NBER WORKING PAPER SERIES

THE EVOLUTION OF BLACK-WHITE DIFFERENCES IN OCCUPATIONAL MOBILITY ACROSS POST-CIVIL WAR AMERICA

Steven N. Durlauf
Gueyon Kim
Dohyeon Lee
Xi Song

Working Paper 32370
http://www.nber.org/papers/w32370

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

We thank participants at the Conference on Inequality and Mobility 2023 and SSHA 2023 for their comments and feedback. This paper has greatly benefited from discussions with Kristina Butaeva, Yoosoon Chang, Jonas Helgertz, Aleksandra Lukina, Joon Park, and Zachary Ward. We thank Ken Suzuki and Jiaowei Gong for their excellent research assistance. Gueyon Kim is grateful for the financial support of the Hellman Fellowship and the UC Santa Cruz COR grant. All authors acknowledge generous support of the Stone Center for Research on Wealth Inequality and Mobility at the University of Chicago. Dohyeon Lee's contribution was prior to joining Amazon and does not represent Amazon's views. James M. and Cathleen D. Stone Foundation are the funders of the Stone Center for Wealth Inequality and Mobility at the University of Chicago, which is the primary source of support for this project. Gueyon Kim also acknowledges financial support of the Hellman Fellowship. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Evolution of Black-White Differences in Occupational Mobility Across Post-Civil War America
Steven N. Durlauf, Gueyon Kim, Dohyeon Lee, and Xi Song
NBER Working Paper No. 32370
April 2024
JEL No. J15,J62,N31,N32

ABSTRACT

This paper studies long-run differences in intergenerational occupational mobility between Black and White Americans. Combining data from linked historical censuses and contemporary large-scale surveys, we provide a comprehensive set of mobility measures based on Markov chains that trace the short-and long-run dynamics of occupational differences. Our findings highlight the unique importance of changes in mobility experienced by the 1940–1950 birth cohort in shaping the current occupational distribution and reducing the racial occupational gap. We further explore the properties of continuing occupational inequalities and argue that these disparities are better understood by a lack of exchange mobility rather than structural mobility. Thus, contemporary occupational disparities cannot be expected to disappear based on the occupational dynamics seen historically.

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

Gueyon Kim
University of California, Santa Cruz
1156 High Street
Santa Cruz, CA 95064
gkim44@ucsc.edu

Dohyeon Lee
Amazon
410 Terry Ave N
Seattle, WA 98109
dohnlee@gmail.com

Xi Song
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104
xisong@upenn.edu

1 Introduction

Disparities in intergenerational mobility between Black and White individuals constitute one of the fundamental dimensions of racial inequality in the United States. Recent research on these disparities has brought attention to how Black-White mobility gaps have evolved over time (Chetty et al., 2020; Collins and Wanamaker, 2022; Jácome et al., 2021). These studies primarily focus on income mobility, measured by individual- or occupation-level income of parents and their offspring. We revisit this growing literature from the perspective of occupational mobility.

While related, occupations and income are inherently different measures of socioeconomic status. Income mobility measures changes in earnings from parents to offspring; occupational mobility focuses on changes in job type, prestige, status, and career path across generations. Importantly, occupational mobility is particularly sensitive to shifts in societal and economic structures. For example, although industrialization may not have drastically altered income percentile rankings, it significantly changed the occupational compositions of farming versus manufacturing work (Long and Ferrie, 2013; Xie and Killewald, 2013). Therefore, we regard the study of occupational mobility as an important complement to studies of income mobility.

This paper examines the long-term trends in racial disparities in *occupational* mobility through a comprehensive set of measurements based on Markov chain models and a vast array of data from linked historical censuses and contemporary large-scale survey data. First, we examine historical patterns in intergenerational occupational mobility between Black and White Americans from the 1860 to the 1970 birth cohorts. Second, we quantify the importance of these mobility dynamics for explaining the observed changes in occupational distribution over time, using the Kitagawa-Oaxaca-Blinder decomposition method and counterfactual experiments. Finally, we assess the persistence of racial inequality using measurements of "memory" embedded in the mobility process and explore the importance of structural changes in occupations in explaining the evolving racial gaps.

Sociologists have long modeled mobility using discrete-time Markov chain models (Prais, 1955; Blumen et al., 1955; Matras, 1975). Building on this line of work, we propose new ways of characterizing mobility using Markov models with time-varying transition matrices. Our stochastic analyses offer several new insights into the existing regression-based analyses. First, we consider both racial differences in the initial occupational distribution and the evolution of intergenerational mobility in jointly shaping racial gaps in mobility outcomes. Leveraging our extensive dataset, we construct six distinct period-specific (or birth-cohort-specific) transition matrices that reflect the changing socioeconomic structures of various periods and regimes, such as the Jim Crow era and the civil rights movement. We then assess which period contributes the most to the narrowing occupational racial gap due to changes in their respective transition processes.

Second, we examine the persistence of racial inequality using measurements of "memory" embedded in the mobility process. These memory measures explore the extent to which parental occupations affect the occupational outcomes of subsequent generations. We assess both the magnitude of this influence across generations and the speed at which it dissipates. In particular, for each birth cohort, we posit that each family, given the parent occupation, converges to a steady-state distribution—a state in which advantages or disadvantages stemming from the initial condition or the transition probabilities disappear. Analyzing how Black and White populations progress toward their own steady states helps elucidate mechanisms through which racial gap in occupational attainments persists.

Third, we introduce a distinction between structural mobility, which reflects a society's evolving occupational demands, and exchange mobility, which stems from individual effort, qualifications, and merit. We investigate the extent to which the observed mobility is attributable to mobility induced by occupational restructuring across generations versus the mobility resulting from status exchange among families that leaves the overall distribution of occupations unchanged. Unlike previous studies using log-linear models to analyze structural and exchange mobility in a mobility table (Goodman, 1968; Hout, 1983), we evaluate the evolution of these two components in a Markov mobility process, assuming the absence of any interventions.

Our results indicate the essential contribution made by the 1940-1950 birth cohort in reducing the racial gap in mobility. In particular, Black men from this birth cohort show significant progress in increasing (decreasing) the probability of upward (downward) mobility and mitigating mobility bottlenecks to gain access to high-skill occupational positions, such as "Proprietors, Managers, and Officials," from families of low-skilled occupations, such as "Semi-Skilled and Unskilled" occupations. Our decomposition exercise further reveals that such improvements in the mobility process largely explain the observed changes in occupational distributions, which exhibit lower degrees of occupational segregation by race. We provide complementary evidence using our counterfactual exercises that the current occupational distribution, characterized by a notable rise in high-skill occupations compared to historical trends, can predominantly be attributed to the mobility experienced by the 1940-1950 birth cohort.

Despite important improvements, racial gaps persist in cross-sectional occupational distributions and intergenerational mobility matrices. Across all measurements, we find evidence that the momentum achieved by the 1940-1950 birth cohort is not maintained in following generations. Specifically, the rate at which mobility patterns improve or occupational segregation declines levels off for subsequent birth cohorts. In fact, as we simulate the mobility process of Black Americans to be the same as Whites (i.e., either they experience the same mobility dynamics or start out at the same initial conditions), we obtain counterfactual distributions that show greater (lower) shares of high-skill (low-skill) occupations than the observed actual distribution, with lower degrees of occupational segregation by race.

Further exploring the properties of the remaining Black-White inequalities, we find that occupational mobility for Black men, compared to that of White men, shows greater memory or greater difficulty moving away from the current conditions toward a steady state—which indicates a more egalitarian distribution with a lower degree of occupational segregation by race compared to the actual—and such memory dissipates at a slower rate. When examining individual memory, Black families positioned at the top and bottom of the occupational hierarchy exhibit stronger memory—namely, a higher tendency to maintain their occupational status over time—than families in the middle. These results suggest that racial disparities in intergenerational mobility vary among occupations and may persist for varying lengths of time. Any racial differences in the structural determinants of occupations between races (e.g., preference parameters and access to information networks) can produce differential memory. Although our results do not point to specific mechanisms, they help constrain what types of heterogeneities are present.

For both groups, the importance of structural factors in explaining mobility patterns is quite limited. Instead, mobility rates are largely explained by exchange mobility. When examining racial differences for the 1940-1950 birth cohort, structural mobility is relatively high for Black men compared to White men, while exchange mobility is low . This finding aligns with the significant impact of the civil rights movement on opportunities available to Black Americans, while also highlighting the importance of anti-discrimination policies in addressing the unequal opportunities that Black Americans encounter in achieving occupational statuses through personal effort and qualifications.

Literature and Contribution

Our work builds on a growing body of literature on historical trends in mobility. The availability of digitized census data and innovative methods for linking censuses across time periods have significantly expanded the scope and scale of research in this area (Abramitzky et al., 2022; Helgertz et al., 2022). Studies by Song et al. (2020) and Collins and Wanamaker (2022) have leveraged these linked censuses with other survey data to assess patterns of mobility over long time horizons. Jácome et al. (2021) and Buckles et al. (2023) delve deeper into the differences of mobility patterns across gender and racial groups. Important contributions also include efforts to improve accuracy of linking methods (Bailey et al., 2020; Olivetti and Paserman, 2015) and measurement of occupations to help address biases in mobility estimates (Ward, 2023).

In complementing these studies, our work constructs a comprehensive dataset combining both the historical censuses from 1850 to 1940 and 10 contemporary surveys from 1962 to 2021 to study the mobility dynamics of post-Civil War America. To construct father-son pairs using the historical data, we reduce false links across census by closely following existing work, test robustness to different linking methods, and apply family reconstitution techniques to reconstruct family lineages.[1] Further combining survey data, which we make extensive adjustments across data sources (weights, variable definitions), we create a sample spanning 170 years with consistent measures of occupations for father-son pairs.

Our work also produces a range of empirical conclusions that augment existing studies on the Black-White gap in mobility. Previous work reveals that intergenerational transmission of income is weaker and more heterogeneous among Black Americans due to lower upward mobility and higher downward mobility compared to White Americans (Bhattacharya and Mazumder, 2011; Mazumder, 2014; Bloome, 2014, 2017; Chetty et al., 2020). More recently, Collins and Wanamaker (2022) examine these long-term mobility disparities, highlighting their significance in explaining the persistence of Black-White inequality.

Our analyses using Markov Chains importantly complement these income-based approaches. By tracing out the evolution of the mobility dynamics over a long time horizon, which extends and enriches earlier occupation-based studies (Blau and Duncan, 1967; Duncan, 1968; Featherman and Hauser, 1976; Hout, 1984), we not only demonstrate the importance of intergenerational transitions in explaining changes in occupational distributions but also identify periods in which the contribution of the mobility channel is more pronounced and show how the patterns and magnitudes differ between Black and White Americans. Our results show that the evolution of both the transitions and the occupational distributions for White men shows a steady and gradual progression, while that of Black men demonstrates a rather dramatic change, particularly for the 1940-1950 birth cohort. Their mobility experiences contribute the most to (i) lowering the mobility barriers from low-status families to high-status occupations, (ii) achieving the current occupational distribution with higher shares of high-status occupations relative to the past, and (iii) reducing the degree of occupational segregation by race.

Finally, our work sheds light on the role of intergenerational mobility in the persistence

[1] See (Willigan and Lynch, 1982; Hammel, 1993; Wrigley et al., 1997) for details.

or lack of convergence in Black-White inequality since the 1960s. Recent studies suggest a stable trend (Lee and Solon, 2009) or no trend in mobility patterns for the whole population of recent cohorts (Chetty et al., 2014; Jácome et al., 2021; Ward, 2023). Black-White convergence in mobility gaps has also stalled for birth cohorts since the 1960s (Bound and Freeman, 1992; Collins and Wanamaker, 2022). Our measurement of *memory* offers a novel perspective on understanding this persistence in racial inequality by focusing on the unequal opportunities *embedded* in the mobility process. Our findings show the mobility process of Black Americans having greater difficulty moving away from their parental conditions—both in terms of the magnitude of parental influences on offspring's mobility opportunities and the rate at which it dissipates—compared to their White counterparts. Further distinguishing between structural versus exchange mobility, we show that an important mechanism through which these racial gaps persist is their differences in exchange mobility.

The rest of the paper is organized as follows. Section 2 discusses the data and sample construction. Section 3 investigates the evolution of Black-White differences in occupational distributions and patterns of and barriers to mobility. Section 4 studies the importance of the mobility process in explaining the observed evolution of occupational distributions. Section 5 examines the persistence of memory of parental occupations in mobility and explores the importance of structural versus exchange mobility. Section 6 concludes.

2 Data and Sample Construction

Our analyses of the mobility experiences of several million unique individuals and their families rely on two main sources of data: (i) historical samples from linked complete count censuses covering the periods of 1850-1940 and (ii) contemporary data from 10 large-scale social surveys that cover periods following the 1960s. Appendix C includes details on the data and variables used in our analyses.

2.1 Historical Data, 1850–1940

We first draw on newly released linkages between full-count US Censuses from 1850 to 1940 in the Census Linking Project (CLP version 2.0) (Abramitzky et al., 2022). These linkage data are then merged with age, occupation, race, and household member information in full-count US Censuses 1850–1940 (Ruggles et al., 2020). The CLP linking method relies on several different automated algorithms to create matched individual records across census years based on variables such as first and last names, year of birth, and state/country of birth. Abramitzky et al. (2020) and Abramitzky et al. (2021) discuss the linking methods used for the construction of CLP in detail.

We focus on the decadal links between 1850-1860, 1860-1870, 1870-1880, 1880-1900, 1900-1910, 1910-1920, 1920-1930, and 1930-1940.[2] We implement links based on "Exact" and "Conservative" matching methods, both of which leverage the ABE algorithm with exact names. The "Exact" linking approach, as opposed to the NYSIIS, reduces false positives (Bailey et al., 2020). The "Conservative" links require individuals to be unique within a 5-year age window, which further reduces errors in matching individuals across different census years. Given our focus on race, we use the link that relies on race as a matching variable. We further merge the CLP data with other variables in the full-count censuses such as age, gender, race, and occupation.

As CLP links individuals' records between censuses but not across generations, we reconstruct family lineages of Black and White Americans by linking family members living in the same households and retrieving records of parents across different census years. This method, often known as the family reconstitution technique, is widely used in historical demography to recreate family histories and events with registry data (Willigan and Lynch, 1982; Hammel, 1993; Wrigley et al., 1997). We construct father-son pairs as follows: we (i) form a panel dataset of individuals by linking their records across different census years; (ii) pinpoint the year in which an individual male resided in the same household as his father and extract the father's historical identification; and (iii) connect the father to different census years to assemble a longitudinal dataset for the father.

Appendix Figure C demonstrates four scenarios that help identify parent-offspring pairs. If a father resided with his son in any recorded year between 1850 and 1930 (e.g., 1920), we then trace the father's information to a later census, provided that his individual-level identifier from 1850 to 1930 (e.g., 1920 HISTID) can be found in our 1860 to 1940 multiyear dataset. This tracing method leads to one of two scenarios: the linked father no longer lived with the son in a later census (scenario A) or the linked father continued to live with the son in a later census (scenario B). If the father lived with the son in any year between 1850 and 1930 but is not linked in our 1860 to 1940 multiyear dataset, we keep the father's occupation from 1850 to 1930 (scenario C). In addition, we include male offspring who lived with their fathers in the same household in childhood but whose fathers cannot be found in a later census (scenario D).

As a robustness check, we also construct an alternative sample using linkages of full-count US Censuses from 1850 to 1940 in the IPUMS Multigenerational Longitudinal Panel (MLP version 1.0) (Helgertz et al., 2022). The MLP linking method relies on a probabilistic algorithm based on individuals' sex, place of birth, birth year, first and last names, and

[2]The 1890 Census data are missing because the original files were lost due to a fire.

household information to find potential matches across census years. Unlike the CLP, the MLP linking technique produces a single linked record for each individual. Helgertz et al. (2022) describes the linking method in detail. We report results using this alternative sample in Section 3.2.1.

2.2 Contemporary Data, 1962-2021

Our analysis of contemporary periods relies on samples from 10 different large-scale social surveys, including General Social Survey (1972–2021), Occupational Changes in a Generation (1962, 1973), National Survey of Families and Households (1987-1988, 1992-1994, 2001-2003), Panel Study of Income Dynamics (1968-2019), National Longitudinal Study of Youth (1979-2012), National Longitudinal Study–Young Men (1966–1981), National Longitudinal Study–Older Men (1966–1990), Survey of Income and Program Participation (1986-1988), Americans View Their Mental Health (1957, 1976), and National Survey of Black Americans (1979-1980, 1987-1988, 1988-1989, 1992). In each survey, we construct age, gender, race, and occupation variables for sons and their fathers. See Section C for a detailed description of each data source.

2.3 Sample Construction

After combining the linked historical censuses and survey data, we restrict the sample to Black and White men born between 1856 and 1975 and aged between 35 and 55 in six snapshot years 1910, 1930, 1950, 1970, 1990, and 2010.[3] Individuals are grouped into six birth cohorts (1856-1875, 1876-1895, 1896-1915, 1916-1935, 1936-1955, and 1956-1975), which are labeled in our analysis by midyears (1860-1870, 1880-1890, 1900-1910, 1920-1930, 1940-1950, and 1960-1970). Birth cohorts from 1860 to 1910 are drawn from historical censuses, whereas cohorts between 1920 and 1970 are obtained from surveys. As the survey data differ by weighting scheme, we create a standardized weight variable by setting the average to one while preserving the variance of the original weights. We do not apply any adjustment to the census data, as these full-count censuses are self-weighted. Finally, we reweight the sample to match the population distributions (by birth cohort, race, and 7-category occupations) using the Decennial Censuses (from 1870 to 2000, except for 1890) and the American Community Survey (ACS, 2010).[4]

In addition to the alternative sample constructed using MLP linkages for the historical censuses, as discussed in Section 2.1, we construct another version to test the sensitivity of our analysis with respect to changes in the data sources from the linked censuses to these

[3]See Table B.2 for the mapping between birth cohorts and snapshot periods.
[4]See Section 2.4 and Table B.1 for details on the occupation categories.

survey data. As the coverage for the 1900-1910 birth cohort overlaps between the two data sources, we examine how the mobility estimates change using the alternate sample based on survey data instead of the censuses for this birth cohort. We find that the results remain qualitatively unchanged (see Section 3.2 for a detailed discussion).

2.4 Measurement of Occupations

For the linked historical censuses, whenever feasible, we maintain multiple observations of occupations in the linked sample and select the modal occupation observed when individuals are aged between 25 and 55. The modal value helps remove measurement errors associated with a snapshot measure of occupations.[5] When no modal value exists, we use the occupation measured in the year closest to age 40 (Haider and Solon, 2006).

We construct occupation variables in survey data in two different ways. In longitudinal studies, in which both offspring and parents are asked to report their main occupations in each survey year, we choose the occupation measured in the year closest to age 40 as the lifetime occupation. In cross-sectional surveys where fathers' information is often not directly observed, we rely on the retrospective information of fathers during the respondents' childhood or adolescence reported by the respondents.[6] Father's information provided in prospective and retrospective surveys may not always match. However, according to the analysis of Jácome et al. (2021), nationally representative longitudinal surveys such as the PSID demonstrate a high consistency rate—over 80%—between the occupations reported by individuals and those their offspring recall.

To construct broad occupational categories in historical censuses, we rely on the 1950 IPUMS harmonized three-digit census occupation codes. For other surveys, with the exception of National Survey of Black Americans, we convert the three-digit census codes for different years to the 1950 census scheme.[7] We then collapse the three-digit occupations into seven broad groups: Professional and Semi-professionals; Proprietors, Managers, and Officials; Clerical; Sales; Semi-Skilled and Unskilled; Craftsmen and Government Services; and Farming).[8] Given the limited number of Non-Manual and Manual occupations before the 20th century, particularly among Black individuals, we reorganize the occupation data into either three broad categories (Non-Manual, Manual, and Farming) or into five distinct categories (Professional and Managers; Clerical and Sales; Semi-Skilled and Unskilled;

[5] Ward (2023) shows that such errors in linked censuses may lead to biased mobility estimates.

[6] If the occupation information was reported multiple times, we choose the mode of the data values. Otherwise, we use the occupations of fathers when the respondents were 15 or 16 years old.

[7] In the National Survey of Black Americans, parental occupation information is only available in 22 occupation categories, which we map to our categorization of occupations as we discuss below.

[8] See Table B.1 for detailed mapping of three-digit Census occupation codes to our categorization of occupations.

Craftsmen and Government Services; and Farming) for various analyses. From a practical perspective, Markov chain models may yield unstable estimates (non-ergodicity) when some cells in the transition matrix have only a few observations or zero entries. Thus, our main exercises rely on these broad definitions of occupational categories.

Several recent studies discussed the importance of constructing separate income measures by race (Collins and Wanamaker, 2022; Jácome et al., 2021). Following this line of work, we develop separate transition matrices by race to explore racial disparities in mobility opportunities (see Section 3 for more details). One may further consider incorporating information on region by separately constructing these matrices by region and race. Given that our method for assessing racial mobility relies on transition matrices in Markov chains, which are sensitive to cells with zero observations, such disaggregation may generate additional measurement issues. The current scope of our work does not focus on understanding differences in the observed occupation categories between Black and White Americans nor on their implications for mobility patterns. Rather, we focus on the type of job characterized by a set of tasks and skills, and therefore, our results mainly reflect the "between-occupation" component of the Black-White gap in mobility rates given an observed occupation category.

