NBER WORKING PAPER SERIES

IMMOBILITY AS MEMORY: SOME NEW APPROACHES TO CHARACTERIZING INTERGENERATIONAL PERSISTENCE VIA MARKOV CHAINS

Lawrence Blume
Neil A. Cholli
Steven N. Durlauf
Aleksandra Lukina

Working Paper 33166
http://www.nber.org/papers/w33166

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2024

All authors thank the Stone Center for Research on Wealth Inequality and Mobility for financial support. Cholli also thanks the Klarman Fellowship at Cornell University for support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Lawrence Blume, Neil A. Cholli, Steven N. Durlauf, and Aleksandra Lukina. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Immobility As Memory: Some New Approaches to Characterizing Intergenerational Persistence via Markov Chains

Lawrence Blume, Neil A. Cholli, Steven N. Durlauf, and Aleksandra Lukina

NBER Working Paper No. 33166
November 2024
JEL No. C10, C50, D30, J6

ABSTRACT

This paper proposes some new measures of intergenerational persistence based on the idea of characterizing the memory of origin in the stochastic process that links the socioeconomic classes of parents and children. We introduce "memory curves" for all future generations given any initial condition of class for a family dynasty, which reveal how initial conditions interact with the transition process between parents and children to create mobility and persistence. We also propose ways to aggregate information across different classes to produce overall characterizations of mobility in the population. To illustrate our measures, we estimate occupational "memory curves" using U.S. survey data. Our findings show that, on average, the memory of initial conditions dissipates largely within three generations, though there is meaningful heterogeneity in mobility rates across dynasties originating from different occupational classes.

Lawrence Blume
Department of Economics
Uris Hall
Cornell University
Ithaca, NY 14850
lb19@cornell.edu

Steven N. Durlauf
University of Chicago
Harris School of Public Policy
1307 E. 60th Street
Chicago, IL 60637
and NBER
durlauf@gmail.com

Neil A. Cholli
Cornell University
109 Tower Rd.
Ithaca, NY 14853
nac85@cornell.edu

Aleksandra Lukina
Harris School of Public Policy
University of Chicago
1307 E. 60th Street
Chicago, IL 60637
lukina@uchicago.edu

A data appendix is available at http://www.nber.org/data-appendix/w33166

1 Introduction

Markov chains continue to be one of the workhorse models for the analysis of intergenerational mobility in the social sciences. Prais (1955) is arguably the first example of this. The continuing importance of this method can be seen from the classic survey of Boudon (1973) and the recent synthesis of Song (2021). For outcomes with a discrete support, transition matrices summarize all information concerning the probabilistic dependence of offspring outcomes on parental background. A challenge in using Markov chain mobility models is reducing the information in a matrix into interpretable statistics.

The purpose of this paper is to propose some novel ways to extract information on intergenerational mobility/persistence from Markov chains. In doing this, we build on the classic idea that underlies much mobility research, namely that a mobile society is one in which the memory of an individual's socioeconomic status fades rapidly as one considers that individual's descendants. Following Prais (1955), a perfectly mobile society is understood this way, specifically, as "a society in which the probability of entering a particular social class is independent of the class of one's father" (p. 59). As such, Prais' idea of origin independence operationalizes the responsibility-sensitive egalitarianism approach offered in Roemer (1993, 1998). Roemer argues that inequalities are unjust if they derive from factors for which an individual is not responsible. Parental status is the canonical example of such a factor. We pursue Prais-Roemer logic by proposing mobility measures that fully characterize the memory of ancestor's status on their descendants.[1] In what follows we focus on occupational status but the methods apply to any outcome with a discrete support.

Our approach is predicated on the following idea. Suppose family dynasties obey a Markov chain for occupation. Dynasties differ from each other in terms of the initial condition π_0. The initial condition π_0 is an ancestor's probability distribution over occupations.[2] Assume that there exists a generation k such that for any dynasty, regardless of the occupation in generation k, there is a non-zero probability of the $(k+1)$-th generation descendant having any of the possible occupations. Under this condition, each dynasty will have the same steady-state distribution, denoted π^*, that describes the probabilities of occupational status for arbitrarily distant descendants and is independent of π_0. At each generation t, $t = 0, 1, 2, ...$, each dynasty will be associated with a generation-specific probability distribution over occupations, denoted π_t, including an initial condition π_0. One can therefore measure the discrepancy between π_t, which depends on π_0, and π^*. This deviation represents the memory of the ancestor's status π_0 in its descendants' status π_t. We call the sequence of these deviations *memory curves*. The deviations converge to zero, and memory of initial condition dissipates. The faster this memory dissipates, the more mobility is observed. Alternatively, one could analyze the discrepancy between generation-specific probability

1. An ancestor is defined as the original or founding member from whom all subsequent members of the dynasty are descended.

2. If π_0 assigns the unit probability to one occupation, this fixes the dynasty's origin. If π_0 is different from a coordinate vector, we refer to a random position of the dynasty in period 0.

distributions π_t and μ_t for two dynasties that started from distinct initial conditions π_0 and μ_0. We propose several measures of mobility based on memory curves and apply them to U.S. survey data to study intergenerational occupational mobility.

Relative to standard uses of Markov chains for measuring intergenerational mobility and measures based on log-linear regression models,[3] our approach possesses several advantages.

First, we characterize the full effects of an ancestor's given occupation on descendants because we analyze the complete probability distribution over occupations for each generation of descendants, avoiding large dimension reductions and consequent loss of information as in regression methods for assessing mobility.

Second, our approach integrates mobility measures over both short and long horizons. We consider the repeated application of the transition matrix to π_0 which creates a contraction in the sense that as t increases, π_t becomes closer to π^*. The changes in the deviations of π_t from π^*, therefore, define mobility rates at *every horizon* from a given initial condition. This method allows us to identify differences in mobility rates between short- and long-run horizons and distinguishes our approach from existing practices of Markov chains that have emphasized long-run (asymptotic) rates of convergence.[4] Similarly, the application of the transition matrix to two probability distributions shrinks the distance between them, which means that opportunities for dynasties of different origins become closer to each other as time elapses. By measuring this contraction, we evaluate the loss of memory in the process over time.

Mobility analysis over long horizons has been controversial. Criticisms of Markov chain approaches such as Duncan (1966) and Pullum (1975) include the argument that the more generations into the future one considers, the less credible the dynamics of a stationary Markov chain become.[5] The decreasing credibility of a stationarity assumption is obviously true, but that does not mean that the long-run behavior of a Markov model is not informative. We regard long-run behaviors of memory curves as predictions only in the absence of socioeconomic changes that induce nonstationarity.[6] As Boudon (1973) argued, mobility measures should be interpreted not "as a prediction of what should actually occur in the more or less remote future, but rather as a way of summarizing the structure" (p. 44) of an empirical transition matrix observed at a given time. In our judgement,

3. We refer to the major class of models used in sociology based on estimation of row, column, and layer effects, which produce log-frequencies of parent-child socioeconomic status pairs. In economics, the standard approach is to measure the intergenerational elasticity of income or earnings.

4. Axiomatic studies of Markov chain mobility have concluded that the second largest eigenvalue modulus of the transition matrix is the appropriate persistence measure for such environments (see Shorrocks (1978), Sommers and Conlisk (1979), and Conlisk and Sommers (1979)). The argument for using the second largest eigenvalue, in fact, has to do with the deviation of π_t from π^* for large t. Basically, this deviation depends on all the eigenvalues of the Markov transition matrix. However, as t grows, the behavior of the deviations is dominated by the second largest eigenvalue, which determines the convergence rate in the long run.

5. At the same time, the Markov chain approach has been more developed in studies of intragenerational mobility, e.g., McFarland (1970), Stewman (1975), Dudel (2021).

6. The analysis of a particular Markov chain estimated based on parent-child data derives what would happen were a society's social and economic structure remain fixed. Such an approach is analogous, for example, to the analysis of a synthetic or fictitious cohort in life table analysis.

it is crucial to understand whether internal mechanisms produce short or long memory of an initial condition, even in the absence of socioeconomic changes that could disrupt the stationarity of the process.

Third, an important feature of our analysis is careful attention to both mobility measures conditional on a specific origin and the aggregation of these origin-specific measures across the population. More precisely, we emphasize the importance of accounting for heterogeneity in mobility curves across different origins. This contrasts with the standard measures of memory used in the Markov chain literature (namely, measures of "mixing"), which focus on "worst-case" initial conditions that in our context bound the rate of mobility for the dynasties in a population.

Fourth, we offer several ways to analyze subgroup differences in opportunity, clarifying the roles of distinct initial conditions and distinct transition mechanisms. In the empirical part of the paper these ideas are illustrated with the Black-White subgroup analysis of intergenerational mobility.

We also argue that our memory curves may be used to give some insights into structural versus exchange mobility.[7] Specifically, the behavior of memory curves characterizes the rate at which structural mobility fades over time.

Finally, we compare the introduced measures of intergenerational persistence with odds (or cross-product) ratios, commonly used in sociology as a measure of mobility. Specifically, we show that odds ratios alone do not capture short-run or long-run memory measures.

The rest of the paper is organized as follows. In Section 2 we briefly describe Markov models for analysis of intergenerational mobility, offering different interpretations of these models. Section 3 explains our motivation for introducing memory curves and outlines the formal technique. Section 4 demonstrates that transition matrices, characterized by the same set of odds ratios, may have very different long- and short-run memory measures. In Section 5, we illustrate our memory curve measures by studying occupational persistence in U.S. survey data. Section 6 concludes.

2 Markov Models for Studying Intergenerational Mobility

We imagine families as dynasties, with one representative member in each generation. Our analysis pursues the idea that intergenerational mobility involves estimating the extent to which a child's occupation is independent of parental background. Our paper's goal is to augment some of the ways that this idea has been operationalized.

We model the link between parental and child occupations, but we assume no further ancestral influence (e.g., "grandparent effects"). Thus, the dynastic process of moving across occupational classes is a Markov chain. It exists on a state space $S = \{1, 2, ..., n\}$, which is the set of all possible occupational classes. The process is characterized by an *initial distribution* of occupations, π_0, and

7. Note that here a different interpretation of Markov chains is required. Distribution π_t needs to be interpreted in population frequency terms. This will be explained in more detail in Section 2.

the *transition probability matrix* P. The (i, j) element of P, P_{ij}, is the probability that a child will pick up occupation j conditional on his parent belonging to occupation i. If the time t parent distribution of occupational classes is π_t, the child distribution is $\pi_{t+1} = \pi_t \cdot P$.

Alternatively, one can start from a joint distribution of parent and child occupational classes which is described by a contingency table where rows and columns correspond to parent and child classes, respectively.[8] The parent distribution is then the row marginal distribution, the child distribution is the column marginal distribution, and the elements of the joint distribution define the conditional probabilities P_{ij} in the usual way. An $n \times n$ contingency table and an $n \times n$ transition probability matrix together with an initial distribution π_0 are two ways of describing the same data.

The Markov approach, however, takes the extra step of assuming that the matrix P is stable over some number of subsequent generations, so that the iterates $\pi_2 = \pi_1 \cdot P = \pi_0 \cdot P^2$, and so on, are meaningful.[9] Under the weak condition that P is *primitive* (i.e., there exists power k such that P^k is a positive matrix), the dynastic process is *ergodic*. This means that π_t converges to a limit, a steady state π^*, and that π^* is independent of π_0. Or equivalently, that P^t converges to a rank-1 matrix P^∞ whose rows are the vectors π^*. The steady state π^* is unique and positive. It is the left eigenvector of P corresponding to the unit eigenvalue, that is $\pi^* = \pi^* \cdot P$ (thus, π^* is also the invariant distribution for P).

The vectors π_t may be interpreted in two different ways:

1. *Lottery interpretation (single dynasty interpretation)*: The initial distribution π_0 may be thought of as a lottery faced by any given dynasty in period 0. The ith row $P(i, \cdot)$ of P is the lottery faced by a dynasty that originates from class i with probability 1, i.e., $P(i, \cdot)$ is the conditional probability distribution of a dynasty's class in the next generation given that it occupies state i in the current generation.

   Convergence of P^t to P^∞ means that conditional probability distributions become closer and closer to each other as time goes. The unconditional probability distribution of the dynasty's status evolves with time according to the law $\pi_t = \pi_{t-1} \cdot P = \pi_0 \cdot P^t$. Convergence of π_t to π^* means that any dynasty, regardless of its initial lottery π_0, faces the same unconditional lottery π^* (with a positive probability of transiting to any given class) in the long run. Therefore, all dynasties eventually have equal opportunities no matter what lottery they started with. Equivalently, *the memory of initial condition* disappears over time.[10]

8. Contingency tables/matrices contain relative frequencies F_{ij} of parent-child occupation pairs where a parent is in occupational class i and the child is in class j.

9. The definition of states does not change either. In practice, we deal with classes harmonized (or "cross-walked") across generations which ensures that occupational classes in one generation are comparable with classes in the next generation. In Section 5, we use a standard sociological technique of harmonizing occupational codes across different generations. A similar procedure is implemented in studies of income mobility when, for example, children's incomes are measured in prices of parental generation. Thus, to be more precise, we assume that the process is Markov over classes harmonized across generations.

10. Similarly, the intergenerational elasticity (IGE), the standard measure of intergenerational income mobility in the economics literature, is widely used for making claims such as "advantages or disadvantages

2. *Population interpretation*: The second interpretation focuses on the evolution of the marginal distribution of a population of dynasties over occupational classes where each dynasty obeys the same Markov chain process. Suppose initial occupation lotteries of N dynasties are independently drawn from a probability distribution π_0 on S. The N processes then evolve independently according to the transition probability P. The empirical distribution π_t^N of N dynasties over occupational classes at time t converges to π_t as N becomes large. Thus, for a large number of dynasties, π_t may be interpreted in *population frequency terms*. A deeper fact is that the deterministic evolution of the sequence π_t is a good predictor of the population sample path of occupation frequencies over any finite time horizon. For all $\epsilon > 0$ and time T there is a population size $N(T)$ such that $\sup_{t \leq T} |\pi_t^N - \pi_t| \leq \epsilon$ for all $N \geq N(T)$. This is to say, over any given time horizon, for a large enough population the sequence of t-period marginal distributions of occupations approximates the sequence of occupation distributions uniformly well.

   In this interpretation convergence of π_t to π^* means that the marginal distribution of the population over classes stabilizes over time. Berger and Snell (1957) refer to mobility patterns in the absence of changes in marginal distributions as pure exchange mobility. Exchange mobility therefore describes how different dynasties move over different classes when the marginal distribution does not change. This definition of exchange mobility has become common in the economic literature (see, e.g., Markandya (1982), Dardanoni (1993, 1995)).[11] We employ the Berger and Snell (1957) concept of exchange mobility to propose several ways to characterize roles of structural and exchange mobility in total mobility.[12]

3 Mobility as Independence of Origin

3.1 Motivation

Much of the sociology literature has been concerned with distinguishing between "forced fluidity" due to changing occupational structure and "social fluidity," or exchange mobility. Our interest is

of ancestors are wiped out in three generations" (Becker and Tomes (1986)) and for comparison of, say, the IGE of 0.6 versus the IGE of 0.4 (Mazumder (2005)). Assuming that log-income follows a first-order autoregressive process across generations, when IGE^t gets close to zero an ancestor's income no longer affects the t-th descendant's income. As described by Mazumder (2005), "If the IGE is 0.6, then on average, it will take the descendants of the family 5 to 6 generations (120 to 150 years) before their income would be within 5% of the national average. In contrast, if the parameter is 0.4, the comparable time will be just 3 generations (75 years)" (p. 235).

11. We acknowledge that such approach to structural and exchange mobility drastically differs from the mainstream sociological methodology. Sociologists have been trying to distill "social fluidity" (exchange mobility) from observed contingency tables for decades. Roughly speaking, social fluidity is the total mobility purified from "forced fluidity" due to changing economic structure. The latter may be captured by changes in marginal distributions, and the leftover would capture "true" social mobility, or social fluidity.

12. See also Appendix B.

different. Our mobility measures are designed to explore occupational dynamics; in particular, how ancestral influences dissipate over time. If the dynamics uncovered by the transition matrix are stable over some period of time, then there will be movement away from an initial occupation or distribution of occupations towards the limit distribution. We aim to measure the motion of the marginal distributions during this period.

Specifically, our analyses focus on the path that leads $\pi_t = \pi_0 P^t$ to π^*, which embodies the memory of initial condition. Perfect mobility requires that all children have the same opportunities regardless of their parents' occupations. This requires that all rows of P are identical. Convergence is in one step—there is no memory of initial condition at all. However, in the general case the matrix products P^t converge to an identical-row matrix, but only in the large-t limit. Thus, the memory of initial conditions may be quite long lived. The memory of an initial condition also depends on its particular value. If the initial condition is $\pi_0 = \pi^*$, convergence is immediate.[13] The farther π_0 is from π^*, the longer the convergence might take. Thus, in the short run the speed of convergence of $\pi_0 P^t$ to π^* depends on both the initial condition π_0 and the speed with which P^t approaches P^∞ (i.e., the rate at which our transition mechanism mixes up dynasties that start at different states).[14]

Alternatively, consider dynasties with different initial conditions π_0 and μ_0. Under our assumptions, $(\pi_0 P^t - \mu_0 P^t) \to 0$ as $t \to \infty$ for any π_0 and μ_0. Under the lottery interpretation this means that dynasties, facing different lotteries in period 0, get closer and closer to each other with time (in terms of lotteries they face in period t). Under the population interpretation this yields that groups which share the same transition mechanism, but are characterized by different initial marginal distributions, eventually converge to each other in terms of marginal distributions.

