Contents

1 Introduction **2**

2 Data **4**

3 Markov chains for measuring mobility **6**

4 Identification of two epochs in the late Qing **7**
4.1 Estimation of the steady state of epoch $\mathcal{E}_1$ . . . 11
4.2 Steady state of epoch $\mathcal{E}_2$ . . . 11
4.3 Evidence of Nonhomogeneity within Epochs . . . 11

5 Mobility measures **13**
5.1 Mover-stayer model . . . 14
5.2 Memory approach . . . 14
5.2.1 Dynamics of societal memory in the Late Qing . . . 16
5.2.2 Class heterogeneity . . . 17
5.3 Movement approach . . . 20
5.3.1 Dynamics of exchange and structural mobility in the Late Qing . . . 21
5.4 Average expected outcomes . . . 23

6 Back to history **24**

7 Conclusion **25**

A CMGPD-LN: detailed **29**

B Shannon entropy: all ages **29**

C Aggregation of transition matrices **30**

D Memory measures in the Mover-Stayer model **30**

E Additional memory measure **31**

F Additional results on memory measures **32**

Intergenerational Mobility in Late Qing Dynasty China

Kristina Butaeva[1], Steven N. Durlauf[1], and Alexander Shapoval[2]

[1] *University of Chicago*
[2] *University of Lodz*

January 10, 2025

Preliminary Draft

Abstract

The paper examines intergenerational mobility during the last 130 years of the Qing Dynasty. We employ Markov chain methods to measure two types of mobility: mobility as movement away from parental status and mobility as the absence of memory in the distribution of status among descendants. We identify two epochs within the 1783–1909 period of the Qing Dynasty with different mobility dynamics. From 1783-1850, a stable mobility process was leading society to a steady state class distribution. The second epoch, 1850-1909, whose beginning may be associated with the Opium Wars and the Taiping Rebellion, exhibits unstable class dynamics which continue to the end of the Qing dynasty. In particular, decades preceding the end of the Qing dynasty exhibit a deterioration of upward mobility, while downward mobility remains relatively stable. These results demonstrate how the footprints of major crises in the late Qing era may be found in mobility dynamics and make clear what type of inequality increases are associated with the dynasty's end.

1 Introduction

This paper studies long run intergenerational mobility in the last 130 years of the Qing dynasty. Using a remarkable population registry data on the Liaoning province developed by Lee and Campbell (2014), we are able to trace out the evolution of social status for family dynasties observed across six distinct generations. Given the importance of social mobility as a dimension of inequality, we characterize significant contours of the evolution of social conditions of the late Imperial era.

This research made use of the CMGPD LN dataset. Preparation of the CMGPD-LN and documentation for public release via ICPSR DSDR was supported by United States Department of Health and Human Services National Institutes of Health Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) R01 HD057175-01A1 "Multi-Generational Family and Life History Panel Dataset" with funds from the American Recovery and Reinvestment Act.

Butaeva and Durlauf thank the James M. and Catherine D. Stone Foundation for financial support. Kristina Butaeva: kbutaeva@uchicago.edu. Steven Durlauf: durlauf@gmail.com. Alexander Shapoval: abshapoval@gmail.com.

In order to richly characterize the mobility dynamics in these data, we use Markov chain methods to model the dependence of offspring social status on parental social status. Markov chain mobility models have a long history in sociology, see Song (2021). We build upon ideas taken from sociology to decompose the observed mobility in this period between structural and exchange components. Structural mobility refers to mobility that is induced by nonstationarity in aspects of a given environment. For our context, there are two possible sources of structural mobility: evolution of the cross-section distribution of social status across cohorts and evolution of the Markov transition matrix that links parents and children probabilistically. Our unique data set allows for an exploration of nonstationarities in mobility that is not possible when one studies a single cohort so that only a single Markov chain transition matrix is estimated.

Our Markov chain approach also allows us to distinguish between two different mobility concepts. First, we measure the strength of the effects of parental status on the conditional probabilities of offspring outcome, which we think of mobility in terms of the memory of parental status in the probability distribution of outcomes for descendants. This implies that mobility is understood as the fading of the memory of parental status in future outcomes. Second, we measure the persistence of the specific status of parents across generations, which leads to the concept of mobility as movement. These capture the primary ways in which mobility and immobility are conceptualized in existing empirical work. One technical contribution of our paper is the demonstration of the mathematical relationship between these distinct mobility concepts. A second contribution is the analysis of nonhomogeneous Markov chains, which means we allow the Markov transition structure to change over time. Nonhomogeneous Markov chains can take many forms. We allow for dependence of the Markov chain transition process on the cross-section density of social classes following an approach developed by (Beshers & Reiter, 1963; Conlisk, 1976; Matras, 1967) as well as for time variation in the transition matrix that occurs across epochs in the data.

Our work contributes to a small literature on intergenerational mobility in Qing dynasty China (S. Chen, 2017; Jiang & Kung, 2021; Keller & Shiue, 2023; Mare & Song, 2014). Our work most closely relates to Shiue (2016), who use different data, class definitions, and statistical tools. This paper employs genealogical records from the Tongcheng County of Anhui Province during the Ming and Qing dynasties from 1450 and 1850. We confirm some findings and add new ones that complement this earlier work.