3 Black-White Differences in Occupational Distributions and Mobility

We begin by examining the evolution of occupational distributions for Black and White men and changing occupational mobility between racial groups using Markov chains for the 1860 through 1970 birth cohorts. Our descriptive analyses rely on traditional measures of occupational segregation indices and transition probabilities, as well as new measures on mobility bottlenecks.

3.1 Occupational Segregation by Race and the Historical Context

The occupational distributions for both Black and White Americans change dramatically between the 1910 and 1970 birth cohorts, driven by a movement of individuals from "Farming" and "Manual" occupations to "Non-Manual" occupations (Figure A.1, top right). Differences in occupational distribution between the two groups reveal important dynamics. Using the segregation index (Duncan and Duncan, 1955), Figure A.2 (in solid navy) shows the dissimilarity in the occupational distribution between the two groups increases between the 1860 and 1910 birth cohorts and declines afterward.[9]

[9]We report results in terms of birth cohorts, which is equivalent to using periods when these individuals were prime-age workers. We show results using our definition of occupations in 3, 5, and 7 categories, which

These trends are consistent with the historical characterization of racial inequality during the late 19th and early 20th centuries. Despite the abolition of slavery by the Emancipation Proclamation in 1863 and the 13th Amendment in 1865, Black Americans continued to endure economic, social, and political hardships that impeded social mobility in the following decades (Black et al., 2015; Collins and Wanamaker, 2014; Derenoncourt, 2022; Logan and Parman, 2017; Margo, 2016).

Occupational segregation between Black and White individuals began to show a decline for the 1920-1930 birth cohort. This decline is further accelerated for the 1940-1950 birth cohort, which can be attributed to pivotal historical events in the mid-twentieth century. The active involvement of Black Americans in World War II ignited a spirit of activism, with many "unwilling to accept the pre-war structure of racial dominance" that characterized the nation (Modell et al., 1989). Civil rights legislation, such as the Civil Rights Act of 1964, which banned discrimination and segregation, and the Voting Rights Act of 1965, which eliminated barriers to Black Americans voting in the South, marked a crucial turning point in the fight against racial discrimination (Donohue and Heckman, 1991; Wright, 2013).

The most recent birth cohort (1960-1970) shows the lowest level of occupational segregation over the entire sample period. The rate of decline over time, however, is slower than that of the previous cohorts.[10] This finding is consistent with prior work that suggests the lack of convergence of Black-White socioeconomic status since the 1960s (Bound and Freeman, 1992). While civil rights legislation represents significant strides in addressing racial disparities, it did not eradicate all forms of discrimination and inequality.

3.2 Occupational Transitions and Barriers in Occupational Mobility

Next, we explore changes in mobility using measures based on Markov chains—the outcome state of the present generation only depends on that of the parent generation but not any of the other preceding generations. Assuming that conditional probabilities linking parents and children are fixed over time and that the states in the transition matrix ($\mathbf{P}$) communicate over time, given an initial condition (μ_0), there exists a unique long-run equilibrium (μ^*) for the distribution of descendants: $\lim_{t\to\infty} \mu_0 \mathbf{P}^t = \mu^*$.[11] To study the evolution of Black-White mobility gap over a long time horizon, we incorporate time-varying transition matrices into standard Markov chain models. Specifically, we estimate regime-specific transition matrices ($\mathbf{P}_t$) to represent different historical regimes, which are used in our analyses in this Section as well as Sections 4 and 5.

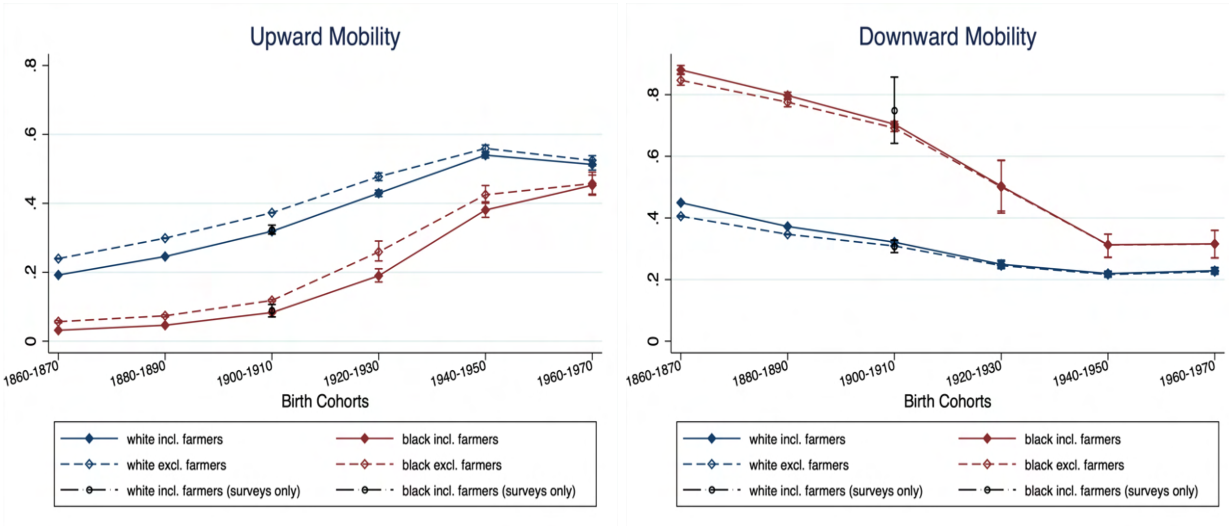
yield qualitatively similar trends in occupational segregation.

[10] As shown in Figure A.2, the trend is more pronounced examining the segregation indices calculated using 5- and 7-category definitions.

[11] See Section D for a detailed description related to Markov chain models.

3.2.1 Trends in Mobility: Period-Specific Transition Probabilities

Figure 1: Upward and Downward Mobility Patterns



Notes: The graphs show upward (left) and downward (right) mobility patterns for White (in navy lines) and Black men (in maroon lines) using 3 occupational categories. The solid graphs with solid diamond symbols show trends in transitions from Farming or Manual to Non-Manual (left) and Non-Manual to Farming or Manual (right). The dashed graphs with hollow diamond symbols show trends in transitions from Manual to Non-Manual (left) and Non-Manual to Manual (right). The long-dashed graphs with hollow circles use survey data and show trends in transitions from Farming or Manual to Non-Manual (left) and Non-Manual to Farming or Manual (right). Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure 1 shows trends in mobility patterns for White (navy lines) and Black men (maroon lines) using period-specific transition matrices.[12] Solid lines in the plots present upward mobility from "Farming" or "Manual" to "Non-Manual" (Figure 1, left) and downward mobility from "Non-Manual" to "Farming" or "Manual" (Figure 1, right). The trends show increasing upward mobility and decreasing downward mobility for both Black and White individuals over time.

Consistent with Duncan et al. (1972), the levels of upward (downward) mobility are greater (smaller) for White than Black men across all birth cohorts. Figure A.3 further shows the log differences in transition probabilities between the two groups, $log(p_{black}) - log(p_{white})$, which captures the mobility patterns of Black men relative to White men. Relative upward mobility rates continue to increase with a notable change from the 1920-1930 and 1940-1950 birth cohorts. By contrast, relative downward mobility increases up to the 1900-1910 birth cohort, significantly declines in the subsequent cohorts, and levels off for the 1960-1970 birth cohort. The slowing convergence in Black-White

[12] See Figures A.5 and A.6 for the full set of transition matrices.

upward (or downward) mobility rates highlights the persistence of racial disparities in intergenerational mobility.

As a robustness check, we also experiment with analyses without farmers (Figure A.3, navy and maroon dashes). The results show a slight increase (decrease) in the level of upward (downward) mobility rates for these birth cohorts. We also conduct a parallel analysis using survey data for the 1900–1910 birth cohort (Figure A.3, black dots). The upward mobility trends are strikingly similar to those presented as the baseline estimates (Figure 1, navy and maroon solid lines), while those for downward mobility show noticeable deviations, partly due to the small sample size issues. Finally, Figure A.4 shows results using the historical sample constructed using MLP instead of CLP, where we find very similar results.

3.2.2 Barriers in Mobility: Bottlenecks

Next, we develop a bottleneck measure of mobility to identify regions within the occupational density that pose significant obstacles to mobility. This concept of the bottleneck, known across the literature by different names, has been widely documented in inequality studies. For example, Fishkin (2014) emphasizes the ethical salience of bottlenecks for inhibiting equality of opportunity. Bottlenecks, such as the gatekeeping role of elite college enrollment, can perpetuate inequalities and limit access to opportunities. We adapt this concept for mobility analyses using the *mean first passage times* in a Markov chain model. Let $f_{ij}^{(n)}$ denote the probability that the first passage time from state i to j is equal to n, which satisfies, $f_{ij}^{(n)} = \sum_{k \neq j} p_{ik} f_{kj}^{(n-1)}$. The mean first passage times, b_{ij}, can be defined as follows:

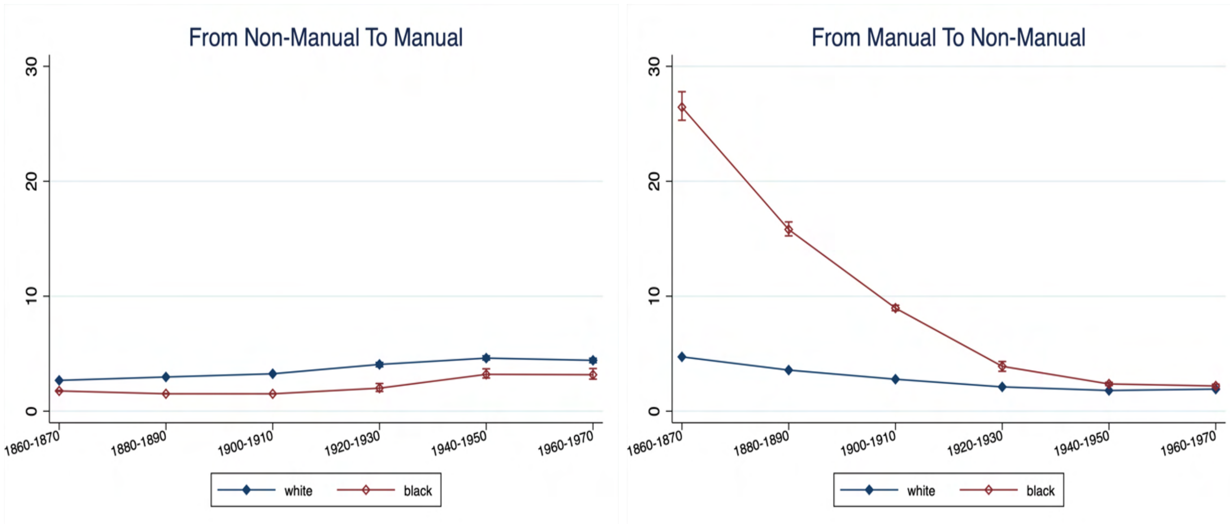
$$b_{ij} = E[f_{ij}^{(n)}] = \sum_{n=1}^{\infty} n f_{ij}^{(n)}. \tag{1}$$

Note that it can be easily verified that b_{ij} satisfies the following system of equations:

$$b_{ij} = 1 + \sum_{k \neq j} p_{ik} b_{kj} \tag{2}$$

If a chain begins at state i, state j can be directly reached from state i with probability $p_{ij} > 0$, and thus the first passage time is equal to 1. With probability p_{ik}, the chain first reaches the state $k \neq j$ and achieves the first passage time $(1+b_{kj})$, since it has taken 1 period without reaching the state j, and it will take the same expected first passage time from state k to j starting the next period. Therefore, $b_{ij} = 1 \cdot p_{ij} + \sum_{k \neq j}(1 + b_{kj}) \cdot p_{ik}$. Expanding the summation and using $p_{ij} + \sum_{k \neq j} p_{ik} = 1$, we obtain $b_{ij} = 1 + \sum_{k \neq j} p_{ik} b_{kj}$.[13]

Figure 2: Barriers to Mobility: Mean First Passage Times



Notes: The graphs show the mean first passage times from Non-Manual to Manual occupations (left) and from Manual to Non-Manual occupations (right) using 3 occupational categories for Black and White men. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure 2 presents the mean first passage times from "Non-Manual" to "Manual" occupations (left) and from "Manual" to "Non-Manual" occupations (right), which are estimated separately for Black and White men. The mean first passage times from "Non-Manual" to "Manual" occupations remain largely stable over time for both racial groups. The levels are lower for Black men than for White men, which is consistent with our earlier finding that Black men have a higher downward mobility rate. By contrast, the estimated mean first passage times from "Manual" to "Non-Manual" occupations reveal a striking difference between the two racial groups. The expected mean first passage times for Black men exceed 20 generations for the initial 1860–1870 cohort, but the values continue to drop over time. The 1960-1970 birth cohort, for example, face mean passage times of less than 5 generations. White men, on the other hand, experience a substantially lower barrier from the beginning, and the estimates remain relatively stable in subsequent periods. As the gap between the two groups shrinks over time, the mean passage times for the 1960-1970 cohort show negligible racial differences.

We further explore these trends by using detailed occupational definitions to identify specific occupation categories in which Black individuals experience significant changes

[13] Let $v_j \equiv b_{ij} | i \neq j = (b_{1j}, \cdots, b_{(i-1)j}, b_{(i+1)j}, \cdots b_{Nj})'$, and let P_{-j} denote the transition matrix with the j-th row and column removed ($(N-1) \times (N-1)$). The above system of equations can be rewritten as $v_j = 1 + P_{-j} v_j$. Rearranging, $v_j = (I_{N-1} - P_{-j})^{-1} \mathbb{1}_{N-1}$. We can then obtain the full set of b_{ij} from the transition matrix P.

in mobility opportunities (Figures A.11 and A.12). Given that some occupations, such as "Clerical" and "Sales," have large mean first passage times due to their small shares in the occupational distributions (Figure A.1), we focus our discussion on mobility barriers faced by individuals from parents in "Semi-Skilled and Unskilled" occupations in achieving professional and managerial positions. Table B.3 shows that the mean first passage times for Black men reduced from 12 to 6 generations between the 1920-1930 and the 1940-1950 birth cohorts. A similar trend but a more dramatic decline (from 43 to 12 generations) was observed for achieving occupations such as "Proprietors, Managers, and Officials." Despite these changes, the mobility barriers remain higher for Black men than White men.

4 The Link Between Intergenerational Transitions and Occupational Distributions

This section studies the importance of intergenerational transitions in explaining the observed evolution of racial gaps in occupational distributions over time. We take two complementary approaches: decomposition exercises and counterfactual experiments.

4.1 Decomposition of Dynamics

We adopt the Kitagawa-Oaxaca-Blinder decomposition method to decompose changes in the occupational distribution from time t to $t+1$ as,

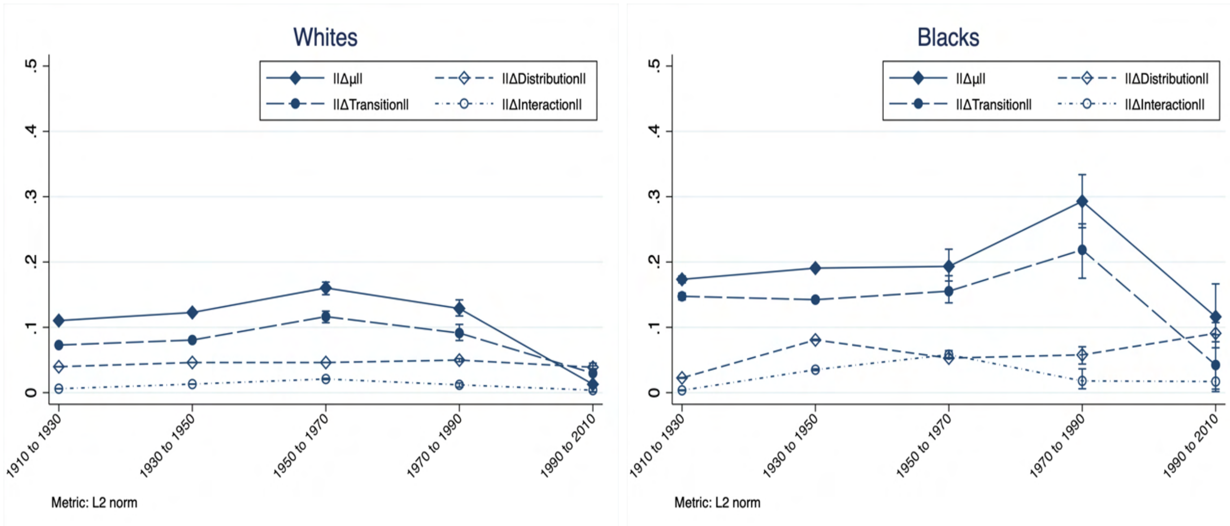
$$\begin{aligned} \mu_{t+1} - \mu_t &= \mu_t \mathbf{P}_t - \mu_{t-1}\mathbf{P}_{t-1} \\ &= \underbrace{(\mu_t - \mu_{t-1})\mathbf{P}_{t-1}}_{\text{distribution effect}} + \underbrace{\mu_{t-1}(\mathbf{P}_t - \mathbf{P}_{t-1})}_{\text{transition effect}} + \underbrace{(\mu_t - \mu_{t-1})(\mathbf{P}_t - \mathbf{P}_{t-1})}_{\text{interaction effect}}. \end{aligned} \tag{3}$$

Figure 3 shows the magnitudes of the overall change in occupational distributions from one period to the next, $\|\mu_{t+1} - \mu_t\|$, and the contribution of each component, $\|(\mu_t - \mu_{t-1})\mathbf{P}_{t-1}\|$, $\|\mu_{t-1}(\mathbf{P}_t - \mathbf{P}_{t-1})\|$, and $\|(\mu_t - \mu_{t-1})(\mathbf{P}_t - \mathbf{P}_{t-1})\|$.[14] The solid line, which delineates the aggregate change in the occupational distribution from one period to another, closely follows the long-dashed line representing the transition effects. This pattern suggests that the transition effects play a more important role than the distribution effects (short-dashed lines) or the interaction effects (dotted lines) in explaining changes in occupational distributions over time. While the results are similar for both racial groups, the magnitude of changes is greater for Black men. In particular, notable changes in both the occupational distribution

[14] As we take (L2) norms, equality no longer holds for Equation (3):
$\|\mu_{t+1} - \mu_t\| \leq \|(\mu_t - \mu_{t-1})\mathbf{P}_{t-1}\| + \|\mu_{t-1}(\mathbf{P}_t - \mathbf{P}_{t-1})\| + \|(\mu_t - \mu_{t-1})(\mathbf{P}_t - \mathbf{P}_{t-1})\|$

and the transitions occur between the 1920-1930 and 1940-1950 birth cohorts.

Figure 3: Aggregate Measurements of Kitagawa-Oaxaca-Blinder Decompositions



Notes: The graphs show the magnitude of the change in occupational distributions ($\|\mu_{t+1} - \mu_t\|$ in solid lines) and the three components ($\|(\mu_t - \mu_{t-1})\mathbf{P}_{t-1}\|$ in short-dashed lines, $\|\mu_{t-1}(\mathbf{P}_t - \mathbf{P}_{t-1})\|$ in long-dashed lines, and $\|(\mu_t - \mu_{t-1})(\mathbf{P}_t - \mathbf{P}_{t-1})\|$ in dotted lines) shown in Equation (3), separately for White (left) and Black men (right) for our sample of birth cohorts. We take L2 norms. Our calculations are based on 3 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Table 1 further shows the results of the decomposition exercise by occupational group for White (Panel A) and Black (Panel B) men.[15] Each row represents (i) changes in the occupational shares from one period to another (ΔOcc Shares) and (ii) the contribution of each component (ΔDistribution, ΔTransitions, and ΔInteraction) to the occupation-specific change in shares. We report the initial shares of each occupation in parentheses in the first column. For example, for White men, the share of "Non-Manual" and "Manual" occupations in the initial occupational distribution (the 1860-1870 birth cohort) are 0.272 (or 27.2 percent) and 0.512 (or 51.2 percent), respectively. For Black men, the corresponding occupational shares are 0.038 (or 3.8 percent) and 0.570 (or 57.0 percent). Column (1) shows that the share of "Non-Manual" occupations increases over time for both racial groups. The magnitude of change for Black men becomes larger than that of Whites between the 1900-1910 and 1920-1930 birth cohorts. In fact, it remains significantly larger across subsequent birth cohorts. A reciprocal pattern holds for "Manual" occupations, where we observe continuous decreases in shares. However, among Black men in the 1900–1910 birth cohort, this share increases, before beginning a decline in all subsequent cohorts.

[15]Due to space constraints and the focus of the current discussion, the results for Farmers are omitted. These results are available upon request.