We define memory curves and memory measures of mobility based on the speed of convergence to the steady state. One can ask questions about the number of generations before a dynasty is near the steady state or the number of generations it takes for descendants from a pair of dynasties with very different initial conditions to have similar distributions of outcomes. There is a literature on "mixing times" which discusses these questions, and we will draw on some of its findings below.[15]

We emphasize that our mobility discussion raises questions for both the *short* and *long run*. Given a suitably chosen norm for measuring the distance between two distributions, one can demonstrate that the next generation's distribution $\pi_{t+1} = \pi_t \cdot P$ will be closer to π^* than the current generation's distribution π_t. Similarly, the application of a transition matrix P to two distributions shrinks the distance between them. The rate at which this contraction occurs may be interpreted as the speed of mobility. One can then ask, for example, how close two dynasties, originating from different classes, become over different numbers of generations. Such questions motivate the next section.

13. If $\pi_0 = \pi^*$ but P is not a perfect mobility matrix, there are differences in conditional probability distributions and, at the same time, the unconditional probability does not change.

14. The long-run convergence rate is governed by the second largest eigenvalue of P and, thus, does not depend on the initial condition π_0.

15. There are a number of formalizations of mixing rates in the time series literature that are used in proofs of the central limit theorem for dependent data. These are not important to our analysis.

3.2 Measuring Memory

In order to study the convergence of π_t to the steady state π^* or convergence between π_t and μ_t, it is necessary to choose a vector norm for measuring distance between two probability distributions. One requirement we impose on the choice of norm is that it ensures that the application of a transition matrix P to π_0 and μ_0 (weakly) shrinks the distance between them.[16] Among the norms that have this property, we work with *the total variation norm* for tractability.[17] For any two distributions μ and ν the total variation norm measuring the distance between them is

$$\| \mu - \nu \|_{TV} = \frac{1}{2} \sum_{1 \leq i \leq n} |\mu_i - \nu_i| = \frac{1}{2} \| \mu - \nu \|_1,$$

where $\| \cdot \|_1$ denotes the l_1-norm. Note that $\| \mu - \nu \|_{TV} \leq 1$ for any probability distributions μ and ν. The distance between μ and ν in the total variation norm $\| \mu - \nu \|_{TV}$ is the largest absolute difference between the probabilities that the distributions μ and ν assign to the same event.[18] Thus, the total variation norm may be thought of as "the sup-probability norm".

Given a norm for measuring distances between distributions, one can construct several characterizations of the memory of status across generations. We begin by outlining a class of measures from the mixing times literature in probability theory which involves "worst-case scenarios", or specific occupation classes that produce the slowest convergence to steady state. We then develop a new class of memory measures that are more suitable for studying intergenerational persistence of actual initial conditions.

3.2.1 Worst-Case Measures

One measure suggested by the mixing times literature involves the *"worst-case origin" for a dynasty* after t periods. It is defined as

$$\begin{aligned} d(t) &= \sup_{\pi_0 : \|\pi_0\|_1 = 1} \| \pi_0 \cdot P^t - \pi^* \|_{TV} = \max_{1 \leq i \leq n} \| e_i \cdot P^t - \pi^* \|_{TV} \\ &= \max_{1 \leq i \leq n} \| P^t(i, \cdot) - \pi^* \|_{TV} = \max_{1 \leq i \leq n} \| P^t(i, \cdot) - P^\infty(i, \cdot) \|_{TV}, \end{aligned} \tag{1}$$

where e_i is the coordinate vector and $P^t(i, \cdot)$ denotes the ith row of matrix P^t, as before.

Under the lottery interpretation, the curve $d(t)$ shows how far the "worst-case dynasty" is from

16. Equivalently, P is (weakly) contractive on the $(n-1)$-dimensional simplex of vectors with entries summing up to zero.

17. Any two norms on a finite-dimensional space are equivalent, and a choice of a specific norm does not make any difference in terms of the trend. However, in the short run it might matter. At the same time, given a norm which would not give us a contraction at a single step we can always find a number of steps at which we would get a contraction (because the trend is the same). We choose the total variation norm not only because it ensures that the application of P is contractive, but because it has a very clear probabilistic meaning that fits the lottery interpretation.

18. See more details related to the total variation norm in Appendix D.1.

the long-run (i.e. steady-state) distribution of chances in period t. The "worst case" is defined based on the initial distribution of ancestral occupational classes. The maximal distance is always (for any t) achieved at a distribution e_i that assigns the unit probability to some state i, i.e., there is a least favorable occupation for an ancestor in terms of memory that is present at t. Note that this least favorable state i may be different for different time periods t.[19] Equivalently, $d(t)$ shows how far the worst-case conditional distribution at step t could be from the steady state distribution π^*. The last equality in Equation (1) is true because π^* represents any row of matrix P^∞.

One can augment this calculation with measures of *"worst-case" distances between members of two dynasties*,

$$\begin{aligned} d'(t) &= \sup_{\pi_0,\mu_0:\ \|\pi_0\|_1=1,\ \|\mu_0\|_1=1} \| \pi_0 \cdot P^t - \mu_0 \cdot P^t \|_{TV} \\ &= \max_{i,j} \| e_i \cdot P^t - e_j \cdot P^t \|_{TV} = \max_{i,j} \| P^t(i,\cdot) - P^t(j,\cdot) \|_{TV} . \end{aligned} \tag{2}$$

Under the lottery interpretation, $d'(t)$ is the maximal distance between the unconditional probability distributions π_t and μ_t for two dynasties originating from different lotteries π_0 and μ_0 at time t. The maximum is achieved with dynasties belonging to different classes at $t = 0$ with probability 1. In other words, $d'(t)$ shows how chances for the t-generation of one dynasty may differ from chances for the t-generation of another dynasty in the worst case. Again, the states i and j that generate $d'(t)$ depend on t, i.e., the dynasties which are the farthest from each other at different t may originate from different pairs of classes. Alternatively, $d'(t)$ represents the worst-case difference in conditional probability distributions for dynasties that belong to different classes at time t. This means that $d'(t)$ characterizes the *association of rows* of matrix P^t. The larger $d'(t)$ is, the more inequality of opportunity is still observed in generation t.

Since $\pi_0 P^t \to \pi^*$ for any π_0 and, as a consequence, $\pi_0 P^t - \mu_0 P^t \to 0$ for any π_0 and μ_0, both $d(t)$ and $d'(t)$ go to zero as time elapses. Obviously, the deviation measures $d(t)$ and $d'(t)$ are connected to each other. Formally, $d(t) \leq d'(t) \leq 2d(t)$. In particular, this means that $d(t)$ and $d'(t)$ are equivalent functions in the long run. Besides, it can be shown that $d(t)$ and $d'(t)$ are both non-increasing functions of t.[20] Also, note that $d(t)$ and $d'(t)$ do not depend on the initial condition π_0 and characterize convergence of P^t to P^∞.[21]

As we have emphasized, convergence dynamics matter for understanding mobility. The faster the functions $d(t)$ and $d'(t)$ decrease, the less persistence is observed. Thus, we are interested in *rates of convergence* as well as deviations. A standard result in linear algebra shows that $|\lambda_2|$ describes the asymptotic convergence rate of the Markov process,[22] and so there is the potential for very different

19. If classes are ordered, it is not necessarily the lowest ranked occupation.

20. See, for example, Levin and Peres (2017).

21. However, for empirical estimates of a transition matrix, the entries depend on the initial condition.

22. Specifically, $P^t = P^\infty + O\left(t^{k_2-1}\,|\lambda_2|^t\right)$, where k_2 is algebraic multiplicity of λ_2 and $O(f(t))$ is a function of t such that there exist constants c_1 and c_2, so that $c_1 f(t) \leq O(f(t)) \leq c_2 f(t)$ for all sufficiently large t. See, for example, Seneta (2006).

convergence behavior before the asymptotic approximation is useful. If algebraic multiplicity of λ_2 equals 1, asymptotic convergence is geometric with the rate $|\lambda_2|$. In particular, it means that in the long run $d(t)$ and $d'(t)$ behave as $|\lambda_2|^t$. Since the long-run convergence is geometric, we employ logarithmic transformations of the deviations, focusing on $\log(d(t))$ and $\log(d'(t))$ rather than on $d(t)$ and $d'(t)$. To study rates of convergence, we analyze the slopes of $\log(d(t))$ and $\log(d'(t))$. In the long run $\log(d(t))$, $\log(d'(t))$ have the slope $\log(|\lambda_2|)$ if algebraic multiplicity of λ_2 is 1.

The mixing times literature suggests one more mobility statistic that characterizes the rate at which two marginal distributions move towards each other when P is applied to them. This statistic is known as *Dobrushin's coefficient*. It is defined as:

$$\delta(P) \;=\; \max_{i,j} \parallel e_i \cdot P - e_j \cdot P \parallel_{TV} \;=\; \max_{i,j} \parallel P(i,\cdot) - P(j,\cdot) \parallel_{TV} = d'(1). \tag{3}$$

Note that $\delta(P) = d'(1) \leq 1$. Dobrushin's coefficient shows the largest possible difference in opportunities (measured by the total variation norm) for children whose parents belong to two different classes and characterizes the association between rows of P itself. It controls the distance between period-t marginal distributions for any initial distributions:[23]

$$\parallel \pi_t - \mu_t \parallel_{TV} = \parallel \pi_0 \cdot P^t - \mu_0 \cdot P^t \parallel_{TV} \leq \; (\delta(P))^t \cdot \parallel \pi_0 - \mu_0 \parallel_{TV} .$$

Thus, Dobrushin's coefficient, or $d'(1)$, itself may be used as a measure of mobility/persistence. The smaller this coefficient is, the faster memory of initial condition fades.

In particular, $\parallel \pi_0 \cdot P - \mu_0 \cdot P \parallel_{TV} \leq \; \delta(P) \cdot \parallel \pi_0 - \mu_0 \parallel_{TV}$. In other words, the application of a transition matrix P to two probability distributions shrinks the distance between them (strictly shrinks if $\delta(P) < 1$). Moreover, since this condition holds with equality for some $\pi_0 = e_i$ and $\mu_0 = e_j$, $i \neq j$, $\delta(P)$ is the exact contraction ratio for P.[24] As a consequence, advancing the chain can only move it closer to its steady state. More precisely, $\parallel \pi_0 \cdot P - \pi^* \parallel_{TV} = \parallel \pi_0 \cdot P - \pi^* \cdot P \parallel_{TV} \leq \parallel \pi_0 - \pi^* \parallel_{TV}$. If π_t is the generation t distribution, the generation $t+1$ distribution $\pi_{t+1} = \pi_t \cdot P$ will be closer to π^* than π_t.

3.2.2 Measuring Memory of Specific Initial Conditions

While $d(t)$ and $d'(t)$ are measures directly suggested from Markov chain theory, they may be further modified for mobility research. Worst-case scenarios for mobility may have some interest, but the objective of measuring mobility is to understand persistence of initial conditions as well as the evolution of inequalities between groups. In this spirit, we develop a set of mobility measures that take into account *actual initial conditions*. Instead of taking the maximum over different initial conditions, we focus on the initial condition π_0 and trace the evolution of π_t and the heterogeneity

23. See Theorem 7.2 in Brémaud (1999), p. 237.

24. Similarly, $d'(t)$ is the contraction ratio for P^t which gives an exact upper bound on how close two distributions get squeezed to each other at step t.

of memory curves experienced by dynasties originating from different occupational classes.

First, we consider the empirical period 0 marginal distribution[25] π_0 and trace the evolution of π_t for a representative dynasty. *The average memory* (AM) *curve*

$$AM(t, \pi_0) \ = \parallel \pi_0 \cdot P^t - \pi^* \parallel_{TV}, \tag{4}$$

represents the deviation between time-t lottery for a representative dynasty and the steady-state lottery π^*. The slope of $AM(t, \pi_0)$ characterizes the rate of convergence to π^* for a representative dynasty. Thus, $AM(t, \pi_0)$ reveals how quickly the population as a whole loses memory of initial condition. Under the population interpretation, $AM(t, \pi_0)$ characterizes the presence of structural mobility, and the slope of $AM(t, \pi_0)$ represents the speed at which structural mobility fades. Note that the convergence rate of $AM(t, \pi_0)$ to zero depends on both the transition matrix P and the initial condition π_0. Obviously, $AM(t, \pi_0) \leq d(t)$ for any t and for any π_0. Thus, $d(t)$ gives an upper bound for the average memory curve.

Our second proposed measure (which is a special case of the previous one) considers how memory fades for a dynasty that originates from a specific class i. This leads to *the individual memory* (IM) *curve*

$$IM(t, i) \ = \parallel e_i \cdot P^t - \pi^* \parallel_{TV} = \parallel P^t(i, \cdot) - \pi^* \parallel_{TV}, \tag{5}$$

which characterizes the "class-specific" deviation. Clearly, $IM(t, i) \leq d(t)$ for any t and any i. The slope of $IM(t, i)$ tells us how quickly i-origin memory is lost. Such curves, plotted for different i, can illustrate different rates of memory decay depending on the occupation of an ancestor.

How might these class-specific mobility trajectories be used to produce an overall mobility measure for dynasties in a society? We propose weighting each $IM(t, i)$ with the fractions of the population in each class at time $t = 0$. This leads to *the average individual memory* (AIM) *curve*

$$AIM(t, \pi_0) = \sum_{i=1}^{n} \parallel e_i \cdot P^t - \pi^* \parallel_{TV} \cdot (\pi_0)_i = \sum_{i=1}^{n} \parallel P^t(i, \cdot) - \pi^* \parallel_{TV} \cdot (\pi_0)_i, \tag{6}$$

which represents an expected value of the deviation from the steady-state distribution in period t, calculated for a dynasty randomly chosen from the population in period $t = 0$. It follows from convexity of the norm that $AIM(t, \pi_0) \geq AM(t, \pi_0)$ for all t and π_0. The slope of $AIM(t, \pi_0)$ shows the rate at which the average (over different origins) memory disappears.

It may also be of interest to track the convergence between two dynasties that represent different subgroups of the population. One can similarly modify $d'(t)$ by changing the worst-case initial states to π_0 and μ_0, where π_0 and μ_0 correspond to two subgroups within the population of ancestors. For example, π_0 and μ_0 could characterize representative dynasties for two distinct racial or ethnic groups or represent specific states e_i and e_j for any $i \neq j$. We define the *subgroup-difference*

25. Choosing different π_0, we could analyze "average" convergence behavior for different subgroups of the population.

memory (SDM) curve as

$$SDM(t, \pi_0, P, \mu_0, P) \ = \parallel \pi_0 \cdot P^t - \mu_0 \cdot P^t \parallel_{TV} . \tag{7}$$

Under the lottery interpretation, this memory curve characterizes convergence between these two subgroups in terms of their opportunities. In particular, it focuses on the role of initial conditions in unconditional-lottery differences between dynasties, assuming they share the same transition mechanism P. Unlike our prior memory curves, we include P as arguments in $SDM(\cdot)$ in order to consider subgroup differences based on different choices of transition matrices; we discuss these variants below. When $SDM(t, \pi_0, P, \mu_0, P)$ is close to zero, dynasties representing the two subgroups face similar lotteries. Thus, one can determine in how many generations the differences in opportunities between these subgroups will disappear. Under the population interpretation, $SDM(t, \pi_0, P, \mu_0, P)$ compares marginal distributions π_t and μ_t for two subgroups that were initially distributed as π_0 and μ_0. Note that the function $SDM(t, \pi_0, P, \mu_0, P)$ is bounded from above by $d'(t)$ for any π_0 and μ_0.

Similar to $\log(d(t))$ and $\log(d'(t))$, when asymptotic approximation becomes useful, all logarithmic transformations of the deviations introduced above, i.e., $\log(AM(t, \pi_0))$, $\log(IM(t, i))$, $\log(AIM(t, \pi_0))$ and $\log(SDM(t, \pi_0, P, \mu_0, P))$, share the same asymptotic slope $\log(|\lambda_2|)$. In the empirical part of the paper we plot only logarithmic transformations and analyze their slopes.

Equation (7) suggests some other ways to study subgroup differences. Suppose the two subgroups, e.g., Black and White dynasties, are associated with distinct transition matrices P_b and P_w and distinct initial conditions $\pi_{0,b}$ and $\pi_{0,w}$. To study the actual difference in opportunities between representative Black and White dynasties, we can analyze $SDM(t, \pi_{0,b}, P_b, \pi_{0,w}, P_w) = = \parallel \pi_{0,b} \cdot P_b^t - \pi_{0,w} \cdot P_w^t \parallel_{TV}$. Note that in the long run, this function converges to $\parallel \pi_b^* - \pi_w^* \parallel_{TV}$, where π_b^* and π_w^* are the steady-state distributions of P_b and P_w, correspondingly. Further, we can contrast $SDM(t, \pi_{0,b}, P_b, \pi_{0,w}, P_w)$ with $SDM(t, \pi_{0,b}, P_b, \pi_{0,w}, P_b)$ and $SDM(t, \pi_{0,b}, P_w, \pi_{0,w}, P_w)$ to examine the role of initial conditions when one subgroup shares the other subgroup's same transition matrix.[26] For example, analyzing $SDM(t, \pi_{0,b}, P_w, \pi_{0,w}, P_w)$ reveals how quickly a representative Black dynasty would "catch up" with a representative White dynasty if the Black dynasty faced the same transition mechanism as the White dynasty. Further, instead of analyzing representative Black and White dynasties, one could compare, e.g., a Black dynasty originating from a low-skilled class i with a White dynasty originating from a high-skilled class j, by replacing $\pi_{0,b}$ and $\pi_{0,w}$ with e_i and e_j, respectively.