Our analysis produces the following primary empirical findings. First, we find strong evidence that the mobility process of the late Qing qualitatively changed in the period of the Opium Wars and the Taiping Rebellion. One can interpret the 1783-1850 era as one epoch, one in which the mobility process appears to be converging, while the 1850-1909 period exhibits different mobility dynamics that revealed instability during the last years of the Qing dynasty.

Second, from the immobility as memory perspective, we find that societal memory decays (mobility increase) during the late Qing dynasty, due primarily to changes in the Markov transition matrix across the cohorts. This result complements Shiue (2016) which documents an increase in intergenerational mobility measured as the slope of the rank-rank regression of father's and son's social classes. Our typology of class structure provides the novel finding that the poorest class of the late-married non-salaried was the least mobile. It took, on average, from three to six-generations for the descendants of individuals from the poorest class to upgrade their social position. At the same time, it took no more than two generations to change social class starting from any other class.

Third, we find that exchange and structural mobility components evolve in different ways. We find low levels of structural mobility in the late Imperial era: the share of movements that changed the distribution of dynasties over classes was around 11% in the first epoch, then it

dropped in the 1870s to 2%, and caught up to the 13% in the 1900s. In contrast, our estimates of the exchange mobility allowing for evolving Markov transition matrices were around 40-45% before the 1850s and then fell to 35% up to the 1900s.
Fourth, upward mobility dominated downward mobility in the late Qing dynasty from 1783 to 1870, while the opposite pattern emerged afterward. Increasing upward mobility compared to downward mobility before 1850 was also found by Shiue (2016). Here, we provide a new finding that average expected outcomes grew up to 1879 and then declined through the 1900s. This is important in understanding the socioeconomic decay of the Qing regime.
The paper is organized as follows. Section 2 describes the unique CMGPD-LN dataset we use in the current study and characterizes our key sample construction procedures. Section 3 introduces existing approaches for studying intergenerational mobility employing Markov chains in the social science literature. Section 4 describes the observed nonstationarity in the Markov process which presumes the identification of two distinct epochs in the late Qing. Section 5 describes the Markov chain measures employed in the current study and presents the results. Section 6 relates the estimated dynamics of intergenerational mobility to the major historical events in late Qing dynasty China. Section 6 concludes.

2 Data

In this study we employ the data from Lee and Campbell (2014) China Multi-Generational Panel Dataset, Liaoning (CMGPD-LN). The CMGPD-LN data contains information of 266,091 individuals in 50,000 residential households between 1749 and 1909. These people belonged to 29 administrative populations, comprising some 700 villages distributed in North, Central, South-Central, and South Liaodong, the central corridor of contemporary Liaoning Province. Figure A1 in Appendix A shows the map of Liaoning communities covered by the dataset. The population records in the CMGPD-LN are transcribed from the Eight Banner population registers preserved in the Liaoning Provincial Archives. Taking into account the limitation posed by the availability of complete job category information for individuals, we have the capability to trace the socio-economic status information for a maximum of 6 generations in our own research. The vast majority (99.9%) of the observations in the data are descendants of Han-Chinese immigrants who moved to the Liaodong area from Shanxi, Hebei, and Shandong provinces in the late seventeenth and early eighteenth centuries (Ding, Guo, Lee, & Campbell, 2004). The Qing government organized these immigrants under the Eight Banner system and the Shengjing Imperial Household Agency, turning them into hereditary laborers of the Imperial Household. As bannermen, they could obtain a wide variety of state employment as artisans, as soldiers, and as civil and military officials. Many of these positions brought with them substantial salaries and a variety of other powers and privileges, including differential access to a banner social security system that protected orphans and widows on one hand, and the aged, the retired and the unemployed on the other hand.

Definition of social status

Based on the existing variables in CMGPD-LN we are able to define five social classes used in our intergenerational mobility analysis. Three out of six categories have higher status and cover those who hold salaried positions: *Artisans*, *Soldiers*, and *Officials*. These people had paid jobs and received remuneration either in the form of salaries or stipends. Salaries at that time might come in the form of money or grain. Since the population in the CMGPD-LN data is mainly rural, holding a salaried position not only indicates the status attainment of the individual under

observation but also means additional income and privileges for the household. Only 2.2% of the individuals aged between 18 and 60 in the data had a salaried position. The remaining 97.8% of individuals who do not hold a position, we call *Non-salaried*. They don't have salaried positions but might receive income from sources other than state salary. To provide a finer delineation within the *Non-salaried* category, we extracted the age at which individuals in this group entered into their first marriage. We established a threshold of 20 years old (B. Chen, Campbell, & Dong, 2018; S. Chen, Campbell, & Lee, 2014) to distinguish between early (*Early-married non-salaried*) and not early marriages (*Late-married non-salaried*). According to the existing literature, only those from rich families at this time could afford to marry early. Therefore we assume that our *Early-married non-salaried* category has a higher status among *Non-salaried*. To set up the hierarchical order of statuses among *Salaried* positions we will use average annual income estimates. The lowest-status category, therefore, is *Artisans*. Their average income is estimated from the data at around 6 taels of silver per year. This class accounts for 0.4% of individuals in the dataset. *Soldiers* are classified as a higher position, because they had an annual income of 24 taels of silver, which is four times higher than that of an artisan. *Soldiers* accounts for 1.41% of individuals in the data. Finally, *Officials* are categorized as the highest class. They have an average income of around 25 taels of silver per year and account for around 0.39% of the individuals in the data.

