A Structural Model of Structural and Exchange Mobility*

Steven Durlauf† Gueyon Kim‡ Dohyeon Lee§ Xi Song¶

Preliminary Draft

This Version: March 2026

Abstract

We propose a novel framework for studying intergenerational occupational mobility, a long-standing topic of importance in both sociology and economics. We build a parsimonious model of occupation choice with shocks to occupation demand and time-varying barriers to transition. We estimate these structural shocks from the data and build two novel measures of mobility that reflect the underlying economic conditions, introducing a concept of counterfactual. We also show how the existing classes of mobility measures fit into our framework. Our measures indicate that the mobility in the US in the 19th century was lower than that in the 20th century, contrary to the findings in the recent literature.

JEL classification: D31, D63, J15, J62, N31, N32
Keywords: Intergenerational Occupational Mobility, Structural vs. Exchange Mobility

*All authors acknowledge the generous support of the Stone Center for Research on Wealth Inequality and Mobility at the University of Chicago.

†Harris School of Public Policy, University of Chicago and James M. and Cathleen D. Stone Foundation. e-mail: durlauf@gmail.com

‡Department of Economics, University of California, Santa Cruz. e-mail: gkim44@ucsc.edu

§Independent scholar. e-mail: dohnlee@gmail.com

¶Department of Sociology, University of Pennsylvania. e-mail: xisong@sas.upenn.edu

1 Introduction

Intergenerational occupational mobility has been an important subject of study in both economics and sociology. In these studies, measurement of mobility plays a critical role as it enables comparison of mobility across different time and space.

However, unlike the case with continuous variables such as income or rank, it is less straightforward to come up with a summary measure for describing the degree of occupational mobility. In particular, one long-standing challenge has been that the "true" mobility patterns are contaminated by the structural changes that make certain occupations grow or shrink over time. Although there exist a few widely used approaches that attempt to deal with this issue, these measures fail to fully address this issue. For example, the "traditional" measures of mobility are originally intended to capture a notion of counterfactual. But this was not possible without an economic model, and the lack of rigor in this approach resulted in major criticisms.

In this paper, we propose a novel framework of occupation choice across generations, which allows for the construction of a counterfactual economy to decompose structural versus exchange mobility. First, we build a parsimonious economic model of occupation choice with both demand and supply side shocks. Second, we show how the existing classes of mobility measures can be encompassed within our framework. Third, using the aggregate level data on the transition patterns and occupational wage structure, we uncover the structural shocks, which in turn are used to construct novel mobility measures that reflect the underlying economic conditions beyond just the descriptive patterns. Lastly, using this novel framework, we revisit the recent controversy on the long-run mobility patterns in the US

Our contributions are two-fold. First, we make a methodological contribution to the long-standing literature on intergenerational mobility through the novel framework. Second, we make an empirical contribution to the recent literature that investigates the long-term trend in intergenerational mobility in the US. While Long and Ferrie (2013) showed evidence that the 19th-century mobility in the US was higher than that of the 20th century, more recent studies have mixed results, as some studies find lower mobility in the 19th century (Abramitzky et al., 2025). Our results indicate that the mobility was lower in the 19th century, and is overall in the increasing trend up until around 1970.

The model consists of the aggregate production function, which governs the labor demand for each occupation, and the individual's occupation choice problem for the labor supply. For simplicity, one time period corresponds to 20 years, and only the individuals in their prime age (35-55) are considered at each point in time, which implies that each

individual works for one period only. This simplifies the problem in at least three aspects: a static discrete choice model for the individuals (and hence no expectation), not having to separately model intra-generational mobility in addition to inter-generational mobility, and not having to aggregate the quantity of labor across overlapping generations. More importantly, this allows full compatibility with the existing sociology literature that relies solely on the mobility table to analyze mobility.

The equilibrium outcome of the model is determined by three sets of exogenous parameters. The first is the occupational distribution in the previous period, or equivalently, that of the fathers of the current generation, which acts as the state variable in this economy. The second is the current occupational demand structure. An occupation with higher demand will pay higher wages and thereby attract more labor, shaping the current distribution with the given elasticity of substitution across occupations. The third is the cost of transition between each pair of occupations, also modeled as time-varying. This captures not only the preferences that make individuals more likely to choose the same occupations as their fathers, but also any institutional barriers or frictions that make the transition to different occupations difficult.

Given these sets of parameters, we classify the mobility measures into two types: those that are based on the pairwise transition costs only, and those based on all information. The former includes most of the widely used measures in sociology, including log-linear, log-multiplicative (i.e., unidiff), and odds-ratio-based. We note that while invariance to the marginal distributions is often deemed a virtue for this class of measures, lack of proper weighting is also a shortcoming. The traditional mobility measure belongs to the latter class. It was already criticized in the 1980s due to its ad hoc nature, but it is still widely used as a descriptive measure thanks to its simplicity.

With this classification in mind, we propose two novel measures of mobility that make use of all information. Our first measure is a direct extension of the traditional exchange mobility. Armed with the ability to perform any counterfactual analysis, we implement the definition of exchange mobility verbatim: "the portion of mobility that would occur in the absence of structural changes." Our second measure captures the average cost of transition, measured in terms of welfare differences with correct weights.

Combining our measures with the linked historical census data and other survey data in the US, we find that mobility has been increasing overall in the long run, and has started to stagnate or even decline since around the 1970s. This finding is contrary to the seminal work by (Long and Ferrie, 2013), which found evidence that the occupational mobility in the US was higher in the 19th century than that in the 20th century. There are two reasons for this discrepancy. First, the data source is different. When we use our data from both

CLP (and MLP) and compute the original Altham statistic, we find that the mobility in the 19th century is still lower, but the gap becomes much smaller. While the Long and Ferrie's data was presumably the state of the art at that time, there has been dramatic advances in the census linking project in the past decade, and thus our data is likely more accurate. Second, as we argue in this paper, appropriate weighting is important for representing the economy-wide mobility, which is absent in the Altham statistic or in the log-linear models. Even if we maintain the Altham statistic, when we apply the weights from the mobility table, we find that the mobility was lower in the 19th century.

1.1 Literature on the mobility measures

The distinction between structural versus exchange mobility dates back to the seminal work of Kahl (1957) in the sociological literature. The concept of exchange mobility, in particular, has been widely used in the sociological literature to analyze the extent of opportunities to improve status based on individuals' merit and effort (Hout and DiPrete, 2006; Treiman and Ganzeboom, 2000). In this section, we provide an overview of the different measurement approaches proposed in historical mobility studies in sociology and revived in a small yet burgeoning body of recent work in economics (Long and Ferrie, 2013; Blume et al., 2024; Durlauf et al., 2024).

Almost all studies in occupational mobility are based on the mobility tables, $n \times n$ matrices that summarize the relationships between the occupational choices of fathers and sons, where each element m_{ij} corresponds to the share of individuals with the father's occupation i and the son's occupation j. Then the marginal distributions for the fathers and the sons can be constructed as $f_i = \sum_k m_{ik}$ and $s_j = \sum_k m_{kj}$, respectively, and the transition probability from i to j can be described as $p_{ij} = m_{ij}/f_i$.

The "traditional" measures of mobility are defined as follows. Total mobility is $1 - \sum_i m_{ii}$, the sum of off-diagonal elements in the mobility table. Structural mobility is defined as $\frac{1}{2}\sum_i \|f_i - s_i\|$ where $\|\cdot\|$ denotes the L^1 norm, to capture the magnitude of changes in the marginal distribution that may result in spurious patterns of mobility. Exchange mobility is then defined as (total mobility) − (structural mobility) ≥ 0, which aims to isolate the "pure" form of mobility after removing the structural mobility.

While the traditional approach has been widely used due to its simplicity, it has been subject to criticism. One standard workaround has been to apply odds-ratio ($\frac{m_{ij}m_{i'j'}}{m_{ij'}m_{i'j}}$) preserving transformations, i.e., arbitrary scaling all elements in each row or column, and calculate the total mobility. By applying these transformations, researchers attempted to adjust for the structural changes without altering the patterns of association. One special case is to match the son's marginal distribution to the father's marginal distribution

(Hazelrigg, 1974), while other variants match some other marginal distributions instead of its own marginal (Hauser et al. (1975); also used in a subset of the results in (Long and Ferrie, 2013)). However, these attempts were criticized for lacking justification (Sobel et al., 1985), and the mainstream sociology literature has shifted to alternative methodologies.

Log-linear models were a popular alternative to the traditional mobility in the 1980s. The basic idea is to regress the log of each element in the mobility table m_{ij} on the row-specific and column-specific dummies to control for the impact of marginal distributions, and investigate the remaining patterns of association using additional dummy variables. The two polar cases would be the full independence (marginal distributions explain everything), and the full saturation (each pairwise transition is distinct after controlling for the marginals). In practice, researchers impose restrictions that certain cells are the same within a single mobility table or across multiple tables (across time or space), focusing on whether the mobility patterns are statistically different. By construction, row and column specific effects are absorbed by the fixed effects and thus do not materially contribute to the properties of the mobility table.

One variant of the log-linear model that gained popularity is the log-multiplicative or uniform difference (i.e., unidiff), originally explored by Goodman and adopted in the sociology literature since Xie (1992).[1] The structure is similar to the log-linear: model the log of each element of the mobility table as the sum of row- and column-specific components and the association component $\lambda_k u_i v_j$, where u_i and v_j are the row and column specific scores that are endogenously determined during the MLE procedure jointly with the fixed effects. It is often estimated across multiple tables, allowing for distinct values of λ_k for each table k. This effectively ranks occupations so that the pairs of occupations with high mobility are assigned similar scores. The resulting degree of association λ_k captures the overall correlation in the joint distribution.

Altham statistic (Altham, 1970) has also been used as an alternative measure of mobility in both economics and sociology (e.g., Long and Ferrie, 2013). It is based on the odds ratio $(\frac{m_{ij}m_{i'j'}}{m_{ij'}m_{i'j}})$ between all possible combinations of two rows (i, i') and two columns (j, j'). It measures the distance between the current mobility table and the "full mobility" benchmark where the son's choice of occupation is independent of the father's, i.e., $m_{ij} = f_i \cdot s_j$ for all i, j. This was the core measure used by Long and Ferrie (2013) in their argument that the mobility in the US was particularly high in the 19th century.

More recently, rank-rank correlation has gained much popularity as a measure of mobility. It is widely used in both the economics and sociology literature with continuous

[1] We do not separately discuss unidiff in this paper because it is an extension of the log-linear approach, sharing the key properties such as invariance to marginal transformations.

variables such as income (Chetty et al., 2013, 2014, 2020). However, even in the US, with rich historical data, neither income nor education data is available before 1940, and typically, the best proxies available are occupation and literacy, which makes the use of rank-rank correlation difficult. Sociology literature has developed scores for each occupation, the most commonly known being Duncan's SocioEconomic Index (SEI) (Duncan, 1961) and Treiman's index (Treiman, 1977). Song et al. (2020) used a variant of this by normalizing the distribution of the ranking index based on the occupational distribution.

Finally, we note a strand of important work building on Berger and Snell (1957), which contrasts (i) mobility that is produced in the process of reaching the steady state of a given Markov chain and (ii) mobility that occurs at the steady state or immobility that would be observed in the absence of this evolution. Blume et al. (2024) also measures these two objects but takes an alternative approach: exchange mobility is defined using an imbalance matrix, capturing 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). Durlauf et al. (2024) proposes measurements based on mixing calculations of Markov Chains to examine long-run trends in exchange versus structural mobility patterns in the US. While this literature has regarded the transition matrix $\mathbf{P}$ as a fundamental object and simply applied it iteratively, the current paper further acknowledges that it is an endogenous object that depends on the current equilibrium wages and hence may vary over time even if the exogenous parameters such as technology and barriers are held constant.

2 Model

In this section, we introduce a theoretical model of how intergenerational associations in occupational choices arise from parental background and equilibrium relative wages. Building on this model, we then propose a decomposition method for distinguishing between structural versus exchange mobility.