There are important antecedents to our approach. The idea of using the speed of convergence towards the steady state (or equivalently, the speed of escape from an initial condition) as a measure of mobility is not new per se. Shorrocks (1978) attributes this idea to Henri Theil; Sommers

26. $SDM(t, \pi_{0,b}, P_b, \pi_{0,w}, P_w) = \parallel \pi_{0,b} \cdot P_b^t - \pi_{0,w} \cdot P_b^t \parallel_{TV}$; $SDM(t, \pi_{0,b}, P_w, \pi_{0,w}, P_w) = \parallel \pi_{0,b} \cdot P_w^t - \pi_{0,w} \cdot P_w^t \parallel_{TV}$. Durlauf et al. (2024) generalize these formulas to allow for non-homogeneous transition probabilities across generations.

and Conlisk (1979) and Conlisk and Sommers (1979) develop measures based on the idea. Those classic papers have ended up advocating the use of the second largest eigenvalue modulus of a transition matrix P, $|\lambda_2|$. As mentioned earlier, while $|\lambda_2|$ describes the asymptotic convergence rate of the process, convergence behavior may be different in the short run before the asymptotic approximation becomes useful. Convergence to the steady-state distribution may be very slow until it becomes asymptotically geometric, or it may be geometric even in the short run, but still the short-run geometric rate may be different from the long-run geometric rate. The usefulness of limiting attention to the second largest eigenvalue depends on how quickly asymptotic geometric convergence kicks in.[27]

Dobrushin's coefficient $\delta(P) = d'(1)$, defined in Equation (3), occurs in passing among a number of other measures discussed in Brumelle and Gerchak (1982).

4 Memory Measures and Odds Ratios

In this section we provide a link between the measures discussed in Section 3 and odds ratios, or cross-product ratios, commonly used in sociology as a measure of mobility (specifically, exchange mobility[28]).

First, we give a formal definition of odds ratios and briefly discuss the procedure of iterative proportional fitting. Second, we demonstrate that odds ratios do not capture our measures of intergenerational persistence. Namely, persistence measures for different transition matrices characterized by the same set of odds ratios can exhibit significant variation.

Odds ratios measure the advantages or disadvantages that the children of parents from one occupational group have compared to the children of parents from another occupational group in obtaining one occupation rather than another. The *odds ratio* of a child whose parent belongs to

27. This is a separate problem which seems to be less studied in the literature.

28. The history of attempts to decompose mobility into structural and exchange components goes back to sociologists of the 1950s (see Boudon (1973) for a review). Goodman (1968, 1969) demonstrated the importance of odds ratios to describing the structure of row-column associations in contingency tables. Since Goodman, the mainstream approach has focused on odds ratios and different types of log-linear or log-multiplicative models closely related to them (see, e.g., Powers and Xie (2008) for a comprehensive description). Sociologists' motivation for using odds ratios as a measure of exchange mobility is based on the mathematical fact of their "margin-invariance" (odds ratios are invariant under row and column transformations). The criticism of odds ratios as a measure of exchange mobility has been rare but persistent. For example, Harrison (1988), using formal vacancy models, shows that when only the availability of positions changes, the odds ratios change too. Thus, he concludes that the availability of positions is controlled not only by the size of marginals, and odds ratios are not free of changes in the occupational structure. Macdonald (2023) presents a whole range of arguments against odds ratios as a measure of solely exchange mobility. In particular, he questions the procedure of iterative proportional fitting. We could add to this criticism the following: in their empirical studies sociologists harmonize occupational structures across generations using so-called "cross-walks". This approach, however, inevitably blurs the changing structure of the job market. Consequently, the observed contingency tables intrinsically possess a structural component. Also, note that odds ratios are well defined only for positive contingency tables.

class i instead of j and who arrives in class m instead of n is p_{im}/p_{in} versus p_{jm}/p_{jn}, that is

$$O_{ij,mn} = \frac{p_{im}/p_{in}}{p_{jm}/p_{jn}} = \frac{p_{im}p_{jn}}{p_{in}p_{jm}}.$$

Note that in sociology, the odds ratios are defined for a contingency table (see, e.g., Powers and Xie (2008)), but it is easy to see that the odds ratios calculated for the corresponding transition matrix are the same. For an $n \times n$ contingency or transition matrix there are $(n-1) \times (n-1)$ independent odds ratios (they correspond to adjacent rows and adjacent columns; in our notation one should choose $j = i+1$ and $n = m+1$), all other odds ratios can be found based on the set of independent odds ratios.

The procedure of *iterative proportional rescaling/fitting* allows one to adjust a contingency table so that it would fit any given marginal distributions while preserving the structure of odds ratios. To the best of our knowledge, the procedure was independently offered by many authors, including Deming and Stephan (1940). Early proof of convergence and uniqueness is presented, e.g., in Fienberg (1970). Using iterative proportional fitting, we construct a variety of contingency matrices and corresponding transition matrices that have the same set of odds ratios.

More specifically, we take a given contingency table, compute the set of independent odds ratios, and keep them fixed in what follows. We then choose two randomly generated distributions as the parental and children's distributions and fit these two distributions using iterative proportional fitting. This process is repeated 100,000 times. In the end, we have 100,000 contingency matrices (and corresponding transition matrices) that share the same set of odds ratios but have different marginals.

In terms of the memory measures, we restrict ourselves to only two: the second-largest eigenvalue modulus, which captures the asymptotic convergence rate, and Dobrushin's coefficient $\delta(P) = d'(1)$, which, on the contrary, describes short-run behavior. We compute these two measures of intergenerational persistence for all transition matrices obtained in the previous step.

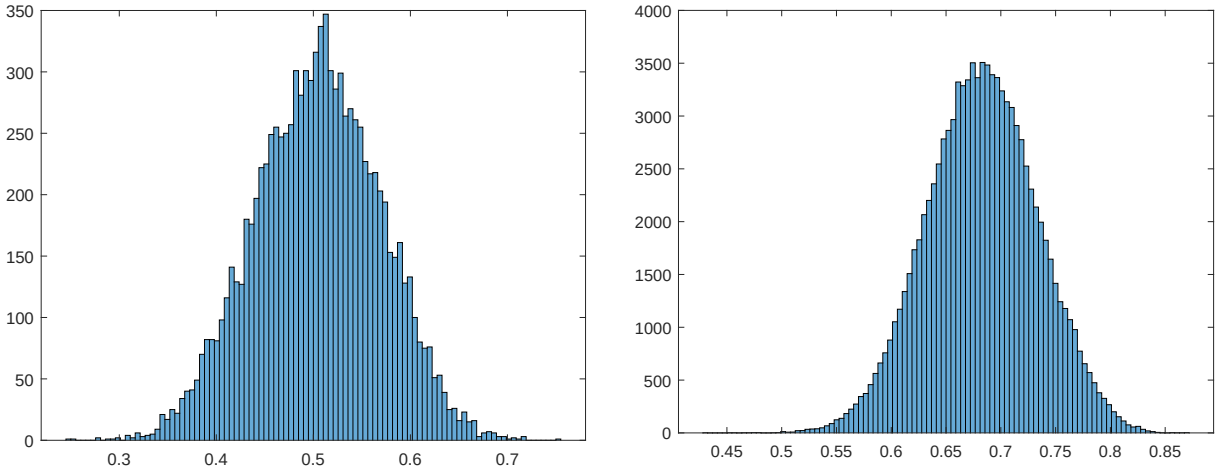
Figures 1 presents the results of our simulations for a contingency table taken from the classic work of Blau and Duncan (1978), Appendix J. The table describes mobility "from father's occupation to 1962 occupation" (p. 496).[29] The corresponding transition matrix is characterized by $|\lambda_2| \approx 0.38$ and $d'(1) \approx 0.66$. Panels (a) and (b) present histograms for $|\lambda_2|$ and $d'(1)$ computed for 100,000 matrices which have the same set of odds ratios. The variance of both characteristics turns out to be very large. Thus, our simulations demonstrate that odds ratios alone do not capture short-run or long-run measures of intergenerational persistence, discussed in Section 3.[30]

29. We exclude from the original table father-son pairs for which either father's or son's occupation is not known.

30. We confirm that we observe large variance for matrices of smaller sizes as well, and even when keeping one marginal distribution fixed and randomizing the other. In this paper we do not study the asymptotic properties of $|\lambda_2|$ and $d'(1)$, or any other characteristics, for matrices that fit random marginal distributions and have a fixed set of odds ratios. We leave such exercises for further research.

Figure 1: Histograms of simulated memory measures, holding fixed the same odds ratios

(a) Second largest eigenvalue modulus $|\lambda_2|$ (b) Dobrushin's coefficient $\delta(P)$



Notes: This figure shows histograms of memory measures for 100,000 simulated contingency tables, where the odds ratios are fixed and the marginals vary. See the main text in Section 4 for details.

5 Empirical Illustration

Having defined a new set of mobility measures, we now apply them to study intergenerational occupational mobility in the U.S. using the National Longitudinal Survey of Youth 1979 (NLSY79) dataset. In short, our empirical estimates of memory curves reveal that the "shirtsleeves-to-shirtsleeves in three generations" adage appears to be true on average (i.e., for a representative dynasty), but this average behavior masks heterogeneity in the experiences of dynasties originating from different occupational classes.

5.1 Data Description

Sample. The NLSY79 is a longitudinal sample of children and their families living in the U.S. that began in 1979, when children ranged between ages 14 and 22. Our main sample is restricted to children belonging to the NLSY's representative cross-sectional sample and supplemental Black and Hispanic sample with at least one birth parent or step-parent in their household. It is further restricted to children who were either residing with their parents or at a temporary residence at some point during the first eight survey years (1979–1986), rather than at a permanent residence independent from their parents.[31] Parents define the initial generation $t = 0$; children define the first generation $t = 1$.

Parent and child occupations are defined as symmetrically as possible to ensure that the transition matrix can best reflect a stationary Markov process (i.e., mobility is not mechanically generated

31. This restriction implies our main analysis sample tends to target younger birth cohorts among the original NLSY79 sample's cohorts. We apply sampling weights to ensure national representativeness.

from differences in the times that occupations are measured in the life cycles of parents and children). These definitions lead to important sample restrictions. Parent occupation is defined as the father's occupation in 1978 or when the child is age 14. If missing, we instead use the mother's occupation; if neither parent is working, then parent occupation is defined as "not employed." For children who are married or cohabiting, we similarly prioritize the male partner's occupation.[32] We focus on children whose occupations are observed when they are cohabiting with their own children (i.e., the dynasty's grandchildren) during adulthood. Since children are between ages 13 and 21 when parents' occupation is reported, we also restrict our attention to years of child occupation measured when the grandchild is between ages 13 and 21. For both generations, we further restrict our attention to observations where occupations are measured between 41 and 48. This leaves us with a sample of over 2,800 parent-child dyads. Approximately 50% and 30% of the dyads are White and Black, respectively; we will focus on these two racial groups as an empirical illustration of subgroup convergence analysis. See Appendix Table A.1 for basic summary statistics and Appendix Table C.1 for a summary of constructing the main sample.

The NLSY79 provides custom marginal sampling weights for cases where samples are constructed using multiple years of survey data. These weights allow our selected sample to be representative of the population of U.S. children born between 1957 and 1964. We apply these sampling weights throughout our analysis.[33]

Occupation classes. The 1970 and 2000 Census occupation codes of parents and children are harmonized across survey years using crosswalks to microclass occupation codes developed by Jonsson et al. (2009). Jonsson et al.'s classification scheme aggregates microclasses to meso- and macroclasses that represent distinct occupational skill and industry types.[34] In our main population analysis, we estimate memory curves using all three class types to examine the implications of employing more versus less aggregated occupation classes. In our Black–White subgroup analysis, we focus on meso- and macroclasses only due to sample size limitations.

Table 1 reports our adaptation of Jonsson et al.'s occupational classification scheme to better suit our data needs. Classes are ordered from primary manual (farming) occupations to increasingly nonmanual occupations with higher complexity and prestige. Notably, we include "not working" as a distinct occupational class, as this arguably captures a distinct notion of social class disadvantage, and rank it as the lowest numbered occupation class. Due to limited sample sizes, we aggregate

32. We use the child's spouse/partner's occupation so long as they are cohabiting with the child. We use partner's occupation only if the partner is the grandchildren's other parent. Prior to 1994, the NLSY79 only collected occupations of the child's spouse. Starting with the 1994 survey, it collected occupations of cohabiting partners. Hence, some years of occupation data may be missing for children with cohabiting partners before 1994.

33. As a result, when weighted, White and Black dyads represent about 80% and 14% of our sample, respectively.

34. Micro-, meso-, and macroclasses refer to different levels of aggregation of occupations in terms of similarity in terms of their locations in the production process and the characteristics required for workers to fulfill them. See Jonsson et al. (2009) for specifics on the aggregation scheme.

Table 1: Occupational classes

Macroclass	Mesoclass	Microclass	Jonsson et al. microclass codes
1. Not working	1.1. Unemployed or not in labor force	1.1.1. Unemployed or not in labor force	99999
2. Primary manual (farming)	2.1. Fishermen, farmers, farm laborers	2.1.1. Fishermen, farmers, farm laborers	25001–25202
3. Non-farming manual	3.1. Service workers	3.1.1. Janitors and cleaners	24311
		3.1.2. Launderers, housekeeping workers, gardeners	24309, 24310, 24312
		3.1.3. Transportation, delivery, and other services	24105, 24302, 24305–24308
		3.1.4. Food service workers	24304
		3.1.5. Protective service workers and guards	24301, 24303
	3.2. Lower manual	3.2.1. Forestry, operatives, and other kindred workers	24209, 24210
		3.2.2. Industrial processors, miners, longshoremen, sawyers, textile workers	24202–24208
		3.2.3. Truck drivers	24201
	3.3. Craft	3.3.1. Engine and machine operators, bricklayers, construction workers	24118–24120
		3.3.2. Specialized service craftsmen (e.g., tailors, plumbers, bakers, welders, painters)	24107, 24112–24117
		3.3.3. Blacksmiths, precious metal workers, other mechanics	24109–24111
		3.3.4. Vehicle mechanics	24108
		3.3.5. Electronics service, electricians, printers	24103, 24104, 24106, 24107
		3.3.6. Foremen and other craftsmen	24101, 24102
4. Routine nonmanual	4.1. Clerical	4.1.1. Postal and mail distribution	13204
		4.1.2. Telephone operators, bookkeepers, office/clerical workers	13201–13203
	4.2. Sales	4.2.1 Sales workers; real estate, insurance, and other agents,	13101–13105
5. Professional-managerial	5.1. Other professions	5.1.1. Hospital attendants, health semiprofessionals	11312, 11313
		5.1.2. Other technical workers (e.g., aircraft pilots, creative artists, ship officers); social workers; workers in religion	11301–11303, 11306–11311
		5.1.3. Teachers, librarians, nursery aides	11304, 11305, 11314
	5.2. Managers and officials	5.2.1. Officials, government, commercial/building managers	11201, 11203, 11204
		5.2.2. Other private managers	11202
	5.3. Classical professions	5.3.1. Engineers and architects	11106, 11109
		5.3.2. Jurists, health professionals, professors, scientists, accountants, journalists	11101–11105, 11107, 11108

Notes: This table reports the occupational classification scheme used in our analysis. We modify the original Jonsson et al. (2009) classification scheme in the following ways: (1) military occupations are combined with protective services (microclass 3.1.5), (2) "not working" is added as a new class category (microclass 99999), (3) proprietors are abstracted away due to data limitations, (4) all primary (farming) occupations are combined as a single mesoclass, and (5) microclasses are defined as aggregates of Jonsson et al.'s original microclasses. The last two modifications are due to limited sample sizes.

Jonsson et al.'s 82 original microclasses into 26 microclasses.

Estimated Markov chains. Figure 2 illustrates the estimated Markov chains (P, π_0) for macro-, meso, and microclasses color-coded as heat maps. Darker red cells indicate lower transition probability estimates while darker blue cells indicate higher transition probability estimates.

Four features can be gleaned from these estimated Markov chains. First, the lower triangles of the transition matrices tend to have lower probabilities than their upper triangles, indicating that children are more likely to transition to non-farming or nonmanual occupations than the converse. In particular, conditional on parents not working or working in manual occupations (macroclasses 1–3), children tend to gravitate toward non-farming manual occupations (macroclass 3), especially craft occupations (mesoclass 3.3); conditional on parents working in routine nonmanual occupations (macroclass 4), about 45% of children transition to professional-managerial occupations (macroclass 5), especially as managers and officials (mesoclass 5.2).

Second, transition matrices indicate that children rarely take up primary manual (farming) occupations (macroclass 2/mesoclass 2.1/microclass 2.1.1). Moreover, given the initial condition reveals only 4% of parents belong to this occupation, the steady-state share belonging to this occupation is tiny, representing only 0.6% of the population. Third, there is particularly high intergenerational persistence among professional-managerial occupations (macroclass 5). The mesoclass transition matrix indicates there is substantial circulation within this macroclass. There is especially high mobility into government, commercial, and building managers (microclass 5.2.1). These two features imply that primary manual and professional-managerial occupations are expected to be outliers when studying individual memory curves.

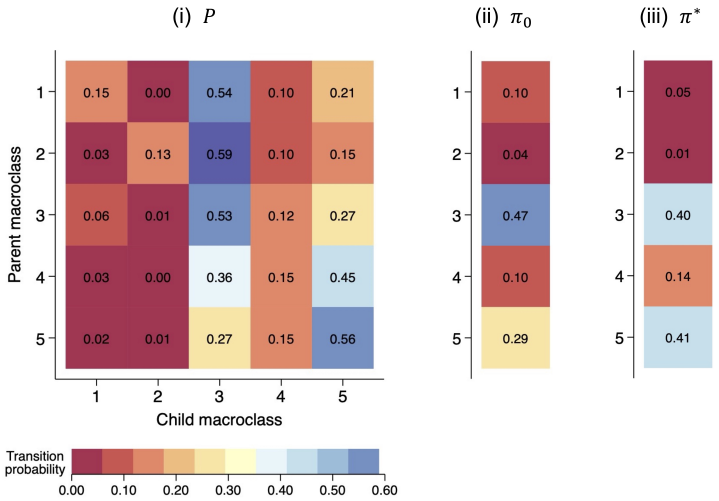
Finally, all transition matrices P possess at least one cell P_{ij} where $P_{ij} = 0$ (or extremely small values). We believe this is likely due to measurement error arising from our relatively small sample size. One may be worried that the presence of zero probabilities may upwardly bias estimates of memory.[35] As a robustness check, we adjust each transition matrix P by imputing $P_{ij} = \epsilon$ whenever $P_{ij} < \epsilon$ for some small $\epsilon > 0$ and adjusting other entries such that $\sum_j P_{ij} = 1$ for each row i. Appendix D.2 provides technical details of a procedure to optimally adjust P so that $P_{ij} > \epsilon$ for all (i, j) while minimizing differences from the original transition matrix.