Table 1: Kitagawa-Oaxaca-Blinder Decompositions by Race

Occupation (Initial Share)	Cohorts (from)	Cohorts (to)	(1) ΔOcc Shares	(2) ΔDistribution	(3) ΔTransitions	(4) ΔInteraction
Panel A: **Whites**						
Non-Manual (0.272)	1860-1870	1880-1890	0.079	0.02	0.056	0.004
			(.078, .08)	(.02, .02)	(.055, .057)	(.003, .004)
	1880-1890	1900-1910	0.096	0.033	0.063	0.001
			(.095, .097)	(.033, .033)	(.062, .064)	(.001, .001)
	1900-1910	1920-1930	0.131	0.037	0.095	-0.002
			(.122, .137)	(.037, .037)	(.087, .102)	(-.003, 0)
	1920-1930	1940-1950	0.095	0.038	0.066	-0.009
			(.087, .104)	(.036, .04)	(.057, .075)	(-.011, -.006)
	1940-1950	1960-1970	0.009	0.028	-0.022	0.003
			(-.001, .019)	(.027, .03)	(-.033, -.011)	(0, .006)
Manual (0.521)	1860-1870	1880-1890	-0.002	0.012	-0.01	-0.005
			(-.003, -.001)	(.012, .013)	(-.011, -.009)	(-.005, -.005)
	1880-1890	1900-1910	-0.024	-0.001	-0.014	-0.01
			(-.025, -.024)	(-.001, -.001)	(-.015, -.013)	(-.01, -.009)
	1900-1910	1920-1930	-0.073	-0.014	-0.046	-0.014
			(-.081, -.066)	(-.014, -.014)	(-.053, -.038)	(-.015, -.012)
	1920-1930	1940-1950	-0.087	-0.031	-0.063	0.008
			(-.096, -.079)	(-.033, -.03)	(-.073, -.055)	(.006, .01)
	1940-1950	1960-1970	-0.009	-0.027	0.02	-0.002
			(-.018, .002)	(-.028, -.025)	(.011, .031)	(-.005, 0)
Panel B: **Blacks**						
Non-Manual (0.038)	1860-1870	1880-1890	0.019	0.001	0.016	0.001
			(.017, .021)	(.001, .002)	(.014, .018)	(.001, .001)
	1880-1890	1900-1910	0.048	0.008	0.035	0.005
			(.045, .05)	(.008, .009)	(.032, .037)	(.004, .006)
	1900-1910	1920-1930	0.156	0.018	0.123	0.014
			(.134, .179)	(.018, .019)	(.105, .144)	(.009, .02)
	1920-1930	1940-1950	0.213	0.044	0.159	0.01
			(.185, .242)	(.034, .053)	(.129, .186)	(0, .023)
	1940-1950	1960-1970	0.083	0.065	0.03	-0.012
			(.049, .119)	(.056, .077)	(-.004, .064)	(-.031, .003)
Manual (0.570)	1860-1870	1880-1890	0.112	0.015	0.095	0.002
			(.107, .116)	(.015, .016)	(.091, .099)	(.001, .002)
	1880-1890	1900-1910	0.104	0.053	0.079	-0.027
			(.101, .108)	(.052, .053)	(.076, .082)	(-.028, -.026)
	1900-1910	1920-1930	-0.055	0.025	-0.033	-0.046
			(-.079, -.033)	(.024, .025)	(-.053, -.013)	(-.052, -.04)
	1920-1930	1940-1950	-0.2	-0.036	-0.15	-0.014
			(-.23, -.171)	(-.046, -.026)	(-.181, -.119)	(-.028, -.004)
	1940-1950	1960-1970	-0.081	-0.063	-0.03	0.012
			(-.117, -.048)	(-.075, -.054)	(-.061, .004)	(-.003, .03)

Notes: The table shows the results of the decomposition exercise for White (Panel A) and Black men (Panel B) across Non-Manual and Manual occupations, where each row represents changes in the occupational shares from one period to another; and how much each component contributes to the occupation-specific change in shares. We report the initial shares of each occupation in the first column in parentheses. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Consistent with Figure 3, Table 1 shows significant changes in occupational shares between the 1920-1930 and 1940-1950 birth cohorts, where the transition effects mainly explain the overall changes in occupational shares for both racial groups. For example, the shares explained by intergenerational transitions for "Non-Manual" ("Manual") occupations are roughly 75 percent (78 percent) and 78 percent (77 percent) for Black and White men, respectively. Note that the observed trends do not persist in later periods. Changes in the occupational shares drop significantly for both racial groups, particularly for White men. Even for Black men, the magnitude of the progress achieved by the 1940-1950 birth cohort does not continue either in terms of increasing the share of "Non-Manual" or in terms of decreasing the share of "Manual" occupations relative to the previous period.[16]

4.2 Counterfactual Dynamics and Distributions

Building on the significance of the intergenerational transition channel, we conduct a series of counterfactual experiments with two goals. First, we investigate the influence of each historical event and its associated transition probabilities on the present occupational disparities between Black and White men. Second, we examine whether the racial gap in occupational distribution narrows if Black men experienced the same mobility dynamics or began with the same initial conditions as White men.

4.2.1 Counterfactual I: Hypothetical Historical Evolution of Transitions

We model the historical evolution of cross-section densities, $\mu_t = \mathbf{P}_{t-1}\mu_{t-1} = \mathbf{P}_{t-1}\mathbf{P}_{t-2}...\mathbf{P}_0\mu_0$, with the counterfactual densities, as follows.

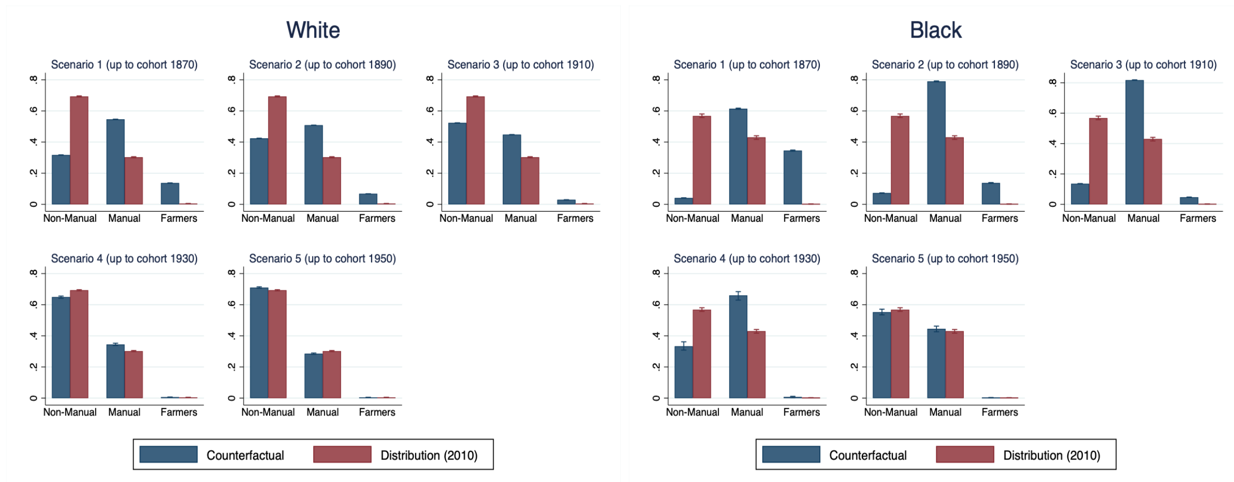
$$\begin{aligned}
\mu_{t|0} &= \mathbf{P}_0\mathbf{P}_0...\mathbf{P}_0\mathbf{P}_0\mathbf{P}_0\mu_0 \\
\mu_{t|1} &= \mathbf{P}_1\mathbf{P}_1...\mathbf{P}_1\mathbf{P}_1\mathbf{P}_0\mu_0 \\
&\vdots \\
\mu_{t|t-2} &= \mathbf{P}_{t-2}\mathbf{P}_{t-2}...\mathbf{P}_2\mathbf{P}_1\mathbf{P}_0\mu_0 \\
\mu_{t|t-1} &= \mathbf{P}_{t-1}\mathbf{P}_{t-2}...\mathbf{P}_2\mathbf{P}_1\mathbf{P}_0\mu_0
\end{aligned} \tag{4}$$

Note that $\mu_{t|t-1} = \mu_t$, by construction. Thus, the counterfactual distributions $\mu_{t|\tau}$ (where $\tau = 0, 1, \cdots, t-2$) represent how the occupational distribution might have evolved if historical events and their consequential changes in intergenerational mobility had not occurred. Furthermore, comparisons of $\mu_{t|t-1} = \mu_t$ and $\mu_{t|\tau}$ (where $\tau = 0, 1, \cdots, t-2$) allow us to

[16] Table B.4 shows results combining both racial groups, and we obtain similar results highlighting the importance of the change observed between the 1920-1930 and 1940-1950 birth cohorts; and the significant contribution of the intergenerational transition channel.

assess the impact of a historical event in period $t = \tau$ on mobility that may have led to an occupational distribution more (dis)similar to the actual distribution.[17]

Figure 4: Counterfactual Occupational Distributions



Notes: The graphs show the counterfactual occupational distributions for White (left) and Black men (right) in five different scenarios generated by sequentially updating the transition matrices over time. Our results are based on 3 occupational categories. For example, in the top left panels, each group's current occupational distributions are contrasted with a counterfactual one where the transition matrix from the 1860-1870 birth cohorts is employed and remains unchanged over time. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure 4 presents the counterfactual occupational distributions for White (left) and Black men (right) in five scenarios, generated by sequentially updating the transition matrices over time. At the top left panel, each group's current occupational distributions are contrasted with a counterfactual distribution in which the transition matrix is assumed to be fixed at the level of the 1860-1870 birth cohort. In this scenario, both Black and White men exhibit a dominant group of farmers and manual workers, which is in stark contrast to the distribution observed for the 1960-1970 birth cohort. At the top middle panels, the transition matrix evolves up to the one from the 1880-1890 birth cohort. The results still suggest a substantial difference from the current distributions. The bottom middle panel incorporates the transition matrix from the 1940-1950 birth cohort and yields a great resemblance to the current distributions.

For White men, the gap between the counterfactual distributions and the distribution observed in 2010 gradually decreases as we update the matrices across different counterfactual scenarios. After including the transition process from the 1900-1910 birth cohorts, the share of "Non-Manual" occupations becomes the highest while the share of "Manual"

[17] Note that this approach is complementary to the Kitagawa-Oaxaca-Blinder decomposition in Section 4.1, which compares μ_{t-1} and $\mu_{t|t-1} = \mu_t$ for each t.

occupations remains high. The share of "Non-Manual" occupations continues to increase as we update the transition matrices. Figures A.13 and A.16 (top panel) show results using detailed occupational definitions. The rising importance of the transition matrix from the 1900-1910 birth cohorts in increasing the shares of "Non-Manual" occupations is mainly driven by the middle-skill occupations in "Clerical" and "Sales" categories, not in "Professionals" and "Managers" that require higher levels of skills or have higher earnings. The inclusion of the transitions from the 1940-1950 birth cohorts leads to an occupational distribution that is similar to the current one, where the shares in "Professionals and Managers" significantly increase relative to other counterfactual scenarios.

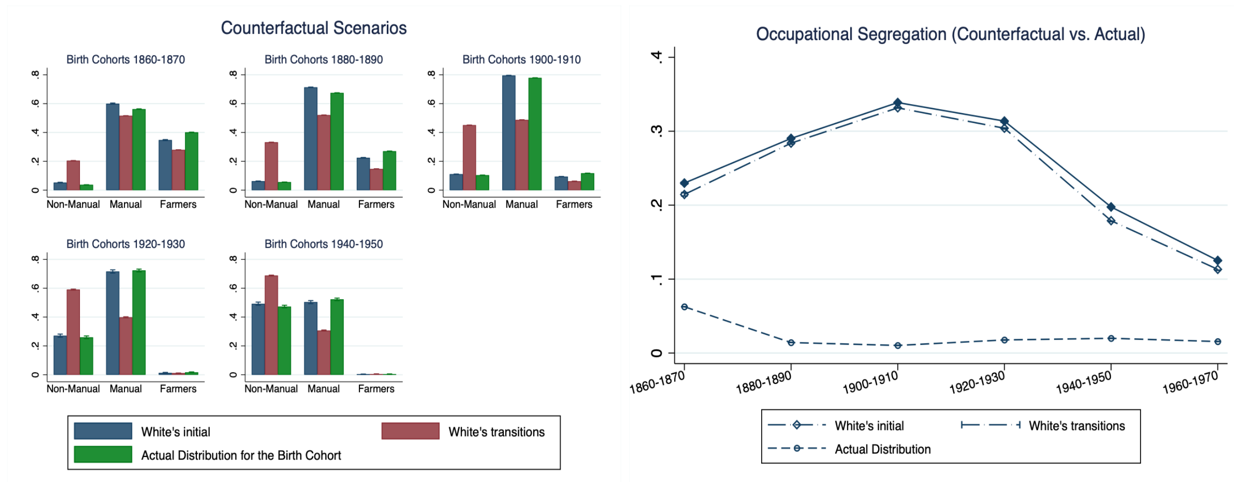
For Black men, the gap between the counterfactual distributions and the current distribution persists and dramatically drops as we introduce the transition process from the 1940-1950 cohorts. The share of "Manual" occupations, in particular, remains a dominant category, even after including the 1910-1920 birth cohorts, which is in contrast with the patterns of Whites. A further analysis using detailed occupational definitions (Figures A.13 and A.16, bottom panel) shows that the overrepresentation of "Manual" occupations is driven by "Semi-Skilled and Unskilled" occupations, whereas the lower share of "Non-Manual" occupations is explained by the lack of "Professionals and Managers," which significantly increases after the 1940-1950 birth cohorts.

4.2.2 Counterfactual II: No Racial Differences in Initial Conditions or Transitions

Next, we conduct additional counterfactual experiments to compare the roles of initial conditions versus the evolution of transitions in explaining Black-White differences in occupational distributions over time. We first consider both racial groups starting at *the same initial condition* but experiencing different mobility dynamics. Next, we consider these groups undergoing *the same mobility dynamics* but facing different initial conditions, The exercises will result in comparing the occupational distribution of White men μ_t^W to the counterfactuals, $\mu_t^{B'} = \mathbf{P}_{t-1}^B \mathbf{P}_{t-2}^B ... \mathbf{P}_1^B \mathbf{P}_0^B \mu_0^W$ and $\mu_t^{B''} = \mathbf{P}_{t-1}^W \mathbf{P}_{t-2}^W ... \mathbf{P}_1^W \mathbf{P}_0^W \mu_0^B$, respectively.

Figure 5 (left) shows the occupational distributions for both counterfactual exercises evaluated across time periods. The blue bars represent scenarios in which Black men begin with the initial conditions of White men; the red bars depict situations in which Black men experience the mobility dynamics of White men; and the green bars represent the actual occupational distribution of Black men in each period. Across all time periods, the share of "Non-Manual" occupations, when Black men undergo White men's transition matrix, exceeds the share obtained from the actual distribution. While smaller in magnitude, beginning at White men's initial conditions results in a slightly higher share of "Non-Manual" occupations than the actual ones. Results based on more detailed occupational definitions (both five and seven categories) are presented in Figures A.14 and A.17, which suggest that the higher shares in "Non-Manual" under these hypothetical scenarios are driven by both "Professional and Managerial" and "Clerical" and "Sales" occupations.

Figure 5: Counterfactual Occupational Distributions of Black Men



Notes: The left graphs contrast the actual occupational distributions of Black men and counterfactual ones where (i) both racial groups start with the same initial condition ($\mu_0^B = \mu_0^W$, in navy) but experience different mobility dynamics; (ii) both racial groups undergo the same mobility dynamics ($\mathbf{P}_t^B = \mathbf{P}_t^W$, in maroon) yet face different initial conditions. The right graph shows the degree of occupational segregation under each scenario. Our results are based on 3 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

We further examine changes in the degree of occupation segregation due to changes in transition probabilities under these two counterfactual scenarios (Figure 5 right). If Black men experience the same mobility dynamics as Whites, the segregation index immediately drops near zero in the second period and persists in the subsequent periods. In contrast, if Black men begin with equal initial conditions, this helps lower the segregation index but only by a small magnitude.

5 Persistence in Inequality: Measurement of Memory

Analyses in the previous sections take a *backward-looking* perspective to highlight the unique importance of changes in the mobility experienced by the 1940–1950 birth cohort in shaping the current occupational distribution and reducing the racial gap. Despite important improvements, Black-White inequality in both the cross-sectional occupation distributions and intergenerational mobility persists, consistent with prior studies stressing the lack of Black-White convergence in economic status (Bound and Freeman, 1992; Collins and Wanamaker, 2022). We now take a *forward-looking* perspective to explore dynamics of

Black-White inequalities implied by the observed transition matrices, drawing on *mixing* properties of Markov chain mobility analyses developed by Blume et al. (2024).

Our analyses examine (i) the *memory* of the mobility system, which refers to how long the influence of one generation lasts on future generations and (ii) how quickly this influence dissipates over time. To better understand the underlying mechanisms, we reformulate the ideas of structural versus exchange mobility in the sociology literature (McClendon, 1977; Sobel, 1983; Sobel et al., 1985) from the perspective of Markov chain models. Specifically, we differentiate between mobility driven by structural shifts in occupations and the swapping of occupational positions among individuals that occurs without changing the overall occupational distribution of society.

5.1 Mixing Analysis

In mixing analysis, the mixing time of a Markov chain is the time until the chain is close to its steady state distribution.[18] Building on this concept, we construct a sequence of mixing curves that measure how the memory of parental occupations at some initial time is preserved in future periods. Because the occupations in our transition matrix all communicate, the occupational density evolves toward a unique steady state in which the advantages or disadvantages of different families, regardless of their initial conditions, will eventually disappear. The difference between the present cross-sectional distribution and the steady state distribution thus serves as a measure of *memory*, that is, the information retained from the initial distribution. When this gap reaches zero (or memory becomes minimal), the current distribution no longer retains any information from the past. Thus, individuals' mobility outcomes become independent of the initial conditions of their families.

More formally, we characterize *aggregate deviations* (AD) as the difference between the cross-sectional densities at time t ($\mu_0 \mathbf{P}^t = \mu_t$) and the steady state ($\mu^*$),

$$AD(\mathbf{P}, \mu_0, t) = \|\mu_0 \mathbf{P}^t - \mu^*\|. \tag{5}$$

Greater values of AD imply stronger memory of the initial conditions and the mobility process in the system. Examining $AD(\mathbf{P}, \mu_0, t)$ over time provides information on the patterns and the rate of convergence toward the steady state.

The AD measure, however, does not capture heterogeneity in the persistence of memory among individuals. The chances for individuals to enter different occupations may vary

[18] Technically speaking, mixing analysis evaluates the memory of stochastic processes based on the dependence of probabilities for sets of random variables realized before some time t and those realized at $t + k$ or later. See Bradley (1986) for an exhaustive treatment.

substantially depending on their family background or for our focus, the occupation of their parents. Thus, we modify AD to evaluate *individual deviations* (ID) by parental occupation group,

$$ID(P, j, t) = \|e_j \mathbf{P}^t - \mu^*\|, \tag{6}$$

where the population-level vector of μ_0 in Equation (5) is replaced with a unit vector e_j, namely, a vector of zeros with a 1 in state (or occupation) j.

To aggregate these individual deviations, we integrate over the initial population density ($\mu_{e,0}$) and derive the total individual memory of the system to solve for the *aggregated individual deviations* (AID), which we characterize as taking a weighted average of the individual mobility curves:

$$AID(P, \mu_0, t) = \int \|e\mathbf{P}^t - \mu^*\| \mu_0. \tag{7}$$

For empirical analyses, we calculate the distances between data points ($\|\cdot\|$) using the L1 norm ($|x_1| + |x_2| + ... + |x_n|$).[19] The difference between the AD and AID measures captures the influence of mobility heterogeneity on the *memory* of inequality in the system.

Finally, we define the *rate* at which memory dissipates, *aggregate mobility* (AM) and *aggregated individual mobility* (AIM), by taking the ratios of the distance to steady state between t and $t+1$,

$$AM(\mathbf{P}, \mu_0, t) = \frac{AD(\mathbf{P}, \mu_0, t+1)}{AD(\mathbf{P}, \mu_0, t)}, \quad AIM(P, \mu_0, t) = \frac{AID(\mathbf{P}, \mu_0, t+1)}{AID(\mathbf{P}, \mu_0, t)}, \tag{8}$$

where AM and AIM are less than one since the distance to steady state decreases in t. As t increases, AM converges to the second-largest eigenvalue of the transition matrix, which governs the asymptotic rate of convergence. Note that *smaller* values of AM and AIM correspond to *faster* convergence and therefore, imply *greater* mobility.

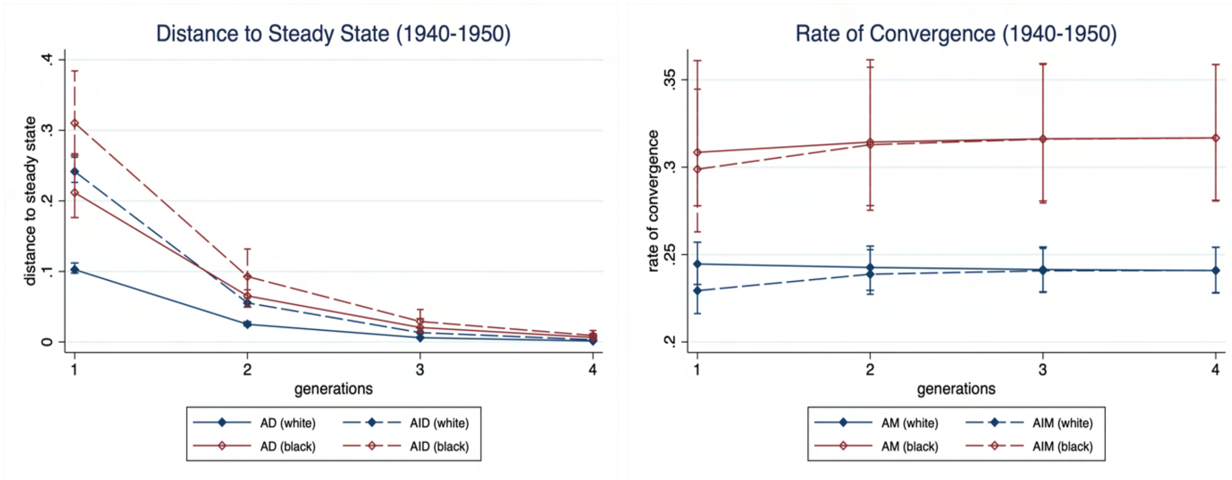
5.2 Memory of Systemic Inequality

We first present results for the 1940-1950 birth cohort, which was instrumental in shaping the current occupational distribution and reducing the racial gap. Figure 6 (left) shows how AD, AID, and ID evolve over each generation t for t = 1, 2, 3, 4. Figure 6 (right) presents the rate at which the mobility process converges to the steady state using AM and AIM.