2.1 Model Environment

We assume that each individual works for one period over one's prime age (e.g., 35 to 55), which is consistent with the empirical studies on intergenerational mobility focusing on occupations or income observed around their prime working age (Mazumder and Acosta, 2015; Haider and Solon, 2006). There are no other decisions made or actions taken before or after the working age, and thus, each individual is assumed to live for one period without overlapping generations.

In each period, aggregate output $\mathbb{Q}_t$ in period t is produced by combining labor from each occupation with constant elasticity of substitution,

$$\mathbb{Q}_t = \left(\sum_k A_{k,t}^{\frac{1}{\gamma}} L_{k,t}^{\frac{\gamma-1}{\gamma}} \right)^{\frac{\gamma}{\gamma-1}}$$

where $L_{k,t}$ is the quantity of labor in occupation k, $A_{k,t} \in (0,1)$ is the exogenous demand shifter for occupation k in period t reflecting the relative importance in aggregate production ($\sum_k A_{k,t} = 1$), and $\gamma > 0$ is the elasticity of substitution across occupations. We will also refer to $A_{k,t}$ as the occupational demand structure, which acts as the source of structural mobility. The labor demand for each occupation j is given by

$$L_{j,t} = A_{j,t} \left(\frac{W_{j,t}}{\mathbb{W}_t} \right)^{-\gamma} \mathbb{Q}_t$$

where $W_{j,t}$ denotes the wage in occupation j which is determined in equilibrium, and $\mathbb{W}_t = \left(\sum_k A_{k,t} W_{k,t}^{1-\gamma} \right)^{\frac{1}{1-\gamma}}$. The total population is fixed and normalized to $\sum_k L_{k,t} = 1$, which is also the total labor supply (no unemployment or non-employment). Therefore, $L_{j,t}$ can also be expressed as,

$$L_{j,t} = \frac{A_{j,t} W_{j,t}^{-\gamma}}{\sum_k A_{k,t} W_{k,t}^{-\gamma}} \tag{1}$$

All individuals are ex ante identical in terms of production capability, and each occupation would pay the same equilibrium wage to all individuals. In our model, differences in wages across occupations arise due to individual preferences or frictions. In the frictionless benchmark, wages in all sectors would be identical in equilibrium, which in turn implies $L_{j,t} = A_{j,t}$, and $\mathbb{Q}_t = (\sum_k A_{k,t})^{\gamma/(\gamma-1)} = 1$.

In each period, individual workers in their prime ages make the occupation choice decision. There are three components that contribute to the utility towards each occupation: wage $W_{j,t}$ in each occupation j, systematic preference $Z_{ij,t}$ towards each occupation j conditional on their parents' occupation i (potentially time-varying),[2] and the idiosyncratic preference shock towards each occupation j, $\widetilde{\epsilon}_{j,t}^{s}$.[3] The systematic component of this utility for a family moving from occupation i to j is given as $u_{ij,t} = \nu \log W_{j,t} + Z_{ij,t}$. Each individual

[2] $Z_{ij,t}$ is intended to capture not only individual's preferences but also all other forces such as institutional barriers and frictions related to changes in occupation across generations. For convenience, all of these components are modeled in the form of transition costs in utility, specific to each ordered pair of occupations. Without loss of generality, our choice of normalization is $Z_{ii,t} = 0$.

[3] We have considered an alternative approach with productivity shocks for both systematic and idiosyncratic components as opposed to preference shocks, and decided to proceed with the preference shocks because [...]. Introducing both is not tractable.

s whose parent had an occupation i makes an occupation choice decision after observing ϵ^s_{jt} to maximize utility.

$$\max_j \left(\nu \log W_{j,t} + Z_{ij,t} + \widetilde{\epsilon}^s_{j,t} \right)$$

$\nu > 0$ governs the importance of the log wage relative to the preference components. $\widetilde{\epsilon}^s_{j,t}$ is a random variable determined as the maximum of a Poisson point process with density $\exp(-e)\,de\,dx$ over $e \in \mathbb{R}$ and $x \in [0, L_{j,t-1}]$ for each occupation j and for each individual s. This models differential "exposure" to each occupation j that impact all individuals in an identical way. For example, since the share of farmers is small today, we don't draw as many (or equivalently as good) random draws for becoming a farmer compared to other more commonly observed occupations. Then it can be shown that $\widetilde{\epsilon}^s_{j,t}$ follows a Gumbel distribution with location parameter $\log L_{j,t-1}$ (proof in the appendix). For convenience, we denote $\epsilon^s_{j,t} \equiv \widetilde{\epsilon}^s_{j,t} + \log L_{j,t-1}$, where $\epsilon^s_{j,t}$ would be i.i.d. Gumbel distribution for all occupation j and individual s.[4] All the subsequent analysis is then identical to the standard logit model. The transition probability from occupation i to j is

$$P_{ij,t} = \frac{\exp(\nu \log W_{j,t} + \log L_{j,t-1} + Z_{ij,t})}{\sum_k \exp(\nu \log W_{k,t} + \log L_{k,t-1} + Z_{ik,t})} = \frac{W^\nu_{j,t} L_{j,t-1} \theta_{ij,t}}{\sum_k W^\nu_{k,t} L_{k,t-1} \theta_{ik,t}} \tag{2}$$

where $\theta_{ij,t} \equiv \exp Z_{ij,t}$, with $Z_{ii,t} \equiv 0$ due to normalization. Hence, occupation-specific labor supply in t is determined by combining the labor supply in period $t-1$ and the transition probability in period t.

$$L_{j,t} = \sum_k P_{kj,t} L_{k,t-1}, \tag{3}$$

or in matrix notation, $\mathbf{L}_t = \mathbf{L}_{t-1} \mathbf{P}_t$ with $\mathbf{L}_t$ as a row vector.

2.2 Equilibrium

The equilibrium in this economy can be fully described by the equations (1), (2), and (3), which govern the labor demand, transition rule, and labor supply, respectively. Taking as given the technology $A_{j,t}$ and the previous distribution of labor $L_{j,t-1}$, there exists an equilibrium relative wage $W_{j,t}$ that equates labor supply and demand in each occupation. Since the excess demand for labor monotonically decreases in relative wage, the existence and uniqueness of the equilibrium can be proved easily.

The parameter γ governs the labor demand elasticity. The parameter ν controls the labor supply elasticity. In particular, ν captures the relative importance between relative

[4] This dependence on $L_{i,t-1}$ not only makes sense, but is also crucial for eliminating a spurious component in our mobility measure due to structural changes in the long run such as decline of farming.

wages and the i.i.d idiosyncratic preference component for labor supply. Holding all else constant, if ν is larger, relative labor supply is more sensitive to relative wages. The set of parameters $A_{j,t}$, which represents production technology, serves as the demand shifter, and $Z_{ij,t}$, the supply shifter.

Consider a negative shock in $A_{j,t}$ which shifts the downward-sloping demand curve to the left while the upward-sloping supply curve remains in place. If $\gamma = 0$, the demand curve would be vertical, and $W_{j,t}$ must absorb the entire shock until $L_{j,t} = \overline{L}_{j,t} \propto A_{j,t}$ holds. If $\nu \to \infty$ (supply curve close to horizontal), only a small decrease in $W_{j,t}$ is needed to reach an equilibrium ($P_{ij,t} \propto W_{j,t}^{\nu}$) because labor supply is highly responsive to wages. If ν is small, $W_{j,t}$ must decrease sufficiently to discourage workers from choosing this occupation based on their preference.

Numerically solving the equilibrium of this model involves finding the set of wages that equates the excess labor demand to zero for each occupation, taking as given the parameters $(A_{j,t}, Z_{ij,t})$ and the state variable $L_{j,t-1}$. This is an exactly identified system of $N-1$ equations with $N-1$ unknowns.[5] For small deviations of $A_{j,t}$, we show in the Appendix how to find analytical solutions using log-linearization, which further elucidates the model mechanisms.

Identification of the unobserved parameters involves the use of equilibrium conditions. In the data, the wages in each occupation and the full mobility table for each pair of occupations are observed. Using this information, we can first identify the current state of the economy encapsulated by the occupation demand $A_{j,t}$ using $A_{j,t} \propto L_{j,t} W_{j,t}^{\gamma}$, normalized using $\sum_j A_{j,t} = 1$. Next, we can identify the current preference for each occupation using $\theta_{ij,t} = \frac{P_{ij,t}}{P_{ii,t}} \cdot (\frac{W_{j,t}}{W_{i,t}})^{-\nu} \cdot (\frac{L_{j,t-1}}{L_{i,t-1}})^{-1}$, where the transition probabilities $P_{ij,t}$ can be easily calculated using the mobility table. Given the flexible set of parameters, we can typically construct the set of parameters that can justify any data.[6]

2.3 Equilibrium allocation of labor

Given the simplicity of the model, it is possible to further manipulate the equations and simplify further, in particular, express the equilibrium without wages. Suppose $\gamma > 0$. With $L_{j,t} = A_{j,t}(W_{j,t}/\mathbb{W}_t)^{-\gamma}$,

$$W_{j,t} = \mathbb{W}_t (\frac{A_{j,t}}{L_{j,t}})^{1/\gamma}$$

[5] Without loss of generality, we normalize $A_{j,t} = W_{j,t} = L_{j,t} = 1$ for one of the occupations, and solve using the remaining $N-1$ occupations.

[6] Zeros in the transition probabilities would give rise to exceptions. If $P_{ij,t} = 0$ for $j \neq i$, $Z_{ij,t} \to -\infty$ and $\theta_{ij,t} = 0$ for this (i,j) only. If some occupation i has none of the sons taking the same occupation as their fathers, $P_{ii,t} = 0$. In this case, $Z_{ij,t}, \theta_{ij,t} \to \infty$ for all j.

As mentioned previously, the mobility table $\mathbf{M}_t = \{m_{ij,t}\}$ refers to the joint distribution of individuals with the father's occupation i and the son's occupation j. Then $m_{ij,t} = L_{i,t-1}P_{ij,t}$ holds. In addition, $\sum_k m_{ik,t} = L_{i,t-1}$ and $\sum_k m_{kj,t} = L_{j,t}$, which simply indicate the relationship between the joint and marginal distributions.

Consider the ratio between an arbitrary element of the mobility table and the diagonal element from the same row, which is also the relative conditional probability from the same father's occupation.

$$\frac{m_{ij,t}}{m_{ii,t}} = \frac{P_{ij,t}}{P_{ii,t}} = \frac{W_{j,t}^{\nu} L_{j,t-1}\theta_{ij,t}}{W_{i,t}^{\nu} L_{i,t-1}} = \left(\frac{A_{j,t}}{A_{i,t}}\right)^{\frac{\nu}{\gamma}} \frac{L_{j,t-1}}{L_{i,t-1}}\theta_{ij,t} \cdot \left(\frac{L_{j,t}}{L_{i,t}}\right)^{-\frac{\nu}{\gamma}} \equiv X_{ij,t} \cdot \left(\frac{L_{j,t}}{L_{i,t}}\right)^{-\rho}$$

with $\rho \equiv \frac{\nu}{\gamma} > 0$, and $X_{ij,t} \equiv \left(\frac{A_{j,t}}{A_{i,t}}\right)^{\frac{\nu}{\gamma}} \frac{L_{j,t-1}}{L_{i,t-1}}\theta_{ij,t}$ combines the direct impact of all the exogenous parameters, with $X_{ii,t} = 1$. The remaining term $\frac{L_{j,t}}{L_{i,t}}$ is an endogenous object to be determined jointly with $\frac{m_{ij,t}}{m_{ii,t}}$.

We can further use $m_{ij,t} = X_{ij,t}\left(\frac{L_{j,t}}{L_{i,t}}\right)^{-\rho} m_{ii,t}$ to characterize the full mobility table using only the diagonal elements and the current marginal distribution $L_{j,t}$. From $\sum_k m_{ik,t} = L_{i,t-1}$,

$$m_{ii,t} = \frac{L_{i,t-1}}{L_{i,t}^{\rho}(\sum_k X_{ik,t} L_{k,t}^{-\rho})} \tag{4}$$

That is, given $L_{j,t}$, $m_{ii,t}$ are fully determined. From $\sum_k m_{kj,t} = \sum_k X_{kj,t}\left(\frac{L_{j,t}}{L_{k,t}}\right)^{-\rho} m_{kk,t} = L_{j,t}$,

$$\sum_k X_{kj,t} L_{k,t}^{\rho} m_{kk,t} = L_{j,t}^{1+\rho} \tag{5}$$

By plugging eq (4) into eq (5), we have a system of N equations that determines $L_{j,t}$. This also suggests a natural iterative method for solving the equilibrium given the exogenous parameters $X_{ij,t}$.