Appendix Figures A.1 and A.2 illustrate macro- and mesoclass Markov chains estimated separately for our White and Black subgroups. In our sample, the White subgroup's parents are disproportionately represented in professional-managerial occupations while the Black subgroup's parents are concentrated in the non-farming manual and "not working" classes, revealing very different initial conditions. In addition, the White transition matrix features higher transitions toward professional-managerial occupations and lower transitions toward "not working" than the Black transition matrix, implying that children's occupational prospects substantially vary by race. Greater

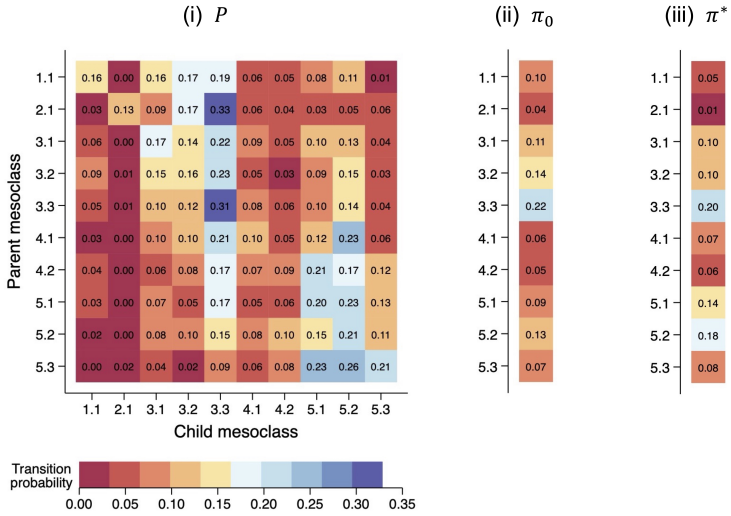
35. This is similar to concerns of estimating intergenerational income elasticities via log-linear regression models with data where some observations possess very low levels of income.

Figure 2: Intergenerational occupation Markov chains (P, π_0) and steady states π^*

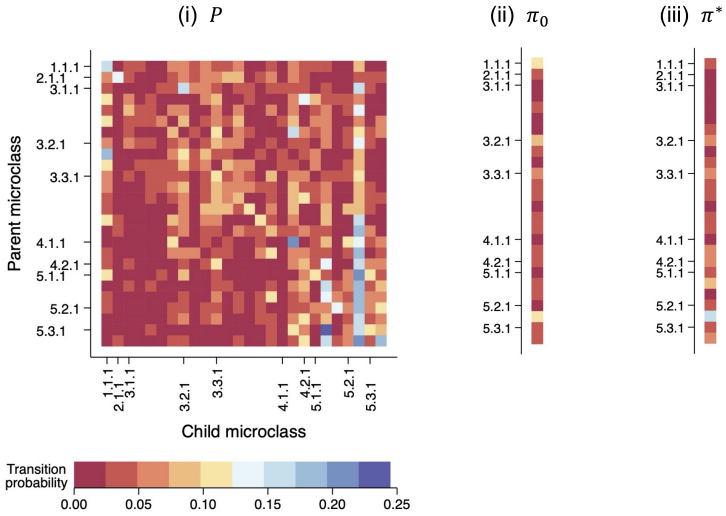
(a) Macroclasses



(b) Mesoclasses



(c) Microclasses



Notes: See Table 1 for definitions of occupational class codes. Panel (c) only labels microclass codes corresponding to class boundaries of mesoclasses for simplicity. All probabilities are estimated using custom NLSY79 sample weights.

downward occupational mobility among the Black subgroup aligns with the existing empirical literature; see, e.g., Durlauf et al. (2024) and the papers they cite.

5.2 Estimating Occupational Class Memory

Having estimated the transition matrices P and parents' marginal distribution over occupation classes π_0, we now illustrate the ideas developed in Section 3. We begin by presenting estimates of $d(t)$, $d'(t)$, and $AM(t, \pi_0)$, defined in Equations (1), (2), and (4), followed by estimates of class-specific $IM(t, i)$, defined in Equation (5). As discussed in Section 3, we focus on rates of convergence. We plot our memory curves in log scale, which helps illustrate the exponential rates of memory decay in the long run and improves the visual clarity of our results.

To conduct statistical inference, we employ a Bayesian analog to the rescaled bootstrap (Ishwaran et al. 2009).[36] This involves re-estimating our memory measures with random, strictly positive weights drawn from a Dirichlet distribution that is more right-skewed than what is conventionally used.[37] This bootstrap method is consistent for nondifferentiable functions such as $d(t)$ and $d'(t)$ and ensures cells or entire rows or columns of bootstrap estimates of P do not vanish.[38] We re-estimate all of our memory measures for 2,000 bootstrap draws. We then construct 95% bias-corrected accelerated (BC_a) point-wise confidence intervals for each estimated memory curve; the BC_a confidence intervals correct any bias and skewness present in the bootstrap distributions (DiCiccio and Efron 1996).[39]

36. The rescaled bootstrap is commonly referred to as the "M-out-of-N" bootstrap, where "M" is the size of the bootstrap sample and "N" is the original sample size.

37. Specifically, weights are drawn from the N-variate distribution Dirichlet$(\lambda \cdot {}^{M}/_{N}, \ldots, \lambda \cdot {}^{M}/_{N})$, where $\lambda > 0$ and $M < N$, instead of Dirichlet$(\lambda, \ldots, \lambda)$ that is customarily employed in the Bayesian bootstrap (Rubin 1981). The Dirichlet weights are constructed by normalizing random weights from the $\Gamma(\cdot, 1)$ distribution; note that $\Gamma(\lambda \cdot {}^{M}/_{N}, 1)$ first-order stochastically dominates $\Gamma(\lambda, 1)$. As Ishwaran et al. (2009) emphasize, the choice of λ and M are crucial. Common choices of λ in the literature are 1 and 4, likely inspired by the results discussed in Weng (1989); a common choice of M is $\zeta\sqrt{N}$ for some scalar $\zeta > 0$. Despite assigning strictly positive weights to all observations, choosing relatively small λ or M is problematic for estimating high-dimensional statistics like P (especially in the case of microclasses), since this assigns extremely small weights to relatively large numbers of observations, leading to qualitatively different convergence properties of the bootstrapped Markov chains. When testing different Dirichlet distributions with the microclass data, we found that choosing (λ, M) such that $\lambda \cdot {}^{M}/_{N} \ll 4$ led to asymmetric bootstrap distributions that often had no coverage of our estimates of $d(t)$ or $d'(t)$ in the original sample. In our main analysis, we choose (λ, M) such that $\lambda \cdot {}^{M}/_{N} = 3.8$.

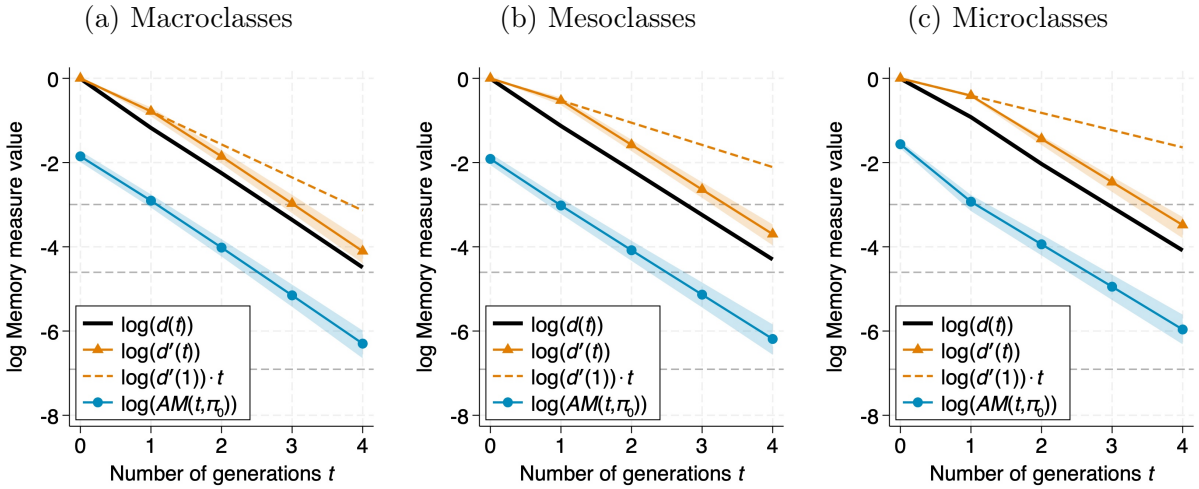
38. Classic bootstrap methods—i.e., drawing (with replacement) samples of parent-child dyads of size N from the original sample—are (1) not suitable in cases of nondifferentiable functions such as maximands (Dümbgen 1993, Andrews 2000) and (2) can lead to bootstrapped transition matrices with empty cells or, even worse, empty rows or columns, leading to Markov chains with different convergence properties or ill-defined memory measures. This issue is especially prominent in the study of microclasses. A common remedy for the first issue is to employ subsampling method such as the rescaled bootstrap (Politis and Romano 1994), which draws bootstrap samples of size $M < N$ such that ${}^{M}/_{N} \to 0$ as $M \to \infty$. A common remedy for the second issue is to employ a Bayesian bootstrap (Rubin 1981), which assigns randomly-drawn Dirichlet$(\lambda, \ldots, \lambda)$ weights to the original sample given concentration parameter λ. Ishwaran et al.'s (2009) proposed Bayesian analog of the rescaled bootstrap combines both methods.

39. Nonparametric methods for estimating the acceleration parameter in BC_a bootstrap confidence inter-

Analysis of $d(t)$, $d'(t)$, and $AM(t, \pi_0)$. Figure 3 plots $\log(d(t))$, $\log(d'(t))$, and $\log(AM(t, \pi_0))$ by the number of generations t, where estimates at $t > 1$ represent projections for their descendants. The three dashed horizontal lines correspond to $\log(0.05)$, $\log(0.01)$, and $\log(0.001)$ and serve as benchmarks for the level of remaining memory.

The asymptotic geometric convergence for $d(t)$, $d'(t)$ and $AM(t, \pi_0)$ kicks in early. This is illustrated by the linearity of all three log-transformations of those for $t \geq 1$. The geometric convergence rate coincides with the second largest eigenvalue of P, $|\lambda_2|$.

Figure 3: Estimates of $d(t)$, $d'(t)$, and $AM(t, \pi_0)$ curves



Notes: This figure plots $d(t)$, $d'(t)$, and $AM(t, \pi_0)$ (defined in Equations 1, 2, and 4) in log scale over generation t. Orange dashed lines $\log(d'(1)) \cdot t$ are also included as reference. Shaded regions are 95% BC_a confidence intervals using 2,000 Bayesian bootstrap replication samples. Confidence intervals are excluded from $\log(d(t))$ for visual clarity but are available upon request. Dashed horizontal lines correspond to $\log(0.05)$, $\log(0.01)$, and $\log(0.001)$.

As discussed in Section 3, $\log(d(t))$ (the black curves) illustrates worst-case-origin convergence to the invariant distribution. Table 2 reports the occupation classes that generate the $d(t)$ and $d'(t)$ estimates. Among macroclasses, primary manual (farming) occupations (class 2) possesses the greatest memory over all periods. But this masks heterogeneity present within other macroclasses. Under more disaggregated occupation classification schemes, classical professions (mesoclass 5.3)—specifically, classical professions such as jurists, health professionals, and professors (microclass 5.3.2)—appear to be the worst-case origin for several generations.

Next, $\log(d'(t))$ (orange curves) characterizes the worst-case convergence between dynasties, originating from different classes. In all three cases, this curve falls within or below 0.05 between generations $t = 3$ and $t = 4$ with 95% confidence. This implies that, by the fourth generation, all dynasties become roughly equivalent in the stochastic sense, irrespective of what class the ancestor

vals rely on the jackknife. In the case of microclasses, we exclude the few jackknife samples that cause any transition probabilities in P to vanish.

Table 2: Occupation classes generating $d(t)$, $d'(t)$

Generation	Macroclass		Mesoclass		Microclass	
t	$d(t)$	$d'(t)$	$d(t)$	$d'(t)$	$d(t)$	$d'(t)$
0	2	(4, 5)	2.1	(5.2, 5.3)	2.1.1	(5.3.1, 5.3.2)
1	2	(2, 5)	5.3	(2.1, 5.3)	3.2.3	(3.2.2, 5.3.1)
2	2	(2, 5)	5.3	(2.1, 5.3)	5.3.2	(3.2.3, 5.3.2)
3	2	(2, 5)	5.3	(2.1, 5.3)	5.3.2	(3.2.3, 5.3.2)
4	2	(2, 5)	5.3	(2.1, 5.3)	5.3.2	(3.2.3, 5.3.2)

Notes: This table reports the occupational classes that generate $d(t)$ and $d'(t)$ (Equations 1 and 2). See Table 1 for definitions of occupation codes.

originated from. Table 2 reports that, for nearly every generation t, the pair of macroclasses (i, j) that generate $d'(t)$ are primary manual (farming) and professional-managerial (corresponding to classes 2 and 5), meaning that these two macroclasses showcase the slowest convergence between dynasties. Turning to mesoclasses, this result is driven by primary manual occupations (2.1) and the classical professions such as engineers, architects, jurists, health professionals, scientists, and accountants (5.3). In the case of microclasses, however, the largest gap is generated by truck drivers (3.2.3) and classical professions unrelated to engineering and architecture (5.3.2). Given classical professions and primary and lower manual occupations possess very different labor market skills, this result appears consistent with the relatively slow exchange of dynasties over the multitude of class boundaries that separate them and small or zero transition probabilities present in the transition matrix.

To assess robustness of these results, Appendix Figures A.3 and A.4 plot histograms of the classes that generate $d(t)$ and $d'(t)$ across the bootstrap samples for $t = 1, 4$. Results for macro- and mesoclasses are very robust: in most bootstrap samples, primary manual and professional-managerial classes bound the speed of convergence. There is relatively greater uncertainty among microclasses.

As mentioned above, the rate of convergence of $d'(t)$ to zero is controlled by $|\lambda_2|$ for $t \geq 2$. However, the rate of contraction during the first step between $t = 0$ to $t = 1$, $d'(1)$, is slower than this asymptotic rate, as showcased by the difference between a linear extrapolation of the first step $\log(d'(1)) \cdot t$ (the dashed orange line) and $\log(d'(t))$ in Figure 3. The first-step rate of convergence is precisely captured by Dobrushin's coefficient $d'(1)$ (recall Equation 3). Values of $d'(1)$ and $|\lambda_2|$ for each occupational class scheme are reported in Appendix Table A.2. The difference in short-run and asymptotic convergence rates is especially stark among microclasses (Panel (c)), as evident from the fanning out between $\log(d'(1)) \cdot t$ and $\log(d'(t))$ for $t \geq 1$.

Turning our attention to $AM(t, \pi_0)$ (blue lines), estimates fall, even in case of microclasses, below or within 0.01 by $t = 3$. Thus, a "representative" dynasty's behavior becomes roughly equivalent to

the behavior of a dynasty randomly chosen from the invariant distribution by the third generation. In particular, this means that average memory—the memory of initial condition for the population as a whole—disappears approximately one generation earlier than worst-case origin memory.

Comparing our results across macro-, meso-, and microclasses, we find that the further we "zoom into" the occupational structure, the longer it takes to lose memory of initial conditions. This is true for both $AM(t, \pi_0)$ and $d(t)$. Similarly, the further we zoom in, the more time it takes for two dynasties to converge to each other. This is additionally reflected by the fact that both $d'(1)$ and $|\lambda_2|$ increase in the number of occupational classes used in the Markov chain.

Individual memories and their average. Figure 4, Panel (a) plots all estimated i-origin memory curves $IM(t, i)$ (see Equation 5) in log-scale for macro-, meso- and microclasses. There is significant spread in mixing times, with some classes reaching a memory level of 0.05 one generation earlier than others. Routine nonmanual occupations (macroclass 4)—and in particular, sales occupations (mesoclass 4.1)—rapidly lose memory over time. Among microclasses, a number of class-specific curves evolve nonlinearly, indicating evolving rates of convergence to steady state. Three such classes, which also possess the least memory by $t = 4$, include not only sales occupations (4.1.2) but also protective service workers (3.1.5) and craftsmen (3.3.6). Note that asymptotic geometric convergence governed by the second largest eigenvalue modulus of P kicks in at different t for different i (for some microclasses after $t = 4$).