Sample construction

Our baseline sample has been constructed in the following steps. First, we choose the most recent, 1909, year as a baseline year. Second, we define our most recent children's cohort as individuals of age 30 to 50 (born between 1859 and 1879) from the baseline data year. Third, following the existing literature (e.g., Bartholomew, 1967), to mitigate the impact of changes in social structure in Markov chain, we subset only older sons among all existing siblings. Forth, searching across all available years of the data, we link the selected individuals to their fathers, grandfathers, and all available subsequent ancestors. To be consistent with the idea of not allowing our Markov chain to expand, among all ancestors in case if we found siblings, we also keep only the older sibling. With this approach, we managed to link 6 subsequent generations in our final sample. The sample size in each generation is 2,279 observations.

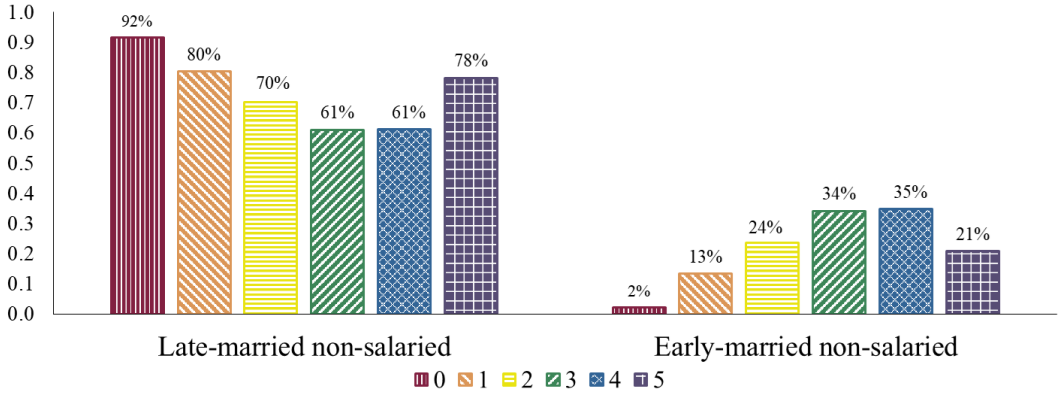
Table 1: Birth years of selected cohorts:

Generations	t	Mean birth year	Most frequent birth year
n	0	1719	1706∼1726
...	1	1750	1735∼1759
...	2	1779	1770∼1787
Granfather	3	1809	1796∼1823
Father	4	1838	1830∼1851
Son	5	1869	1859 ∼1875

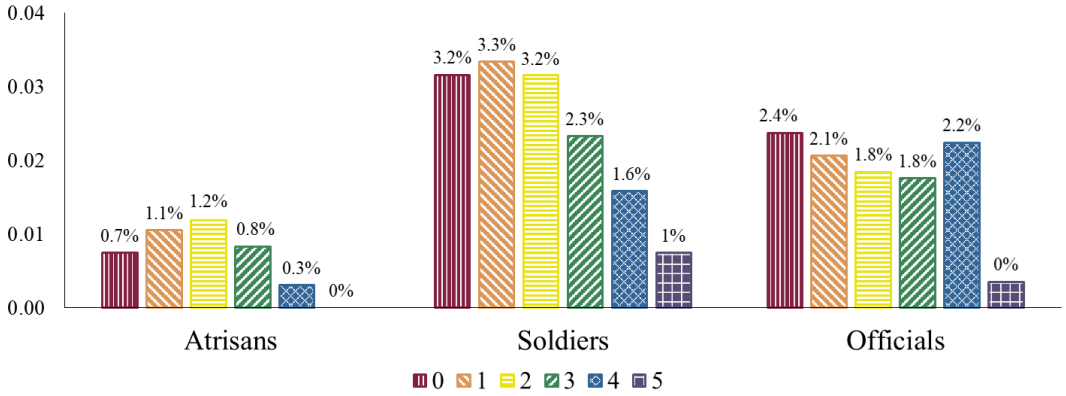
Table 1 shows mean birth year and the bulk of the distribution of of birth years among selected cohorts presented across six-generations in our data. Figure 1 trace changes in distribution of social classes across studied cohorts.

Figure 1: Distribution of social status among generations $t = 0, ..., 5$.

(a) Non-salaried.



(b) Salaried.