Note that for the determination of $m_{ij,t}$, $X_{ij,t}$ serves as the sufficient statistic that summarizes the impact of all the exogenous components besides the state variable $L_{i,t-1}$. Eliminating the $L_{i,t-1}$ component from $X_{ij,t}$, we can say that $\left(\frac{A_{j,t}}{A_{i,t}}\right)^{-\rho}\theta_{ij,t}$ is the sufficient statistic besides $L_{i,t-1}$.

This finding, in fact, shows the central role of wages in identification between the technology $A_{j,t}$ and the preference $Z_{ij,t}$. Given a mobility table $\{m_{ij,t}\}$ and the sufficient statistics $L_{i,t-1}$ and $X_{ij,t}$, we can impose an arbitrary set of values for $A_{j,t}$ and still find the residual values of $Z_{ij,t}$ that can justify the values of $X_{ij,t}$, implying a lack of identification.

Odds ratio Approach Odds ratio is an important approach that has been considered in the mobility literature, and also the key measure of mobility in Long and Ferrie (2013). Now that we have fully characterized the mobility table using the model components, we can proceed to derive the odds ratio. Odds ratio is defined for each set of two rows (i, i') and two columns (j, j') from the given mobility table. Using $m_{ij} = m_{ii} X_{ij} \left(\frac{L_j}{L_i}\right)^{-\rho}$,[7]

$$Odds_Ratio_{i,i',j,j'} = \frac{P_{ij} P_{i'j'}}{P_{ij'} P_{i'j}} = \frac{m_{ij} m_{i'j'}}{m_{ij'} m_{i'j}} = \frac{X_{ij} X_{i'j'}}{X_{ij'} X_{i'j}} = \frac{\theta_{ij} \theta_{i'j'}}{\theta_{ij'} \theta_{i'j}} = \exp\left(Z_{ij} + Z_{i'j'} - Z_{ij'} - Z_{i'j}\right)$$

where the last equation follows from the definition of $X_{ij} = \left(\frac{A_j}{A_i}\right)^{\rho} \frac{L_{j,t-1}}{L_{i,t-1}} \theta_{ij}$. In essence, anything that is additively separable in log would be canceled out, and only the "nonlinear interaction" terms survive.

It turns out that the odds ratio would contain only the preference-related components based on our framework. Given the simplicity of our model, this is a natural outcome—the impact of technology or the previous allocation of labor affects the equilibrium only in a separable manner. This also gives us the benefit of being able to address and incorporate the odds ratio based approaches within our framework.

2.4 Construction of the counterfactual economy

Given our objective of studying the impact of structural changes as encapsulated by the technology parameters $A_{j,t}$, we focus on the impact of altering $A_{j,t}$ while taking as given the preference parameters $Z_{ij,t}$ as identified from the data. The values of $A_{j,t}$ and $Z_{ij,t}$ constructed are fully consistent with the current data $W_{j,t}$ and $L_{ij,t}$, taking as given $L_{i,t-1}$. Using this simple model, we proceed to construct counterfactual economies using different values of demand shifter $\widetilde{A}_{j,t}$.

We focus on the case with $\widetilde{A}_{j,t} = A_{j,t-1}$. This effectively asks the question of how the labor allocation would have changed had the demand shifters not changed from the previous generation. This question is precisely aligned with the core concept of structural and exchange mobility in the sociology literature. The counterfactual patterns of mobility arising under $\widetilde{A}_{j,t} = A_{j,t-1}$ correspond to the net mobility after controlling for the structural changes from $A_{j,t-1}$ to $A_{j,t}$. By taking the total mobility under this counterfactual economy, we measure the share of sons with occupations different from those of their fathers in the absence of structural shocks. This is arguably a valid measure of mobility comparable across different economies, and in particular, coincides with what the sociology literature wanted to capture as the exchange mobility.

[7] The time subscripts are omitted except in $L_{i,t-1}$.

Construction of the counterfactual in period t requires the knowledge of parameters $Z_{ij,t}$ for the current period and $A_{i,t-1}$ for the previous period. The former requires the current mobility table $\mathbf{M}_t$ and the current wages $W_{j,t}$, and the latter additionally requires wages in the previous period $W_{j,t-1}$.

If $\gamma = 0$, $L_{j,t} = \widetilde{A}_{j,t} = A_{j,t-1} = L_{j,t-1}$. This implies that we can simply normalize each row of the mobility table to match the marginal distribution of the son's occupations to that of the parents. Notably, the counterfactual can be constructed without using the wage data, which can be convenient when the data is not available.

3 Aggregate Measures of Mobility

3.1 Conceptual framework

An economy is fully determined by the following exogenous set of objects: $\{L_{i,t-1}, A_{j,t}, Z_{ij,t}\}$. $\{L_{i,t-1}\}$ governs the row-specific impact that reflects history dependence in the presence of mobility costs, $\{A_{j,t}\}$ governs the column-specific impact in the current period, and $\{Z_{ij,t}\}$ is the component that captures pairwise mobility costs.[8] In this section, we define the classes of mobility measures based on which of this information set is used.

Among the three, the most fundamental object related to mobility is $\{Z_{ij,t}\}$. If $Z_{ij,t} = 0$ for all i, j, there is no barrier across occupations, dubbed "perfect mobility" in sociology. If $Z_{ij,t} \ll 0$ for all i, j, the barrier is so large that everyone would maintain the same occupation as parent's ($j = i$), which implies a highly immobile society. Indeed, the most parsimonious measures of mobility in sociology (e.g., Altham and log-linear) are based precisely on $Z_{ij,t}$ only. Thus, we posit that any measure of mobility should include the notion of $Z_{ij,t}$.

If we want a measure of mobility that reflects the current and/or previous occupational distribution, we need to include the information contained in $\{L_{i,t-1}\}$ or $\{A_{j,t}\}$. However, they are not indispensable, as in the examples of Altham and log-linear approaches. Inclusion of $\{L_{i,t-1}\}$ or $\{A_{j,t}\}$ has the benefit of correctly reflecting the occupational distribution, which is a natural approach for aggregation. One shortcoming is that a direct comparison across different economies (either across time or across countries) with vastly different occupational structures would be difficult. Exclusion of $\{L_{i,t-1}\}$ and $\{A_{j,t-1}\}$ has the opposite cost and benefit. It is highly comparable across any society, at the cost of ignoring the weights entirely. Note that inclusion of just one of the two would lose the benefit (ad hoc weights on one side) while facing the similar cost (not comparable on one side) and thus would not be recommended.

[8] Again, while $A_{j,t} \neq L_{j,t}$ in general, our model shows how $A_{j,t} = L_{j,t}$ can be encompassed as a special case.

Based on the above discussion, we define the first class of mobility measures as those that involve $\{Z_{ij,t}\}$ only, and thus independent of the occupational distributions. We define the second class of mobility measures as those that involve $\{L_{i,t-1}, A_{j,t}, Z_{ij,t}\}$, which incorporates the current and previous occupational distribution.

3.2 Existing measures of mobility

The traditional measures of mobility are direct transformations of the mobility table itself, as explained above in section 1.1. Total mobility is $1 - \sum_i m_{ii}$, the sum of off-diagonal elements in the mobility table. Structural mobility is defined as $\frac{1}{2}\sum_i \|f_i - s_i\|$ where $\|\cdot\|$ denotes the L^1 norm, to capture the magnitude of changes in the marginal distribution that may result in spurious patterns of mobility. Exchange mobility is then defined as (total mobility) − (structural mobility) ≥ 0. It is evident that these depend on $\{L_{i,t-1}, A_{j,t}, Z_{ij,t}\}$ since they are not invariant to the marginal distributions.

Altham statistic is defined as an aggregation of the relative log odds ratio between two given mobility tables $(\mathbf{M}, \mathbf{Q})$ for all possible combinations of two rows and two columns.

$$d(\mathbf{M}, \mathbf{Q}) = \Big(\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{i'=1}^{N} \sum_{j'=1}^{N} \Big(\log \frac{m_{ij} m_{i'j'}}{m_{ij'} m_{i'j}} - \log \frac{q_{ij} q_{i'j'}}{q_{ij'} q_{i'j}} \Big)^2 \Big)^{\frac{1}{2}}$$

A widely used special case is when one matrix is the case of independence or "full mobility" where each joint probability is equal to the product of the two marginal distributions, denoted by $\mathbf{I}$. Then the log odds ratio of $\mathbf{I}$ is zero for all combinations, and thus

$$d(\mathbf{M}, \mathbf{I}) = \Big(\sum_{i,j,i',j'} \Big(\log \frac{m_{ij} m_{i'j'}}{m_{ij'} m_{i'j}} \Big)^2 \Big)^{\frac{1}{2}} = \Big(\sum_{i,j,i',j'} \Big(Z_{ij} + Z_{i'j'} - Z_{ij'} - Z_{i'j} \Big)^2 \Big)^{\frac{1}{2}}$$

where the second equation follows from eq ??. In addition, it can be shown (e.g., Bouchet-Valat, 2022) that this value is proportional to the sum of square of double-demeaned log of the mobility table.

$$d(\mathbf{M}, \mathbf{I}) \propto \Big(\sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_{ij}^2 \Big)^{\frac{1}{2}}$$

where $\lambda_{ij} = \log m_{ij} - \overline{\log m_{i\cdot}} - \overline{\log m_{\cdot j}} + \overline{\log m_{\cdot\cdot}} = Z_{ij} - \overline{Z}_{i\cdot} - \overline{Z}_{\cdot j} + \overline{Z}$.

In the log-linear approach, the mobility under fully saturated model can be characterized by the residuals from the regression of $\log m_{ij}$ on the row and column specific fixed effects. As a result, the mobility from i to j is the double-demeaned residual of $\log m_{ij}$, which is just the λ_{ij} defined above.

Log-multiplicative or unidiff approach can be viewed as a special case of the log-linear approach with row and column fixed effects. From the data generation process perspective, it effectively assumes that Z_{ij} follows a specific association pattern along a particular ordering of the occupations. From the estimation perspective, it imposes such a restriction to the data while ignoring the residual components net of this overall association.

3.3 Our proposed measures of mobility

In this section, we propose two novel measures of aggregate mobility. The first measure is more closely aligned with traditional mobility, applied on top of the counterfactual economy that uses $\widetilde{A}_{j,t} = A_{j,t-1}$. The second measure is derived from the model, which aims to measure the average transition cost. It is measured in terms of the utility or welfare differences which is the most natural unit in the current context.

3.3.1 Traditional-based measure

In the first measure, we maintain the definition of total mobility as the sum of off-diagonal elements in the original mobility table. Next, we define exchange mobility as total mobility applied to the counterfactual mobility table that takes $\widetilde{A}_{j,t} = A_{j,t-1}$ as opposed to the actual $A_{j,t}$. This would measure the total share of individuals that would have maintained their parent's occupation in the absence of structural changes. Lastly, we define structural mobility as the difference between total mobility and exchange mobility. The structural mobility may take negative values.

This measure implements the concept of exchange mobility in exactly the way that sociologists wanted to achieve since the 1970s, within their framework. Since these traditional measures of mobility are still widely being used, and give the minimal data requirement for calculating the counterfactual (occupation level wages), our proposed measure can serve to complement the existing studies to a large extent.