For both racial groups, AD and AID converge to the steady state after four genera-

[19] See Blume et al. (2024) for more details.

Figure 6: Aggregate Deviations and Convergence Rates (1940-1950 Birth Cohort)



Notes: The left graph shows our measurements of memory AD (solid lines) and AID (dashed lines) for each racial group. Larger deviations to steady state correspond to greater memory. The right graph shows convergence rates AM (solid lines) and AIM (dashed lines). Smaller AM and AIM correspond to *faster* convergence and therefore, imply *greater* mobility. We use transition matrices for 5-category occupations. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

tions.[20] When examining differences between racial groups, both the levels of AD and AID are greater for Black men compared to White men for any t. In terms of the rate of convergence to the steady state, Black men show higher AM (solid lines) than White men, which correspond to lower mobility rates (Figure 6, right). That is, the opportunity for subsequent generations of the 1940-1950 birth cohort to move across occupations, estimated by iterating the transition matrix (P) over generations, show that Black men face more significant mobility challenges than White men. They are less likely to move away from the conditions of their parents toward a state of equal opportunity (indicated by greater distance to reach a steady state or higher AD and AID) and have greater difficulty in changing occupations (indicated by lower mobility rates or higher AM and AIM).

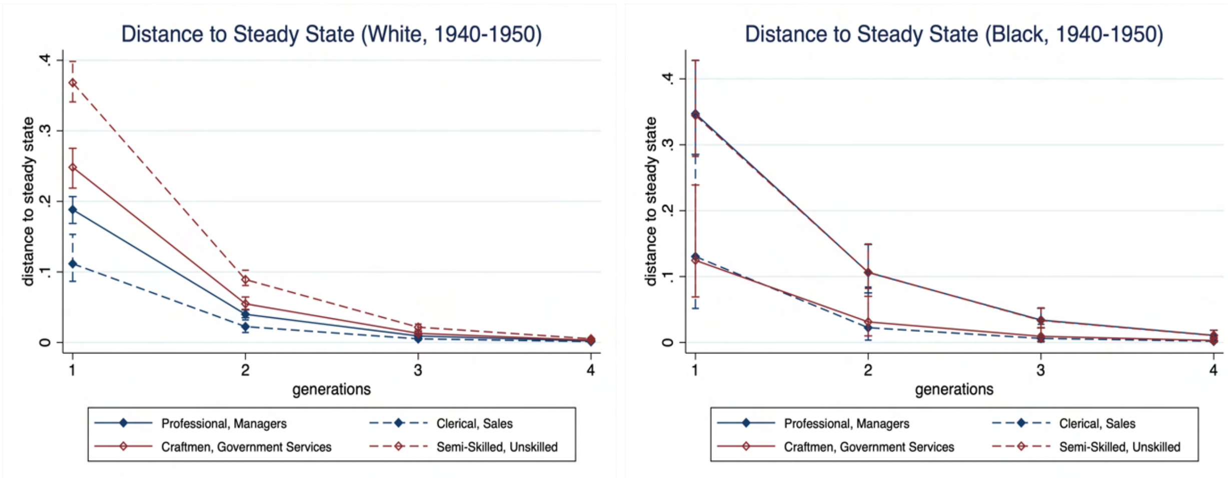
Figure 7 shows results for ID, which examines heterogeneity across parental occupational groups (j). The patterns vary across j for both racial groups.[21] For White men, the parental group that contributes most to the memory of the system at $t = 1$ is the "Semi-Skilled and Unskilled" occupational group. For Black men, the strongest memory at $t = 1$ is observed for both "Semi-Skilled and Unskilled" and "Professional and Manager" occupations. The results indicate that the mobility regime of Black men is both rigid at

[20]Note that the level of AID (dashed lines) is greater than AD (solid lines) for any generation t (Figure 6, left). In Appendix D, we compare the magnitudes of the two measures and show that the inequality always holds by construction.

[21]For better illustration purposes, we omit the farmer category in the graph.

the bottom and top of the system, whereas White men show stronger memory only for low-status occupations.

Figure 7: Individual Deviations (1940-1950 Birth Cohort)

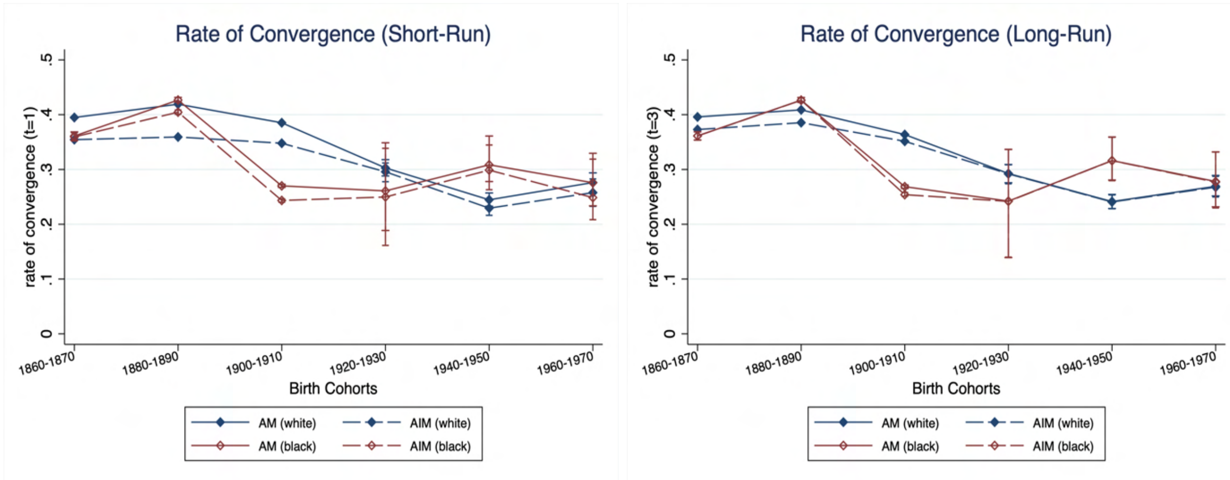


Notes: The graphs show our measurements of individual deviations ID for White men (left) and Black men (right). Larger deviations from the steady state correspond to greater memory. We use transition matrices for 5 occupational categories (excluding farmers). Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Next, we examine the trends in AM and AIM across different birth cohorts. Results in Figure 8 (left) show an overall downward trend for White men, while no obvious trend for Black men is apparent. When comparing the differences between racial groups, AM is lower for White men than Black men for the 1940-1950 and later birth cohorts, although the differences for the 1960-1970 birth cohort are statistically insignificant. The opposite holds for the earlier 1860-1870 and 1900-1910 birth cohorts. Thus, for younger cohorts, the aggregate rate of convergence is faster for White men than for Black men. The AIM measure that takes into account the heterogeneity across individuals across parental occupations shows qualitatively similar trends. Figure 8 (right) examines the long-term patterns of AM and AIM by calculating mobility rates ($t = 3$). Although the values of AM and AIM converge by design, the overall short-run dynamics are preserved.

Figure 9 further compares the occupational distributions observed for each birth cohort group to their respective steady-state distributions separately derived for Black and White men (μ^{B*} and μ^{W*}). We focus on studying how the state of equal opportunity compares to the actual distribution—for example in terms of high- and low-skill occupation shares or the degree of segregation by race. This allows us to interpret our results on the memory of the mobility system, as well as examine the rate at which it dissipates in relation to the inequality in the occupational distributions between Black and White men.

Figure 8: Evolution of the Rate of Convergences (All Birth Cohorts)



Notes: The graphs show AM and AIM in the short-run (left) and those derived in the long-run (right). Smaller AM and AIM correspond to *faster* convergence, and therefore, imply *greater* mobility. We use transition matrices for 5-category occupations. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

For White men (Figure 9, top panel), moving toward the steady state entails compositional shifts that increase the share of high-skill occupations. The share of "Professional and Managers" is always higher at the steady state than the actual, and the share of "Semi-Skilled and Unskilled" is lower at the steady state across all periods except for the 1860-1870 and 1880-1890 birth cohorts. For Black men, in contrast, the steady-state distributions have different implications. The share of "Professional and Managers" at the steady state shows minimal differences compared to the actual distribution up to the 1920-1930 birth cohort and shows significantly higher shares for the 1940-1950 and 1960-1970 birth cohorts. The share of "Semi-Skilled and Unskilled" is strictly higher at the steady state up to the 1900-1910 birth cohort and becomes lower for subsequent cohorts. For Black men, reaching the steady state has similar implications as White men for later periods; however, for earlier periods, the steady state implies a shift toward a distribution with higher shares in low-skill and farming occupations. As we compare the Black-White gap in occupational distributions, we observe a higher level of occupational segregation by race at the steady state than actual for up to the 1900-1910 birth cohort, while the opposite holds for the 1940-1950 and 1960-1970 birth cohorts (Figure A.2). Opportunities to maintain or to move into higher occupation status improved significantly for Black men compared to past cohorts, and therefore, the steady-state distribution indicates an improvement in occupational segregation.

Finally, we relate our discussion of steady states to our results using the measures

Figure 9: Occupational Distribution at Steady States

Notes: In each graph, the bars in navy show the occupational distributions observed in each period, which correspond to years when each birth cohort of interest reaches its prime age. The bars in maroon show the occupational distributions at steady states using the transition matrices of each birth cohort. Our results are based on 5 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

of AM and AIM (Figure 8) where we show that the rate of convergence for Black men is relatively faster for our earlier birth cohorts and slows down over time. The results collectively suggest that it is relatively easier for Black men to move away from the circumstances of the early to mid-twentieth century; however, the steady state at which "equal opportunities" are guaranteed does not indicate "better" circumstances, as we see a higher share of low-skill jobs and greater degrees of occupational segregation by race. The more recent cohorts, while facing a "better" distribution at steady state with lower degrees of occupational segregation, converge rather slowly.

5.3 Within and Between Racial Group Inequality

Next, we examine within- and between-group inequality using various deviation measures in mixing analysis defined in Equation (6). Specifically, for each racial group, we calculate the difference in the occupational distributions of descendants, comparing individuals who have parents in two distinct occupations,

$$ID^{wn}(P, j, k, t) = \|e_j\mathbf{P}^t - e_k\mathbf{P}^t\|. \tag{9}$$

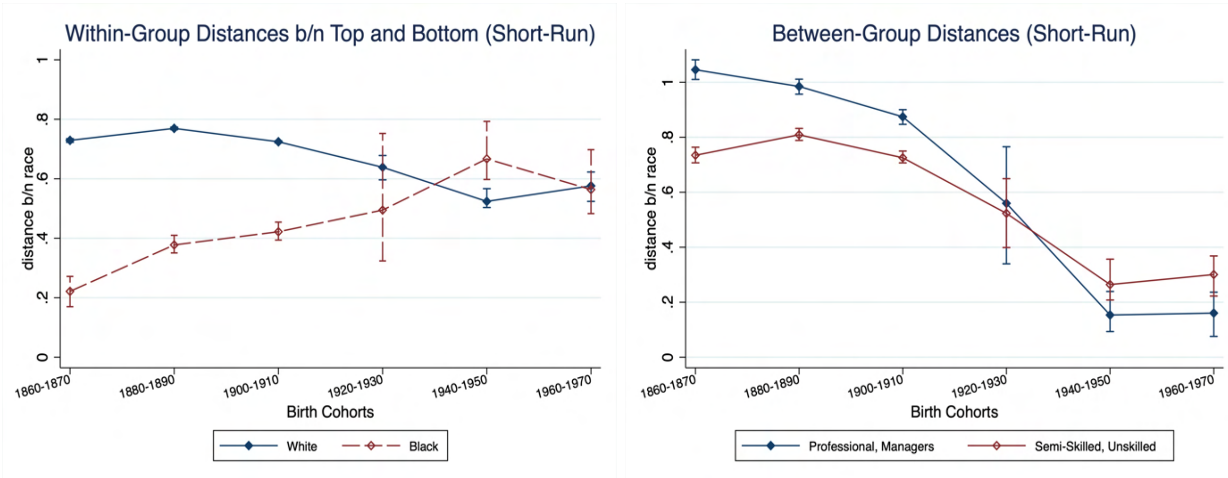
The measure evaluated at $t = 1$ reveals differences in the predicted occupational distribution among *offspring* within the same racial group. Examining the differences between high- and low-status occupations, for example, captures the within-group variance in the opportunities to sort across occupations faced by offspring.

We also examine the between-group inequality by comparing individuals from different racial groups who have parents in the same occupations,

$$ID^{bn}(P_W, P_B, j, t) = \|e_j\mathbf{P}_W^t - e_j\mathbf{P}_B^t\|. \tag{10}$$

For example, investigating the between-group variance for "Professional, Managers" parents provides insights into how the opportunities to maintain or move into different occupations for the offspring generation differ by race.

Figure 10: Within and Between Group Convergences (Short-Run)



Notes: The graphs show the within-race-group deviations between "Professional and Managers" and "Semi-Skilled and Unskilled" groups (left); and the between-race-group deviations for "Professional and Managers" and "Semi-Skilled and Unskilled" groups separately (right). We use transition matrices for 5 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure 10 (left) shows the within-group differences for families of high- and low-status occupations in occupations, namely "Professional and Managers" and "Semi-Skilled and Unskilled." We find that the trends are similar to the dynamics of the rate of convergences shown in Figure 8 using AIM, which reflect within-group heterogeneity in mobility across families. For Black Americans, the within-race difference continues to increase, though we observe a decrease for the most recent cohort. The opposite holds for White men. Our results echo recent findings studying within-variance between the two racial groups, which highlight similar contrasting trends between Black and White Americans for younger cohorts (Chetty et al., 2023).

Figure 10 (right) separately shows the between-group differences in their job opportunities for families of "Professional and Managers" and "Semi-Skilled and Unskilled" groups. We show that the trend closely follows the dynamics in the degree of occupational segregation (Figure A.2), which is initially high but dramatically falls for the 1940-1950 and 1960-1970 birth cohorts. Our measurements further highlight that, while the between-group difference falls by a greater magnitude for the high-status compared to the low-status group over time, the downward trend does not continue, but rather slightly increases for the 1960-1970 birth cohort. Black-White convergences have improved significantly, driven by the 1940-1950 birth cohort, but recent trends still show a non-negligible Black-White gap in their opportunities for offspring to sort into high-status occupations or to move away from low-status ones (Collins and Wanamaker, 2022).

5.4 Structural Versus Exchange Mobility

Finally, we develop measures to distinguish between structural and exchange mobility.[22] Structural mobility refers to shifts in occupational distributions across generations driven by aggregate changes in the economy. Due to changes in discriminatory regimes in history, the amount of structural influence on mobility may vary by race and play a significant role in Black-White gaps in mobility outcomes. Exchange mobility, on the contrary, is often considered a zero-sum game, as the overall occupational distribution does not vary. A society with a higher level of exchange mobility is generally considered more open and fluid, providing opportunities for individuals to improve their status through their own merit and effort (McClendon, 1977; Sobel, 1983; Sobel et al., 1985).

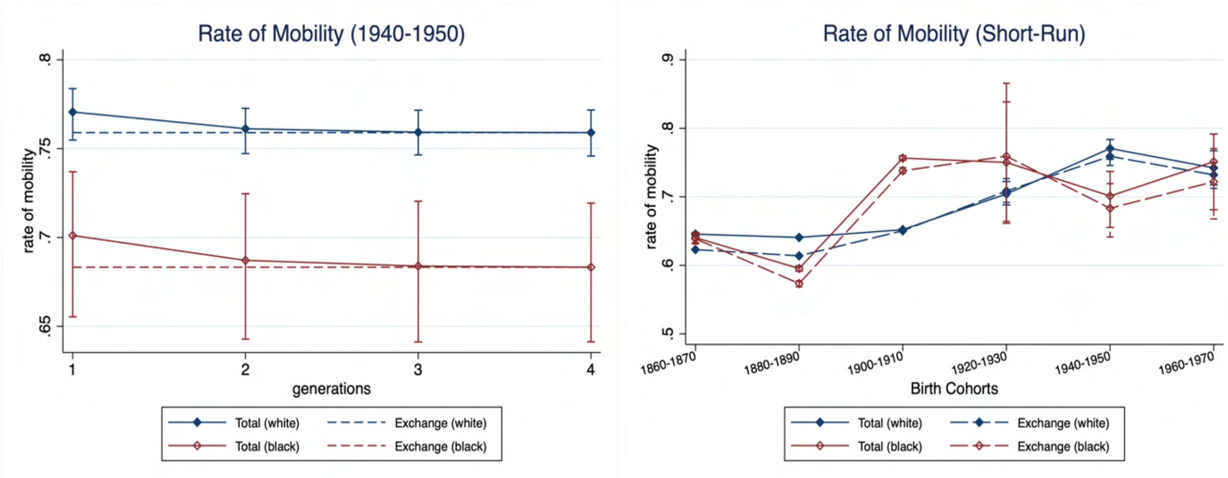
Our measurement approach builds on Berger and Snell (1957) who conceptually distinguish between exchange and structural mobility by contrasting mobility rates when the cross-section occupational distribution is at the steady state versus when it is not. Specifically, we measure exchange mobility as the amount of mobility at the steady state where we observe no structural mobility in comparison to the historically observed mobility associated with a population that is not at its steady state. Thus, the process of approaching a steady state can be viewed as generating a reduction in structural mobility, while the level of exchange mobility remains constant. Using our measures of memory based on mixing curves, we define the rate of mobility measured using $AIM(P, \mu_0, t)$ near steady state as the rate of *exchange mobility*. Then, the overall rate of mobility calculated using $AIM(P, \mu_0, t)$ examined at each time t, net of exchange mobility, gives the rate of *structural mobility*, $SM(P, \mu_0, t)$.[23]

Figure 11 (left) shows the mobility rates for the 1940-1950 birth cohort. For interpretative purposes, the graphs are based on calculations using 1-$AIM(P, \mu_0, t)$, and thus. larger values indicate higher mobility rates. Solid lines represent aggregate mobility rates, and dashed horizontal lines indicate exchange mobility rates. The difference between the two captures structural mobility rates, which become minimal as we iterate the mobility process over generations. For both racial groups, the importance of structural factors in explaining mobility patterns of the 1940-1950 birth cohort is limited (Figure 11, left), which accounts for less than 5 percent of the total mobility rates from $t = 1$ onward. A

[22]The conceptual distinction between structural and exchange mobility dates back to the seminal work of Kahl (1957), where he identified "four causes of movement" in social mobility: changes due to technology, population dynamics, immigration, and the individual shifts that occur as some people fall in social standing, thereby creating opportunities for others to rise.

[23]Note that Blume et al. (2024) takes an alternative approach to define exchange mobility using an imbalance matrix, which captures the difference between the inflow and outflow of each occupation (or class). The imbalances disappear as the process reaches steady state (or as structural mobility becomes zero).

Figure 11: Structural versus Exchange Mobility



Notes: The left graph shows mobility rates for the 1940-1950 birth cohorts. For interpretation purposes, our plots are based on calculations using 1- $AIM(P, \mu_0, t)$, so that larger values indicate greater mobility rates. Solid lines represent aggregate mobility rates; dashed horizontal lines indicate the level of exchange mobility rates. The right graph shows the aggregate and exchange mobility examined at $t = 1$ for our sample of birth cohorts. The difference between the solid and dashed lines, for each racial group, captures the level of structural mobility rates. We use transition matrices for 5 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

comparison between racial groups shows that the level of structural mobility is greater for Black men than White men while the opposite holds for exchange mobility. In other words, the experiences of Black men relative to White men differ in two main ways: Black men's occupational mobility is more influenced by aggregate changes in the economy, and embed lower opportunities to improve their status through their own merit and effort.

Figure 11 (right) compares the aggregate and exchange mobility rates at $t = 1$ for all birth cohorts in our sample. Again, the differences between the solid and dashed lines for each racial group capture the magnitude of structural mobility rates. We still find limited importance for structural mobility across all cohorts, which is more relevant for White than Black Americans. This finding aligns with the significant impact of the civil rights movement in transforming opportunities for Black Americans. The overall mobility rate for White men, largely driven by exchange mobility, demonstrates an upward trend for the 1880-1900 birth cohort onward but levels off for the 1960-1970 birth cohort. Black men, on the contrary, exhibit a notable increase in mobility rates among the 1900-1910 birth cohort relative to preceding generations. Following several periods of relative stability, this trajectory reveals a subsequent uptick in both exchange and structural mobility for the 1960-1970 birth cohort.