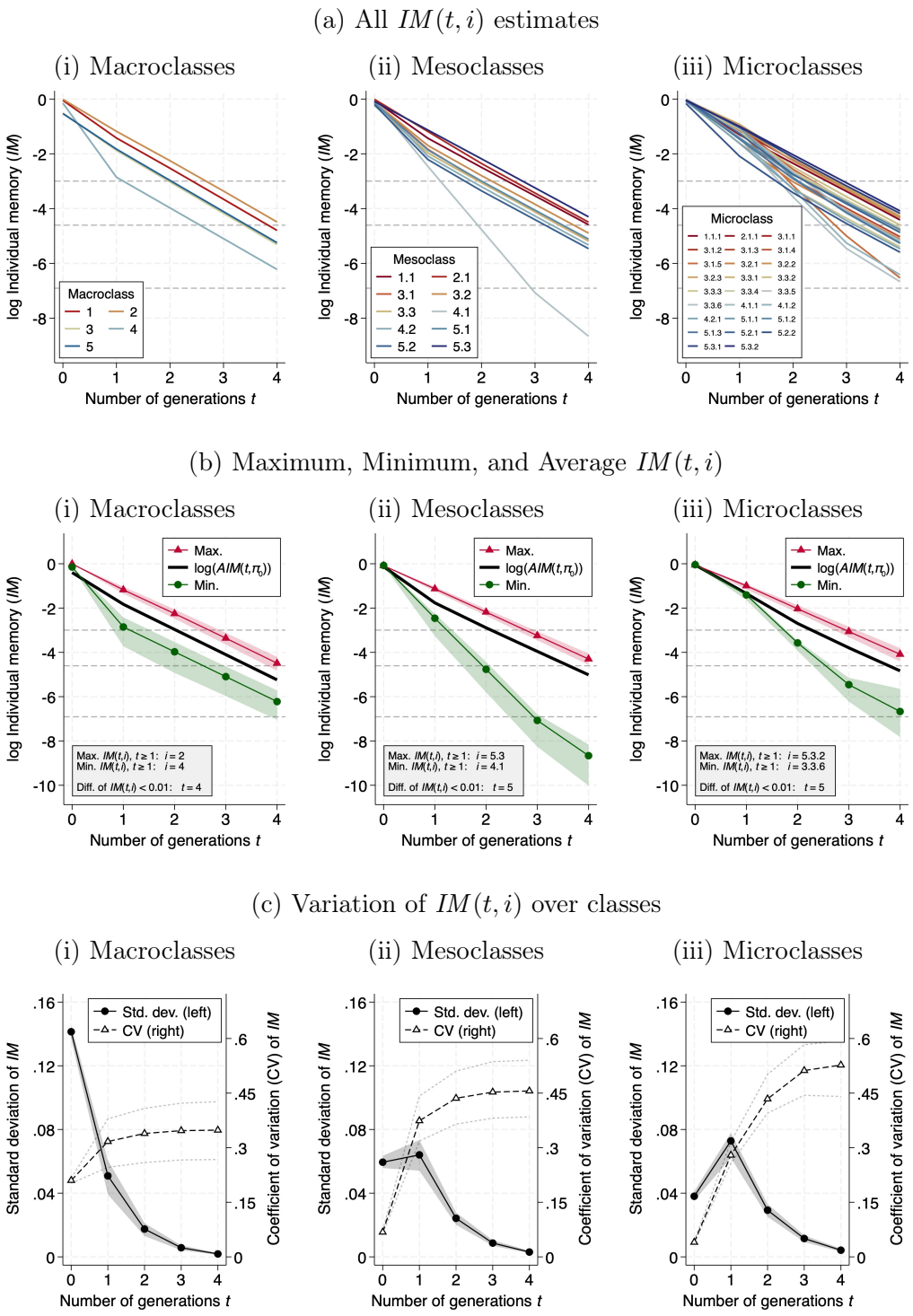
Panel (b) highlights the individual memory curves with maximum and minimum average memory between $1 \leq t \leq 4$ together with the average individual memory curve $AIM(t, \pi_0)$ defined in Equation (6).[40] The figures report the classes corresponding to the maximum and minimum $IM(t, i)$ curves and two statistics on their degree of convergence with each other. It takes 4–5 generations for dynasties originating from these classes to be within 0.01 deviation of each other.

To illustrate and summarize the degree of heterogeneity in the $IM(t, i)$ estimates, Panel (c) plots measures of variation across classes i at each generation. For $t > 1$, the *absolute* degree of variation across $IM(t, i)$, as measured by the standard deviation, monotonically declines in t and reaches close to 0 by $t = 4$ for all aggregation levels of occupations. However, the *relative* degree of variation, captured by the coefficient of variation, shows clearer differences across occupation schemes. The degree of heterogeneity in i-origin memories increases in the number of occupational classes considered and takes significantly longer to plateau. This leads to slower aggregate mixing rates.

Robustness to adjusting zeros in transition matrices. A consistent theme that emerges in our results is that the more disaggregated occupation schemes exhibit relatively longer memory. Could this simply be driven by the larger number of zeros in the transition matrix P that, in turn,

40. In many cases, maximal average memory corresponds to $d(t)$, but note that $d(t)$ can be generated by different classes at different t. Appendix Figure A.5 provides histograms of classes associated with maximum and minimum average memory estimated over the bootstrap samples.

Figure 4: Individual (class-specific) memory curves



Notes: Panel (a) plots estimates of $IM(t,i)$ (defined in Equation 5) for all occupation classes i in log scale over generations t. See Table 1 for definitions of occupational class codes. Panel (b) plots the mean $AIM(t,\pi_0)$ (defined in Equation 6) and $IM(t,i)$, for classes i with the largest and smallest mean memory (i.e., $\log(IM(t,i))$ value) for $t \geq 1$. The gray textbox reports the maximum and minimum classes and the times in which their difference in IM estimates falls below 0.01. Dashed horizontal lines correspond to $\log(0.05)$, $\log(0.01)$, and $\log(0.001)$. Panel (c) plots measures of variation of $IM(t,i)$ across classes i: the standard deviation and the coefficient of variation. Shaded regions in all panels and the dotted lines in the bottom panels represent 95% BC_a confidence intervals using 2,000 Bayesian bootstrap replication samples. Confidence intervals are excluded from $\log(AIM(t,\pi_0))$ for visual clarity but are available upon request.

may be an artifact of measurement error? To assess this, we replicate our analysis using adjusted versions of P where all zero or extremely small probabilities are replaced with some minimum threshold value $\epsilon > 0$.[41] Appendix Figure A.6 plot estimates of $d(t)$, $d'(t)$, and $AM(t, \pi_0)$ for different choices of ϵ. We find that, for $0 < \epsilon \leq 0.01$, all macro- and mesoclass estimates lie within the 95% confidence interval of the original curves. Microclass memory curves are highly robust for small ϵ and only decay somewhat more rapidly when setting $\epsilon = 0.01$, a relatively large perturbation for the sparse matrix P. Thus, we believe our results are not driven by potential measurement error.

Black–White subgroup analysis. We now turn to subgroup differences between racial subgroups in our sample. We begin by studying $d(t)$, $d'(t)$ and $AM(t, \pi_0)$ for Black and White samples separately. Appendix Figure A.7 reproduces Figure 3 using P_b together with $\pi_{0,b}$ and P_w together with $\pi_{0,w}$. Overall, we observe substantially faster dissipation of memory among Black dynasties compared to White dynasties. Thus, we observe greater equality of opportunity within the Black subgroup. This, however, does not provide any information about differences in opportunities per se between Black and White dynasties. Our analysis of marginal distributions reveals that $\pi_{t,w}$ first-order stochastically dominates $\pi_{t,b}$ among parents ($t = 0$) and several subsequent generations.[42] The same relation is true for the steady states π^*_w and π^*_b. This means that a representative White dynasty faces a better lottery within the considered occupational hierarchy than a representative Black dynasty.

Figure 5, Panel (a) studies convergence between the representative dynasties of the Black and White subgroups by plotting variants of Equation (7). Applying the population transition matrix P to both the Black and White subgroup initial conditions $\pi_{0,b}$, $\pi_{0,w}$ (black lines), we conclude that differences between representative Black and White dynasties fall to 0.01 within about three generations. Convergence between dynasties occurs even more rapidly when applying the Black transition matrix P_b to both initial conditions (blue lines), indicating opportunities are better equalized between the Black and White subgroups under the Black transition mechanism. Unsurprisingly, applying the White transition matrix P_w to $\pi_{0,b}$ and $\pi_{0,w}$ (orange lines) results in nearly identical dynamics as applying the population transition matrix P, given that Whites make up the majority of the population and thus disproportionately contribute to P.[43] The actual difference in opportunities for a representative Black and a representative White dynasty is illustrated with $SDM(t, \pi_{0,b}, P_b, \pi_{0,w}, P_w)$ (dashed red lines). The steady-state difference in probability distributions over occupations $\| \pi^*_b - \pi^*_w \|_{TV} = 0.32$ for mesoclasses and is approached almost immediately and is significantly higher than the other subgroup difference curves at $t = 1$. This indicates the power-

41. See Appendix D.2 for details on how we obtain an optimally-adjusted P.

42. Here we rely on the ordering of occupation classes according to skills and assume that occupations which require higher skills are more prestigious and desirable. Probability distribution π_w stochastically dominates probability distribution π_b if $(\pi_w)_1 + (\pi_w)_2 + ... + (\pi_w)_k \leq (\pi_b)_1 + (\pi_b)_2 + ... + (\pi_b)_k$ for any $k = 1, 2, ..., n-1$, where $(\pi_w)_i$ is the i-th entry of π_w and n is the number of classes, as before.

43. Since π^*_w first-order stochastically dominates π^*_b, applying P_w to $\pi_{0,b}$ and $\pi_{0,w}$ yields "leveling up" the playing field for Black and White dynasties, while applying P_b—"leveling down".

ful role that differences in transition mechanisms faced by the Black and White subgroups play in perpetuating subgroup inequality.

Analogous to the maximum and minimum $IM(t,i)$ analysis, we can go beyond analyzing representative dynasties by studying Black and White dynasties originating from specific classes that maximizesubgroup differences. Panel (b) of Figure 5 plots differences with initial conditions (e_i, e_j) where classes (i,j) maximize $SDM(t, e_i, P_b, e_j, P_w)$ for $0 \leq t \leq 4$—in other words, produce the greatest intergenerational inequality between racial subgroups—in the case of distinct transition matrices P_b and P_w. For both macro- and mesoclasses, these classes correspond to higher-skilled professional occupations for the White subgroup and lower-skilled occupations for the Black subgroup.[44] If the low-skilled Black dynasty and the high-skilled White dynasty share the same transition matrix, either P_b or P_w, they become roughly equivalent in the stochastic sense only by the fourth generation for both macro- and mesoclasses (orange and blue lines). When race-specific transition matrices are applied (dashed red lines), convergence to the steady-state difference $\| \pi_b^* - \pi_w^* \|_{TV}$ is much slower relative to the case of representative dynasties. Thus, the experiences of a low-skilled Black dynasty and a high-skilled White dynasty are even more profoundly different.

6 Conclusions

This paper develops a framework for analyzing the intergenerational persistence of socioeconomic status based on the memory of origin in the stochastic processes that characterize intergenerational links. We draw on concepts from the theory of Markov chains to investigate the rates at which the memory of ancestors' occupations diminishes across generations, equalizing opportunities for descendants—a property known as mixing. Relative to the existing literature on measuring intergenerational mobility, we offer some new ways to extract information on intergenerational persistence through a new class of memory curve measures. The curves provide a unified way of understanding mobility over short versus long horizons. In addition to memory measures used in the Markov chain literature which involve worst-case origin, we develop memory measures for both individual dynasties and for the aggregate population. Proposals are also put forward to characterize disparities between subgroups.

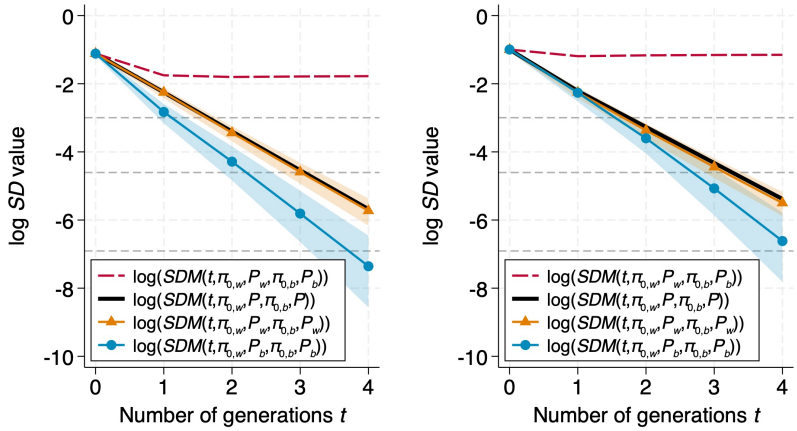
We empirically illustrate our memory curve measures by studying occupational mobility in U.S. survey data. We explore the behavior of our memory curves at three distinct levels of occupational aggregation. Overall, we find relatively rapid convergence between dynasties originating from different classes. On average, memory of origin disappears by approximately the third generation. The worst-case-origin memory disappears one generation later. Among macroclasses, primary manual (farming) occupations possess the greatest memory. This class, coupled with professional-managerial

44. For macroclasses, these initial occupations correspond to professional-managerial for the White subgroup and "not working" for the Black subgroup; for mesoclasses, they respectively correspond to classical professions and clerical occupations.

Figure 5: Black–white subgroup-difference memory (SDM) curves

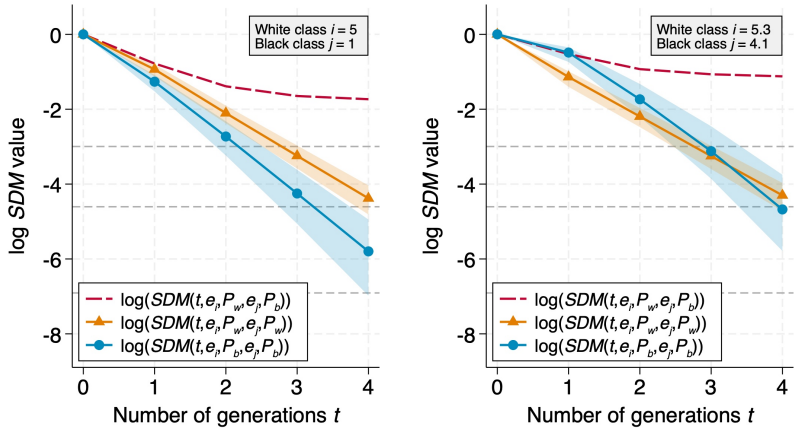
(a) Representative dynasties $(\pi_{0,w}, \pi_{0,b})$

(i) Macroclasses (ii) Mesoclasses



(b) Maximizing class-specific dynasties (e_i, e_j)

(i) Macroclasses (ii) Mesoclasses



Notes: This figure plots differences between the Black and white subgroups based on variants of Equation (7) involving different choices of transition matrices and initial state vectors. Lines within a given plot use the same pair of initial state vectors applied to different pairs of transition matrices. Panel (a) uses $(\pi_{0,w}, \pi_{0,b})$ as the initial state vectors. Panel (b) employs (e_i, e_j) where (i, j) are the modal classes that generate $\max_{i,j} SDM(t, e_i, P_w, e_j, P_b)$ for $t \leq 4$; the maximizing classes are reported in the gray textboxes. Dashed horizontal lines correspond to $\log(0.05)$, $\log(0.01)$, and $\log(0.001)$. Shaded regions represent 95% BC_a confidence intervals using 2,000 Bayesian bootstrap replication samples. Confidence intervals are excluded from select lines for visual clarity but are available upon request.

occupations, produce the worst-case convergence between two dynasties originating from different classes. At the more disaggregated microclass level, however, the strongest memory is associated with classical professions, such as jurists, health professionals, and professors. The worst-case gap between two dynasties of different origins is generated by truck drivers and classical professions.

When plotting individual, or class-specific, memory curves, however, a significant degree of heterogeneity is observed, particularly in the case of microclasses. At this level we also discover qualitatively different types of behavior among individual memories, including very different rates at which these memories dissipate. Delving deeper into the occupational structure reveals more heterogeneity in class-specific memories, longer times to escape from initial conditions, and a greater average memory of origin. Thus, the more one "zooms into" the occupational structure, the more inequality of opportunity is found.

Finally, we analyze the memory curves of Black and White subgroups and study the memory of differences between these subgroups. We find that the Black transition matrix equalizes opportunity for dynasties of different origins more rapidly than the White transition matrix, indicating faster loss of memory of origin within the Black subgroup. At the same time, a representative White dynasty is more likely to occupy a high-skilled occupation compared to a representative Black dynasty. We find differences in race-specific transition matrices play a substantial role in subgroup inequality. If the Black and White subgroups shared the same transition matrix, the difference between representative Black and White dynasties would largely vanish within three generations.

In terms of future work, applying our proposed measures to larger datasets may prove fruitful for uncovering short- versus long-run mobility dynamics with greater precision and examining heterogeneity for larger numbers of occupational classes. Natural methodological extensions include multivariate generalizations of mobility calculations, such as the joint consideration of occupation, income, and education. Such approaches will allow for the analysis of feedbacks between different socioeconomic characteristics. Here the latent class approach (e.g., Bavaro and Tullio (2023)) provides interesting directions to pursue. In our judgment, a multivariate approach would likely be essential in the identification of bottlenecks to socioeconomic progress for specific population groups.

References

Andrews, Donald W. K. 2000. "Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space." *Econometrica,* 399–405.

Bavaro, Michele, and Federico Tullio. 2023. "Intergenerational Mobility Measurement with Latent Transition Matrices." *The Journal of Economic Inequality* 21 (1): 25–45.

Becker, Gary S., and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4 (3, Part 2): S1–S39.

Berger, Joseph, and James L. Snell. 1957. "On the Concept of Equal Exchange." *Behavioral Science* 2 (2): 111–118.

Blau, Peter M., and Otis Dudley Duncan. 1978. *The American Occupational Structure.* The Free Press (Reprint of the Ed. published by Wiley, New York).

Boudon, Raymond. 1973. *Mathematical Structures of Social Mobility.* Jossey-Bass.

Brémaud, Pierre. 1999. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues.* Springer.

Brumelle, S. L., and Y. Gerchak. 1982. "The Structure of Indices of Social Mobility and Inheritance." *Journal of Mathematical Sociology* 8 (2): 251–264.

Conlisk, John, and Paul M. Sommers. 1979. "Eigenvector Status Proxies in Markov Chain Mobility Models." *Sociological Methods & Research* 8 (2): 159–178.

Dardanoni, Valentino. 1993. "Measuring Social Mobility." *Journal of Economic Theory* 61:372–394.

———. 1995. "Income Distribution Dynamics: Monotone Markov Chains Make Light Work." *Social Choice and Welfare* 12:181–192.

Deming, W Edwards, and Frederick F Stephan. 1940. "On a Least Squares Adjustment of a Sampled Frequency Table when the Expected Marginal Totals are Known." *The Annals of Mathematical Statistics* 11 (4): 427–444.