Notes: These figures illustrate the distribution of social classes in every generation of the six-generations panel dataset separately for (a) non-salaried and (b) salaried.

3 Markov chains for measuring mobility

The discrete nature of our social class categories means that Markov chain transition matrices are the natural language for measuring intergenerational mobility. When one studies long horizons, it is natural to question whether this transition matrix is nonstationary, see (Duncan, 1966a, 1966b; Pullum, 1975) for discussion. We will address stationarity below.

Formally, for n social classes in a society, we study the matrix with typical element P_{ij} which defines the probability that the process under study will move to the state j in the next step, given it is currently in the state i. For a **homogeneous (stationary) Markov chain (HMC)**, the transition matrix P is independent of time t. Then, the transition from the initial state of a parent to the state of a descendant for any subsequent generation is determined:

$$\mu^t = \mu^{t-1} \cdot P = \mu^0 \cdot P^t. \tag{1}$$

where P is the $n \times n$ transition matrix consists of transition probabilities P_{ij}, μ^0 is the initial distribution of social classes, μ^t is the distribution of classes among every subsequent generation t. If the transition matrix is primitive, i.e. one can eventually reach any state from any other stare,

then the distribution μ^t converges to a unique μ^∞ as $t \to \infty$. The distribution μ^∞ is called the invariant or steady state distribution. The transition matrix P^t converges to P^∞, the so-called limiting matrix P^∞. This limiting matrix consists of identical rows equal to the stationary distribution μ^∞, implying that all descendants of parents from initial classes formulated in μ^0 will face identical probabilities of transiting to any given class. The process of moving toward such an origin-independent distribution of transiting probabilities is commonly referred to as "mixing" in the probability literature.

For a **nonhomogeneous (non-stationary) Markov chain (NHMC)**, the transition probabilities p_{ij} vary over time t. We allow two forms of nonhomogeneities in the process. First, we follow Conlisk (1976), in studying Markov transition matrices such that P depends on the population distribution of individuals across classes μ, $P = P(\mu)$, but does not depend on time explicitly. This setting presumes that the existing shares of parents in each social class affect opportunities (lotteries) for a given descendant to change his class. We validate this assumption with the data in the next section and find it to be plausible. Second, we allow this interactive Markov chain structure to change across epochs, by which we mean we allow for low-frequency changes. We find that the intergenerational mobility process for the last 130 years of the Qing was governed by two distinct Markov processes for the two epochs 1783-1850 and 1850-1909. Hence, we measure distinct $P = P(\mu)$ functions within each epoch. Within an epoch, the transition matrix for the first generation of children can be derived from the transition matrix observed in the data and the distribution of classes among parents if both generations belong to the same epoch: $P(\mu^1) = P\left(\mu^0 P(\mu^0)\right)$. In general, the transition from a class of a parent to a class of a descendant in the NHMC, in parallel to (1), taking the form:

$$\mu^t = \mu^{t-1} \cdot P(\mu^{t-1}) = \mu^0 \cdot \prod_{k=1}^{t} P(\mu^{k-1}). \tag{2}$$

4 Identification of two epochs in the late Qing

In this section, we employ the full CMGPD-LN data set that link the distribution of classes of fathers $\mu_F(y)$ to the distribution of classes of their sons $\mu_S(y)$:

$$\mu_S(y) = \mu_F(y) \cdot P(\mu_F(y)), \tag{3}$$

where y is every year of available cross-section data. We also consider subsequent generations of fathers and sons from our six-generation panel dataset. In the six-generation panel, we call $\mu^t = \mu_F^t$ distribution of fathers, and $\mu^{t+1} = \mu_S^t$ distribution of sons for every generation $t = 0, \ldots, 5$[1]. The pairs of fathers and sons, (μ_F^t, μ_S^t) are associated with specific years of cross-section data y to compare the results obtained from the six-generation panel and the cross-section data. The notation $\mu_F^t(y)$, $\mu_S^t(y)$ formalizes this association. We match t to y using the mean birth year of sons from μ_S^t, $t = 1, \ldots, 5$ (Table 1), with the mean birth year of sons at every cross-section year of the data $\mu_S(y)$. The resulting match is presented in Table 2.

In general, multiplication by the matrix $P(\mu)$ in (3) maps the Euclidean space of n-dimensional vectors into itself. The identification of the exact dependence of P on μ is impeded by the lack of available data. To simplify the problem, we assume that the transition matrix depends on the distribution of classes through a one-dimensional function $E(\cdot)$: $P(\mu) = P(E(\mu))$. Following, Brumelle and Gerchak (1982), we employ the Shannon entropy as the function, so that