3.3.2 Welfare-based measure

Now we turn to our second measure, which is more closely tied to the model. Consider the ex ante expected utility of an individual with parents' occupation i before the realization of an idiosyncratic shock (which is also the population average of the ex post utility):

$$V_{i,t} = \log \sum_{k} W_{k,t}^{\nu} L_{k,t-1} \theta_{ik,t}$$

In our model, the core object related to exchange mobility is $\theta_{ij,t}$. The "full mobility" benchmark can be represented by $Z_{ij,t} = 0$ or equivalently $\theta^0_{ij,t} = 1$, which implies $V^0_{i,t} = \log \sum_k \widetilde{W}^{\nu}_{k,t} L_{k,t-1}$, independent of i. Note that we include the impact of equilibrium adjustment in wages, and hence $\widetilde{W}_{j,t}$ is used instead of $W_{j,t}$.[9] The object of our interest is $U_{i,t} = V^0_{i,t} - V_{i,t}$ for each i, which captures the loss in welfare compared to the full mobility benchmark, and its average over population $U_t = E[U_{i,t}]$. To delineate this effect of adjustment in wages, we also define the utility while holding the wages fixed, $V^W_{i,t} = \log \sum_k W^{\nu}_{k,t} L_{k,t-1}$.

The relative utility due to the existence of frictions or barriers $\{\theta_{ij,t}\}$ conditional on i is:

$$V_{i,t} - V^0_{i,t} = \log \frac{\sum_k W^{1/\nu}_{k,t} \theta_{ik,t}}{\sum_k W^{1/\nu}_{k,t} \theta^0_{ik,t}} = \log \frac{W^{1/\nu}_{i,t}}{\sum_k L_{k,t-1} W^{1/\nu}_{k,t}} - \log \frac{P_{ii,t}}{L_{i,t-1}}$$

This uses the solution to the individual's maximization problem $\theta_{ij,t} = \frac{P_{ij,t}}{P_{ii,t}} (\frac{W_{j,t}}{W_{i,t}})^{-1/\nu} L_{j,t-1}$, which is satisfied in equilibrium. This welfare difference can be divided into two parts: the wage in occupation i relative to an (unweighted) average, and the probability of choosing the same occupation i as the father. The first term can be positive or negative, whereas the second term is negative if there is persistence in occupational transition ($p_{ii,t} > \frac{1}{N}$). For convenience, we define $U_{i,t} = V^0_{i,t} - V_{i,t}$, which is typically positive, so that higher values indicate higher persistence or lower mobility, a direction also consistent with that of Altham's statistic.

Finally, we take the weighted average using the marginal distribution of the father's occupation, $L_{k,t-1}$.

$$\begin{aligned} U_t = E[U_{i,t}] &= \sum_k L_{k,t-1} \cdot (V^0_{k,t} - V_{k,t}) \\ &= \sum_k L_{k,t-1} \log \frac{P_{kk,t}}{L_{k,t-1}} + \left(\log \sum_k L_{k,t-1} W^{1/\nu}_{k,t} - \sum_k L_{k,t-1} \log W^{1/\nu}_{k,t} \right) \end{aligned}$$

The first term captures the average persistence of remaining in the same occupation. It depends only on the mobility table and does not depend on wages or other parameter values. The second term is the Jensen's inequality term. It is positive (equality iff all wages are identical), and tends to be larger when the variance of wage is larger.

The current welfare analysis turns out to be particularly convenient because it does not require solving the entire model. The first term requires only a mobility table, whereas the

[9] Since the wages are in relative terms and thus not comparable across different equilibria, we normalize by the CES price index $\mathbb{W}_t$ to make the wages comparable.

other two terms require the lagged occupation share and the current occupational wages, along with an estimated or calibrated value of ν (γ is not needed). Even if wage data is not available, we may simply take the first term while assuming that the other two terms are constant. The first term alone is more theoretically grounded than the traditional mobility (sum of off-diagonals), even if they use the same amount of information.

Extending the concept from the traditional mobility measures, we define *total* mobility as the value of U_t that uses the actual wages $W_{j,t}$. We define *exchange* mobility U_t^X as the value of U_t using the wages $\widetilde{W}_{j,t}$ under the counterfactual economy simulated using $\widetilde{A}_{j,t} = A_{j,t-1}$. *Structural* mobility is then defined as $U_t^S = U_t - U_t^X$, which may take negative values. Note that γ would have an impact on the exchange mobility through the use of counterfactual wages.

4 Data and Sample Construction

To bring the model to the data, two inputs are required: mobility tables recording father–son occupational transitions $M_{ij,t}$, and occupation-specific wages $W_{j,t}$. The mobility tables are constructed from the linked census and survey data, adjusted by using the current occupation distribution to minimize sampling biases. Further details are provided below in this section. Wage data are drawn from the Decennial Censuses (1940–2000) and the American Community Survey (2010, 2019). For periods prior to 1940, when individual-level wage data are not available in the census, we impute occupation-specific wages using auxiliary information.[10]

Our data construction builds on Durlauf et al. (2024), which assembles a comprehensive dataset of father–son pairs spanning the post-Civil War era to the present. We summarize the key elements of this construction below and refer the reader to that paper for further details on the linking procedures, survey harmonization, and occupational coding.

4.1 Data Sources

The analysis requires intergenerational data on occupations that cover both the historical and contemporary periods. Two distinct types of data sources serve this purpose.

For the historical period (birth cohorts 1856–1915), we draw on linked full-count US Censuses from 1850 to 1940. Our primary linkages come from the Census Linking Project (CLP version 2.0) (Abramitzky et al., 2022), which uses several automated algorithms to match individual records across census years on the basis of names, birth year, and state or country of birth (see Abramitzky et al., 2020, 2021, for details on the linking methods).

[10] We discuss the sensitivity of our results to the wage imputation procedure in the Online Appendix.

We implement decadal links between successive censuses—excluding 1890—using the "Exact" and "Conservative" matching methods, both of which rely on the ABE algorithm with exact names and, in the conservative variant, require individuals to be unique within a five-year age window. The "Exact" linking approach reduces false positives relative to the NYSIIS-based approach (Bailey et al., 2020). Because our analysis also examines racial differences, we use the link variant that conditions on race as the matching variable. The linkage data are then merged with age, sex, race, occupation, education, region of residence, and household member information in the full-count US Censuses 1850–1940 (Ruggles et al., 2020).

We supplement the CLP-based sample with two alternatives: the IPUMS Multigenerational Longitudinal Panel (MLP version 1.0) (Helgertz et al., 2022), which relies on a probabilistic algorithm incorporating household information, and the Census Tree data (Buckles et al., 2023), which leverages genealogical records from FamilySearch.org together with machine learning refinements. Individual linking rates are highest in the Census Tree data and lowest in the CLP, with the MLP falling in between; the same ordering holds for the intergenerational matching rates.[11]

For the contemporary period (birth cohorts 1916–1975), we combine 10 large-scale social surveys: General Social Survey (1972–2021), Occupational Changes in a Generation (1962, 1973), National Survey of Families and Households (1987–2003), 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 the National Survey of Black Americans (1979–1992).[12] Following Song et al. (2020) and Jácome et al. (2021), we collect father–son information on age, sex, race, occupation, education, and region of residence from each survey, restricting attention to Black and White men aged 35–55.[13] When combining surveys, we normalize the sampling weights so that their sum within each survey equals the corresponding sample size, preserving within-survey relative weights while ensuring

[11] Compared to CLP, individual linking rates are higher when using either the MLP or Census Tree data. For White men, rates range from 54 to 70 percent with MLP and 78 to 86 percent with the Census Tree. For Black men, the rates are between 27 and 41 percent with MLP and 55 and 66 percent with the Census Tree. When examining the probability of matching father-son pairs with valid occupational information, the rates are much lower overall but still higher than with CLP. With MLP (Census Tree), they range between 14 and 19 (24 and 45) and between 4 and 5 (10 and 21) percent for White and Black men, respectively (Table B.3).

[12] See Online Appendix D for a detailed description of each data source.

[13] We exclude the Wisconsin Longitudinal Survey (1957–2011) as the survey does not provide a nationally representative sample. We also exclude the American National Election Studies (1948–2016) because the occupational definitions used for fathers' occupations cannot be harmonized with our occupational groupings. Since our analysis focuses on men, we do not include National Longitudinal Study–Young Women, National Longitudinal Study–Mature Women, and National Fertility Survey in our sample.

that surveys with larger samples contribute proportionally more information. This normalization is performed separately by race to account for the oversampling of certain groups in some surveys. Our approach closely aligns with that of Jácome et al. (2021), which also combines large-scale survey data to study income-based mobility trends for both Black and White men and women. Since our analysis further combines the linked censuses with these surveys, the resulting dataset extends the birth cohort coverage from the post-Civil War period to the present.

4.2 Construction of the Intergenerational Sample

We construct father–son pairs from the linked censuses in three steps. First, we create a master "person-link" file that tracks all individuals ever linked across any pair of censuses, alongside a separate "intergen-link" file that identifies all father–son relationships observed in the raw census data. Second, beginning from each census snapshot, we merge individuals with the person-link file to track them over time and then attach the intergen-link file to identify their fathers. When a father is observed in multiple census years, we retain the most recent observation, under the assumption that the most recent father–son link is the most accurate for the current census year.[14] Third, we merge again with the person-link file to track each father's occupational trajectory. The resulting sample of father–son pairs with valid occupational information yields intergenerational matching rates of roughly 6–9 percent for White men and about 2 percent for Black men.[15]

We partition the sample into birth cohorts defined by 10-year intervals spanning from 1800–1810 to 1969–1979. Each cohort is anchored to a benchmark year in which the individuals are observed during their prime working age. Specifically, for each benchmark year, we include individuals born within a 20-year window such that they are aged between 35 and 55 at the time of observation. For example, the benchmark year 1900 includes individuals born between 1846 and 1865, assigned to the 1850–1860 birth cohort; the benchmark year 2010 includes those born between 1956 and 1975, assigned to the 1960–1970 cohort. This construction yields 18 overlapping cohorts observed at benchmark years from 1850 to 2019, providing a finer temporal resolution than the 20-year intervals used

[14] If multiple "intergen-links" are found for the same individual at different points in time, it is possible that different fathers are linked to the same individual (as identified by "person-link"). It could be that the fathers are the same person but failed to be linked (false negative), or also that the sons are not the same person but linked by error (false positive). The current approach avoids all this complexity with the assumption that the most recent observation of the father-son link is the most accurate, at least for the current census year.

[15] Table B.3 compares linking and matching rates by different linking methods. CLP yields individual linking rates of 23–30 percent for White men, which are generally more than twice those for Black men (10–15 percent). The intergenerational linking rates are less than half of these figures for both groups.

in **?**.[16]

We construct an *original mobility table* recording the count of individuals in each father-occupation-by-son-occupation cell for each cohort. We then adjust this table to match the contemporaneous occupational distribution among prime-age men (35–55) of the same race, drawn from the Decennial Censuses (1910, 1930, 1950, 1970, and 1990) and the American Community Survey (ACS, 2010 and 2019). Specifically, denoting each cell count as m_{ij}, where i indicates the father's occupation and j the son's, the adjusted values are $\widehat{m}_{ij} = m_{ij}/(\sum_k m_{kj}) \cdot g_j$, where g_j is the occupational distribution among the prime-age men of the current race. This reweighting ensures that $\sum_i \widehat{m}_{ij} = g_j$ holds, thereby correcting for the possibility that the probability of observing a linked father varies across cells of the mobility table. Similar approaches have been applied in prior studies (e.g., Ward, 2023; Jácome et al., 2021).

4.3 Occupational Classification

We measure occupations using the 1950 Census three-digit occupation codes and group them into three nested classification schemes. Our main analyses employ a 5-category scheme (Professional and Managers; Clerical and Sales; Skilled Manual; Semi-Skilled and Unskilled Manual; Farming). Our classification approach is consistent with the Erikson–Goldthorpe–Portocarero (EGP) class scheme widely used in the sociological literature (Erikson and Goldthorpe, 1992), which focuses on differences in task and skill requirements across occupational categories and therefore does not impose an explicit ranking. This aligns with our analyses based on Markov chains, which do not require ranked categories.[17]

For individuals observed in multiple census or survey years, we assign the modal occupation recorded between ages 25 and 55. The modal value helps remove measurement errors associated with a snapshot measure of occupations.[18] When the mode is not unique, we use the occupation observed closest to age 40 as the lifetime occupation (Haider and Solon, 2006). In longitudinal surveys, where both offspring and parents report their main occupations in each survey year, we apply the same procedure. In cross-sectional surveys, where fathers' occupations are typically reported retrospectively by the respondent (usually recalling the father's occupation at age 15 or 16), we rely on these retrospective reports.