DiCiccio, Thomas J., and Bradley Efron. 1996. "Bootstrap Confidence Intervals." *Statistical Science* 11 (3): 189–228.

Dudel, Christian. 2021. "Expanding the Markov Chain Toolbox: Distributions of Occupation Times and Waiting Times." *Sociological Methods & Research* 50 (1): 401–428.

Dümbgen, Lutz. 1993. "On Nondifferentiable Functions and the Bootstrap." *Probability Theory and Related Fields* 95:125–140.

Duncan, Otis D. 1966. "Methodological Issues in the Analysis of Social Mobility." In *Social Structure and Mobility in Economic Development,* 51–97. Chicago: Aldine.

Durlauf, Steven N, Gueyon Kim, Dohyeon Lee, and Xi Song. 2024. *The Evolution of Black-White Differences in Occupational Mobility Across Post-Civil War America.* Technical report. National Bureau of Economic Research.

Fienberg, Stephen E. 1970. "An Iterative Procedure for Estimation in Contingency Tables." *The Annals of Mathematical Statistics* 41 (3): 907–917.

Goodman, L. A. 1968. "The Analysis of Cross-Classified Data: Independence, Quasi-independence, and Interactions in Contingency Tables with or without Missing Entries: Ra Fisher Memorial Lecture." *Journal of the American Statistical Association* 63 (324): 1091–1131.

———. 1969. "How to Ransack Social Mobility Tables and Other Kinds of Cross-Classification Tables." *American Journal of Sociology* 75 (1): 1–40.

Harrison, Roderick J. 1988. "Do Odds Ratios Really Control for the Availability of Occupational Positions in Status Contingency Tables?" *European Sociological Review* 4 (1): 65–79.

Ishwaran, Hemant, Lancelot F. James, and Mahmoud Zarepour. 2009. "An Alternative to the m out of n Bootstrap." *Journal of Statistical Planning and Inference* 139 (3): 788–801.

Jonsson, Jan O., David B. Grusky, Matthew Di Carlo, Reinhard Pollak, and Mary C. Brinton. 2009. "Microclass Mobility: Social Reproduction in Four Countries." *American Journal of Sociology* 114 (4): 977–1036.

Levin, David A., and Yuval Peres. 2017. *Markov Chains and Mixing Times.* Vol. 107. American Mathematical Soc.

Macdonald, Kenneth. 2023. "The Marginal Adjustment of Mobility Tables, Revisited."

Markandya, Anil. 1982. "Intergenerational Exchange Mobility and Economic Welfare." *European Economic Review* 17 (3): 307–324.

Mazumder, Bhashkar. 2005. "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data." *Review of Economics and Statistics* 87 (2): 235–255.

McFarland, David. 1970. "Intragenerational Mobility as a Markov Process: Including a Time-Stationary Markovian Model that Explains Observed Declines in Mobility Rates over Time." *American Sociological Review* 35 (3): 463–476.

Politis, Dimitris N., and Joseph P. Romano. 1994. "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions." *The Annals of Statistics,* 2031–2050.

Powers, Daniel, and Yu Xie. 2008. *Statistical Methods for Categorical Data Analysis.* Emerald Group Publishing.

Prais, Sig J. 1955. "Measuring Social Mobility." *Journal of the Royal Statistical Society* 118 (1): 56–66.

Pullum, Thomas W. 1975. *Measuring Occupational Inheritance.* Amsterdam: Elsevier.

Roemer, John E. 1993. "A Pragmatic Theory of Responsibility for the Egalitarian Planner." *Philosophy & Public Affairs,* 146–166.

Roemer, John E. 1998. *Equality of Opportunity.* Harvard University Press.

Rubin, Donald B. 1981. "The Bayesian Bootstrap." *The Annals of Statistics,* 130–134.

Seneta, Eugene. 2006. *Non-negative matrices and Markov chains.* Springer Science & Business Media.

Shorrocks, Anthony F. 1978. "The Measurement of Mobility." *Econometrica* 46 (5): 1013–1024.

Sommers, Paul M., and John Conlisk. 1979. "Eigenvalue Immobility Measures for Markov Chains." *Journal of Mathematical Sociology* 6:253–276.

Song, Xi. 2021. "Multigenerational Social Mobility: A Demographic Approach." *Sociological Methodology* 51 (1): 1–43.

Stewman, Shelby. 1975. "Two Markov Models of Open System Occupational Mobility: Underlying Conceptualizations and Empirical Tests." *American Sociological Review* 40 (3): 298–321.

Weng, Chung-Sing. 1989. "On a Second-Order Asymptotic Property of the Bayesian Bootstrap Mean." *The Annals of Statistics,* 705–710.

Appendix

A Supplemental Tables and Figures

Table A.1: Summary statistics

	Mean	S.D.
Demographics		
Female	0.50	0.50
Year of birth	1960.52	2.26
Non-Hispanic white	0.80	0.40
Non-Hispanic Black	0.14	0.34
Hispanic	0.06	0.24
Parent macroclasses		
1. Not working	0.10	0.30
2. Primary manual (farming)	0.29	0.45
3. Non-farming manual	0.10	0.30
4. Routine nonmanual	0.45	0.50
5. Professional-managerial	0.04	0.19
Child macroclasses		
1. Not working	0.04	0.20
2. Primary manual (farming)	0.41	0.49
3. Non-farming manual	0.13	0.34
4. Routine nonmanual	0.41	0.49
5. Professional-managerial	0.01	0.09
Number of observations	2,847	

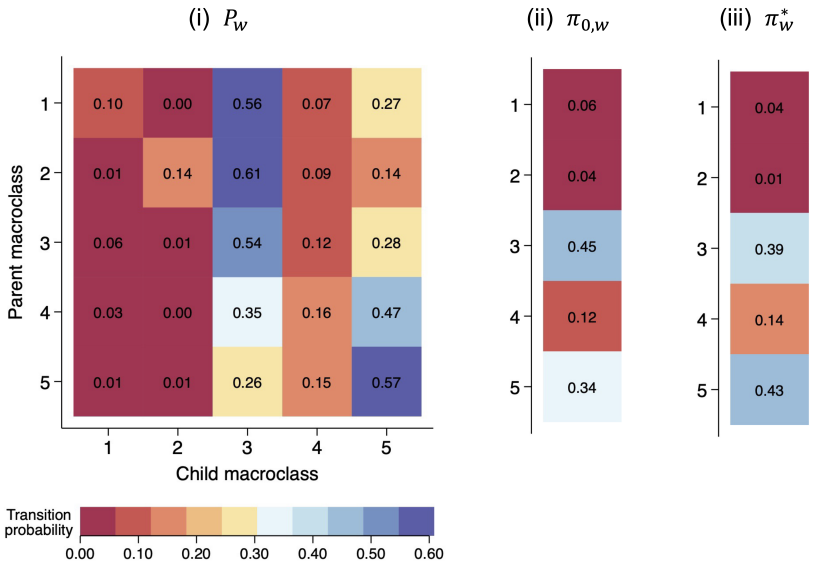
Notes: All statistics apply custom NLSY79 sampling weights. See Table 1 for more details on the Jonsson et al. (2009)-based occupation classification scheme.

Table A.2: $d'(1)$, defined in Equation 3, and second largest eigenvalue modulus $|\lambda_2|$ for macro-, meso- and microclasses

	No. classes	$d'(1)$	$\lvert\lambda_2\rvert$
Macroclasses	5	0.456	0.320
Mesoclasses	10	0.590	0.348
Microclasses	26	0.664	0.360

Figure A.1: Markov chains by racial group: Macroclasses

(a) White



(b) Black

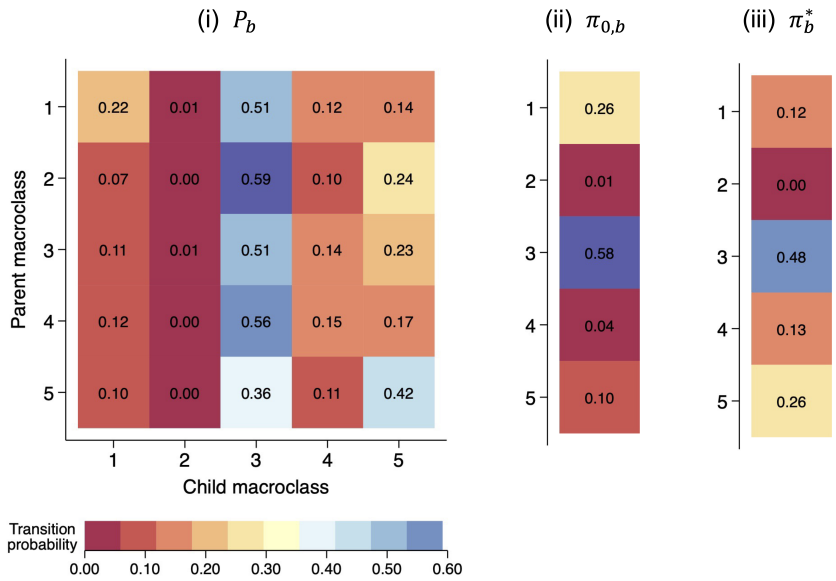
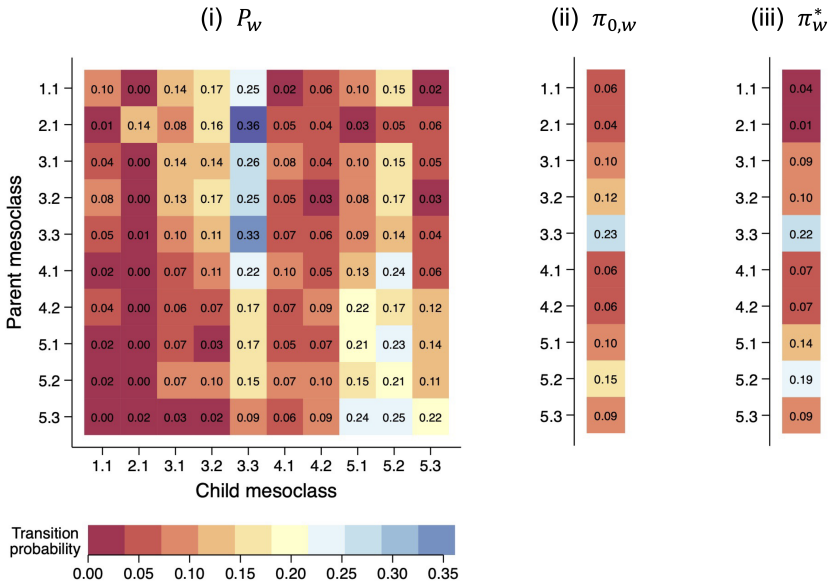


Figure A.2: Markov chains by racial group: Mesoclasses

(a) White



(b) Black

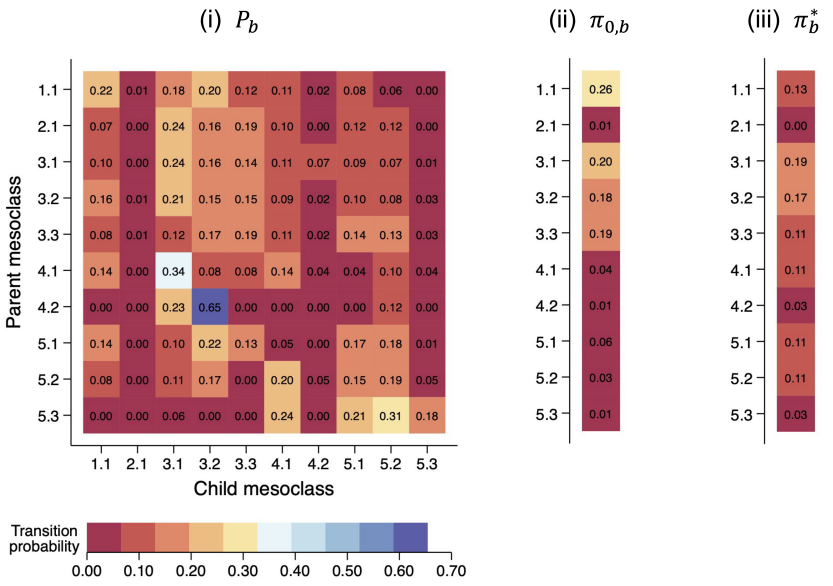
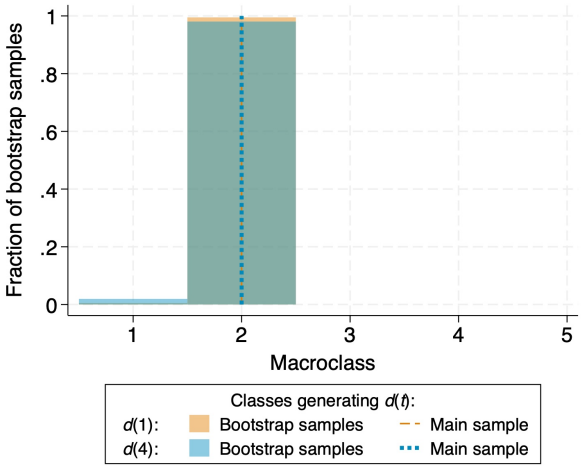
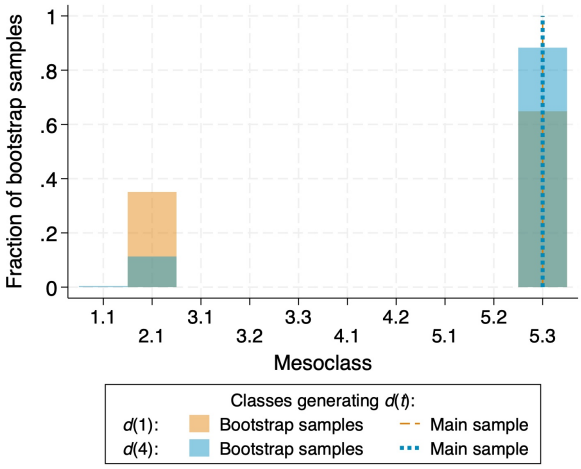


Figure A.3: Histograms of classes generating $d(t)$ for $t = 1, 4$ over bootstrap samples

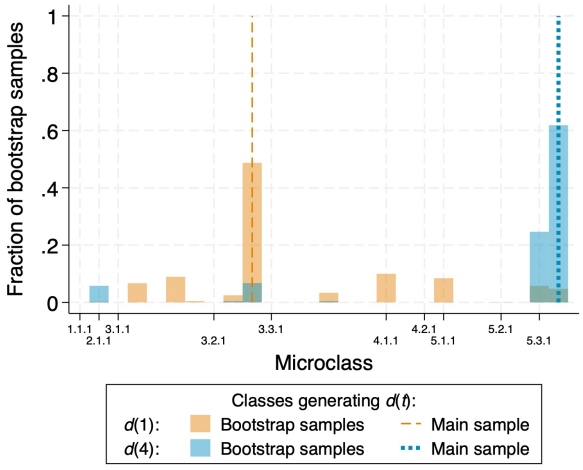
(a) Macroclasses



(b) Mesoclasses



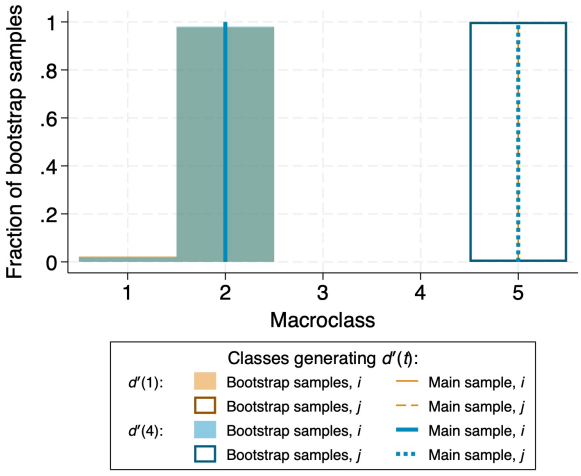
(c) Microclasses



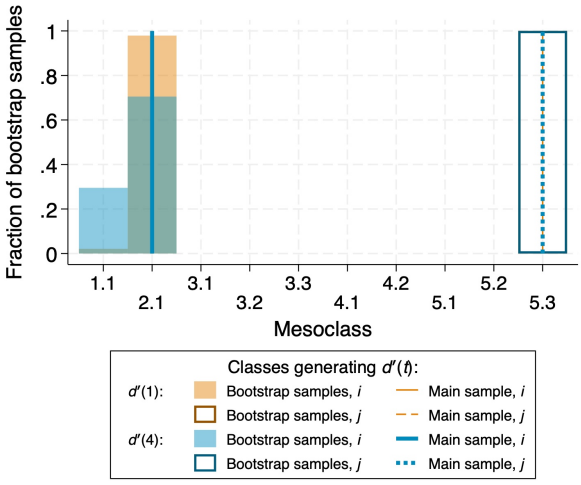
Notes: This figure plots histograms of the occupational classes $i^\star$ that generate $d(t)$ (i.e., $i^\star = \arg\max_i IM(t, i)$) across 2,000 Bayesian bootstrap replication samples for $t = 1, 4$. The vertical dashed lines indicate the classes that generate $d(t)$ in the main sample.