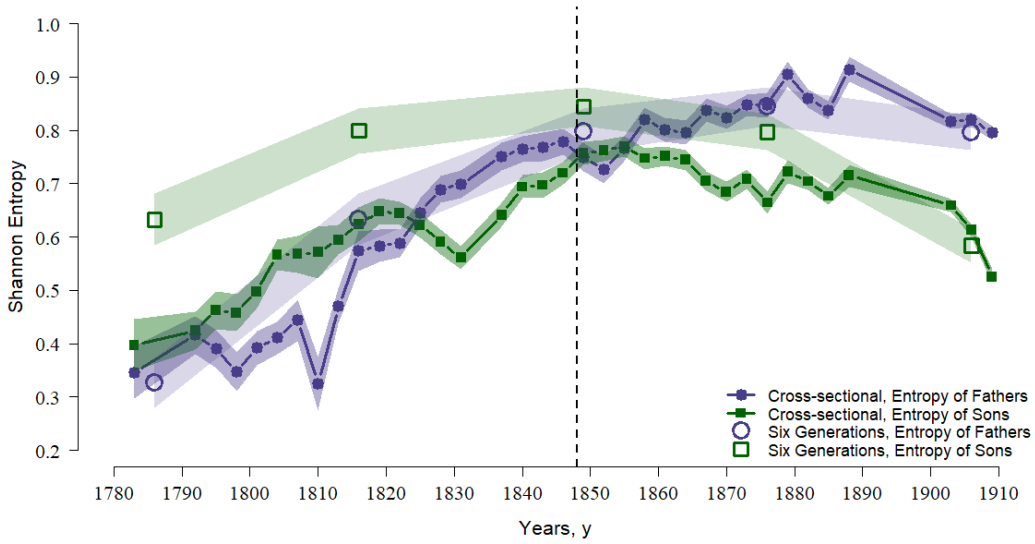
[1] Generations $t = 1, ..., 5$ can be referred to as both fathers and sons, therefore $\mu_F^t = \mu_S^{t-1}$ for $t = 0, \ldots, 5$.

Table 2: Matching of generations t to the cross-section data years y.

Generation t	Mean birth year from μ_S^t	Mean birth year from $\mu_S(y)$	cross-section data year y
1	1750	1749	1786
2	1779	1778	1816
3	1809	1810	1849
4	1838	1837	1876
5	1869	1868	1906

$E(\mu) = -\sum_{i=1}^{n} \mu_i \ln(\mu_i)$. Shannon entropy measures the degree of inequality between the size of the social classes. Maximal cross-section inequality is attained when individuals belong to a single class, which corresponds to a minimal value of the entropy index (0). Inequality completely disappears if classes hold equal shares, which conforms to entropy attaining its maximal value of $n \ln n$. Following Theil (1972), Brumelle and Gerchak (1982) used entropy as a proxy for the amount of uncertainty held by a social distribution to construct a mobility measure. In contrast, we use entropy to identify the nonhomogeneity of the Markov chain within two epochs in the late Qing dynasty China and to describe their steady states.

Figure 2: Dynamics of the Shannon entropy inequality measure



Notes: This figure plots the dynamics of the entropy of fathers $E(\mu_F)$ and the entropy of sons $E(\mu_S)$ separately for the cross-section and six-generations data. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

Figure 2 illustrates the dynamics of the estimated Shannon entropy among fathers $E(\mu_F)$ (green color) and sons $E(\mu_S)$ (purple color). For each cross-section year y from 1783 to 1909, we plot the Shannon entropy of fathers $E(\mu_F(y))$ in filled circles, and the entropy of sons $E(\mu_S(y))$ in filled squares. Open circles on Figure 2 depict Shannon entropy of fathers from the six-generations panel data $E(\mu_F^t(y))$, and open squares show entropy among their sons $E(\mu_S^t(y))$. The results from the six-generation panel suggest that the Shannon entropy of fathers increased from around

0.3 in 1783s to 0.8 in 1880s and then had been slightly decreasing until 1909[1]. The entropy of sons attained the maximum earlier, in the middle of the 19th century, and then also decreased. While the entropy of both fathers and sons grew, the entropy of sons was higher than the entropy of fathers before the 1950s, and then the gap between them shrank, while the process evolved to a steady state. We claim that as soon as the entropy of sons exhibited a steady decrease, the Markov process switched to a different regime. Consider, for example, the values of $E(\mu_S)$ within the range 0.65–0.7 in the 1880s. If the mechanism remained the same, the entropy of their sons would be defined as follows. One has to find these values 0.65–0.7 among the entropy of fathers from the 1830s and the matching values of the entropy among their sons. The latter values are larger, belonging to the range 0.75–0.8. However, nothing similar to this increase of $E(\mu_F)$ from 0.65–0.7 to $E(\mu_S)$ from 0.75–0.8 was observed after 1850. On the contrary, the values 0.65–0.7 of the entropy were not observed in $E(\mu_F)$ after 1860. Moreover, the entropy decreased from fathers to sons after 1880. This observed change in the mechanism underlying the dynamics of the mobility suggests the presence of two epochs $\mathcal{E}_1$ and $\mathcal{E}_2$ corresponding to the intervals before and after this change at 1850.