[16] See Table B.2 for the complete mapping between benchmark years, birth year ranges, and birth cohort definitions.

[17] However, interpreting our results together with wage and skills information of occupations helps connect our findings with the existing literature. We include these discussion points in the relevant sections of our paper.

[18] Ward (2023) shows that such errors in linked censuses may lead to biased mobility estimates. Mazumder and Acosta (2015) articulates the importance of constructing "primary occupations" to mitigate concerns related to transitory effects.

While fathers' information provided in prospective and retrospective surveys may not always match, Jácome et al. (2021) shows that nationally representative longitudinal surveys such as the PSID demonstrate over 80% consistency rates between the occupations reported by individuals and those their offspring recall.

We note that the "Farming" group, constructed from the 1950 Census 3-digit occupation codes 100 (farmers, both owners and tenants), 123 (farm managers), and 810–840 (farm laborers), combines landowners, tenants, and laborers who differ significantly in economic and social status. Previous studies have attempted to address this heterogeneity by introducing finer subgroups based on land ownership (Collins and Wanamaker, 2022; Collins and Margo, 2011) or by constructing income measures that include non-wage and self-employment income (Collins and Wanamaker, 2022; Ward, 2023; Jácome et al., 2021).[19] In our setting, further subdivision is impractical because it produces cells with zero transitions in the Markov chain, particularly among recent cohorts of Black men (see discussions in Durlauf et al. (2024)). As a robustness check, we reclassify farm laborers into the Manual group.

5 Structural versus Exchange Mobility in the US Since 1850

5.1 Estimation

In section 2.2, we showed how the parameters $A_{j,t}$ and $Z_{ij,t} \equiv \log \theta_{ij,t}$ can be mapped from the data, taking as given (γ, ν).[20] We now discuss how we estimate these remaining structural parameters (γ, ν). The parameter ν governs the elasticity of occupational choice with respect to wages—that is, the responsiveness of transition probabilities to cross-occupation wage differentials—while γ governs the elasticity of labor demand.

We estimate ν using the log-odds specification implied by the model. For each origin occupation i and destination $j \neq i$, the transition probabilities satisfy

$$\log \frac{P_{ij,t}}{P_{ii,t}} - \log \frac{L_{j,t-1}}{L_{i,t-1}} = \nu \cdot \log \frac{W_{j,t}}{W_{i,t}} + \overline{Z}_t + \widetilde{Z}_{ij,t}, \tag{6}$$

where $W_{j,t}$ denotes occupational wages, $\overline{Z}_t$ captures the average friction associated with moving to a different occupation in period t, and $\widetilde{Z}_{ij,t}$ is a residual reflecting origin–

[19] Collins and Margo (2011) relies on the home ownership information available in historical censuses to proxy for farm ownership given the close correspondence in the data. Note that, in our context, since we treat our occupations as unranked, this issue is less concerning compared to income-based mobility studies that use occupations to construct proxies for income rank or status.

[20] To reiterate, $A_{j,t} = \frac{L_{j,t} W_{j,t}^{\gamma}}{\sum_k L_{k,t} W_{k,t}^{\gamma}}$, and $\theta_{ij,t} = \frac{P_{ij,t}}{P_{ii,t}} \cdot (\frac{W_{j,t}}{W_{i,t}})^{-\nu} \cdot (\frac{L_{j,t-1}}{L_{i,t-1}})^{-1}$.

destination specific preference heterogeneity. With 5 occupations, estimating equation (6) yields $\hat{\nu} = 0.437$, varying in the range of around $(0.3,\ 0.8)$ depending on the number of occupations. These estimates indicate that transition probabilities are moderately responsive to cross-occupation wage differentials, consistent with meaningful but not frictionless occupational mobility. Note that ν can be estimated directly from the data $\{P_{ij,t}, W_{j,t}\}$ before estimating $\{A_{j,t}, Z_{ij,t}\}$.

For the demand elasticity γ, we adopt a calibration approach rather than direct estimation, examining the sensitivity of our results across three specifications that span a range of plausible values: $\gamma = 0$ (perfect complements, in which occupational demand does not respond to wages), $\gamma = 1$ (Cobb-Douglas, our baseline), and $\gamma = 2$ (moderate substitutability across occupations). The Cobb-Douglas case serves as a natural benchmark, as it implies that occupational expenditure shares are constant and that demand adjusts proportionally to wage changes. As to be explained below, the main findings on long-run trends in exchange mobility are qualitatively robust across all three specifications, though the estimated magnitude of the structural component varies with γ.

We now apply the structural framework developed in Section 2 to the long-run mobility data described in Section 4. Our analysis proceeds in three steps. We first document long-run trends in exchange mobility for White Americans under our baseline measurement, comparing these to estimates obtained using traditional approaches. We then examine welfare-based measures that quantify the cost of mobility frictions in utility terms. Finally, we extend the analysis to Black Americans and characterize Black–White differences in exchange mobility over the full sample period.

5.2 Exchange Mobility Among White Americans

Figure 1 presents the main results for White American men from 1850 to 2019, using the 5-category occupational classification under the baseline Cobb-Douglas specification ($\gamma = 1$). The figure displays three series: total mobility, measured as the sum of off-diagonal elements of the observed mobility table (navy long-dashed line); and exchange mobility under our model-based framework with $\gamma = 1$ (navy solid line) and $\gamma = 0$ (orange solid line). The gap between total mobility and each exchange mobility series captures the structural component implied by the respective specification.

Two patterns emerge. First, exchange mobility—measured as the sum of off-diagonal elements of the counterfactual mobility table—increases gradually over the entire period and stabilizes beginning in the 1970s. Second, the gap between total mobility and exchange mobility, which captures the structural component, is substantial during the 1950s through the 1970s and narrows considerably in recent decades. This pattern is consistent with

the large-scale shift out of agriculture during mid-century, generating a sizable structural component that subsequently diminishes as the occupational distribution stabilizes.

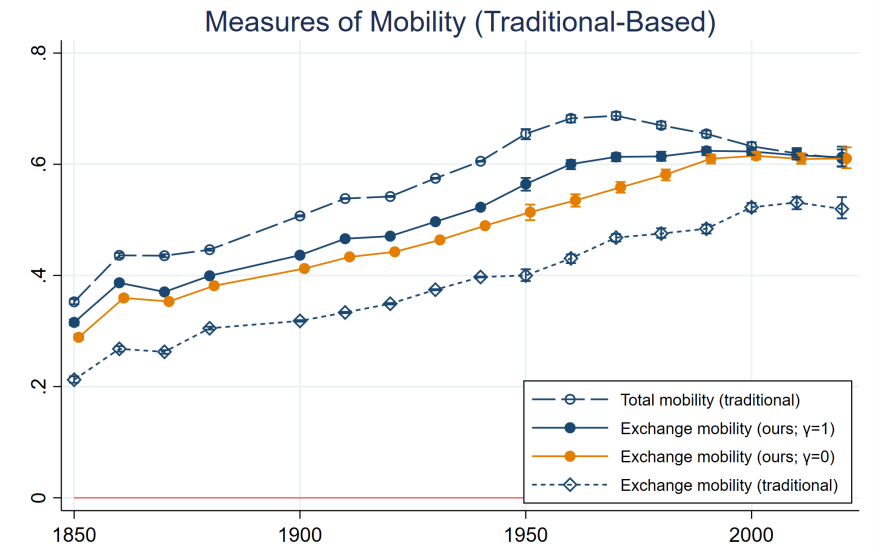


Figure 1: Long-run Trends in Exchange Mobility (White Americans, 1850-2019)

Notes: This figure plots total mobility and model-based exchange mobility for White American men across birth cohorts from 1850 to 2019, using the 5-category occupational classification. Total mobility is the sum of off-diagonal elements of the observed mobility table (in navy long-dashed line). Exchange mobility is computed from the counterfactual economy in which occupational demand shifters are held constant ($\tilde{A}_{j,t} = A_{j,t-1}$), under two specifications of the demand elasticity: $\gamma = 1$ (Cobb-Douglas, baseline; in navy solid line) and $\gamma = 0$ (perfect complements, in orange solid line). Exchange mobility under the traditional approach (in navy short-dashed line) is computed by subtracting the traditional measure of structural mobility, $\frac{1}{2}\sum_i |f_i - s_i|$, from total mobility, where f_i and s_i denote the marginal distributions of fathers' and sons' occupations. Error bars represent 95% confidence intervals constructed via bootstrap.

The stabilization of exchange mobility since the 1970s suggests that, after controlling for aggregate occupational restructuring, the underlying fluidity of intergenerational occupational transitions has plateaued for White men over the past half-century. This result stands in contrast to the continued increase in total mobility observed through the 1970s, which our framework attributes primarily to structural forces.

The qualitative pattern is similar under $\gamma = 0$, which assumes that occupational demand does not respond to wages in the aggregate production function. This specification nests the approach proposed by Hazelrigg (1974) and provides a particularly useful benchmark for comparisons spanning the historical periods, since it does not require occupation-specific wage data—an important consideration given that wage information for periods prior to 1930 must be imputed from the 1940 Census.

5.3 Comparison with Existing Approaches

To situate our findings within the existing literature, we compare our exchange mobility estimates with those obtained using two alternative approaches: the traditional residual method and the Altham statistic used in Long and Ferrie (2013).

Traditional measures. The traditional approach defines structural mobility as $\frac{1}{2}\sum_i |f_i - s_i|$, where f_i and s_i denote the marginal distributions of fathers' and sons' occupations, and computes exchange mobility as the residual after subtracting this quantity from total mobility. Figure 1 overlays these traditional estimates in a navy short-dashed line. Three differences are noteworthy. First, the traditional approach attributes a substantially larger share of total mobility to structural forces. Second, while exchange mobility follows a qualitatively similar increasing trajectory under both approaches, the traditional measure does not exhibit the trendless pattern that our framework identifies for recent decades. Third, the traditional residual method lacks a structural interpretation—it is a purely statistical decomposition that does not specify the economic mechanisms through which occupational restructuring affects mobility. Our approach, by contrast, models the counterfactual equilibrium that would obtain if occupational demand had remained unchanged, allowing wages and labor supply to adjust endogenously.

Comparison with Long and Ferrie (2013). The work is closest to ours in scope, as it also examines long-run trends in occupational mobility in the US. However, the two studies differ in both data and methodology, and reach different conclusions: Long and Ferrie (2013) finds that long-run mobility has declined, while our analysis shows that exchange mobility has increased over time. Two sources of difference merit discussion.

The first concerns the underlying data. Our data construction draws on rich father-son linkages made available through automated algorithms such as the CLP and MLP, which provide substantially richer coverage than was available to their work. To assess the role of data differences, we apply our framework to the mobility tables reported in their work and find that exchange mobility increases over time even using their data (Figure 2, orange solid line), suggesting that the divergence in conclusions is driven primarily by methodological differences rather than data.