6 Concluding Remarks

This study investigates the evolution of Black-White inequality over a long time horizon, focusing on the racial gap in *occupational* mobility using an extensive dataset that combines linked historical census data from 1850 to 1940 and 10 contemporary large-scale survey data from the 1960s to the present. Our findings highlight the significance of mobility dynamics achieved through the 1940-1950 birth cohort in shaping the current occupational distribution and reducing the racial gap. Despite important improvements, the racial gap in cross-sectional occupational distributions and intergenerational mobility persists. Furthermore, the momentum achieved through the 1940-1950 birth cohort dissipates in later cohorts. The mobility process of Black men compared to that of White men shows greater path dependence, or difficulty moving away from current conditions; this memory endures over time. The disadvantages in opportunities are attributed to the lack of exchange mobility based on individual effort and merit rather than structural factors.

Our work emphasizes the ongoing challenges faced by Black individuals in the United States in achieving upward mobility and the importance of addressing historical legacy and mobility barriers to promote greater equality of opportunity. The stochastic models help address long-standing questions regarding the consequences of institutional changes for intergenerational mobility rates, patterns, and group differences in subsequent decades. These findings illuminate the impact of intergenerational mobility on the enduring legacy of racial disparity within systems. They further offer insight into projecting convergences in future economic opportunities, thereby guiding policymakers in identifying the type of policy interventions required to reduce Black-White inequality.

References

Abramitzky, Ran, Leah Boustan, Katherine Eriksson, James Feigenbaum, and Santiago Pérez, "Automated Linking of Historical Data," *Journal of Economic Literature*, 2021, *59* (3), 865–918.

_ , _ , _ , **Myera Rashid, and Santiago Pérez**, "Census Linking Project: 1850-1860 Crosswalk," Technical Report, Harvard Dataverse 2022.

_ , **Roy Mill, and Santiago Pérez**, "Linking Individuals Across Historical Sources: A Fully Automated Approach," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2020, *53* (2), 94–111.

Bailey, Martha J., Connor Cole, Morgan Henderson, and Catherine Massey, "How Well Do Automated Linking Methods Perform? Lessons from US Historical Data," *Journal of Economic Literature*, December 2020, *58* (4), 997–1044.

Berger, J. and J. L. Snell, "On the Concept of Equal Exchange," *Journal of Behavioral Science*, 1957, *2*, 111–118.

Bhattacharya, Debopam and Bhashkar Mazumder, "A Nonparametric Analysis of Black–White Differences in Intergenerational Income Mobility in the United States," *Quantitative Economics*, 2011, *2* (3), 335–379.

Black, Dan A., Seth G. Sanders, Evan J. Taylor, and Lowell J. Taylor, "The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South," *American Economic Review*, February 2015, *105* (2), 477–503.

Blau, Peter M and Otis Dudley Duncan, *The American Occupational Structure.*, John Wiley & Sons, 1967.

Bloome, Deirdre, "Racial Inequality Trends and the Intergenerational Persistence of Income and Family Structure," *American Sociological Review*, 2014, *79* (6), 1196–1225.

—, "Childhood Family Structure and Intergenerational Income Mobility in the United States," *Demography*, 2017, *54* (2), 541–569.

Blume, L., N. Cholli, S. Durlauf, and A. Lukina, "Mobility as Memory: Some New Approaches to Characterizing Intergenerational Dependence via Markov Chains," 2024. Working Paper.

Blumen, Isadore, Marvin. Kogan, and Philip. J. McCarthy, *The Industrial Mobility of Labor as a Probability Process*, Ithaca, New York: Cornell University Press, 1955.

Bound, John and Richard B. Freeman, "What Went Wrong? The Erosion of Relative Earnings and Employment Among Young Black Men in the 1980s," *Quarterly Journal of Economics*, 1992, *107* (1), 201–232.

Bradley, Richard C, "Basic Properties of Strong Mixing Conditions," in Ernst Eberlein and Murad S. Taqqu, eds., *Dependence in Probability and Statistics: A Survey of Recent Results*, Springer, 1986, pp. 165–192.

Buckles, Kasey, Joseph Price, Zachary Ward, and Haley E.B. Wilbert, "Family Trees and Falling Apples: Historical Intergenerational Mobility Estimates for Women and Men," Working Paper 31918, National Bureau of Economic Research November 2023.

Chetty, Raj, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter, "Race and Economic Opportunity in the United States: An Intergenerational Perspective," *Quarterly Journal of Economics*, 2020, *135* (2), 711–783.

—, —, Patrick Kline, Emmanuel Saez, and Nicholas Turner, "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility," *American Economic Review*, 2014, *104* (5), 141–47.

— , **Will Dobbie, Sonya Porter, and Crystal Yang**, "Growing Class Gaps, Shrinking Race Gaps: Economic and Sociological Mechanisms Underlying Recent Trends in Intergenerational Mobility," Technical Report, Working Paper 2023.

Collins, William J. and Marianne H. Wanamaker, "Selection and Economic Gains in the Great Migration of African Americans: New Evidence from Linked Census Data," *American Economic Journal: Applied Economics*, January 2014, *6* (1), 220–52.

— **and** — , "African American Intergenerational Economic Mobility since 1880," *American Economic Journal: Applied Economics*, July 2022, *14* (3), 84–117.

Derenoncourt, Ellora, "Can You Move to Opportunity? Evidence from the Great Migration," *American Economic Review*, February 2022, *112* (2), 369–408.

Donohue, John J. and James Heckman, "Continuous Versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks," *Journal of Economic Literature*, 1991, *29* (4), 1603–1643.

Duncan, O., D. Featherman, and B. Duncan, *Socioeconomic Background and Achievement*, Seminar Pressr, 1972.

Duncan, Otis D., "Inheritance of Poverty or Interitance of Race?," in Moynihan Daniel P., ed., *Understanding Poverty*, New York, NY: Basic Books, 1968, pp. 85–110.

Duncan, Otis Dudley and Beverly Duncan, "A Methodological Analysis of Segregation Indexes," *American Sociological Review*, 1955, *20* (2), 210–217.

Featherman, David L and Robert M Hauser, "Changes in the Socioeconomic Stratification of the Races, 1962-73," *American Journal of Sociology*, 1976, *82* (3), 621–651.

Fishkin, J., *Bottlenecks: A New Theory of Equal Opportunity*, Oxford University Press, 2014.

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

Haider, Steven and Gary Solon, "Life-Cycle Variation in the Association between Current and Lifetime Earnings," *American Economic Review*, 2006, *96* (4), 1308–1320.

Hammel, Eugene A., "Incomplete Histories in Family Reconstitution: A Sensitivity Test of Alternative Strategies with Historical Croatian Data.," in David S. Reher and Roger Schofield, eds., *Old and New Methods in Historical Demography*, New York, NY: Oxford University Press, 1993, pp. 125–144.

Helgertz, Jonas, Joseph Price, Jacob Wellington, Kelly J. Thompson, Steven Ruggles, and Catherine A. Fitch, "A New Strategy for Linking US Historical Censuses: A Case Study for the IPUMS Multigenerational Longitudinal Panel," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 2022, *55* (1), 12–29.

Hout, M., "Occupational Mobility of Black Men: 1962-1973," *American Sociological Review*, 1984, *49* (3), 308–322.

Hout, Michael, *Mobility Tables*, Beverly Hills: Sage Publications 1983., 1983.

Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu, "Mobility for All: Representative Intergenerational Mobility Estimates Over the 20th Century," Working Paper 29289, National Bureau of Economic Research September 2021.

Kahl, Joseph Alan, *The American Class Structure*, New York: Rinebart, 1957.

Lee, Chul-In and Gary Solon, "Trends in Intergenerational Income Mobility," *Review of Economics and Statistics*, 11 2009, *91* (4), 766–772.

Logan, Trevon D. and John M. Parman, "The National Rise in Residential Segregation," *Journal of Economic History*, 2017, *77* (1), 127–170.

Long, Jason and Joseph Ferrie, "Intergenerational Occupational Mobility in Great Britain and the United States Since 1850," *American Economic Review*, 2013, *103* (4), 1109–1137.

Margo, Robert A., "Obama, Katrina, and the Persistence of Racial Inequality," *Journal of Economic History*, 2016, *76* (2), 301–341.

Matras, Judah, *Social Inequality, Stratification, and Mobility*, Englewood Cliffs: Prentice-Hall, 1975.

Mazumder, Bhashkar, "Black-White Differences in Intergenerational Economic Mobility in the United States," *Economic Perspectives*, 2014, *38* (1), 1–37.

McClendon, McKee J, "Structural and Exchange Components of Vertical Mobility," *American Sociological Review*, 1977, *41* (1), 56–74.

Modell, John, Marc Goulden, and Sigurdur Magnusson, "World War II in the Lives of Black Americans: Some Findings and Interpretation," *Journal of American History*, 1989, *76* (3), 838–848.

Olivetti, Claudia and M. Daniele Paserman, "In the Name of the Son (and the Daughter): Intergenerational Mobility in the United States, 1850-1940," *American Economic Review*, August 2015, *105* (8), 2695–2724.

Prais, S. J, "Measuring Social Mobility," *Journal of the Royal Statistical Society. Series A (General)*, 1955, *118* (1), 56–66.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek, "IPUMS USA: Version 9.0," Minneapolis, MN: IPUMS 2020.

Sobel, M., "Structural Mobility, Circulation Mobility and the Analysis of Occupational Mobility: A Conceptual Mismatch," *American Sociological Review*, 1983, *4*, 721–727.

Sobel, Michael E, Michael Hout, and Otis Dudley Duncan, "Exchange, Structure, and Symmetry in Occupational Mobility," *American Journal of Sociology*, 1985, *91* (2), 359–372.

Song, Xi, Catherine G Massey, Karen A Rolf, Joseph P Ferrie, Jonathan L Rothbaum, and Yu Xie, "Long-Term Decline in Intergenerational Mobility in the United States Since the 1850s," *Proceedings of the National Academy of Sciences*, 2020, *117* (1), 251–258.

Ward, Zachary, "Intergenerational Mobility in American History: Accounting for Race and Measurement Error," *American Economic Review*, December 2023, *113* (12), 3213–48.

Willigan, J. Dennis and Katherine A. Lynch, *Sources and Methods of Historical Demography*, New York, NY: Academic Press, 1982.

Wright, Gavin, *Sharing the Prize: The Economics of the Civil Rights Revolution in the American South*, Harvard University Press, 2013.

Wrigley, Edward Anthony, Ros S. Davies, James E. Oeppen, and Roger S. Schofield, *English Population History from Family Reconstitution 1580–1837*, Cambridge, UK: Cambridge University Press, 1997.

Xie, Yu and Alexandra Killewald, "Intergenerational occupational mobility in Great Britain and the United States since 1850: Comment," *American Economic Review*, 2013, *103* (5), 2003–2020.

Appendix

FOR ONLINE PUBLICATION

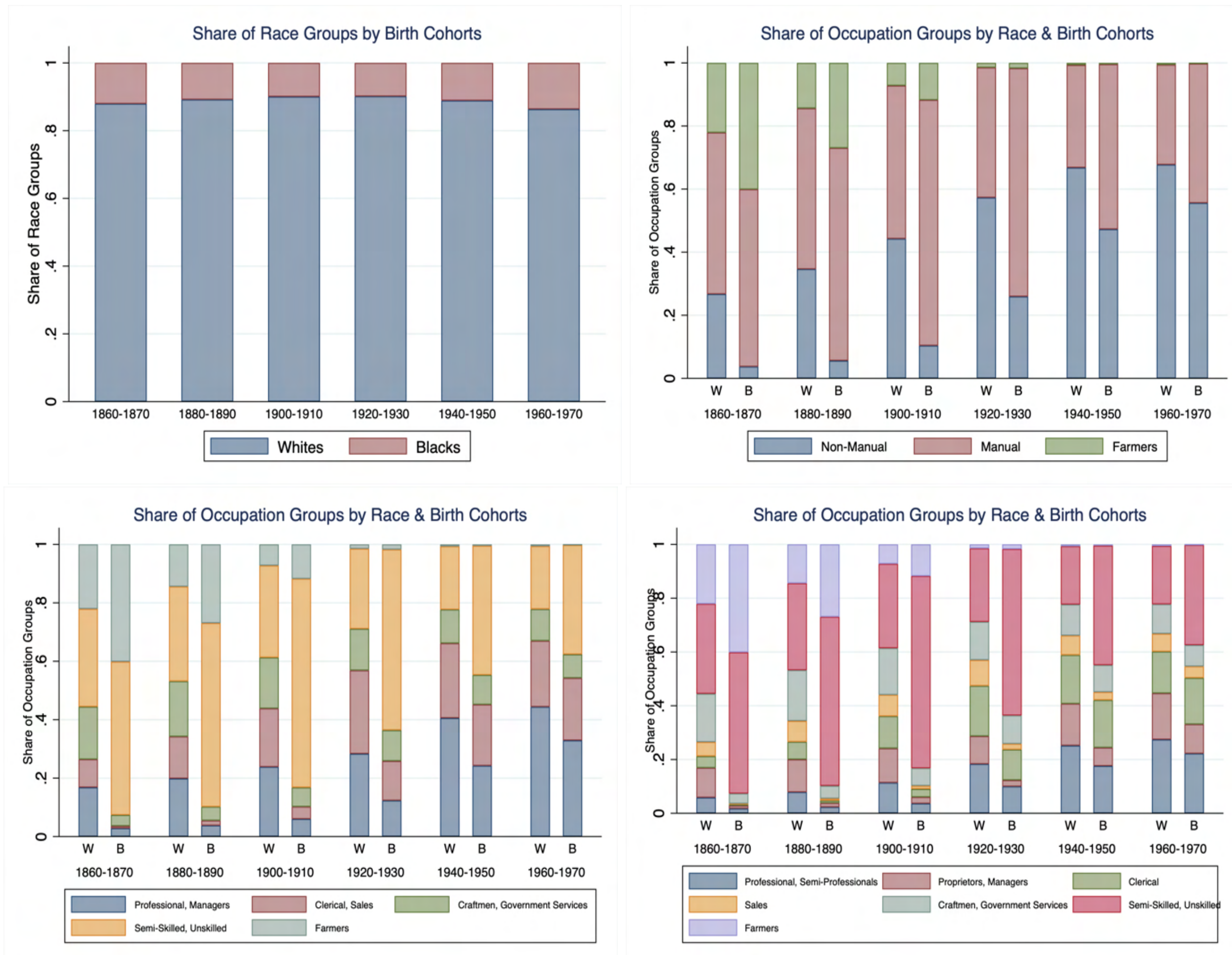
The Evolution of Black-White Differences in Occupational Mobility Across Post Civil War America

Steven Durlauf Gueyon Kim Dohyeon Lee Xi Song

Appendices

A Additional Figures

Figure A.1: Distribution of Race and Occupations

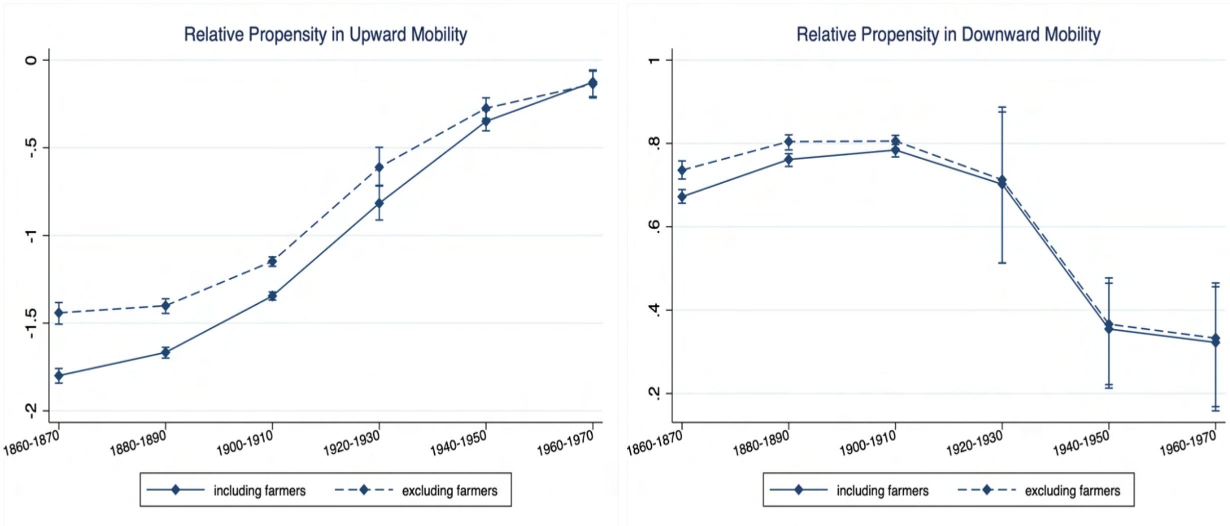


Notes: The graphs show the distribution of race (top, left) and the distribution of race and occupation (top right and bottom) for our sample of birth cohorts.

Figure A.2: Trends in Occupational Segregation by Race

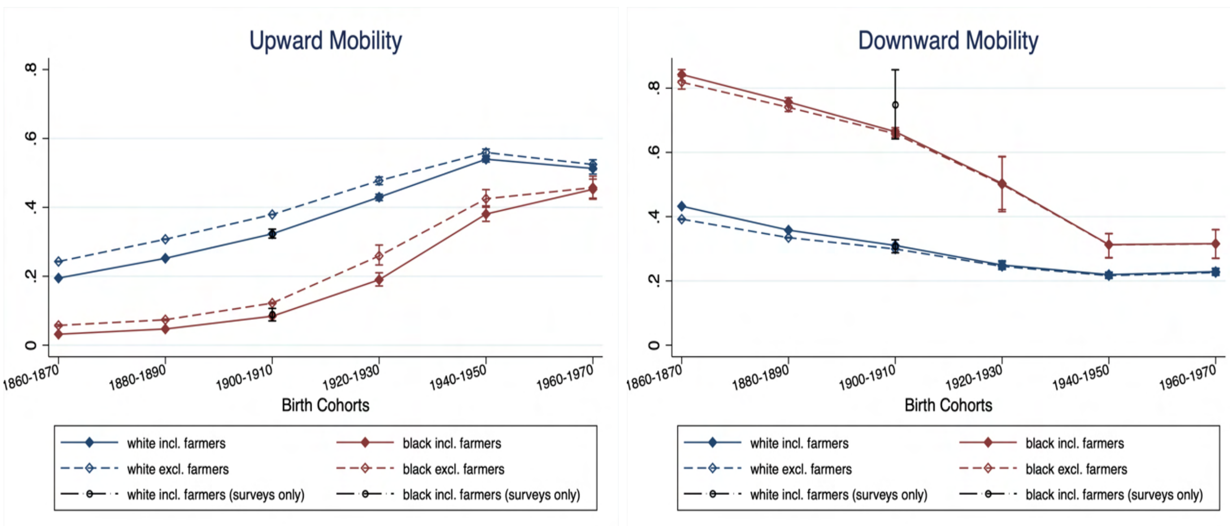
Notes: The graphs show trends in occupational segregation by race using (i) the actual distributions observed in each period—which correspond to years when each birth cohort of interest reaches its prime age—; and (ii) the occupational distributions at steady states using the transition matrices of each birth cohort.