We provide additional evidence on the existence of two epochs with our cross-section dataset. The dynamics of the entropy of fathers (filled purple circles in Figure 2) fully agree with the pattern discussed above (and shown in the light purple). The steady decrease of the entropy of sons after 1850 (filled green squares) is also clearly seen on the graph. These observations support the existence of two epochs. We note a sudden drop in the entropy of sons occurred in the 1810s but believe this is due to sampling variation associated with our use of a truncated cross-section due to age restrictions. When we consider the full dataset without age restrictions, the drop disappears (Figure B1 in Appendix B). The entropy of the steady state distribution of fathers and sons coincide, $E(\mu_F) = E(\mu_S)$ in the 1850s (in the six-generation dataset) and in 1815-1855 (in the cross-section dataset). Thus, based on the change in the entropy patterns after 1850 discussed above and the proximity of the entropy of fathers to the entropy of sons around 1850s, we conclude that the transition between the epochs occurred around 1850 and that the steady state of the first epoch is characterized by the value of the entropy located near $E(\mu_F(y)) = E(\mu_S(y)) \approx 0.7$ if estimated with the cross-section dataset.

We further validate the existence of the two epochs in the late Qing Dynasty with a statistic denoted d', known as the Dobrushin contraction coefficient (Dobrushin, 1956) in the the Markov chain literature:

$$d'(P,t) = \max_{i,j} \|e^i \cdot P^t - e^j \cdot P^t\| = \max_{i,j} \frac{1}{2} \sum_{k=1}^{n} |P^t_{ik} - P^t_{jk}|, \tag{4}$$

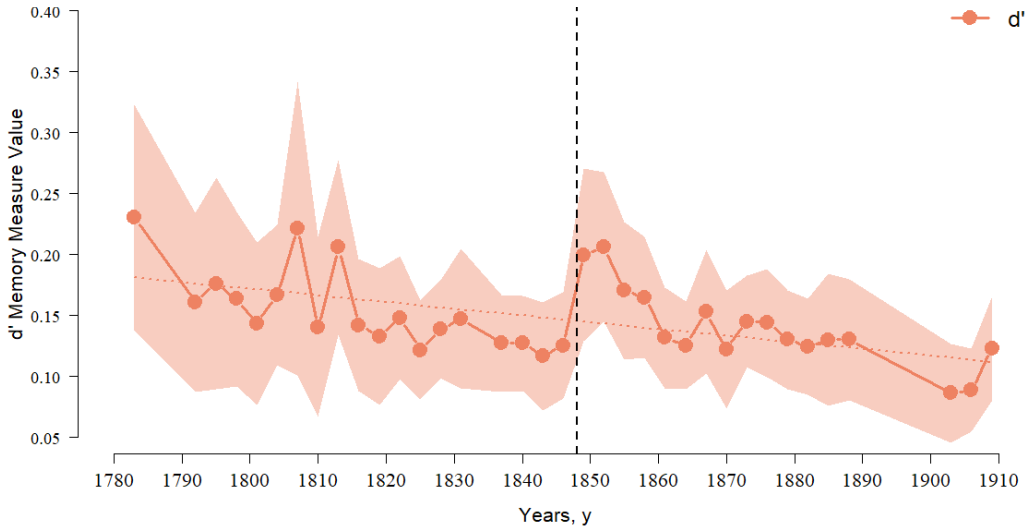
where e^i is an indicator vector ($e^i_j = 1$ if $j = i$ and $e^i_j = 0$ otherwise) and P^t_{ik} denotes the i^{th} row of the matrix P^t[2]. This measure asses the maximum distance between the distribution of chances to reach any social class for descendants of every pair of initial classes of fathers. The distribution of chances to reach any social class is defined at every row of the transition matrix P and will be referred to as "lottery" here and after. In other words, the Dobrushin coefficient d' measures the contraction rate of matrix P^t by giving the upper bound of the distance between its' rows at every transition step t. At the steady state, the limiting transition matrix P^∞ consists of similar rows (lotteries) μ^∞. Therefore, at the steady state $d' = 0$. Applying the d' measure for the cross-section dataset, we focus only on the first step toward the steady state from the distribution of fathers

[1] It should be noted that for all the estimated indices here and below we consider point estimates from the six-generations dataset as our baseline estimates.

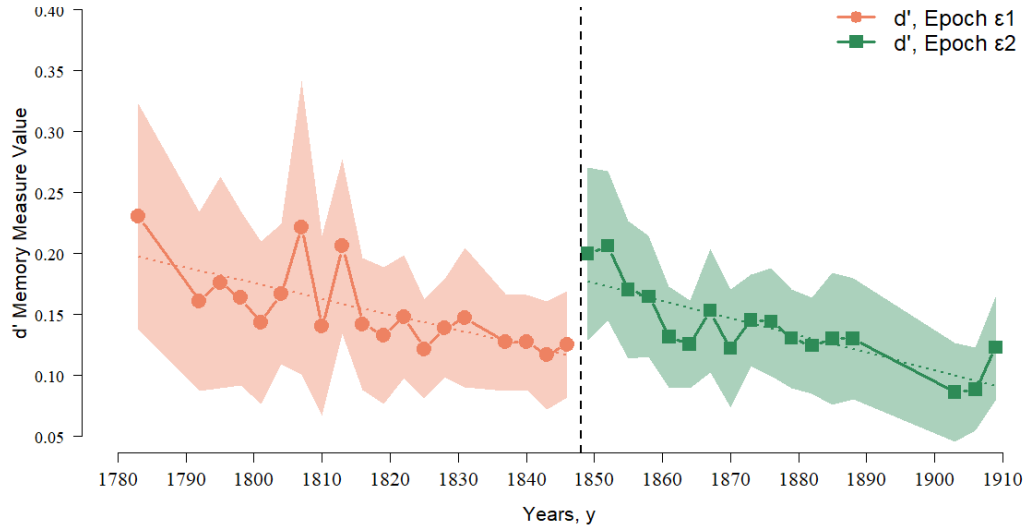
[2] This equation, and all the mobility measures introduced later are formulated for the HMC case. To switch to the NHMC, P^t should be replaced with the $\prod_{k=1}^{t} P(\mu^{k-1})$.