The second concerns how structural components are addressed. The Altham statistic, which is employed as their primary measure, is based on odds ratios and is therefore invariant to marginal transformations of the mobility table. This property effectively absorbs any changes in the occupational distribution, but it is not transparent about what, precisely, is being controlled for—and it precludes separate assessment of the magnitude of structural

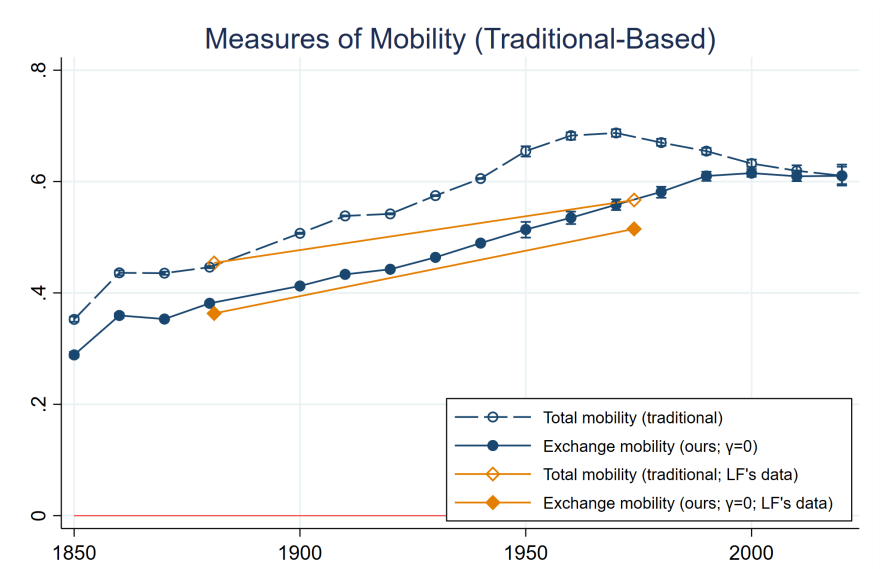


Figure 2: Long-run Trends in Exchange Mobility (cp. Long and Ferrie, 2013)

Notes: This figure plots total mobility and model-based exchange mobility for White American men across birth cohorts from 1850 to 2019, using the 5-category occupational classification. Total mobility is the sum of off-diagonal elements of the observed mobility table (in navy long-dashed line). Exchange mobility is computed from the counterfactual economy in which occupational demand shifters are held constant ($\tilde{A}_{j,t} = A_{j,t-1}$) with $\gamma = 1$ (Cobb-Douglas, baseline; in navy solid line). The orange solid line plots exchange mobility estimated by applying the same model-based framework ($\gamma = 1$) to the mobility tables reported in Long and Ferrie (2013). Error bars represent 95% confidence intervals constructed via bootstrap.

mobility. Our model-based approach, by contrast, explicitly specifies the counterfactual: we hold occupational demand shifters constant and solve for the equilibrium that would obtain in the absence of structural change, allowing us to quantify both the exchange and structural components.

5.4 Welfare-Based Measures

We next present our mobility estimates using the welfare measure developed in Section 3.3. This measure quantifies immobility as the average welfare loss experienced by individuals in the current economy relative to a perfectly mobile benchmark in which all mobility frictions are eliminated ($Z_{ij,t} = 0$). As derived in Section 3.3, this welfare gap decomposes into two terms: a mobility component, $\sum_i m_i \log(P_{ii}/m_i)$, which captures the welfare cost of intergenerational persistence in occupational attainment; and a term reflecting the variance of occupational wages.

Figure 3 presents the welfare-based results. Again, we focus on White Americans using 5-category occupations. Total immobility—measured as the full welfare gap—is represented in solid lines, and the mobility component—the first term as we decompose

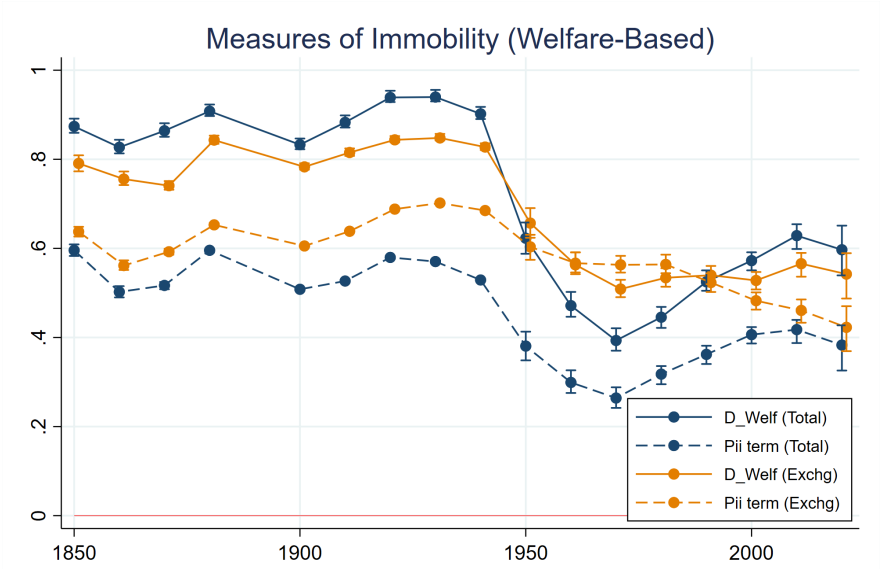


Figure 3: Long-run Trends in Exchange Mobility (Welfare-based measure)

Notes: This figure plots welfare-based measures of immobility for White American men across birth cohorts from 1850 to 2019, using the 5-category occupational classification. Immobility is measured as the average welfare loss relative to a perfectly mobile economy in which $Z_{ij,t} = 0$. Navy lines correspond to estimates based on the observed mobility table; orange lines correspond to estimates based on the counterfactual mobility table in which occupational demand shifters are held constant ($\tilde{A}_{j,t} = A_{j,t-1}$, $\gamma = 1$). For each color, the solid line plots the full welfare gap, and the dashed line isolates the mobility component, $\sum_i m_i \log(P_{ii}/m_i)$, which captures the welfare cost of intergenerational persistence net of the occupational wage variance term. Error bars represent 95% confidence intervals constructed via bootstrap.

total immobility—is shown in dashed lines. The results in navy are based on the actual mobility table, whereas those in orange are based on the counterfactual mobility table.

Focusing on total immobility (navy lines), we find that the estimates decrease until the 1970s and then increase from the 1980s onward. The overall trend, including this reversal, remains similar, even when isolating the mobility component of this measure (in the navy dashed line). When we isolate exchange mobility by holding occupational demand constant (orange lines), however, the mobility component of the welfare gap continues to decline over the full sample period, confirming that exchange immobility has decreased monotonically once structural forces are netted out. This result provides a welfare-grounded counterpart to the finding from the traditional-based measure that exchange mobility has increased over time.

The welfare-based measure offers two advantages over standard approaches. First, it aligns conceptually with existing measures in the sociological literature—such as the log-linear and log-multiplicative models—that assess mobility relative to the independence case, but grounds the comparison in a micro-founded utility framework. Second, it

provides a quantitative interpretation of the welfare cost of mobility frictions, moving beyond purely descriptive summaries of transition probabilities.

5.5 Black-White Differences

We extend the analysis to Black American men and examine racial differences in exchange mobility. Figure 4 presents the results.

For Black Americans, total mobility increases over the full sample period, with the rate of increase being particularly rapid during the 1950s through the 1970s. Our decomposition reveals that structural mobility accounts for a nontrivial share of this increase—and a larger share than for White Americans. This finding is consistent with the disproportionate impact of the mid-century transformation of the agricultural economy on Black workers, who were heavily concentrated in farming occupations in the early periods of our sample.

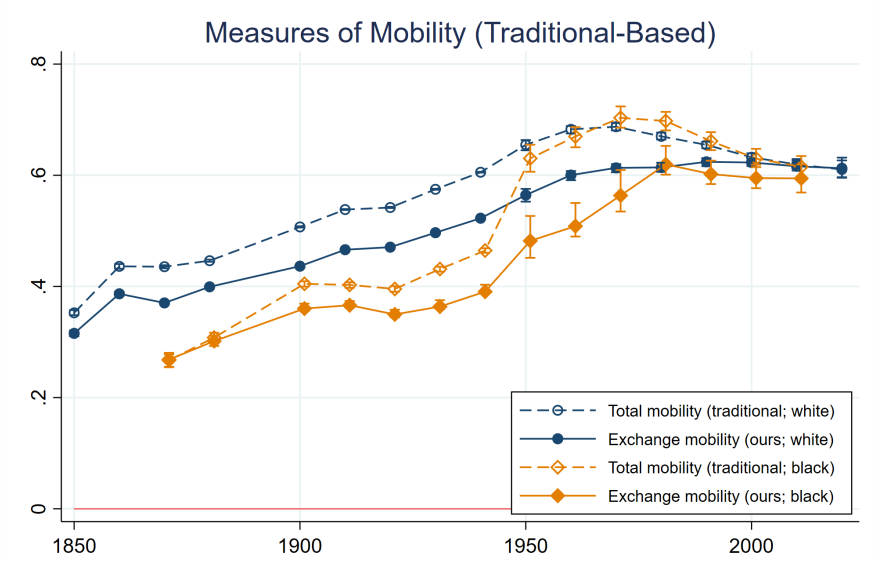


Figure 4: Black-White Differences in Long-run Trends in Exchange Mobility

Notes: This figure plots total mobility and model-based exchange mobility separately for White and Black American men across birth cohorts from 1850 to 2019, using the 5-category occupational classification. Navy lines correspond to White Americans; orange lines correspond to Black Americans. For each group, total mobility is the sum of off-diagonal elements of the observed mobility table (dashed lines), and exchange mobility is computed from the counterfactual economy in which occupational demand shifters are held constant ($\tilde{A}_{j,t} = A_{j,t-1}$) with $\gamma = 1$ (Cobb-Douglas, baseline; solid lines). The gap between the dashed and solid lines of each color captures the structural component for the respective racial group. Error bars represent 95% confidence intervals constructed via bootstrap.

Exchange mobility for Black men also increases over time, but follows a somewhat different trajectory than for Whites. It rises steadily through the 1980s and then levels off, remaining at a persistently lower level than exchange mobility for White men. This gap

implies that, even after controlling for the structural forces associated with occupational restructuring, Black men face greater frictions in intergenerational occupational transitions than their White counterparts. The persistence of this gap into recent decades—when the structural component has largely dissipated—underscores that contemporary racial disparities in occupational mobility cannot be attributed solely to differences in the pace of structural change.

The welfare-based analysis reinforces these patterns. The results are shown in Figure 5. For Black men, the total welfare gap (relative to the perfectly mobile benchmark) decreases until the 1970s and increases from the 1980s to the present. The welfare gains observed during the 1950s through the 1990s—which reflect a temporary period of low intergenerational persistence—are driven primarily by structural components associated with the mass movement of Black workers out of agricultural occupations. Controlling for these structural forces, the welfare cost of mobility frictions for Black men declines gradually over the full sample period, indicating that exchange mobility has improved in welfare terms, albeit from a substantially higher baseline of immobility than for White men.

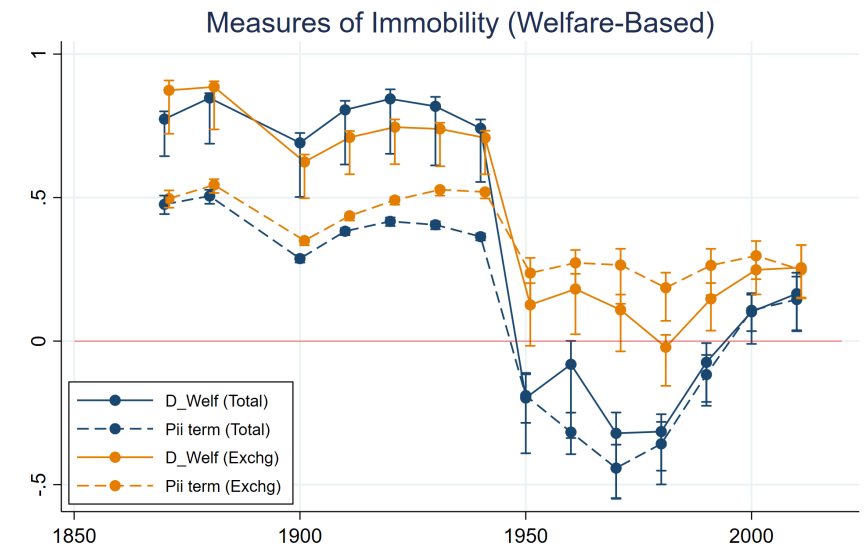


Figure 5: Long-run Trends in Exchange Mobility (Welfare-based, Black Americans)

Notes: This figure plots welfare-based measures of immobility for Black American men across birth cohorts from 1850 to 2019, using the 5-category occupational classification. Immobility is measured as the average welfare loss relative to a perfectly mobile economy in which $Z_{ij,t} = 0$. Navy lines correspond to estimates based on the observed mobility table; orange lines correspond to estimates based on the counterfactual mobility table in which occupational demand shifters are held constant ($\tilde{A}_{j,t} = A_{j,t-1}$, $\gamma = 1$). For each color, the solid line plots the full welfare gap, and the dashed line isolates the mobility component, $\sum_i m_i \log(P_{ii}/m_i)$, which captures the welfare cost of intergenerational persistence net of the occupational wage variance term. Error bars represent 95% confidence intervals constructed via bootstrap.