Figure A.3: Relative Propensities in Mobility (3 categories)



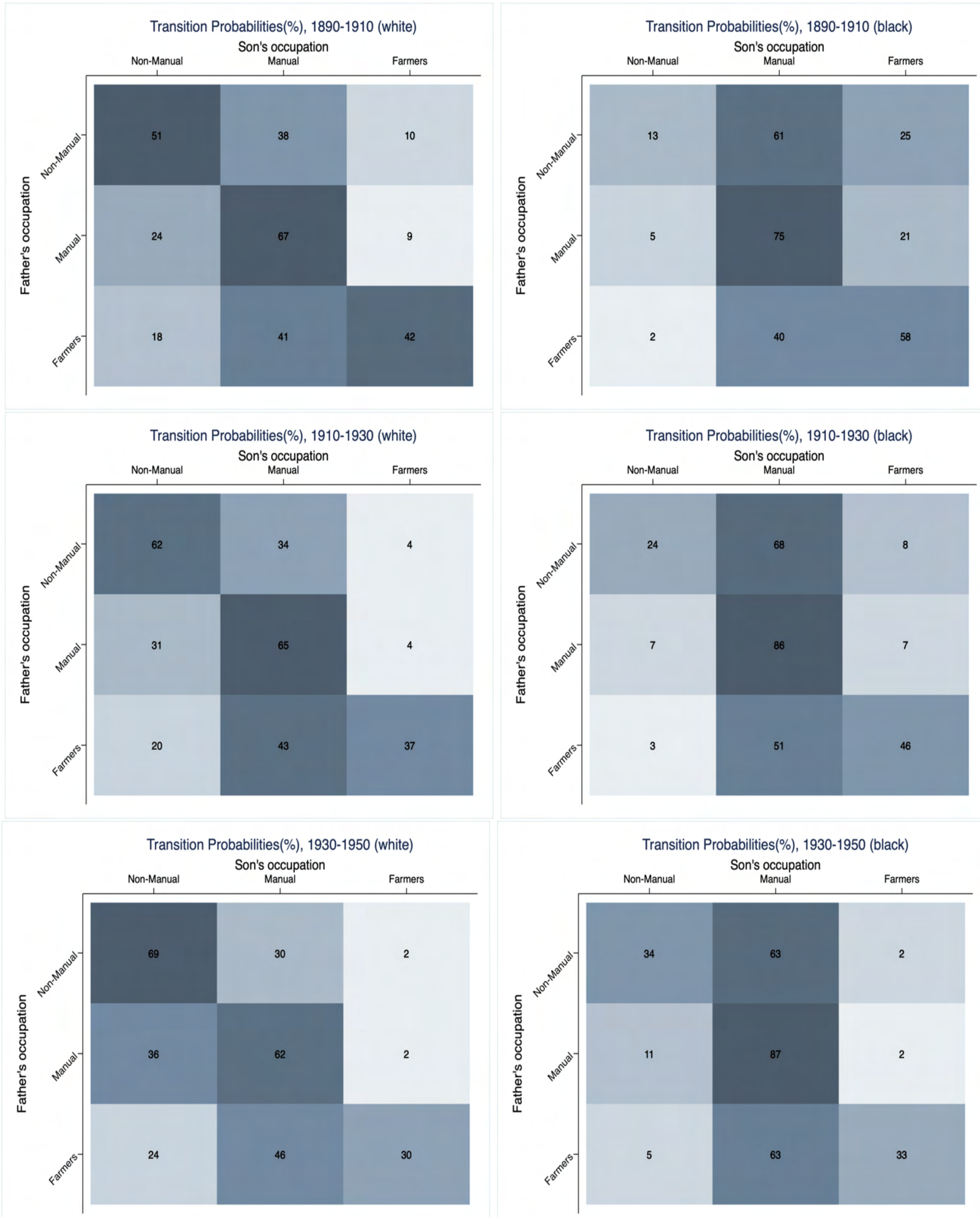
Notes: The graphs show trends in the Black-White log differences in transition probabilities ($log(p_b) - log(p_w)$) for upward (left) and downward (right) mobility. The solid graphs show trends in transitions from Farming or Manual to Non-Manual (left) and Non-Manual to Farming or Manual (right). The dashed graphs show trends in transitions from Manual to Non-Manual (left) and Non-Manual to Manual (right). Our results are based on 3 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure A.4: Mobility Patterns (3 categories) with MLP



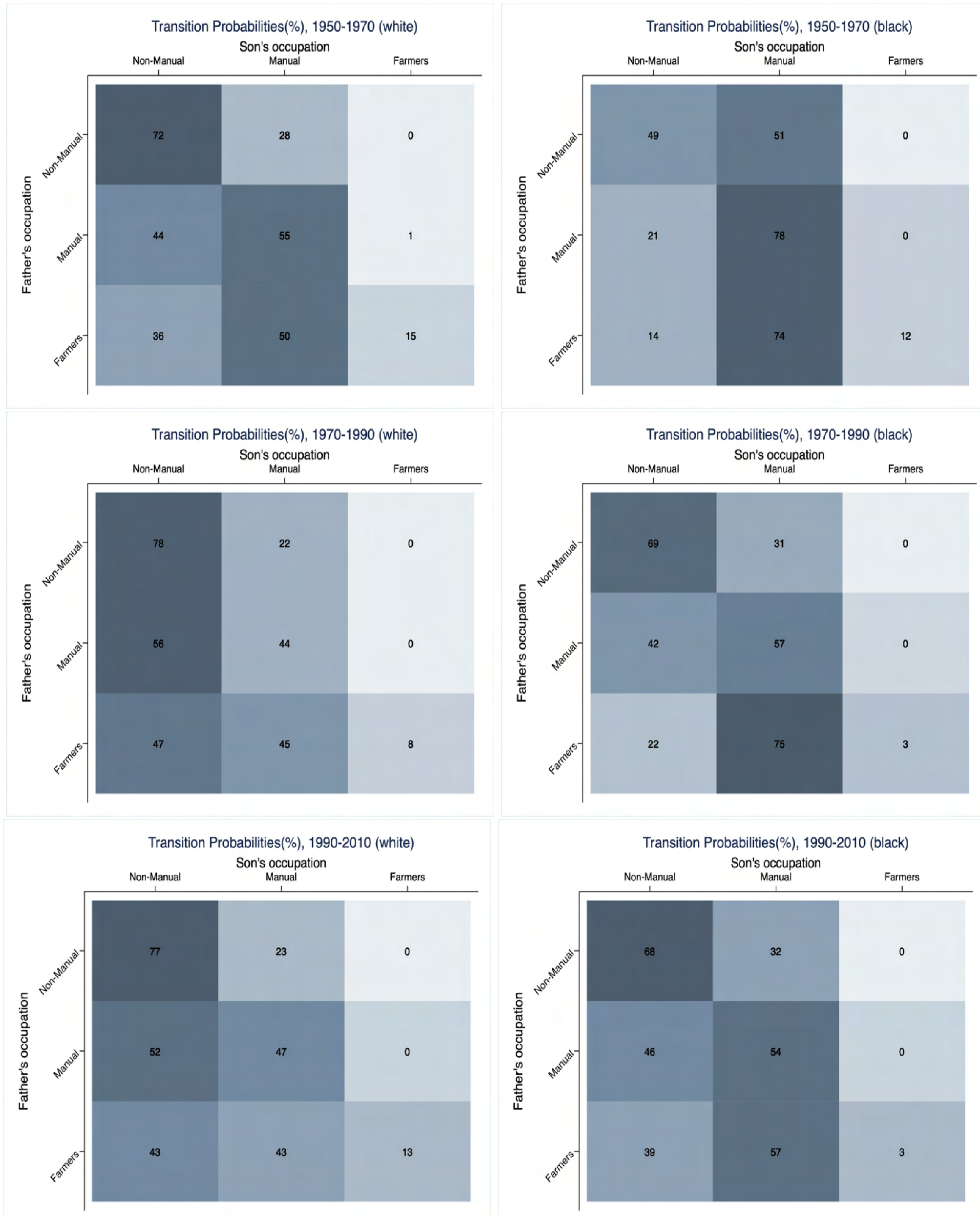
Notes: The graphs show upward (left) and downward (right) mobility patterns for White (in navy lines) and Black men (in maroon lines) using MLP for the historical sample. The solid graphs with solid diamond symbols show trends in transitions from Farming or Manual to Non-Manual (left) and Non-Manual to Farming or Manual (right). The dashed graphs with hollow diamond symbols show trends in transitions from Manual to Non-Manual (left) and Non-Manual to Manual (right). The long-dashed graphs with hollow circles use survey data and show trends in transitions from Farming or Manual to Non-Manual (left) and Non-Manual to Farming or Manual (right). Our results are based on 3 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure A.5: Transition Probabilities by Occupation Groups: 3 Categories



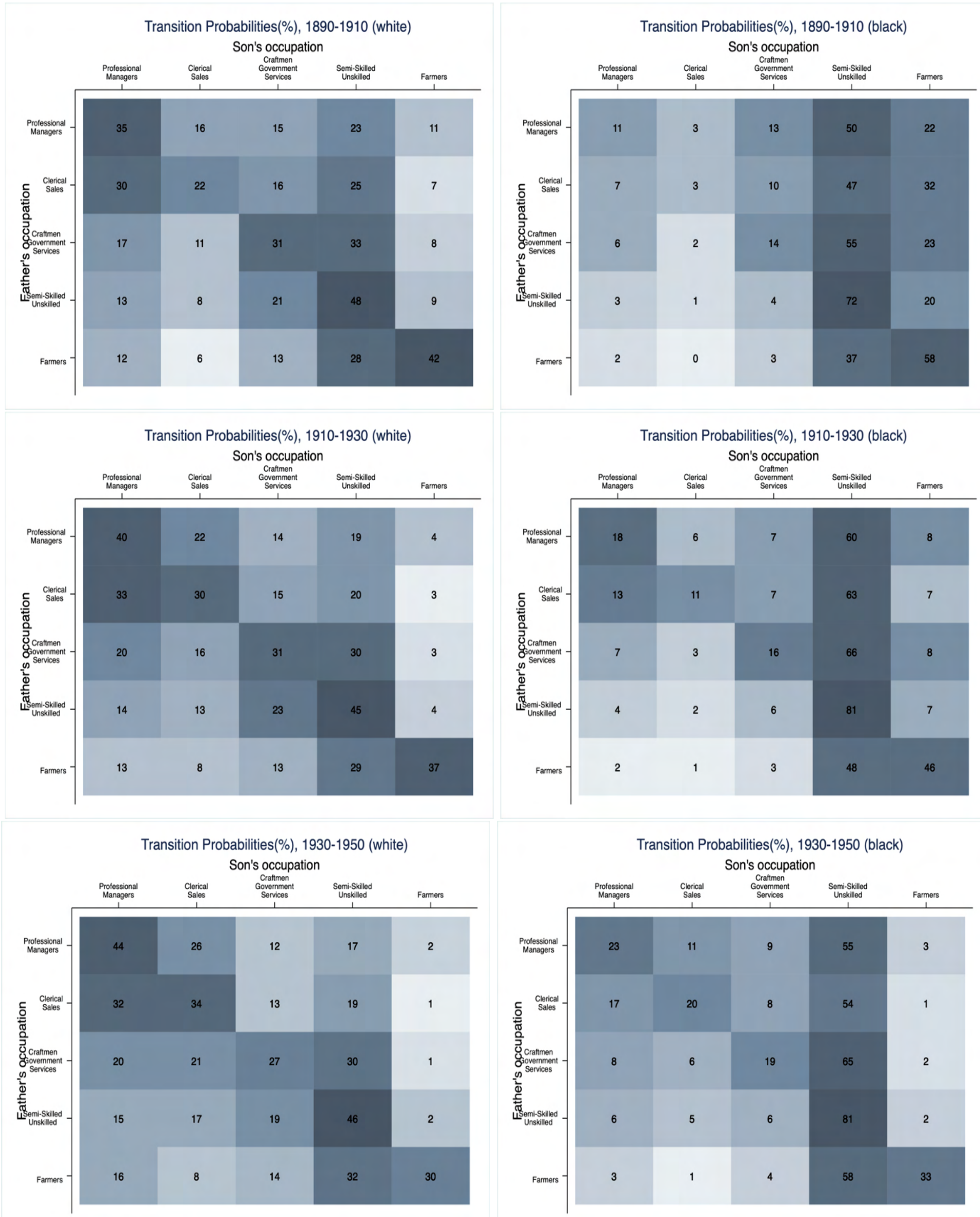
Notes: The graphs show the transition probabilities for the 1860-1910 birth cohorts using 3 occupational categories: White population (left) and Black population (right).

Figure A.6: Transition Probabilities by Occupation Groups: 3 Categories



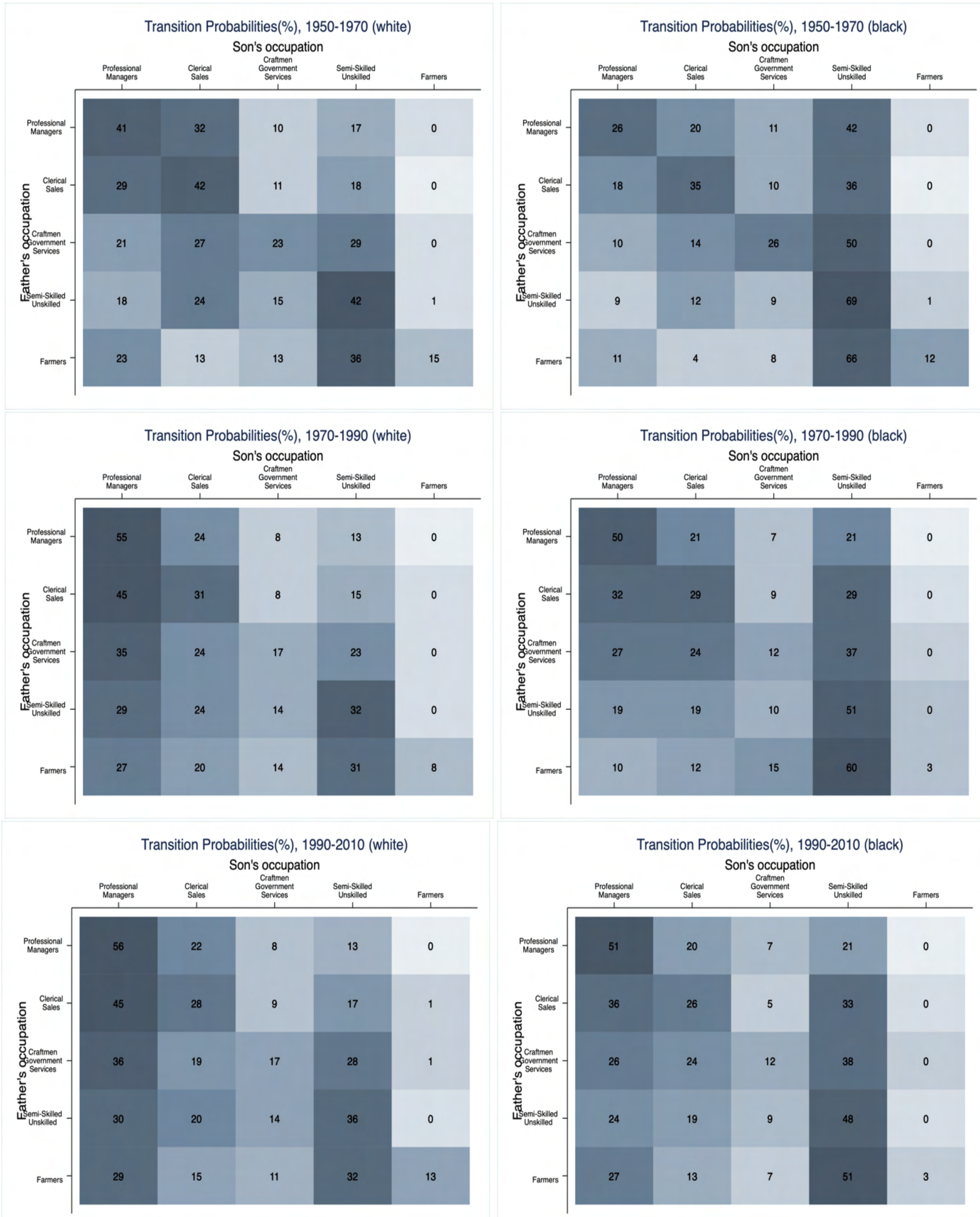
Notes: The graphs show the transition probabilities for the 1920-1970 birth cohorts using 3 occupational categories: White population (left) and Black population (right).

Figure A.7: Transition Probabilities by Occupation Groups: 5 Categories



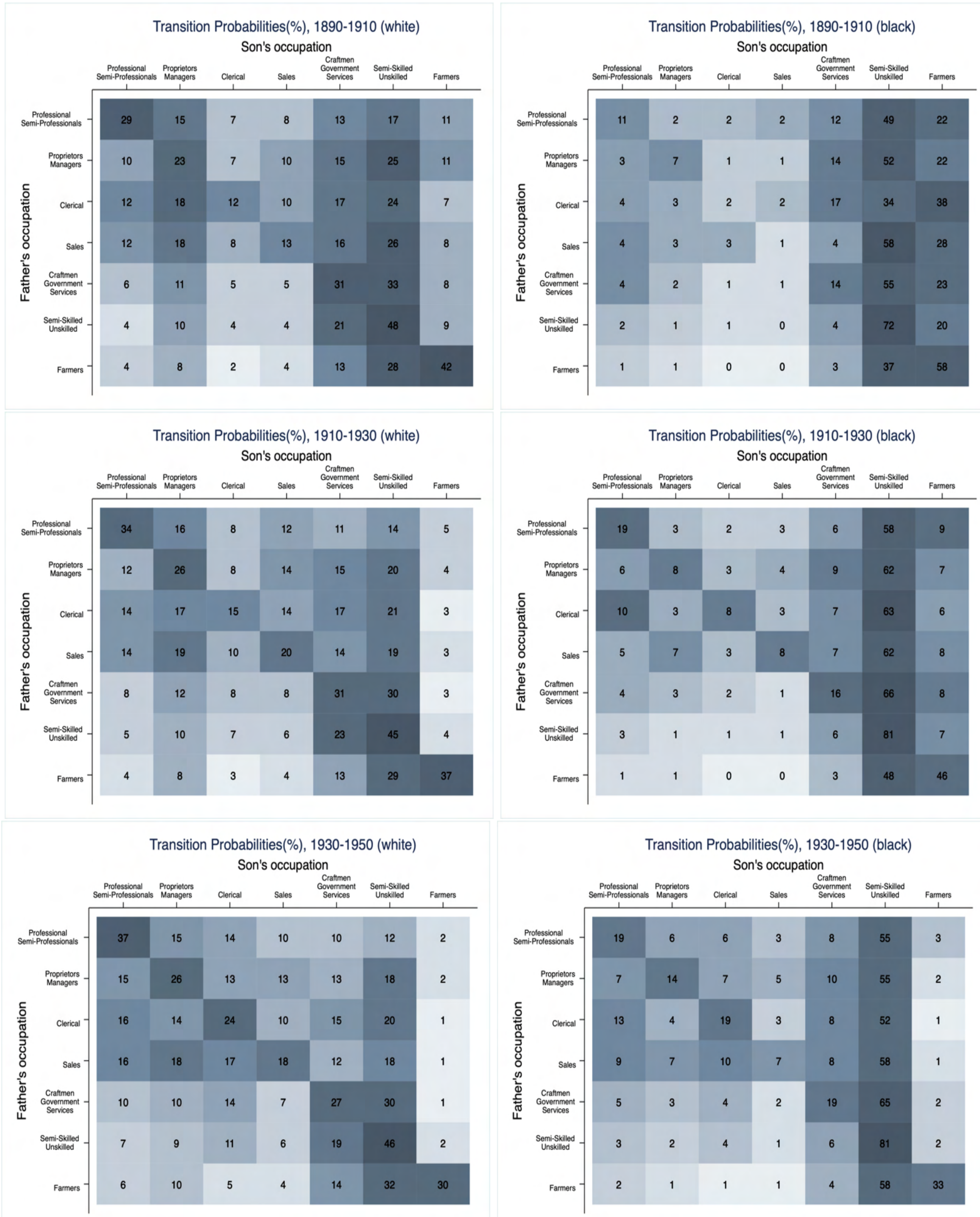
Notes: The graphs show the transition probabilities for the 1860-1910 birth cohorts using 5 occupational categories: White population (left) and Black population (right).

Figure A.8: Transition Probabilities by Occupation Groups: 5 Categories



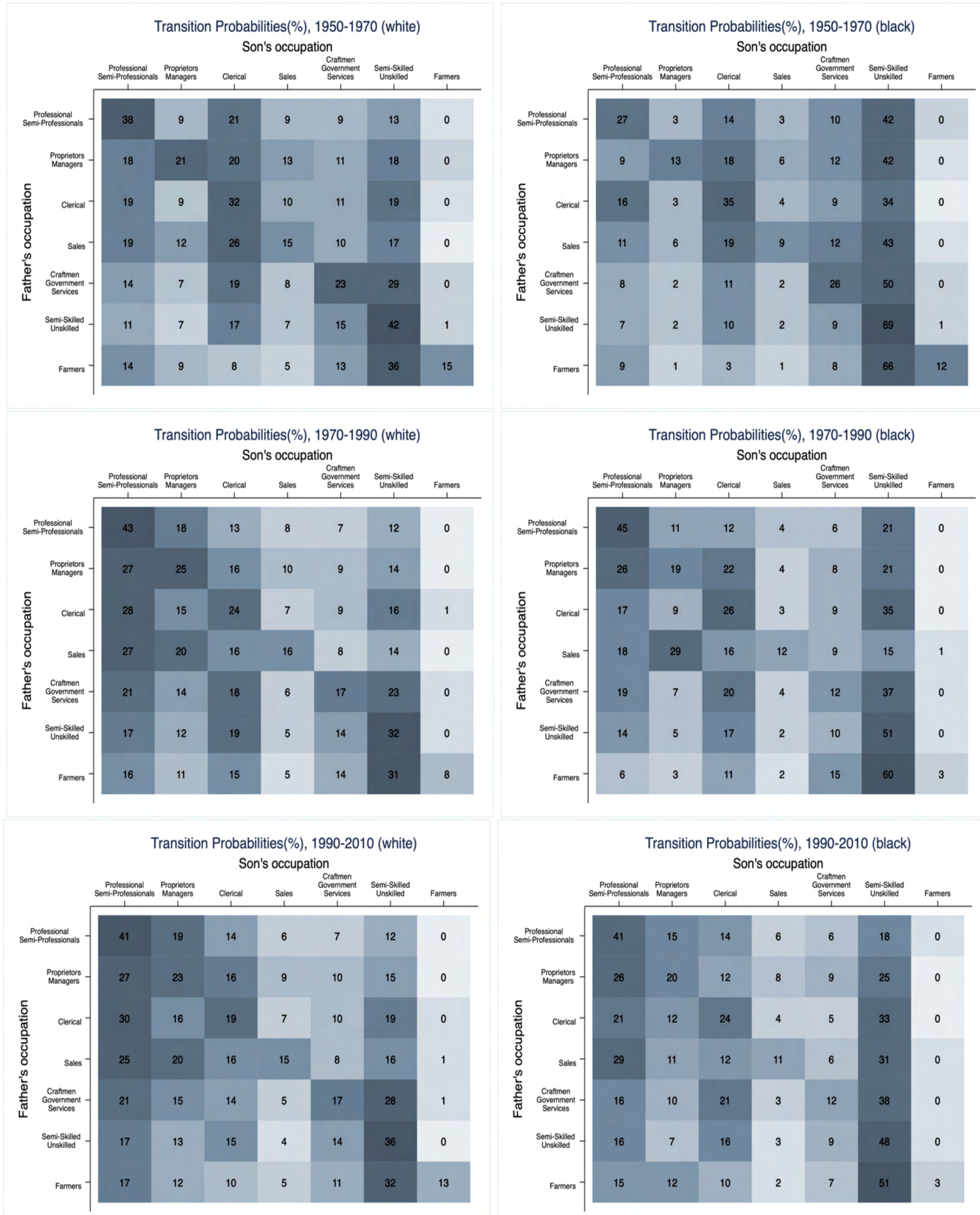
Notes: The graphs show the transition probabilities for the 1860-1910 birth cohorts using 7 occupational categories: White population (left) and Black population (right).