Figure 3: Dynamics of d'

(a) Single Epoch.



(b) Two Epochs.



Notes: This figure plots the evolution of $d'(P(y))$ memory measure from eq. (5) assuming a single versus two epochs. Dashed lines represents the OLS fit. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

$\mu_F(y)$ observable at the year y to the distribution of sons $\mu_S(y)$ observable at the same year. Therefore, the notation $P = P(\mu(y))$ is used, which can be reduced to $P(y)$. Thus, we rewrite (4):

$$d'(P(y)) = \max_{i,j} \|e^i \cdot P(y) - e^j \cdot P(y)\| = \max_{i,j} \frac{1}{2} \sum_{k=1}^{n} |P_{ik}(y) - P_{jk}(y)|, \tag{5}$$

Figure 3 exhibits the dynamics of $d'(P(y))$ for every cross-section year y. We can clearly see that the regime changed around the 1850s. To prove this visual observation, we run OLS for a simple linear regression model relating $d'(P(y))$ to years y. The OLS fit is represented by the dashed line

in each sub-figure. Under the assumption of a single epoch (Figure 3a), the estimated $R^2 = 0.32$. The introduction of two epochs with the threshold at 1949 (Figure 3b) improves the model fit for both epoch $\mathcal{E}_1$ ($R^2 = 0.49$) and epoch $\mathcal{E}_2$ ($R^2 = 0.67$). We assign the year 1949 to the threshold between epochs in all graphs, meaning the transition between them occurred in a neighborhood of this year.

4.1 Estimation of the steady state of epoch $\mathcal{E}_1$

We estimate the steady state $\mu^{\infty\varepsilon_1}$ of the first epoch again employing both cross-section and panel six-generations data. With the cross-section data, we propose to measure $\mu^{\infty\varepsilon_1}$ processing observable pairs $(\mu_F(y), \mu_S(y))$ such that the pairs of the corresponding entropy $(E(\mu_F(y)), E(\mu_S(y)))$ are located in a neighborhood of the point $(0.7, 0.7)$. First, we find a (single) matrix T that transforms all chosen distribution vectors of fathers $\mu_F(y)$ as close as possible to the distribution of their children $\mu_F(y) \cdot P(y)$. This procedure of "aggregation" of the selected transition matrices $P(y)$ is described in detail in the Appendix C. With our choice of a neighborhood, we deal with the vectors corresponding to years from 1840 to 1852. Second, by iterating the obtained transition matrix T, we estimate its steady state and consider the result as an estimate of the steady state $\mu^{\infty\varepsilon_1}$ of the epoch $\mathcal{E}_1$.
With the six-generations data, a transition between the epochs is associated with the interval between the transition matrices three and four but closer to the third one. From Figure 2, it follows that point estimates of entropy are larger (especially for sons) in the six-generation dataset compared to the cross-section data, but they exhibit similar dynamics. Therefore, we assume the intersection of $E(\mu_F^t)$ and $E(\mu_S^t)$ is likely to happen after the transition described by the third matrix but before the fourth matrix. Consequently, we estimate the steady state $\mu^{\infty\varepsilon_1}$ by the "aggregation" (see above) of the transition matrices three and four and iterating the corresponding aggregated transition matrix T up to the steady state. The resulting estimated first-epoch steady states $\mu^{\infty\varepsilon_1}$ from both cross-section and six-generation data will be applied to the analysis of the exchange mobility in NHMC, which we discuss in more detail in Section 5.3.