5.6 Summary of Findings

Three principal findings emerge from the analysis. First, exchange mobility among White Americans has increased over the long run but has stabilized since the 1970s, a pattern that is obscured in measures of total mobility by the confounding effects of mid-century occupational restructuring. Second, welfare-based measures confirm this trajectory and further reveal that the recent increase in total immobility is attributable to rising occupational wage inequality rather than to a decline in the fluidity of intergenerational transitions. Third, Black Americans exhibit lower exchange mobility than Whites throughout the sample period; while exchange mobility has improved for both groups, the racial gap persists into the contemporary era and cannot be explained by differential exposure to structural change. These findings contribute to the ongoing debate on long-run mobility trends in the United States (Long and Ferrie, 2013; Song et al., 2020; Ward, 2023; Jácome et al., 2021) by offering the first structural decomposition of total mobility into exchange and structural components grounded in an economic model.

6 Conclusion

This paper introduces a structural framework for decomposing intergenerational occupational mobility into its structural and exchange components. Although this decomposition has been a central objective of the sociological mobility literature since at least the 1970s, prior approaches have relied on purely statistical methods that lack an explicit economic interpretation of what is being held constant.

Our framework addresses this limitation through a parsimonious model of occupational choice in which labor demand, wages, and intergenerational transition probabilities are jointly determined in equilibrium. By holding occupational demand shifters constant and solving for the resulting counterfactual equilibrium, we obtain a transparent and economically grounded measure of exchange mobility. We further introduce a welfare-based measure that quantifies the cost of mobility frictions in utility terms.

A key feature of our approach exploits information in occupational wages—data that the existing literature on mobility tables has not utilized—to separately identify the roles of occupational demand and supply-side frictions. We show, however, that in the limiting case where the elasticity of substitution across occupations is zero, wage information is not required, and the resulting mobility patterns are qualitatively similar to those obtained under more realistic parameterizations. This special case nests existing approaches and provides a useful benchmark for historical periods where wage data must be imputed.

Applying this framework to US data spanning from 1850 to the present, we find that

exchange mobility has increased since the late nineteenth century and stabilized around the 1970s. This result speaks to what Abramitzky et al. (2025) identify as an open question in the literature—whether long-run mobility in the US has been rising or falling—and aligns with more recent evidence favoring an increasing trend. Our decomposition clarifies that much of the apparent decline in total mobility in recent decades reflects the exhaustion of structural forces, particularly the transition out of agriculture, rather than a reversal in the fluidity of intergenerational occupational transitions.

We acknowledge several simplifying assumptions, most notably the absence of overlapping generations—a restriction that was necessary to maintain compatibility with the mobility table framework used throughout the literature. A companion paper relaxes this assumption by introducing a two-period overlapping generations structure. More broadly, our model can be viewed as a stylized version of the structural labor models in the tradition of Dix-Carneiro (2014), adapted to the intergenerational setting. Extensions that incorporate worker heterogeneity in productivity or endogenous human capital investment are natural directions for future work.

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.

_ , Leah Platt Boustan, and Tamar Matiashvili, "Intergenerational Mobility over Two Centuries," Working Paper 33330, National Bureau of Economic Research January 2025.

_ , 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.

Altham, Patricia M. E., "The Measurement of Association of Rows and Columns for an r × s Contingency Table," *Journal of the Royal Statistical Society: Series B (Methodological)*, 1 1970, *32* (1), 63–73.

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.

Blau, Peter M., Otis Dudley Duncan, David L. Featherman, and Robert M. Hauser, "Occupational Changes in a Generation, 1962 and 1973," Inter-university Consortium for Political and Social Research (ICPSR). May 1994.

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

Buckles, Kasey, Adrian Haws, Joseph Price, and Haley E.B. Wilbert, "Breakthroughs in Historical Record Linking Using Genealogy Data: The Census Tree Project," Working Paper 31671, National Bureau of Economic Research September 2023.

Bumpass, Larry L. and James A. Sweet, "National Survey of Families and Households, Wave 2: 1992–1994, [United States]," Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. 2018.

_ **and** _ , "National Survey of Families and Households, Wave 3: 2001–2003, [United States]," Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. 2018.

_ **,** _ **, and Vaughn R. A. Call**, "National Survey of Families and Households, Wave 1: 1987–1988, [United States]," Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. 2017.

Bureau of Labor Statistics, U.S. Department of Labor, "National Longitudinal Survey of Young Men, 1966–1981 (Rounds 1–12)," Center for Human Resource Research (CHRR), Ohio State University. 1992.

_ , "National Longitudinal Survey of Older Men, 1966–1990 (Rounds 1–13)," Center for Human Resource Research (CHRR), Ohio State University. 1999.

_ , "National Longitudinal Survey of Youth 1979 cohort, 1979–2012 (Rounds 1–25)," Center for Human Resource Research (CHRR), Ohio State University. 2014.

Bureau of the Census, U.S. Department of Commerce, "Survey of Income and Program Participation (SIPP), 1986–1988 Panel [dataset]," Inter-university Consortium for Political and Social Research (ICPSR). 2008. Public use data files.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber, "Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities," *NBER Macroeconomics Annual*, 2013, *27*, 1–56.

_ **, 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.

Collins, William J. and Marianne H. Wanamaker, "African American Intergenerational Economic Mobility since 1880," *American Economic Journal: Applied Economics*, July 2022, *14* (3), 84–117.

_ **and Robert A. Margo**, "Race and Home Ownership from the End of the Civil War to the Present," *American Economic Review*, May 2011, *101* (3), 355–59.

Davern, Michael, Rene Bautista, Jeremy Freese, Pamela Herd, and Stephen L. Morgan, "General Social Survey, 1972–2021," NORC at the University of Chicago. 2022.

Dix-Carneiro, Rafael, "Trade Liberalization and labor market dynamics," *Econometrica*, 2014, *82* (3), 825–885.

Duncan, Otis Dudley, "A Socioeconomic Index for All Occupations," in Albert J. Reiss, Jr., ed., *Occupations and Social Status*, New York: Free Press, 1961, pp. 109–138.

Durlauf, Steven N, Gueyon Kim, Dohyeon Lee, and Xi Song, "The Evolution of Black-White Differences in Occupational Mobility Across Post-Civil War America," Working Paper 32370, National Bureau of Economic Research April 2024.

Erikson, Robert and John H Goldthorpe, *The Constant Flux: A Study of Class Mobility in Industrial Societies*, Oxford University Press, USA, 1992.

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

Hauser, Robert M., Peter J. Dickinson, Harry P. Travis, and John N. Koffel, "Structural Changes in Occupational Mobility Among Men in the United States," *American Sociological Review*, 1975, *40* (5), 585–598.

Hazelrigg, Lawrence E., "Partitioning Structural Effects and Endoge Nous Mobility Processes in the Measurement of Vertical Occupational Status Change," *Acta Sociologica*, 1974, *17* (2), 115–139.

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, Michael and Thomas A DiPrete, "What We Have Learned: RC28's Contributions to Knowledge About Social Stratification," *Research in social stratification and mobility*, 2006, *24* (1), 1–20.

Jackson, James S. and Harold W. Neighbors, "National Survey of Black Americans, Waves 1–4, 1979–1980, 1987–1988, 1988–1989, 1992," Inter-university Consortium for Political and Social Research (ICPSR). November 2005.

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.

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.

Mazumder, Bhashkar and Miguel Acosta, "Using Occupation to Measure Intergenerational Mobility," *The Annals of the American Academy of Political and Social Science*, 2015, *657* (1), 174–193.

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

Social Research Center, "Panel Study of Income Dynamics, public use dataset," Institute for Social Research, University of Michigan. 2020.

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.

Treiman, Donald J., *Occupational Prestige in Comparative Perspective*, New York: Academic Press, 1977.

Treiman, Donald J and Harry BG Ganzeboom, "The Fourth Generation of Comparative Stratification Research," in Stella R Quah and Arnaud Sales, eds., *The International Handbook of Sociology*, London: Sage, 2000, pp. 122–50.

Veroff, Joseph, Elizabeth Douvan, and Richard Kulka, "Americans View Their Mental Health, 1957 and 1976: Selected Variables," Inter-university Consortium for Political and Social Research (ICPSR) [distributor]. November 2005.

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

Appendix

FOR ONLINE PUBLICATION

A Structural Model of Structural and Exchange Mobility

Steven Durlauf Gueyon Kim Dohyeon Lee Xi Song

Appendices

A 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	Skilled Manual	Skilled Manual	500-594, 762, 773, 781, 782
	Semi-Skilled, Unskilled Manual	Semi-Skilled, Unskilled Manual	595, 600-690, 700-790, 910-970
Farming	Farming	Farming	100, 123, 810-840

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)
1850	1796–1815	1800–1810
1860	1806–1825	1810–1820
1870	1816–1835	1820–1830
1880	1826–1845	1830–1840
1900	1846–1865	1850–1860
⋮	⋮	⋮
2000	1946–1965	1950–1960
2010	1956–1975	1960–1970
2019	1965–1984	1969–1979

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: Matching Rates by Linking Methods

Linking Method	Year	Whites				Blacks			
		Linked	IG Linked			Linked	IG Linked		
			Any	w/Occ$_s$	w/Occ$_{sf}$		Any	w/Occ$_s$	w/Occ$_{sf}$
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
CLP	1910	0.244	0.097	0.090	0.071	0.138	0.036	0.034	0.029
	1920	0.273	0.095	0.088	0.065	0.134	0.032	0.030	0.023
	1930	0.303	0.108	0.101	0.074	0.142	0.033	0.031	0.024
	1940	0.232	0.125	0.118	0.089	0.101	0.040	0.037	0.026
MLP	1910	0.647	0.184	0.173	0.154	0.405	0.057	0.055	0.047
	1920	0.669	0.178	0.166	0.142	0.389	0.054	0.052	0.043
	1930	0.694	0.195	0.184	0.157	0.385	0.056	0.054	0.044
	1940	0.547	0.216	0.207	0.181	0.279	0.064	0.060	0.049
Census Tree	1910	0.810	0.323	0.303	0.270	0.657	0.147	0.140	0.120
	1920	0.823	0.311	0.291	0.243	0.635	0.147	0.141	0.116
	1930	0.854	0.409	0.386	0.339	0.644	0.191	0.184	0.160
	1940	0.782	0.501	0.480	0.447	0.551	0.240	0.225	0.205

Notes: The table reports various definitions of linking rates for each census year from 1910 onward for each of the three linking methods (CLP, MLP, and Census Tree). This starting point corresponds to our sample's earliest cohort, born between 1860 and 1870, who were aged 35 to 55 in 1910. Note that the linked censuses 1850-1900 are also used for the linking process. We focus on Black and White men in their prime age (35 to 55) in each census year for constructing these rates. Columns (1) and (5) show the probability of men linked across different censuses; Columns (2) and (6) show the probability of men matched to their fathers; Columns (3) and (7) show the probability of men with valid occupation information matched to their fathers; and Columns (4) and (8) show the probability of men with valid occupation information matched to their fathers with valid occupation information.