Figure A.9: Transition Probabilities by Occupation Groups: 7 Categories



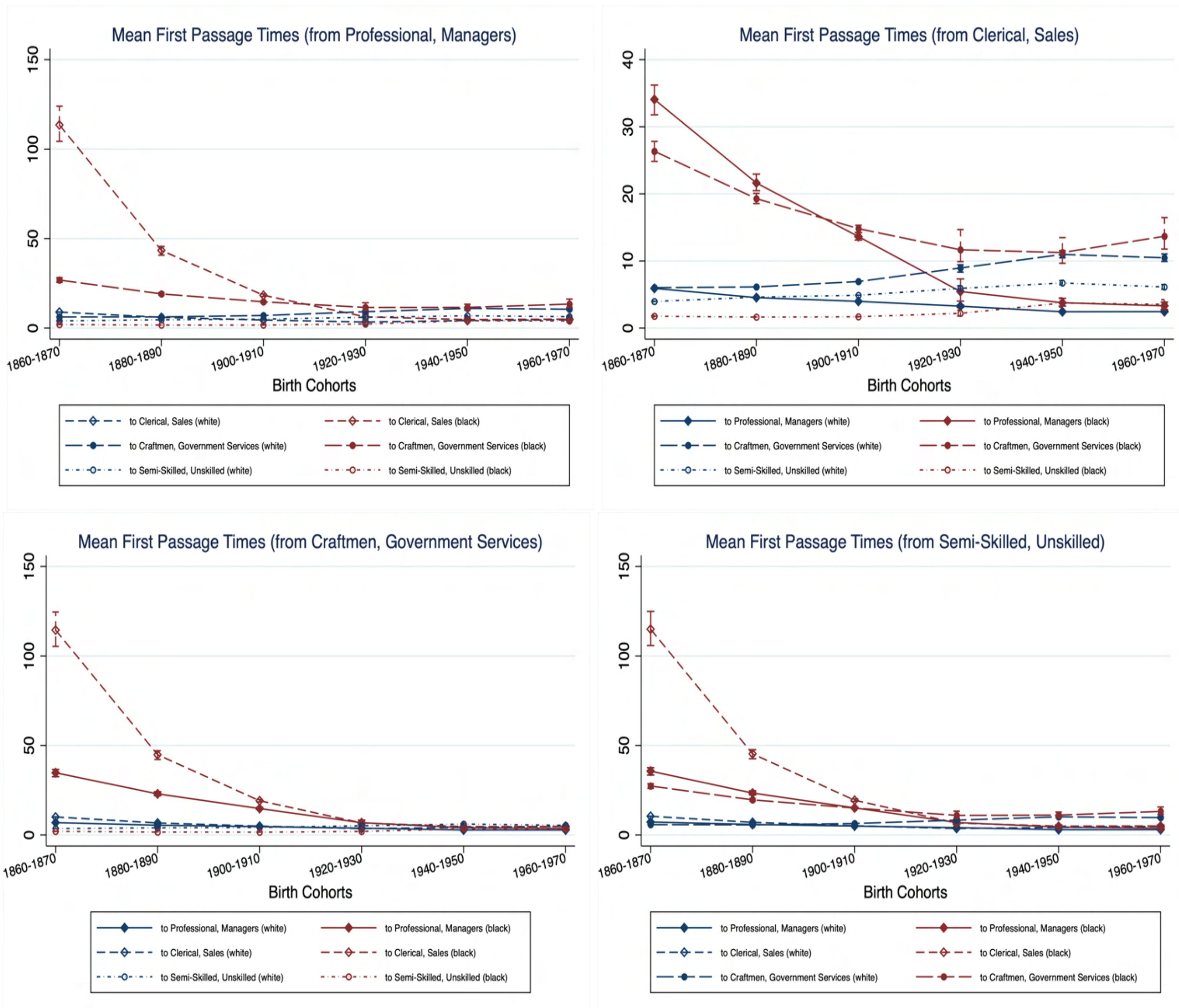
Notes: The graphs show the transition probabilities for 1880-1940 using 7 occupational categories: White population (left) and Black population (right).

Figure A.10: Transition Probabilities by Occupation Groups: 7 Categories



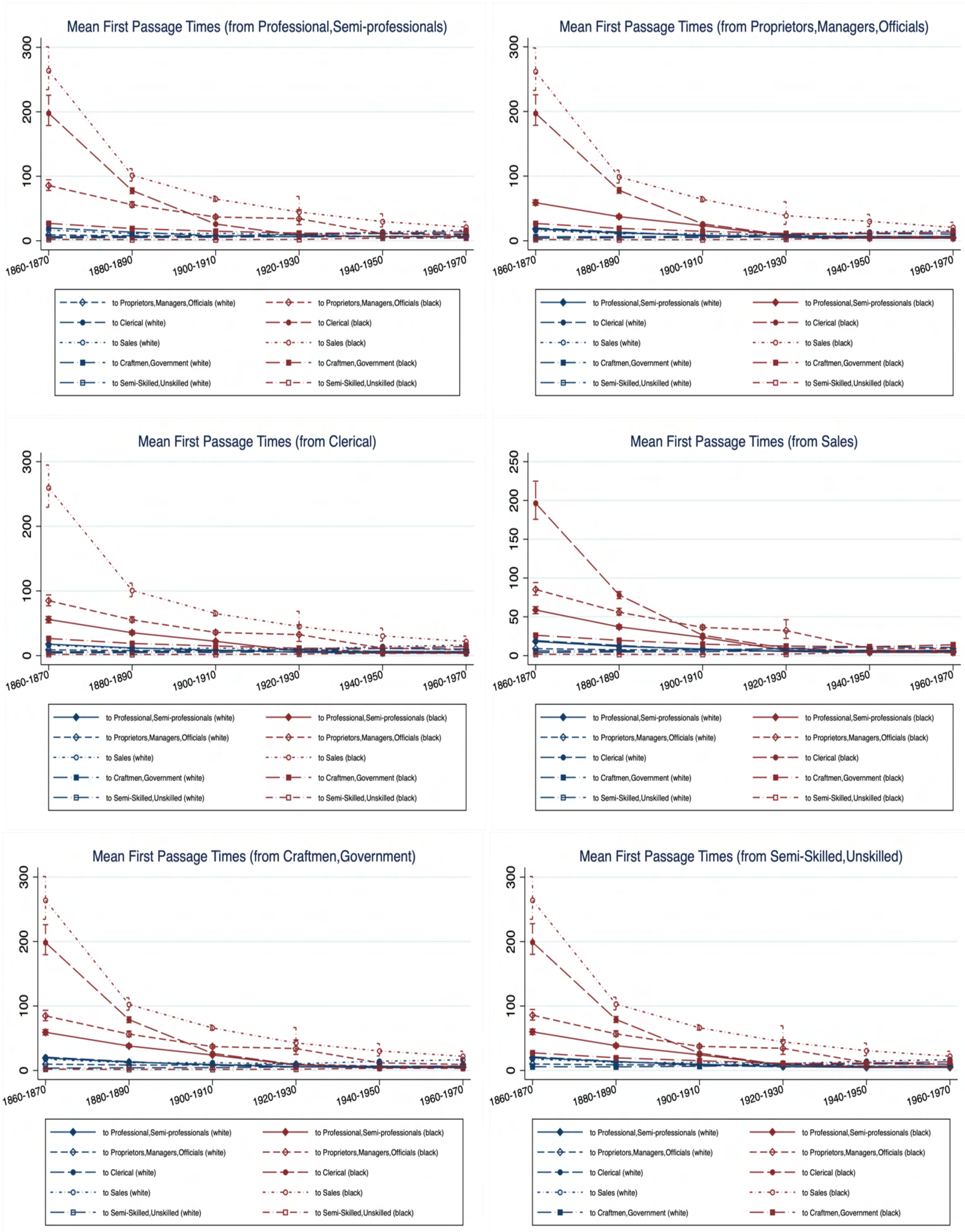
Notes: The graphs show the transition probabilities for the 1920-1970 birth cohorts using 7 occupational categories: White population (left) and Black population (right).

Figure A.11: Mean First Passage Times: 5 categories except Farmers



Notes: The graphs show the mean first passage times from Professional and Managerial occupations to others (top left); Clerical and Sales occupations to others (top right); Craftsmen and Government Services occupations to others (bottom left); Semi-Skilled and Unskilled occupations to others (bottom right) estimated using period-specific transition matrices for Black and White men. Our results are based on 5 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure A.12: Mean First Passage Times: 7 categories except Farmers

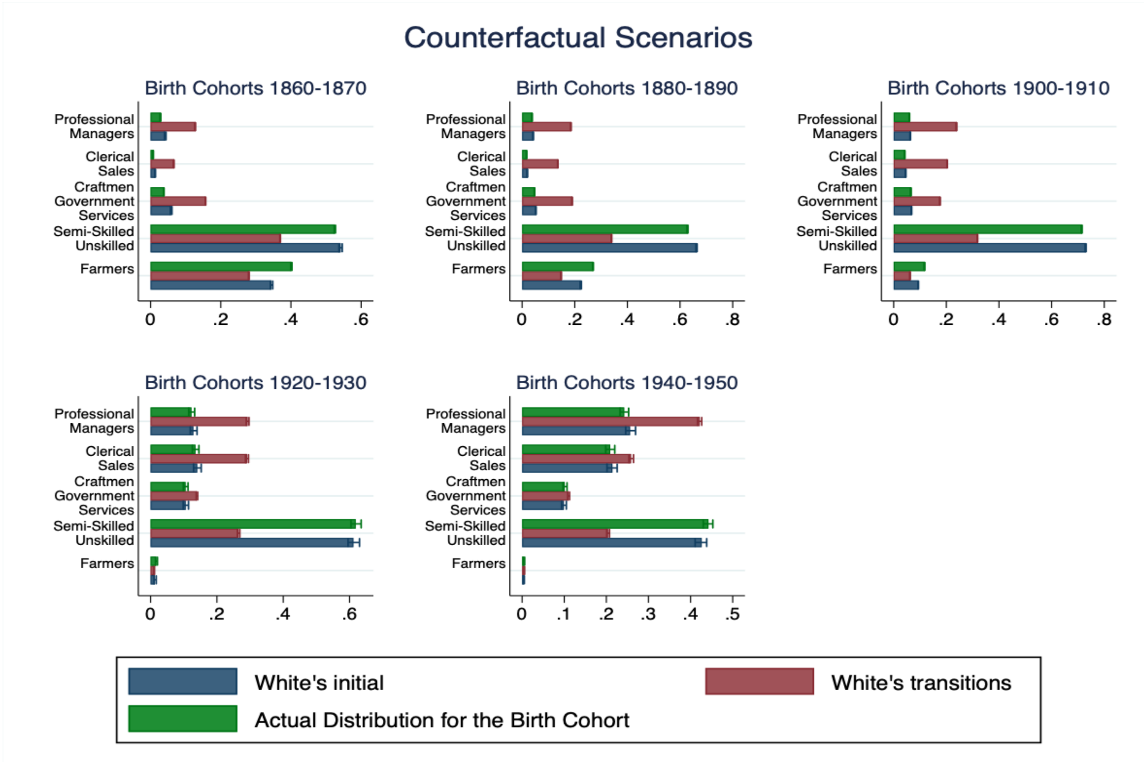


Notes: The graphs show the mean first passage times from Professional and Semi-professional occupations to others (top left); Proprietors, Managers, Officials to others (top right); Clerical occupations to others (middle left); Sales occupations to others (middle right); Craftsmen and Government Services (bottom left); and Semi-Skilled and Unskilled occupations to others (bottom right) estimated using period-specific transition matrices for Black and White men. Our results are based on 7 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure A.13: Counterfactual Occupational Distributions: 5 categories

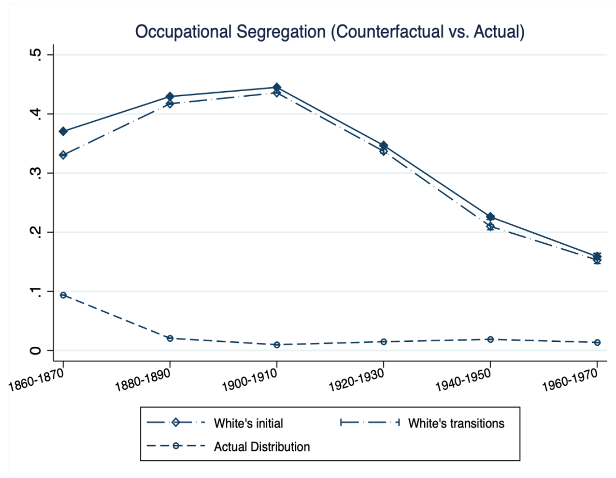
Notes: The graphs show the counterfactual occupational distributions for White (top) and Black men (bottom) in five different scenarios generated by sequentially updating the transition matrices over time. For example, in the top left panels, each group's current occupational distributions are contrasted with a counterfactual one where the transition matrix from the 1910 period—the 1860-1870 birth cohorts—is employed and remains unchanged over time. Our results are based on 5 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure A.14: Counterfactual Occupational Distributions (Black Men): 5 categories



Notes: The graphs compare the actual occupational distributions of Black men and counterfactual ones where (i) both racial groups start with the same initial condition ($\mu_0^B = \mu_0^W$) but experience different mobility dynamics; (ii) both racial groups undergo the same mobility dynamics ($\mathbf{P}_t^B = \mathbf{P}_t^W$) yet face different initial conditions. Our results are based on 5 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure A.15: Counterfactual Occupational Distributions (Black Men): 5 categories

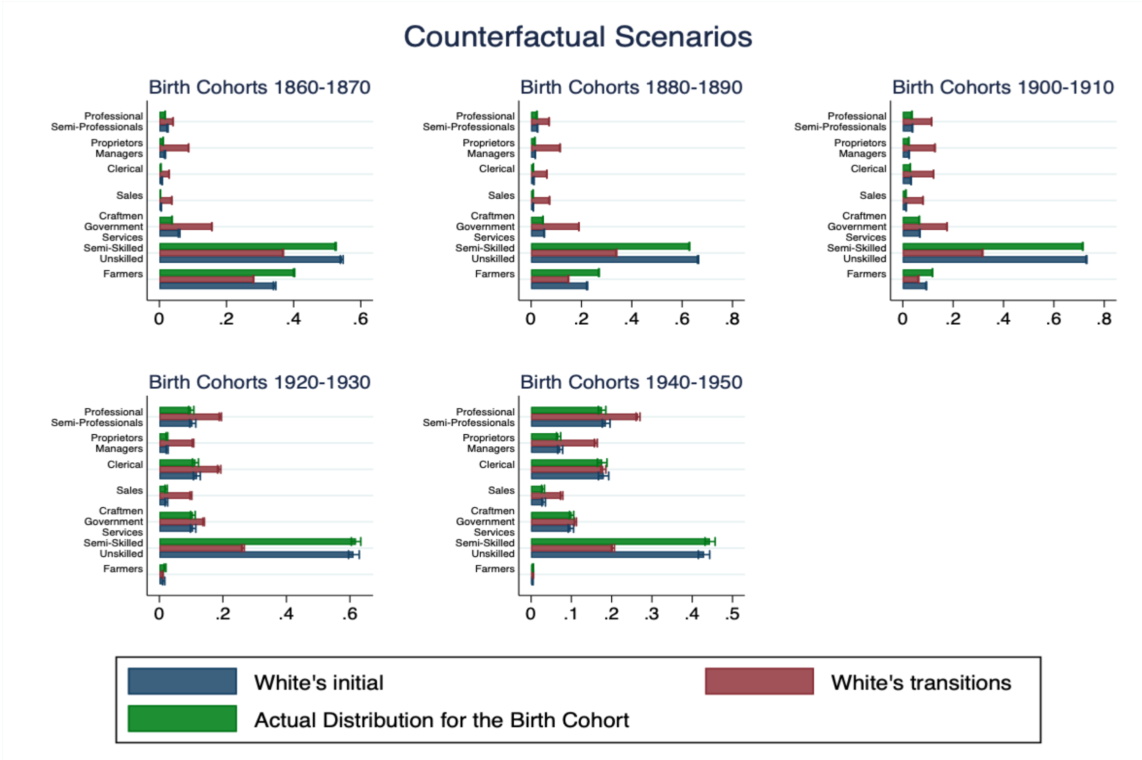


Notes: The graphs compare the degree of occupational segregation under each scenario: (i) both racial groups start with the same initial condition ($\mu_0^B = \mu_0^W$) but experience different mobility dynamics; (ii) both racial groups undergo the same mobility dynamics ($\mathbf{P}_t^B = \mathbf{P}_t^W$) yet face different initial conditions. Our results are based on 5 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure A.16: Counterfactual Occupational Distributions: 7 categories

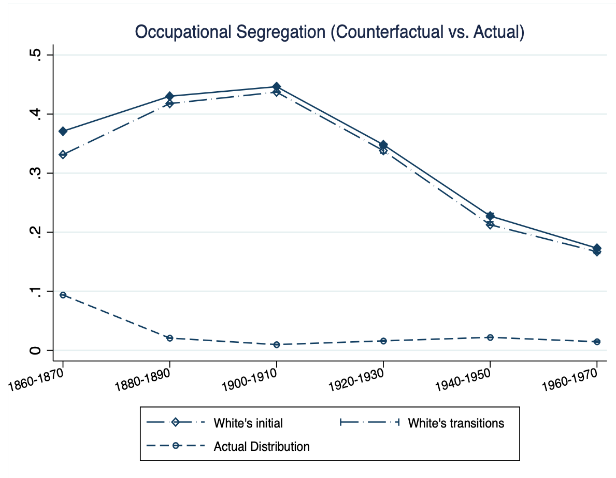
Notes: The graphs show the counterfactual occupational distributions for White (top) and Black men (bottom) in five different scenarios generated by sequentially updating the transition matrices over time. For example, in the top left panels, each group's current occupational distributions are contrasted with a counterfactual one where the transition matrix from the 1910 period—the 1860-1870 birth cohorts—is employed and remains unchanged over time. Our results are based on 7 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure A.17: Counterfactual Occupational Distributions (Black Men): 7 categories



Notes: The graphs compare the actual occupational distributions of Black men and counterfactual ones where (i) both racial groups start with the same initial condition ($\mu_0^B = \mu_0^W$) but experience different mobility dynamics; (ii) both racial groups undergo the same mobility dynamics ($\mathbf{P}_t^B = \mathbf{P}_t^W$) yet face different initial conditions. Our results are based on 7 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

Figure A.18: Counterfactual Occupational Distributions (Black Men): 7 categories



Notes: The graphs compare the degree of occupational segregation under each scenario: (i) both racial groups start with the same initial condition ($\mu_0^B = \mu_0^W$) but experience different mobility dynamics; (ii) both racial groups undergo the same mobility dynamics ($\mathbf{P}_t^B = \mathbf{P}_t^W$) yet face different initial conditions. Our results are based on 7 occupational categories. Bootstrapped 95 percent confidence intervals based on 200 draws are included.

B Additional Tables

Table B.1: Occupational Groupings

Occupation Groups (3)	Occupation Groups (5)	Occupation Groups (7)	Census 1950 Occupation Codes (269)
White-Collar	Professional & Managerial	Professional, Semi-professionals	0-99
		Proprietors, Managers, Officials	200-290
	Clerical & Sales	Clerical	300-390
		Sales	400-490
Blue-Collar	Craftmen, Government Services	Craftmen, Government Services	500-594, 762, 773, 781, 782
	Semi-Skilled, Unskilled	Semi-Skilled, Unskilled	595, 600-690, 700-790, 810-840, 910-970
Farming	Farming	Farming	100, 123

Notes: The table shows the mapping of the 3-digit Census occupation codes to our categorization of occupations (3, 5, and 7).

Table B.2: Period and Birth Cohort Definitions

Period (Benchmark Year)	Birth Years (Data)	Birth Cohorts (Definition)
1910	1856-1875	1860-1870
1930	1876-1895	1880-1890
1950	1896-1915	1900-1910
1970	1916-1935	1920-1930
1990	1936-1955	1940-1950
2010	1956-1975	1960-1970

Notes: The table shows how our definitions of periods and birth cohorts of individuals correspond to one another. For each snapshot year we consider, we focus on those aged between 35 and 55 to back out the birth years of interest. Then, we assign decadal definitions of birth cohorts.

