p_L p_H \left[d\sigma_0^{\tau-1} + d\sigma_\tau^{n-1}\right]. \tag{26}$$

Note that $\sigma_0^{\tau-1} + \sigma_\tau^{n-1} = 1$ (this is a sum of probabilities), and thus $d\sigma_0^{\tau-1} + d\sigma_\tau^{n-1} = 0$. We thus conclude that

$$dVar(p) = (p_H - p_L)\left[\sigma_\tau^{n-1} dp_H - \sigma_0^{\tau-1} dp_L\right] > 0, \text{ when } p_H > p_L.$$

■