Figure A.4: Histograms of classes generating $d'(t)$ for $t = 1, 4$ over bootstrap samples

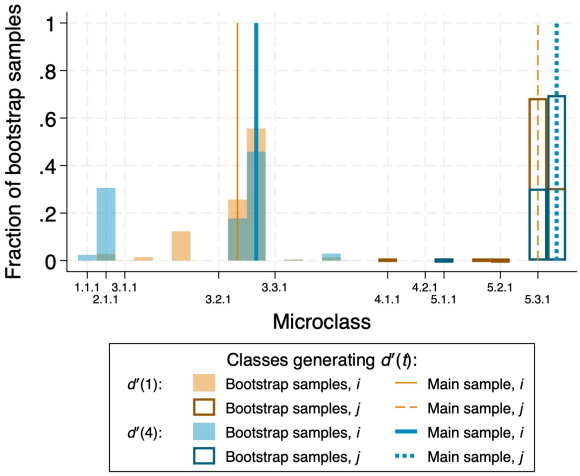
(a) Macroclasses



(b) Mesoclasses



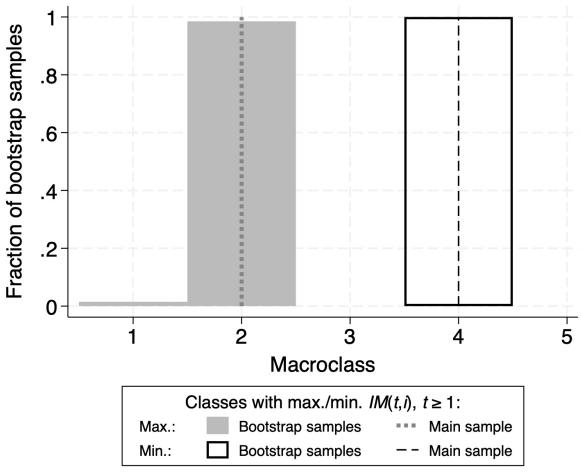
(c) Microclasses



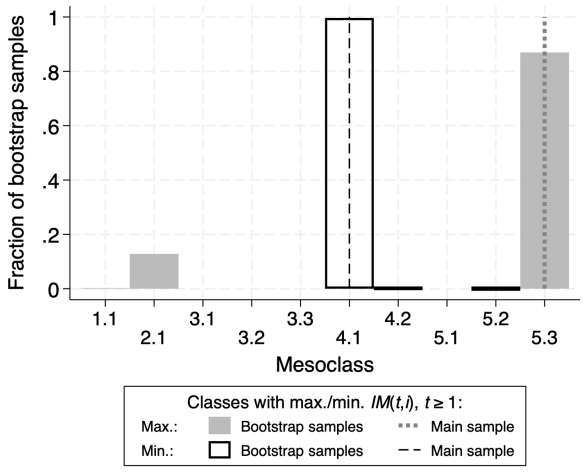
Notes: This figure plots histograms of the occupational classes $(i^\star, j^\star)$ that generate $d'(t)$ (i.e., $(i^\star, j^\star) = \arg\max_{(i,j)} d'(t,i,j)$ where $d'(t,i,j) = \| e_i \cdot P^t - e_j \cdot P^t \|$) across 2,000 Bayesian bootstrap replication samples for $t = 1, 4$. The vertical dashed lines indicate the classes that generate $d'(t)$ in the main sample.

Figure A.5: Histograms of classes with maximum and minimum $IM(t,i)$ for $t \geq 1$ over bootstrap samples

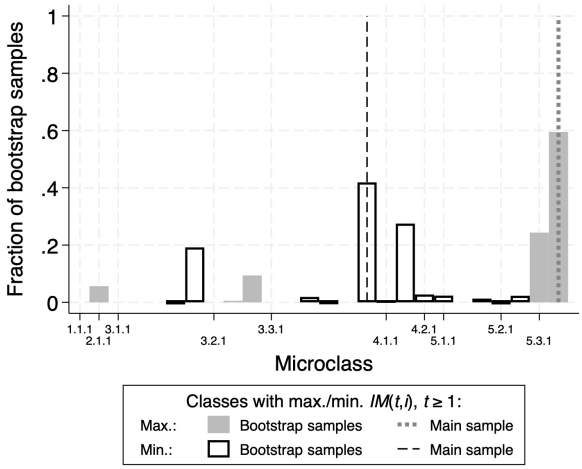
(a) Macroclasses



(b) Mesoclasses

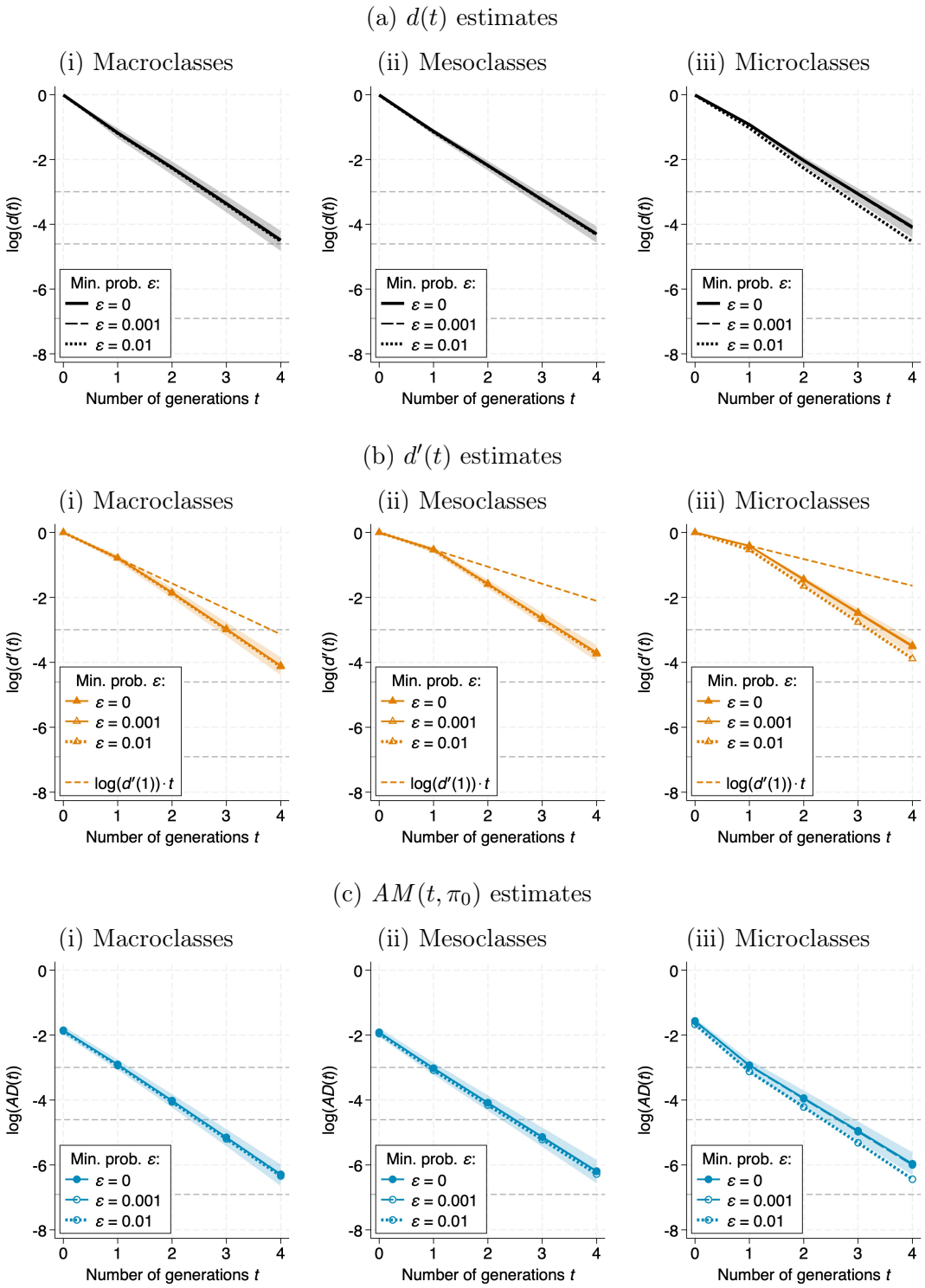


(c) Microclasses



Notes: This figure plots histograms of the occupational classes (i_{max}, i_{min}) that possess the maximum and minimum class-specific memory IM between $1 \leq t \leq 4$ (i.e., $i_{max} = \arg\max_i \sum_{t=1}^{8} \log(IM(t,i))$ and $i_{min} = \arg\min_i \sum_{t=1}^{8} \log(IM(t,i))$) across 2,000 Bayesian bootstrap replication samples for $t = 1, 4$. The vertical dashed lines indicate the classes that generate the maximum and minimum IM curves in the main sample.

Figure A.6: Robustness of estimates to adjusting small probabilities in transition matrix

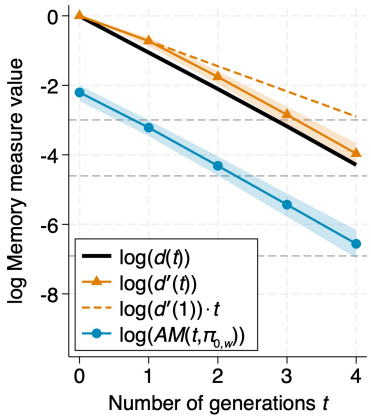


Notes: This figure compares memory measure estimates when adjusting the transition matrix P to have minimum probability ϵ. The original memory measure estimates ($\epsilon = 0$) is included as a reference. Shaded regions are 95% BC_a confidence intervals of the original memory measure estimates using 2,000 Bayesian bootstrap replication samples. See the notes of Figure 3 for details on the confidence intervals of Panels (a) and (b).

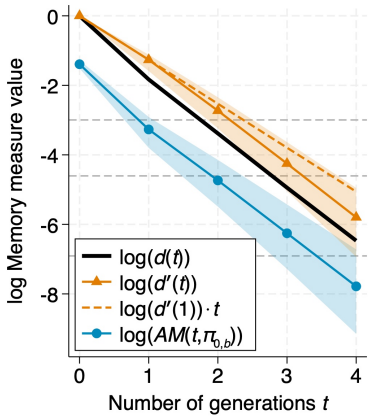
Figure A.7: Estimates of $d(t)$, $d'(t)$, and $AM(t, \pi_0)$ curves, by racial subgroup

(a) Macroclasses

(i) White

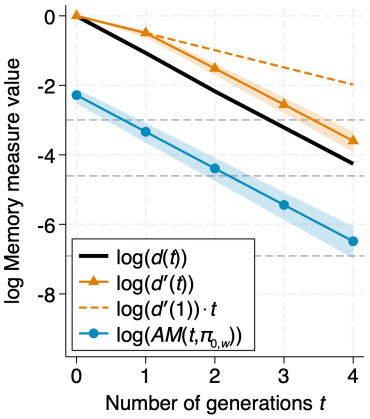


(ii) Black

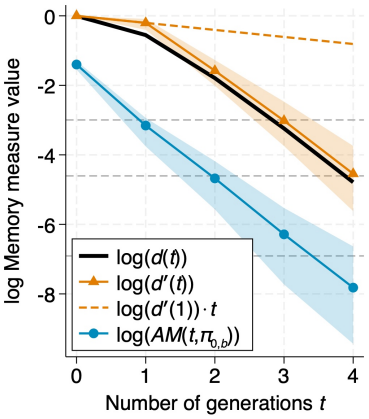


(b) Mesoclasses

(i) White



(ii) Black



Notes: This figure plots $d(t)$, $d'(t)$, and $AM(t, \pi_{0,r})$ (defined in Equations 1, 2, and 4) in log scale over generation t by racial subgroup $r \in \{w, b\}$. All race-specific memory measures apply their respective race-specific transition matrix P_r. Orange dashed lines $\log(d'(1)) \cdot t$ are also included as reference. Shaded regions are 95% BC_a confidence intervals using 2,000 Bayesian bootstrap replication samples. Confidence intervals are excluded from $\log(d(t))$ for visual clarity but are available upon request. Dashed horizontal lines correspond to $\log(0.05)$, $\log(0.01)$, and $\log(0.001)$.

B Structural and Exchange Mobility in Markov Models

B.1 Additional Formalization

The population interpretation of Markov models suggests a straightforward approach to formalizing the exchange versus structural distinction. The steady state distribution of dynasties over classes is such distribution that, for each class, the inflow of dynasties into this class equals the outflow of dynasties from this class. If this was not true, the cross section distribution would necessarily change. If the population is not in the steady state and structural mobility is present, there must exist at least two classes for which there are imbalances between the inflows and outflows into the classes. One metric for structural mobility may therefore be based on the vector of such imbalances, the magnitude of structural mobility is measured by a norm applied to the vector. Over time the marginal distributions converge to the steady state, so the norm of this vector will converge to 0, meaning that structural mobility disappears.

To be more precise, consider inflows and outflows into the first class observed in period $t+1$. If in generation t all N dynasties are distributed over n classes as $(N_1, N_2, ..., N_n)$, the (expected) inflow of dynasties into the first class may be computed as $N_2 \cdot P_{21} + N_3 \cdot P_{31} + ... + N_n \cdot P_{n1} = (0, N_2, N_3, ..., N_n) \cdot P(\cdot, 1)$, where $P(\cdot, 1)$ is the first column of matrix P. The (expected) outflow of dynasties from the first class is $N_1 \cdot (1 - P_{11})$ ($N_1 \cdot P_{11}$ stayed in same first class). The difference between the inflow and the outflow into class 1 is equal to

$$N_1 \cdot (P_{11} - 1) + (0, N_2, N_3, ..., N_n) \cdot P(\cdot, 1) = -N_1 + (N_1, N_2, ..., N_n) \cdot P(\cdot, 1).$$

Likewise, the difference between the inflow and the outflow for class i, $1 \leq i \leq n$, is $-N_i + (N_1, N_2, ..., N_n) \cdot P(\cdot, i)$.

$$\begin{aligned} imb_{vector} &= (-N_1 + (N_1, N_2, ..., N_n) \cdot P(\cdot, 1), ..., -N_n + (N_1, N_2, ..., N_n) \cdot P(\cdot, n)) = \\ &= (N_1, N_2, ..., N_n) \cdot P - (N_1, N_2, ..., N_n) = N_{t+1} - N_t, \end{aligned}$$

where N_t is the period-t distribution of the population over the classes. If we move from absolute numbers of dynasties to fractions of the total number of dynasties, we end up with $imb_{vector} = \pi_{t+1} - \pi_t$, where all imbalances are measured in fractions of the population.

One can also evaluate structural mobility at a more disaggregated level by comparing the observed "communication" between any two classes with the steady-state communication between same classes. For each pair of classes (i, j) one can compute the fraction of dynasties which move from i to j in the absence of structural mobility (i.e. in the steady state): $\pi_i^* \cdot P_{ij}$. And then, in presence of structural mobility, we can compare this same fraction computed as $(\pi_t)_i \cdot P_{ij}$ with pure

exchange fraction $\pi_i^* \cdot P_{ij}$. Thus, one can calculate the following imbalance matrix:

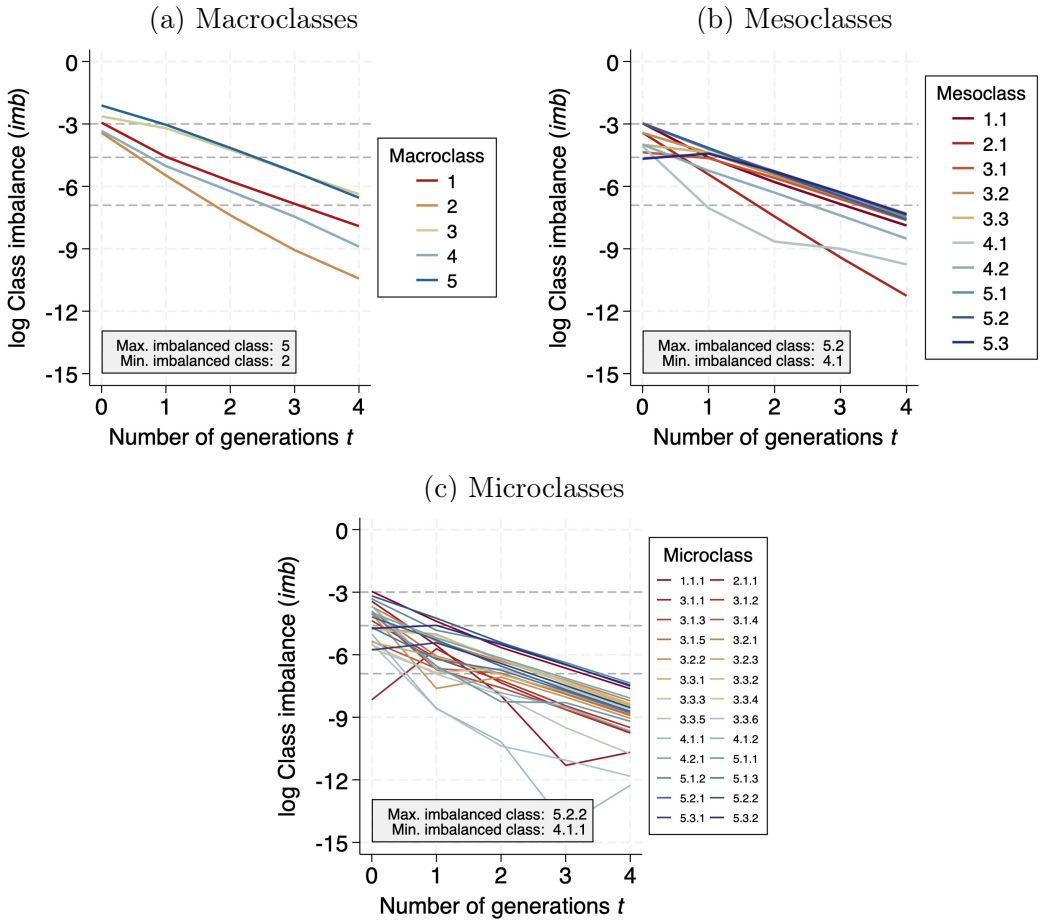
$$imb_{matrix} = \{((\pi_t)_i - \pi_i^*) \cdot P_{ij}\}_{1 \leq i,j \leq n}.$$

This "Imbalance" matrix may be a better object to focus on because it contains the maximal possible information. As a single number metric we may consider its norm.