Table B.3: Mean First Passage Times: Semi-Skilled and Unskilled

Birth Cohorts	Professional Semi-profs	Proprietors, Managers	Clerical	Sales	Craftmen, Government	Farmers
Panel A: **Whites**						
1860-1870	17.56	9.07	20.17	18.15	5.46	10.98
1880-1890	12.43	8.24	13.31	11.37	5.38	25.72
1900-1910	9.61	8.78	7.92	12.14	6.44	59.09
1920-1930	6.94	12.21	5.33	11.6	8.21	197.95
1940-1950	4.61	6.72	6.1	14.32	10.23	239.87
1960-1970	4.63	6.39	6.67	16.68	9.63	251.36
Panel B: **Blacks**						
1860-1870	52.95	91.02	207.36	331.89	27.69	4.89
1880-1890	37.02	59.77	67.93	99.9	17.84	14.69
1900-1910	24.55	38.46	26.15	64.11	15.03	47.41
1920-1930	12.07	43.06	9.09	37.37	10.5	236.7
1940-1950	6.15	12.27	6.01	30.66	11.28	454.39
1960-1970	5.41	9.6	6.51	22.08	13.5	621.59

Notes: The table shows the mean first passage times from Semi-Skilled and Unskilled to other occupations. We use the 5-category occupation definition.

Table B.4: Kitagawa-Oaxaca-Blinder Decomposition of occupational distributions

Occupation (Initial Share)	Period (from)	Period (to)	(1) ΔOcc Shares	(2) ΔDistribution	(3) ΔTransitions	(4) ΔInteraction
Non-Manual (0.175)	1910	1930	0.081	0.017	0.058	0.005
	1930	1950	0.085	0.034	0.049	0.002
	1950	1970	0.107	0.040	0.073	-0.006
	1970	1990	0.131	0.038	0.098	-0.006
	1990	2010	0.014	0.032	-0.019	0.002
Manual (0.454)	1910	1930	0.005	0.009	0.001	-0.005
	1930	1950	-0.018	-0.002	-0.009	-0.007
	1950	1970	-0.052	-0.013	-0.031	-0.008
	1970	1990	-0.117	-0.027	-0.090	0.000
	1990	2010	-0.081	-0.063	-0.030	0.012

Notes: The table shows the results of the decomposition exercise across Non-Manual and Manual occupations, where each row represents (i) changes in the occupational shares from one period to another (Δ Occ Shares) and (ii) how much each component (ΔDistribution, ΔTransitions, and ΔInteraction) contributes to the occupation-specific change in shares. We use the 3-category occupation definition.

C Data Sources

General Social Survey (GSS) is a large-scale, cross-sectional survey that has been implemented since 1972 by the National Opinion Research Center (NORC) at the University of Chicago. The survey was conducted annually in most years before 1994 and changed to a biennial basis thereafter. It is designed to be representative of all adults in US households aged 18 and older. The data have been widely used in studies on societal changes in individual attributes and attitudes and population composition. Our analysis relies on data from the years 1972 to 2018 and applies sample weight (wtssall) to account for the multistage sampling design. We keep oversamples of Black respondents in GSS 1982 and 1987. Respondents were asked to report their own occupation and education at the time of the survey and their father's occupation and education while they "were growing up." The data are publicly available and can be downloaded from the GSS website: `http://gss.norc.org/get-the-data/`.

Occupational Change in a Generation (OCG) I and II. OCG I is a large-scale survey of U.S. social mobility, implemented as a supplement to the March 1962 Current Population Survey (CPS). The target population of OCG I was U.S. males aged 20 to 64 in the civilian, noninstitutional population. We applied a sample weight using V24 to account for the multistage sampling design (see OCG I codebook Introduction p.3 for a description of the construction of sample weight). Occupational Changes in a Generation II (OCG II) was collected in 1973 as a supplement to the March 1973 Current Population Survey. It was designed as a strict replication of OCG I, but also incorporated some new questions about social background and career development. In both OCG I & II, Respondents were asked to report their own occupations at the time of the survey and their fathers' occupations when the respondent was 16 years old. Individuals were also asked to report their own education at the time of the survey and their fathers' education when the respondent was 16 years old. The data are publicly available and can be downloaded from the OCG website: `http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/6162`.

National Survey of Families and Households (NSFH) was a nationally representative household survey designed to provide information on family life. It was conducted in 3 waves: 1987–1988, 1992–1994, and 2001–2002. Households were randomly selected from 1,700 selection units that resulted from 17 enumeration districts within each of the 100 primary sampling units. For the first wave of data collection, a primary respondent at least 19 years of age was randomly selected from each household to participate in a personal interview. We used data from waves 1 to 3 and applied sample weight (WEIGHT) to account for the multistage sampling design. Respondents were asked to report their own occupations in each wave and their fathers' occupations when the respondent was 16 years old. Respondents were also asked to report their own education and their fathers' education when the respondent was 16 years old in the 1987–1988 survey. The data are publicly available and can be downloaded from the NSFH website: `http://www.ssc.wisc.edu/nsfh/home2.htm`.

Panel Study of Income Dynamics (PSID) carried out at the Survey Research Center at the University of Michigan, is a longitudinal panel household survey in the US. The survey collects extensive economic, social, and health items on individuals and their household members over the life course and across generations. Begun in 1968, the PSID follows over 18,000 household members from roughly 5,000 US families. The study interviewed individuals annually until 1997 and biennially thereafter. All original 1968 PSID respondents and their offspring are considered to carry the PSID "gene" and thus are permanent PSID respondents. Their demographic and socioeconomic information, such as age and occupation, is gathered in each wave of the PSID survey and can be linked across years. The PSID project also provides a "Family Identification Mapping System" (FIMS) tool designed to link family members across generations. The data consist of 2 distinct samples: SRC (Survey Research Center) is a nationally representative household sample based on a stratified, multistage selection of the civilian noninstitutional population of the US; SEO (Survey of Economic Opportunity) is a national sample of low-income families with household heads under age 60 in 1968. Occupation is collected for household heads and their spouses in all PSID waves. We first harmonized all the occupation variables by converting them to the 1950 Census scheme and then chose individuals' occupations measured in the year closest to age 40 as the lifetime occupation. That is, if occupation at age 40 was unavailable, we then looked for occupation at ages 41, 39, 42, 38, and so forth. For individuals whose fathers were not PSID respondents and thus were not directly observed, we relied on the question that asked household heads to report the father's occupation in waves 1997–2019. If the father's occupation was reported multiple times, we chose the mode of the data values. The level of education was asked for household heads in all PSID waves. For individuals whose fathers were not PSID respondents and thus were not directly observed, we relied on the question that asked household heads to report the father's education. We chose the highest level of education for individuals and their fathers across waves. The data are publicly available and can be downloaded from the PSID website: `https://simba.isr.umich.edu/data/data.aspx`.

National Longitudinal Survey of Youth 1979 (NLSY79) was another longitudinal cohort study conducted by the Bureau of Labor Statistics under the United States Department of Labor. The project followed the lives of a nationally representative sample of 12,686 respondents born between 1957 and 1964. The respondents were aged 14—22 when first interviewed in 1979 and revisited annually until 1994 and biennially thereafter. We used data from 1979 to 2012 and applied sample weight (R0216100) to account for the multistage sampling design. Respondents were asked to report their occupations in each survey year and their fathers' (or stepfathers') longest jobs in 1978 during the first wave. Individuals were also asked to report their own education in each survey year and their fathers' (or stepfathers') education in 1979. We relied on respondents' highest education across survey years. The data are publicly available and can be downloaded from the NLS website: `http://www.nlsinfo.org/investigator/pages/login.jsp`.

National Longitudinal Survey of Young Men (NLS–YM) was another dataset among the several longitudinal cohort studies conducted by the Bureau of Labor Statistics under the United States Department of Labor between 1966 and 1981. The project began with a nationally representative sample of 5,225 American males aged 14 to 24 in 1966 and was discontinued in 1981 when the respondents were 29 to 39, at the time of their last interview. Respondents were surveyed annually between 1966 and 1971, and in the subsequent years of 1973, 1975, 1976, 1978, 1980, and 1981. The data have been used to study the education and labor market experiences of men during their early careers, such as their educational attainment and expectations, school-to-work transitions, labor market attachment, and activities related to crime, delinquency, and school discipline. We used data from the years 1966 to 1981 (12 rounds) and applied the sample weight (R0000200) to account for the multistage sampling design. Respondents were asked to report their current or the last occupations and those of their fathers in each survey year between 1966 and 1969. To make the data comparable with other datasets used in the analysis, we relied on respondents' occupations reported in the last wave of the survey and their fathers' occupations reported in 1966. If the father's information was missing in the first wave, occupation in the next available year was used. Respondents were asked to report their highest levels of education in each survey year and those of their fathers in 1966. We relied on respondents' highest education ever reported across survey years. The data are publicly available and can be downloaded from the NLS website: `http://www.nlsinfo.org/investigator/pages/login.jsp`.

National Longitudinal Survey of Older Men (NLS-OM) is one of a set of longitudinal cohort studies conducted by the Bureau of Labor Statistics under the United States Department of Labor from 1966 to 1990. The project began with a nationally representative sample of 5,020 American men aged 45 to 59 in 1966 and was discontinued when respondents were 69 to 83 at the time of their last interviews in 1990. Respondents were surveyed annually between 1966 and 1969, and in the subsequent years of 1971, 1973, 1975, 1976, 1978, 1980, 1981, 1983, and 1990. The final interview was conducted with both living older male respondents and widows or other family members of deceased respondents. The data have mainly been used to study labor market activities of middle-aged men and men close to retirement, such as their work and unemployment experiences, retirement planning, work and health insurance, and job evaluation and satisfaction. We used data from the years 1966 to 1976 (8 rounds) and applied the sample weight variable R0000200 to account for the multistage sampling design. Respondents were asked to report their own occupations in each survey year, and in the year 1966, they were asked about their fathers' occupations when the respondents were 15 years old. Respondents who did not live with their fathers at age 15 were asked to report household heads' occupations. The mobility analysis relies on respondents' occupations reported in the year 1966. If this information was missing, occupation in the next available year was used. The data are publicly available and can be downloaded from the NLS website: `http://www.nlsinfo.org/investigator/pages/login.jsp`.

Survey of Income and Program Participation (SIPP), conducted by the United States Census Bureau, is a longitudinal panel survey that is designed to provide income of individuals and households and their participation in income transfer programs. Other topics include education, occupation, family dynamics, health insurance, childcare, and food security. The SIPP survey design is a continuous series of national panels, with the sample size ranging from approximately 14,000 to 52,000 interviewed households. The original goal was to have all panels participate for a 32-month period (8 waves), with each panel randomly divided into one of 4 rotation groups. Each rotation group is interviewed in a separate month. Four rotation groups constitute one wave of interviewing. At each interview, respondents provide information covering the 4 months since the previous interview. The first interview began in October 1983 for the 1984 panel. Subsequent panels began interviews in February of each year. We used data from survey years 1986, 1987, and 1988 because the information about fathers' occupations and education was collected only in a topical module in wave 2 during these years. We applied the sample weight ($FNLWGT_5$) to account for the multistage sampling design and kept respondents aged 25 to 64. Respondents were asked to report their own occupations during each survey year as well as their fathers' occupations when the respondent was 16 years old. Respondents were also asked to report their own education during each survey year as well as their fathers' education when the respondent was 16 years old. The data are publicly available and can be downloaded from the SIPP website: `http://www.census.gov/sipp/`.

Americans View Their Mental Health was sponsored by the US Congress Joint Commission on Mental Illness and Health, and the first survey was conducted in 1957 by the Survey Research Center at the University of Michigan. A sample of Americans were interviewed on marriage, parenthood, employment, and general social relationships. The second survey was implemented in 1976 with the sponsorship of the National Institute of Mental Health. All the variables used in our project are taken from the publicly available data set (Americans View Their Mental Health, 1957 and 1976: Selected Variables (ICPSR 7949)). The dataset includes variables for respondents' occupation (V133) and respondents' father's occupation (V256) from the 1976 survey round. Both variables code occupation in the 1970 US Census Classification System. We recode the 1970 Census occupation codes into the 1950 Census occupation codes using the IPUMS crosswalk. Race and ethnicity variable in the dataset (V244) classifies respondents into three groups (White, Black, and others). We use the same grouping and only recode the group numbers (1 for White, 2 for Black, and 3 for others).

National Survey of Black Americans is a series of surveys conducted in 4 waves (1979-1980, 1987-1988, 1988-1989, and 1992). The survey's target population is Black US citizens 18 years of age or older. Our project uses Wave-1 variables from a publicly available data set (National Survey of Black Americans, Waves 1-4, 1979-1980, 1987-1988, 1988-1989, 1992 (ICPSR 6668)). Respondent's occupation variable (V307) classifies respondents' occupation into the 1970 Census 3-digit codes. We recoded this variable into the 1950 Census occupation codes using the IPUMS crosswalk. The data set also includes variables for the occupation of the respondents' father (V1471) and mother (V1477). The parents' occupation variables assign two-digit codes and have substantially fewer categories than the 1970 US Census occupation classification. We mapped these variables to a new occupation variable (occ_b) that we constructed by collapsing the 1950 Census occupation code.

Figure C.1: Father–Son Pair Construction in Historical Samples, 1850–1940

Data sources: IPUMS 1850–1940 linked full-count Censuses (MLP); Census Linking Project 1850–1940 linked full-count Censuses (CLP). IPUMS full-count 1940 Census.
Note: This plot shows four linking scenarios in our sample using historical U.S. full-count Censuses. In Scenario (A), parents and offspring lived in the same household in one of the Census years during 1850–1930 when the offspring was a child, and they were observed in different households in a later census when the offspring became an adult. In Scenario (B), parents and offspring lived in the same household during the offspring's childhood and adulthood. In Scenario C, parents and offspring lived in the same household when the offspring was an adult, and no childhood records were available. In Scenario D, parents and offspring lived in the same household in one of the Census years during 1850–1930 when the offspring was a child, and only the child can be linked to a later census.

D Theory and Proofs

D.1 Markov Chain Mobility Models

Suppose that we observe f_i fathers in occupation i and s_j sons in occupation j. The transition matrix $\mathbf{P}$ that transforms the distribution of fathers into the distribution of sons satisfies, $s_j = \sum_{i=1}^{I} f_i \cdot p_{Y_2=j|Y_1=i}$ for $j = 1, 2, ..., J$, where $p_{Y_2=j|Y_1=i}$ denotes the probability that the son of a father in occupation i ends up in occupation j. Assuming that mobility rates are fixed over time, the distribution of descendants, namely the expected proportion of men in various occupations after t generations, is, $\mu_{\mathbf{t}} = \mu_{\mathbf{0}} \cdot \mathbf{P}^t$.

If the transition matrix is *regular* (irreducible, positive recurrent, and aperiodic), as time progresses, the process will "forget" its initial distribution and converge to a unique equilibrium distribution,

$$\lim_{t\to\infty} \mu_{\mathbf{0}} \cdot \mathbf{P}^t = \mu^* \tag{11}$$

where μ^* is called the invariant or steady-state vector of the Markov process. At the aggregate level, the invariant measure describes a cross-sectional density of occupations that, once achieved, stops evolving and hence reaches a steady state. At the individual level, the measure describes the common probability density to which descendants of families converge, in accordance with the described mobility process. These properties indicate how, in the short run, the initial distribution of ancestors influences future generations, but the influence diminishes as time passes. In the long run, the descendant distribution is only determined by the transition matrix $\mathbf{P}$.

D.2 Mixing Analysis: Example

To illustrate our measures of mixing calculations (AD, AID), this section provides a simple example. Let $P_\alpha = \alpha I + (1-\alpha)\frac{ii'}{N}$, where I is the identity matrix with dimension N, i is an N-dimensional column vector consisting of 1's, and $\alpha \in [0,1)$. Note that $\frac{ii'}{N}$ is the transition matrix that makes the next period's distribution equal to i', since $\mu_0 \cdot \frac{ii'}{N} = (\mu_0 i) \cdot \frac{i'}{N} = \frac{i'}{N}$, where the last equation uses $\sum_k \mu_{0,k} = 1$ for any distribution μ_0. In other words, P_α is a combination of maintaining the original distribution with probability α and fully randomizing with probability $(1-\alpha)$.

Claim: $P_\alpha \cdot P_\beta = P_{\alpha\beta}$

Proof:

$$P_\alpha \cdot P_\beta = \Big(\alpha I + (1-\alpha)\frac{ii'}{N}\Big)\Big(\beta I + (1-\beta)\frac{ii'}{N}\Big)$$
$$= \alpha\beta I + \Big(\alpha(1-\beta) + \beta(1-\alpha) + (1-\alpha)(1-\beta)\Big)\frac{ii'}{N}$$
$$= \alpha\beta I + (1-\alpha\beta)\frac{ii'}{N}$$

where the second equation uses the fact that $\frac{ii'}{N} \cdot \frac{ii'}{N} = \frac{1}{N} \cdot i \cdot (\frac{i'i}{N}) \cdot i' = \frac{1}{N} ii'$.

A corollary of this is that $P_\alpha^t = P_{(\alpha^t)}$. That is, $P_\alpha^t = \alpha^t I + (1-\alpha^t)\frac{ii'}{N}$. Also, $\lim_{t\to\infty} P_\alpha^t = \frac{ii'}{N}$, which implies that the steady state distribution is $\mu^* = \frac{i'}{N}$.[24]

Now to characterize the measures related to mixing using this $P = P_\alpha$,

$$(AD) = \|\mu_0 P^t - \mu^*\|$$
$$= \|\alpha^t \mu_0 + (1-\alpha^t)\frac{i'}{N} - \frac{i'}{N}\|$$
$$= \alpha^t \|\mu_0 - \frac{i'}{N}\|$$

$$(AID) = \sum_k m_k \|e_k P^t - \frac{i'}{N}\|$$
$$= \sum_k m_k \alpha^t \|e_k - \frac{i'}{N}\|$$
$$= \alpha^t \|e_1 - \frac{i'}{N}\|$$

where the last equation is due to the symmetry across k, and $\sum_k m_k = 1$. Note that (AM) depends on μ_0 but (AIM) doesn't,[25] and they both converge to zero at rate α. Finally,

$$(AID) - (AD) = \alpha^t \Big(\|e_1 - \frac{i'}{N}\| - \|\mu_0 - \frac{i'}{N}\|\Big) \geq 0$$

D.3 Mixing Analysis: Bounds

Let $\mu_0 = (m_1, \cdots, m_N)$ where m_j's are scalar, and $P^t = (r_1; \cdots; r_N)$ where r_j's are row vectors for which the sum of N elements is equal to 1. Let e_j denote the row vector with j-th

[24] Note that $\alpha \neq 1$ is important because $\alpha = 1$ will result in $\mu^* = \mu_0 \neq \frac{i'}{N}$. This is because $\lim_{t\to\infty} \lim_{\alpha\to 1} \alpha^t = 0$, but $\lim_{\alpha\to 1} \lim_{t\to\infty} \alpha^t = 1$, resulting in a discontinuity at $\alpha = 1$.

[25] That (AIM) does not depend on μ_0 is an artifact of the symmetry assumed in this example in terms of P, and is clearly not a general property.

element equal to 1 and 0 otherwise. Then the following equations hold:

$$e_j P^t = r_j$$

$$\mu_0 = \sum_k m_k e_k$$

$$\mu_0 P^t = \sum_k m_k e_k P^t = \sum_k m_k r_k$$

Then

$$AD = \|\mu_0 P^t - \mu^*\| = \left\| \sum_k m_k r_k - \mu^* \right\| \in [0, M]$$

$$ID = \|e_j P^t - \mu^*\| = \|r_j - \mu^*\| \in [0, M]$$

$$AID = \sum_k m_k \|e_k P^t - \mu^*\| = \sum_k m_k \|r_k - \mu^*\| \in [0, M]$$

where M is the upper bound of the norm, e.g., $\|(1,0,0) - (0,1,0)\| = \ say, \ 2$.

Examining the difference between AID and AD,

$$\begin{aligned} AID - AD &= \\ &= \sum_k m_k \| \underbrace{r_k - \mu^*}_{v_k} \| - \left\| \sum_k m_k r_k - \mu^* \right\| \\ &= \sum_k m_k \|v_k\| - \left\| \sum_k m_k v_k \right\| \\ &\geq 0 \ \text{ by triangular inequality (Minkowski)} \end{aligned}$$

For the bound in the other direction,

$$
\begin{aligned}
AID - AD &= \sum_j m_j \Big(\|r_j - \mu^*\| - \Big\| \sum_k m_k r_k - \mu^* \Big\| \Big) \\
&\leq \sum_j m_j \Big\| r_j - \sum_k m_k r_k \Big\| \text{ by triangular inequality (Minkowski)} \\
&= \sum_j m_j \Big\| \sum_k m_k (r_j - r_k) \Big\| \\
&\leq \sum_{j,k} m_j m_k \|r_j - r_k\| \text{ by triangular inequality (Minkowski)} \\
&= \sum_{j \neq k} m_j m_k \|r_j - r_k\| \ (\because \|r_j - r_j\| = 0) \\
&\leq \sum_{j \neq k} m_j m_k \cdot M \ \text{ (M is the upper bound of the norm, e.g., 2)} \\
&= M \Big(\underbrace{(\sum_j m_j)^2}_{=1} - \underbrace{\sum_j m_j^2}_{\geq \frac{1}{N}(\sum_j m_j)^2 \text{ by Cauchy-Schwarz inequality}} \Big) \\
&\leq (1 - \frac{1}{N}) M = (1 - \frac{1}{N}) \cdot 2
\end{aligned}
$$