B Model Appendix

B.1 Log-Linear Approximation

Our approach to solving the model uses log-linearization by assuming a small change in A_t. Specifically, we start with the observed economy in t with $A_{j,t}$ that is consistent with the equilibrium. Next, we consider a deviation from this economy denoted by $A^*_{j,t}$, with $A^*_{j,t} = A_{j,t} \exp \alpha_{j,t} \approx A_{j,t}(1+\alpha_{j,t})$, with $\alpha_{j,t} = \log \frac{A^*_{j,t}}{A_{j,t}} \approx \frac{A^*_{j,t}}{A_{j,t}} - 1$. Equations (1), (2), and (3) can be log-linearized as,

$$\alpha_{j,t} = x_{j,t} - \sum_k A_{k,t} x_{k,t} \quad (x_{j,t} \equiv \gamma w_{j,t} + m_{j,t}) \tag{7}$$

$$p_{ij,t} = \frac{1}{\nu}\Big(w_{j,t} - \sum_k P_{ik,t} w_{k,t}\Big) \tag{8}$$

$$m_{j,t} = \sum_k \gamma_{kj,t} p_{kj,t} \tag{9}$$

where $\gamma_{ij,t} \equiv \frac{P_{ij,t} M_{i,t-1}}{M_{j,t}}$ is the probability that the parent's occupation was i conditional on the child's occupation being j. In Equation (7), since the absolute level of wages acts as numeraire, we can arbitrarily normalize it so that $\alpha_t = x_t \equiv \gamma w_t + m_t$ holds, which describes the demand for labor.

Combining (8) and (9), and noting that $\sum_k \gamma_{kj,t} = 1$, and further using the matrix notation for P and γ,

$$m_t = \frac{1}{\nu}(I - \gamma_t' P_t) w_t \equiv \frac{1}{\nu}(I - H_t) w_t \tag{10}$$

which summarizes the relative labor supply curve. Also, each element $h_{ij,t}$ of H_t is a weighted average of $P_{ik,t}$ where the weights $\gamma_{lk,t}$ add up to 1. The sum of each row i of H_t is

$$\sum_k h_{ik,t} = \sum_k \Big(\sum_l \gamma_{lk,t}\Big) P_{ik,t} = \sum_k P_{ik,t} = 1$$

In turn, the sum of each row of $(I - H_t)$ is 0. Note that due to the built-in normalization, $\sum_k P_{ik,t} = 1$, $\sum_k M_{k,t} = 1$, and $\sum_k A_{k,t} = 1$ hold, which translate to $\sum_k P_{ik,t} p_{ik,t} = 0$, $\sum_k M_{k,t} m_{k,t} = 0$, and $\sum_k A_{k,t} \alpha_{k,t} = 0$ respectively.

Assuming that the parameters ν and γ are known, the counterfactual economy can be obtained by the following steps. First, we find the levels of the aggregate variables $(M_t, W_t), (M_{t-1}, W_{t-1})$, as well as P_t, γ_t and construct the weights for each occupation: $A_{j,t} =$

$\frac{M_j W_j^\gamma}{\sum_k M_k W_k^\gamma}$ and similarly for $A_{j,t-1} = A^*_{j,t}$. Second, we construct the deviation from the current economy, $\alpha_{j,t} = \log \frac{A^*_{j,t}}{A_{j,t}} \approx \frac{A^*_{j,t}}{A_{j,t}} - 1$. Third, We use the inverse mapping between α_t and w_t to back out w_t where $w_t = \frac{1}{\nu}\Big((1+\gamma\nu)I - H\Big)^{-1}\alpha_t$.[21] Finally, we adjust the transition probability (and the mobility table) using w_t:

$$P^*_{ij,t} = \frac{P_{ij,t} \exp \frac{w_{j,t}}{\nu}}{\sum_k P_{ik,t} \exp \frac{w_{k,t}}{\nu}}. \tag{11}$$

This counterfactual transition probability P^*_t is the exchange mobility that we wanted to extract from the observed transition matrix P_t.

B.2 Derivation of the welfare

$$V^s_i = \max_j \Big(\log W_j + Z_{ij} + \Big[\nu \log m_j\Big] + \nu \epsilon^s_j \Big)$$

where $m_j \equiv L_{j,t-1}$ captures the differential exposure to each occupation when drawing ϵ^s_j, microfounded by the Poisson point process with exponential density.

$$\begin{aligned}
\overline{V}_i &= \log \sum_j \exp \frac{\log W_j + Z_{ij} + \nu \log m_j}{\nu} \\
&= \log \sum_j W_j^{1/\nu} \exp \frac{Z_{ij}}{\nu} \cdot m_j \\
&= \log \sum_j W_j^{1/\nu} \theta_{ij}
\end{aligned}$$

Relative transition rule:

$$\begin{aligned}
\frac{P_{ij}}{P_{ii}} &= \Big(\frac{W_j}{W_i}\Big)^{1/\nu} \frac{\theta_{ij}}{\theta_{ii}} \\
\frac{\theta_{ij}}{\theta_{ii}} &= \frac{P_{ij}}{P_{ii}} \Big(\frac{W_j}{W_i}\Big)^{-1/\nu}
\end{aligned}$$

Normalization: $Z_{ii} = 0$, $\theta_{ii} = m_i$

[21] If $\gamma = 0$, this inversion is not possible since $(I - H)$ is not full rank. In that case, we can exclude an arbitrary row and a column before taking the inverse. The resulting w_t can be interpreted as the relative wage with respect to the omitted occupation. If $\gamma > 0$, this inversion is possible (with probability 1), but with a caveat that the level (numeraire) was arbitrarily chosen so as to satisfy $\gamma w_t + m_t = \alpha_t$.

$$\begin{aligned}\overline{V}_i &= \log \sum_j W_j^{1/\nu} \Big(\frac{P_{ij}}{P_{ii}}\Big)\Big(\frac{W_j}{W_i}\Big)^{-1/\nu} m_i \\ &= \log \frac{m_i W_i^{1/\nu}}{P_{ii}} \sum_j P_{ij} \\ &= \log \frac{m_i W_i^{1/\nu}}{P_{ii}}\end{aligned}$$

Aggregation:

$$\begin{aligned}E[\overline{V}_i] &= \sum_i m_i \log \frac{m_i W_i^{1/\nu}}{P_{ii}} \\ &= \sum_i m_i \log W_i^{1/\nu} - \sum_i m_i \log \frac{P_{ii}}{m_i}\end{aligned}$$

Counterfactual:

$$\overline{V}_i^0 = \overline{V}_i\Big|_{Z_{ij}=0 \Leftrightarrow \theta_{ij}=m_j} = \log \sum_j m_j W_j^{1/\nu} = E[\overline{V}_i^0] \text{ (indep. of } i)$$

$$E[\overline{V}_i^0] - E[\overline{V}_i] = \underbrace{\left(\log \sum_i m_i W_i^{1/\nu} - \sum_i m_i \log W_i^{1/\nu}\right)}_{\text{Gap in the Jensen's inequality}} + \sum_i m_i \log \frac{P_{ii}}{m_i}$$

Compared to the previous version, the covariance term disappeared as intended, and the P_{ii} related term changed from $\sum_i m_i \log \frac{P_{ii}}{1/N}$ to $\sum_i m_i \log \frac{P_{ii}}{m_i}$. Note that since $P_{ii} = \frac{c_{ii}}{m_i}$, $\frac{P_{ii}}{m_i} = \frac{c_{ii}}{m_i^2}$. If the son's distribution is also the same as the father's (m_i), this ratio would measure the deviation from independence along the diagonal, which might be more appealing than the previous result.

C Measurement Appendix: Altham Statistic

Altham statistic measures the sum of relative "log odds ratio" between two mobility tables for all possible combinations of two rows and two columns.

$$d(\mathbf{P},\mathbf{Q}) = \Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N} \Big(\log \frac{p_{ij}p_{lm}}{p_{im}p_{lj}} - \log \frac{q_{ij}q_{lm}}{q_{im}q_{lj}} \Big)^2 \Big)^{\frac{1}{2}}$$

A widely used special case is when one matrix is the case of independence or "full mobility" where each joint probability is equal to the product of the two marginal distributions, denoted by **I**. Then the log odds ratio of **I** is zero for all combination, and thus

$$d(\mathbf{P},\mathbf{I}) = \Big(\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=1}^{N}\sum_{m=1}^{N} \Big(\log \frac{p_{ij}p_{lm}}{p_{im}p_{lj}} \Big)^2 \Big)^{\frac{1}{2}}$$

It can be shown (e.g. Bouchet-Valat, 2022) that this value is proportional to the sum of square of double-demeaned log of the mobility table.

$$d(\mathbf{P},\mathbf{I}) \propto \Big(\sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_{ij}^2 \Big)^{\frac{1}{2}}$$

where $\lambda_{ij} = \log p_{ij} - \overline{\log p_{i\cdot}} - \overline{\log p_{\cdot j}} + \overline{\log p_{\cdot\cdot}}$. Therefore, elements of the log mobility table that have particularly large values after double demeaning contribute to larger values of Altham statistic.

		1880, US (%), Son				1973, US (%), Son			
		W	S	U	F	W	S	U	F
Father	White collar	2.7	1.6	0.5	2.2	19.9	6.2	1.6	0.1
	Skilled	4.1	8.3	1.1	4.6	*18.0*	19.3	3.8	0.2
	Unskilled	1.5	2.0	1.2	1.7	5.5	7.9	2.1	0.2
	Farmer	8.8	10.7	6.4	42.4	4.8	6.5	1.8	2.0
Double-demeaned log ($= \lambda_{ij}$)									
Father	White collar	0.52	-0.16	-0.24	-0.11	**0.96**	-0.09	-0.03	**-0.83**
	Skilled	0.04	0.58	-0.37	-0.25	*0.18*	0.35	0.14	-0.67
	Unskilled	-0.13	-0.01	0.51	-0.38	-0.25	0.22	0.28	-0.25
	Farmer	-0.43	-0.41	0.11	**0.73**	**-0.89**	-0.48	-0.38	**1.75**
$(\sum \lambda_{ij}^2)/N^2$		0.37				0.65			
Altham's statistic		11.9				20.8			

The table illustrates how much each cell contributes to the Altham's statistic, using

an example from Long and Ferrie (2013). The largest element in the bottom left panel is marked bold (0.73), and the elements larger than this (in absolute value) are marked bold in the bottom right panel. In 1973 US, W to W and F to F are particularly high, and W to F and F to W are particularly low, compared to other pairs of occupations as well as to 1880 US.

These extreme values in the relative odds result in the much higher Altham's statistic than the left panel. Note that these values make huge contribution to the Altham's statistic even though their shares in the original mobility table are small. By construction, Altham's statistic is driven by the outliers of the double demeaned probabilities. Existence of outliers in the table, either particularly positive or particularly negative, makes the Altham's statistic larger, which in turn is typically interpreted as a sign of larger distance from independence (full mobility) and thus of lower degree of mobility.

In contrast, modest values that show evidence of high mobility by being close to the product of the two marginal distributions are not able to bring down the Altham's statistic. For example, S to W in 1973 is as large as 18% (italicized) of the entire population, and is also close to the product of two marginal distributions (20.0%). This can be taken as an evidence of sizable mobility, and resulted in a double demeaned log value as small as 0.18. However, it is not able to bring down the already high value of Altham's statistic driven by the outliers.

D Data Appendix: Sources for Survey Data

General Social Survey (GSS) (Davern et al., 2022) 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 (Blau et al., 1994). 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 (Bumpass et al., 2017), 1992–1994 (Bumpass and Sweet, 2018a), and 2001–2002 (Bumpass and Sweet, 2018b). 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 respon-

dent 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 (Social Research Center, 2020). 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 biannually 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) (Bureau of Labor Statistics, U.S. Department of Labor, 2014) 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) (Bureau of Labor Statistics, U.S. Department of Labor, 1992) 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) (Bureau of Labor Statistics, U.S. Department of Labor, 1999) 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 (Bureau of the Census, U.S. Department of Commerce, 2008). 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/`. These files are also accessible in STATA format from the NBER SIPP data archive page: `https://www.nber.org/research/data/survey-income-and-program-participation-sipp`.

Americans View Their Mental Health (Veroff et al., 2005) 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). The data are freely available to data users at the ICPSR member institutions: `https://www.icpsr.umich.edu/web/ICPSR/studies/3503` and `https://www.icpsr.umich.edu/web/ICPSR/studies/7948`

National Survey of Black Americans (Jackson and Neighbors, 2005) 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 codes. The data are publicly available and can be downloaded from: https://www.icpsr.umich.edu/web/NACDA/studies/6668.