B.2 Empirical Results

Illustrating Section B and disappearance of structural mobility as time evolves, we report imbalance vectors computed for different classes and different time steps (see Figure B.1) together with imbalance matrices for macro- and mesoclasses (see Tables B.1 and B.2). From Figure B.1, the maximum and the minimum average imbalances over the period from $t = 0$ to $t = 4$ are achieved from the following occupational classes: for macroclasses, professional-managerial (5) and primary manual (farming) occupations (2); for mesoclasses, managers and officials (5.2) and clerical (4.1); for microclasses, other private managers (5.2.2) and postal and mail distribution (4.1.1). By the fourth generation, all such imbalances essentially disappear and we begin observing pure exchange mobility patterns.

Figure B.1: Measuring structural mobility through class-specific imbalances



Notes: This figure plots class-specific imbalances $imb(j)$ (defined in Appendix B) for all classes over generation t in log scale. The gray textbox reports classes j with the maximum and minimum average $imb(j)$ for $t \geq 1$. Dashed horizontal lines correspond to log(0.05), log(0.01), and log(0.001). See Table 1 for definitions of macro- and mesoclass codes.

Table B.1: Imbalance matrices: Macroclasses

(a) $t = 0$

Parent class	Child class 1	2	3	4	5
1	0.008	0.000	0.028	0.005	0.011
2	0.001	0.004	0.019	0.003	0.005
3	0.005	0.001	0.038	0.009	0.020
4	−0.001	0.000	−0.013	−0.005	−0.016
5	−0.002	−0.001	−0.032	−0.018	−0.067

(b) $t = 1$

Parent class	Child class 1	2	3	4	5
1	0.002	0.000	0.006	0.001	0.002
2	0.000	0.001	0.002	0.000	0.001
3	0.003	0.000	0.022	0.005	0.011
4	−0.000	0.000	−0.002	−0.001	−0.003
5	−0.001	−0.000	−0.013	−0.007	−0.027

(c) $t = 2$

Parent class	Child class 1	2	3	4	5
1	0.000	0.000	0.002	0.000	0.001
2	0.000	0.000	0.000	0.000	0.000
3	0.001	0.000	0.008	0.002	0.004
4	−0.000	0.000	−0.001	−0.000	−0.001
5	−0.000	−0.000	−0.004	−0.002	−0.009

(d) $t = 3$

Parent class	Child class 1	2	3	4	5
1	0.000	0.000	0.001	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.003	0.001	0.001
4	−0.000	0.000	−0.000	−0.000	−0.000
5	−0.000	−0.000	−0.001	−0.001	−0.003

(e) $t = 4$

Parent class	Child class 1	2	3	4	5
1	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.001	0.000	0.000
4	−0.000	0.000	−0.000	−0.000	−0.000
5	−0.000	−0.000	−0.000	−0.000	−0.001

Notes: This table reports imbalance matrices, defined as $\{(\pi_{t,i} - \pi_i^*) \cdot P_{ij}\}_{1 \leq i,j \leq n}$, by generation t. See Table 1 for definitions of class codes.

Table B.2: Imbalance matrices: Mesoclasses

(a) $t = 0$

Parent class	Child class 1.1	2.1	3.1	3.2	3.3	4.1	4.2	5.1	5.2	5.3
0	0.008	0.000	0.008	0.009	0.010	0.003	0.002	0.004	0.006	0.001
2.1	0.001	0.004	0.003	0.005	0.011	0.002	0.001	0.001	0.002	0.002
3.1	0.001	0.000	0.002	0.002	0.003	0.001	0.001	0.001	0.002	0.001
3.2	0.003	0.000	0.005	0.005	0.007	0.002	0.001	0.003	0.005	0.001
3.3	0.001	0.000	0.002	0.002	0.006	0.001	0.001	0.002	0.003	0.001
4.1	−0.000	0.000	−0.002	−0.002	−0.004	−0.002	−0.001	−0.002	−0.004	−0.001
4.2	−0.001	0.000	−0.001	−0.002	−0.003	−0.001	−0.002	−0.004	−0.003	−0.002
5.1	−0.002	0.000	−0.004	−0.003	−0.009	−0.003	−0.003	−0.010	−0.012	−0.006
5.2	−0.001	−0.000	−0.004	−0.005	−0.008	−0.004	−0.005	−0.008	−0.011	−0.006
5.3	0.000	−0.000	−0.000	−0.000	−0.001	−0.001	−0.001	−0.002	−0.002	−0.002

(b) $t = 1$

Parent class	Child class 1.1	2.1	3.1	3.2	3.3	4.1	4.2	5.1	5.2	5.3
1.1	0.002	0.000	0.002	0.002	0.002	0.001	0.000	0.001	0.001	0.000
2.1	0.000	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000
3.1	0.001	0.000	0.002	0.001	0.002	0.001	0.000	0.001	0.001	0.000
3.2	0.001	0.000	0.002	0.002	0.003	0.001	0.000	0.001	0.002	0.000
3.3	0.001	0.000	0.001	0.002	0.004	0.001	0.001	0.001	0.002	0.000
4.1	−0.000	0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
4.2	−0.000	0.000	−0.000	−0.000	−0.001	−0.000	−0.000	−0.001	−0.001	−0.001
5.1	−0.001	0.000	−0.001	−0.001	−0.003	−0.001	−0.001	−0.003	−0.003	−0.002
5.2	−0.000	−0.000	−0.001	−0.002	−0.002	−0.001	−0.002	−0.002	−0.003	−0.002
5.3	0.000	−0.000	−0.000	−0.000	−0.001	−0.001	−0.001	−0.003	−0.003	−0.003

(c) $t = 2$

Parent class	Child class 1.1	2.1	3.1	3.2	3.3	4.1	4.2	5.1	5.2	5.3
1.1	0.000	0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000
2.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.1	0.000	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.000	0.000
3.2	0.000	0.000	0.001	0.001	0.001	0.000	0.000	0.000	0.001	0.000
3.3	0.000	0.000	0.001	0.001	0.002	0.000	0.000	0.000	0.001	0.000
4.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.2	−0.000	0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
5.1	−0.000	0.000	−0.000	−0.000	−0.001	−0.000	−0.000	−0.001	−0.001	−0.001
5.2	−0.000	−0.000	−0.000	−0.000	−0.001	−0.000	−0.000	−0.001	−0.001	−0.001
5.3	0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.001	−0.001	−0.001

(d) $t = 3$

Parent class	Child class 1.1	2.1	3.1	3.2	3.3	4.1	4.2	5.1	5.2	5.3
1.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.3	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000
4.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.2	−0.000	0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
5.1	−0.000	0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
5.2	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
5.3	0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000

(e) $t = 4$

Parent class	Child class 1.1	2.1	3.1	3.2	3.3	4.1	4.2	5.1	5.2	5.3
1.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3.3	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4.2	−0.000	0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
5.1	−0.000	0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
5.2	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000
5.3	0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000	−0.000

Notes: This table reports imbalance matrices, defined as $\{(\pi_{t,i} - \pi_i^*) \cdot P_{ij}\}_{1 \leq i,j \leq n}$, by generation t. See Table 1 for definitions of class codes.

C Data Appendix

C.1 Constructing the Sample

Table C.1 summarizes the sampling scheme used to construct our main samples. Steps 1 and 2 are commonly taken by the mobility literature that uses the NLSY79 sample in order to observe both parent and child income or occupations during adulthood. Step 3 ensures that at least one parental figure (birth or step-parent) is present in household to ensure consistent definitions of parent and child occupations.

Steps 4 and 5 drop individuals with missing occupation data to retrieve our final occupation samples. More than 5,000 observations are dropped from our occupation sample, namely since parent occupations are observed for only one or two years in the data.

C.2 Constructing Occupation Variables

Parent occupation: We use the qualifying father's and mother's occupation reported in 1978 and retrospectively when the child is age 14.

- We first adjust the father and mother's occupation separately using the available data. We prioritize the 1978 survey year variable over the retrospective age 14 variable. Next, we prioritize the father's occupation over the mother's occupation unless the father is not working or the father's occupation is missing. If neither the father or mother is working, we assign the parent occupation to "not working"; if no occupational data is observed, then it is coded as missing. "Not working" status is elicited from the NLSY family background questionnaire.

Table C.1: Overview of sampling scheme

Step	Description of step	N	% Base sample
0	Base sample	12,686	100.0
1	Drop economically-disadvantaged white and military samples	9,763	77.0
2	Drop observations without Household Interview Versions A or B	8,507	67.1
3	At least one parental figure in household, 1979–1986	8,122	64.0
Income Sample			
4	Drop missing parent income after age restrictions	4,767	37.6
5	Drop missing child income after age restrictions	2,839	22.4
Occupation Sample			
4	Drop missing parent occupation after age restrictions	4,470	35.2
5	Drop missing child occupation after age restrictions	2,752	21.7

- Parent occupations are based on the 1970 Census codes. We map these Census occupation codes to our adaptation of Jonsson et al.'s microclass codes.

Child occupation: We use the NLSY's constructed variable providing the occupation of the child or spouse/partner's current or most recent job.

- Occupations are reported since the calendar year *prior* to the survey year.
- The NLSY elicits occupations of the current or *most recent* spouse/partner. To remove the possibility of picking up occupations of previous (divorced or separated) spouses/partners, we only use this data whenever a spouse or partner is cohabiting in the household. Before 1994, occupations were elicited from *only* spouses; after 1994, occupations of partners were also elicited. Therefore, we only can use occupations of the child's spouse/partner before 1994 if the child is married.
- When the child has a spouse or cohabiting partner, we prioritize the male's occupation unless the male is not working or the male's occupation is missing. If neither the male or female is working, we assign the child occupation to "not working"; if no occupational data is observed, then it is coded as missing. "Not working" is defined as having worked 0 weeks in the previous calendar year, either due to unemployment or nonparticipation in the labor force.
- Before 2002, occupations are based on the 1970 Census codes; thereafter, they are based on the 2000 Census codes. We map these Census occupation codes to our adaptation of Jonsson et al.'s microclass codes.
- After assigning one harmonized occupation code for the child's household per year, we find the modal value across all survey years (subject to all age restrictions described in the main text).

D Technical Appendix

D.1 The Total Variation Norm

The total variation norm is coming from the Markov chains literature, and it has some very clear interpretation. If we have two distributions μ and ν over the same state space S (in our case it is the set of all classes), then the distance between them in the total variation (TV) norm is defined as $\| \mu - \nu \|_{TV} = \max_{A \in \Omega} |\mu(A) - \nu(A)|$, where Ω is the σ-algebra on S. The probabilistic interpretation is consistent with the lottery interpretation. In this case $\| \mu - \nu \|_{TV}$ is the maximum difference between the probabilities assigned to a single event by distributions μ and ν. Namely, we go over all events $A \in \Omega$, i.e. in our framework we should consider events of a single (randomly chosen) dynasty being in all classes and in all possible unions of those classes to find such class or a union of classes that would give the maximum difference between the probabilities of being in this set of classes "under" μ and ν. For example, $\| e_i P^t - e_j P^t \|_{TV}$ is the maximum difference between the probabilities of dynasties with ancestors in classes i and j being in some set of classes in period $t+1$. Alternatively, $\| \mu - \nu \|_{TV}$ may be computed as follows:

$$\| \mu - \nu \|_{TV} = \frac{1}{2} \sum_{1 \leq i \leq n} |\mu_i - \nu_i| = \frac{1}{2} \| \mu - \nu \|_1 = \sum_{i:\mu_i \geq \nu_i} |\mu_i - \nu_i| = \sum_{i:\nu_i \geq \mu_i} |\mu_i - \nu_i|,$$

where $\| \cdot \|_1$ denotes the l_1-norm. These expressions allow us to interpret the TV norm in marginal population interpretation. First we identify a common part of distributions μ and ν, i.e. distribution θ such that $\theta_i = \min\{\mu_i, \nu_i\}$ and then we sum up fractions of the population in μ or in ν (meaning that those fractions equal each other) which are over this common distribution θ. In other words, $\| \mu - \nu \|_{TV}$ is "overconcentration" of the population in μ or ν relative to the common part of μ and ν. Please see Levin and Peres (2017), pp. 47-48 for details.

D.2 Adjusting Transition Matrices with Small Probabilities

Let P be a $K \times K$ transition matrix where some cells $P_{ij} = 0$. Consider the problem

$$\arg\min_{Q \in \mathcal{P}} \| Q - P \|_{mat} \quad \text{s.t.} \quad Q_{ij} \geq \epsilon > 0; \quad \sum_{j=1}^{K} Q_{ij} = 1, \ \ \forall i = 1, \ldots, K, \tag{8}$$

where $\mathcal{P}$ is the set of well-defined transition matrices and the matrix norm operator $\| \cdot \|_{mat}$ is defined as

$$\| M \|_{mat} = \sum_{i=1}^{K} \sum_{j=1}^{K} |M_{ij}|. \tag{9}$$

Thus, the optimization problem can be rewritten as

$$\underset{\{Q_{ij}\,|\,i,j=1,\ldots,N\}}{\arg\min} \sum_{i=1}^{K}\sum_{j=1}^{K} |Q_{ij} - P_{ij}| \quad \text{s.t.} \quad Q_{ij} \geq \epsilon > 0; \quad \sum_{j=1}^{K} Q_{ij} = 1, \;\; \forall i = 1, \ldots, K. \tag{10}$$

Since the choice of Q_{ij} values are independent across rows i, this problem is equivalent to the composite of K separate optimization problems corresponding to each row: For each i,

$$\underset{\{Q_{ij}\,|\,j=1,\ldots,N\}}{\arg\min} \sum_{j=1}^{K} |Q_{ij} - P_{ij}| \quad \text{s.t.} \quad Q_{ij} \geq \epsilon > 0; \quad \sum_{j=1}^{K} Q_{ij} = 1. \tag{11}$$

There are three possible cases for a given row i:

1. $P_{ij} \geq \epsilon$, $\forall j = 1, \ldots, N$. Then the solution $\{Q^*_{ij}\,|\,j = 1, \ldots, N\}$ of (11) is $Q^*_{ij} = P_{ij}$.

2. Either $P_{ij} < \epsilon$ or $P_{ij} \geq \epsilon + \frac{M_i}{K - z_i}$), where $M_i = \sum_{j=1}^{K} (\epsilon - P_{ij}) \cdot \mathbb{1}\{P_{ij} < \epsilon\}$ (i.e., the cumulative perturbations required to satisfy only the minimum constraint) and $z_i = \sum_{j=1}^{K} \mathbb{1}\{P_{ij} < \epsilon\}$ (i.e., the number of elements smaller than ϵ in the ith row of P). Define $\Delta_{ij} = Q_{ij} - P_{ij}$. Note that, for values j such that $P_{ij} < \epsilon$, we must have $\Delta_{ij} \geq \epsilon - P_{ij}$. Without the adding-up constraint, the optimal solution would simply set $Q_{ij} = \epsilon$ whenever $P_{ij} < \epsilon$, which implies $\sum_{j=1}^{K} \Delta_{ij} = \sum_{j=1}^{K} |\Delta_{ij}| = \sum_{j=1}^{K} (\epsilon - P_{ij}) \cdot \mathbb{1}\{P_{ij} < \epsilon\} = M_i$. However, with the adding-up constraint, we must satisfy $\sum_{j=1}^{K} \Delta_{ij} = 0$, which immediately implies that minimum vector norm value is $\sum_{j=1}^{K} |\Delta_{ij}| = 2M_i$. Hence, any solution $\{Q^*_{ij}\,|\,j = 1, \ldots, N\}$ that attains this minimum value is an optimal solution of (11). One such solution[45] is $Q^*_{ij} = \epsilon$ for j s.t. $P_{ij} < \epsilon$ and $Q^*_{ij} = P_{ij} - \frac{M_i}{K - z_i}$ for j s.t. $P_{ij} \geq \epsilon + \frac{M_i}{K - z_i}$. Note that, in the latter case, $Q^*_{ij} \geq \epsilon + \frac{M_i}{K - z_i} - \frac{M_i}{K - z_i} = \epsilon$.

3. $P_{ij} < \epsilon$ and $P_{ik} \in (\epsilon,\, \epsilon + \frac{M_i}{K - z_i})$ for some $j, k = 1, \ldots, K$, where M_i, z_i are defined as before. The solution concept is similar to before, except now the minimum constraint $Q_{ij} \geq \epsilon$ would bind for elements j s.t. $P_{ij} \in (\epsilon,\, \epsilon + \frac{M_i}{K - z_i})$. One possible solution is the following algorithm (in what follows, suppress the index i to reduce notational burden):

   (a) Set all $q^*_j = \epsilon$ whenever $p_j < \epsilon$. Initialize $k = 0$.

   (b) Define $p_{(j)}$ as the jth order statistic (i.e., $p_{(1)} \leq \cdots \leq p_{(K)}$); note that $p_{(j)} > \epsilon$ for $j \geq z$ and $p_{(z+1)} \in (\epsilon, \epsilon + \frac{M_i}{K - z})$. Set $q^*_j = \epsilon$ s.t. $p_j = p_{(z+1)}$ and define $n = |\{j\,|\,p_j = p_{(z+1)}\}|$. Set k to $k + n$.

45. Note that this proposed solution is *not unique*. Using the L^1-norm will always leads to a family of possible solutions; the solution Q^* defined above provides one method for preventing large changes occurring for any given element p_{ij} by attempting to undo the perturbations evenly over all remaining non-zero elements of P. Using alternative norms may lead to unique solutions; we employ the L^1-norm for the sake of consistency.

(c) Define $\delta_{(z+k)} = n \cdot (p_{(z+k)} - \epsilon)$. Then the total magnitude of perturbations that remain to be distributed across the remaining row elements to satisfy the adding-up constraint is $S_{(z+k)} = M_i - \delta_{(z+k)}$. Check whether $p_{(z+k+1)} \geq \epsilon + \frac{S_{(z+k)}}{K-z-k}$.

- If true, then define $q_j^* = p_j - \frac{S_{(z+k)}}{K-z-k}$ for all remaining entries of the row vector.
- If false, then proceed to Step (b) except set $M_i = S_{z+k}$.

The solutions recovered from each row can be combined to attain the final solution Q^*.