4.2 Steady state of epoch $\mathcal{E}_2$

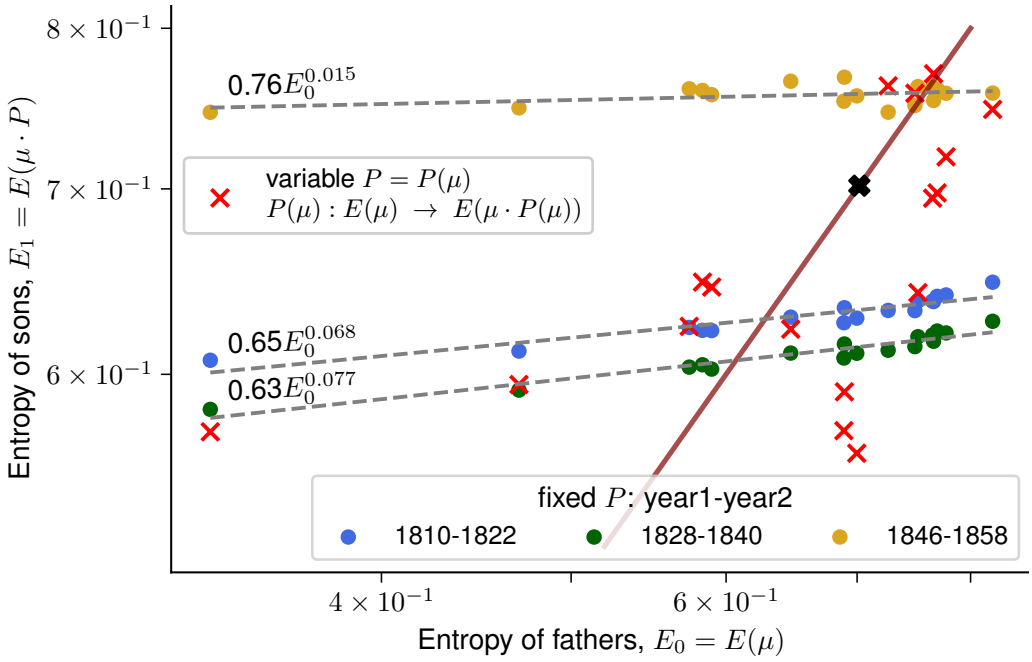
The different method is required to assess the steady state $\mu^{\infty\varepsilon_2}$ in the second epoch $\mathcal{E}_2$. To recall, the dynamics of the entropy of fathers and sons after the 1850s became diverging: the entropy of fathers stabilizes at around 0.8, and the entropy of sons dropped from 0.8 in the cross-section data and from 0.7 in the six-generations panel data in 1850 to around 0.6 in 1909 in both datasets. The drop in inequality among sons was mainly driven by the increase of the already prevailing class of non-salaried from around 94 to 97%. If such a tendency continued, the salaried class would totally disappear. However, the collapse of the Qing Dynasty terminated this trend. Therefore, we are not able to observe the converging pattern in reaching the second steady state, and we are using the latest available transition matrix in both cross-section and six-generations datasets to define $\mu^{\infty\varepsilon_2}$ as a counterfactual steady state.

4.3 Evidence of Nonhomogeneity within Epochs

The nonhomogeneity of a Markov chain can be revealed by considering the iterations $\mu^t \cdot P(\mu^t)$ as a function of μ^t, for some vectors μ^0, $t = 0, 1, \ldots$. The consideration requires the usage of the dependence $\mu \cdot P(\mu)$ on distribution vectors μ. With the Shannon entropy, we simplify the consideration reducing the dimensionality from five (the number of classes in μ) to one when exploring the implicit dependence $E\big(\mu(y) \cdot P(\mu(y))\big)$ on $E(\mu(y))$ as y varies, where $\mu(y) = \mu_F(y)$

are the classes of fathers in the cross-section data. This graph, $E(\mu(y) \cdot P(\mu(y)))$ vs. $E(\mu(y))$ is compared with graphs $E(\mu(y) \cdot P)$ vs. $E(\mu(y))$, where the matrix P is constant. Due to the nature of our data, this comparison will be limited to the cross-section data from the epoch $\mathcal{E}_1$.

Figure 4: Shannon entropy of fathers and sons under fixed and changing P.



Notes: This figure plots the entropy of sons $E_1 = E(\mu_S(y))$ vs. the entropy of fathers $E_0 = E(\mu_F(y))$ with (i) the actual transition matrix $P(\mu(y))$ observed together with the distribution of fathers $\mu(y)$ and (ii) with three fixed transition matrices that "average" the matrices assigned to the years described in the legend. Yellow line is a 45° line where $E_1 = E_0$. Dashed lines represent the OLS fit to the dependencies obtained with constant transition matrices.

Red crosses on Figure 4 display the pairs $\big(E(\mu(y) \cdot P(\mu(y))), E(\mu(y))\big)$ with years y of the cross-section data varying from 1783 to 1858. We compare these values with three proxies of the HMC dynamics related to observations. Namely, we chose three 12-year intervals (avoiding the time interval at the beginning, where the data was incomplete). For each interval, the procedure of "aggregation", mentioned above and explained in detail in the Appendix C, is applied for four transition matrices prescribed to this interval. The resulting matrix P is used to display the dependence $E_1 = E(\mu(y) \cdot P)$ on $E_0 = E(\mu(y)$, where $\mu(y) = \mu_F(y)$ are all observable distributions of fathers in 1783–1858. Each of three dependencies E_1 on E_0 (in orange, blue, and green circles) obtained with the constant transition matrix is accurately approximated by a power-law function shown in dashed lines (they are OLS for the $(\log(E_0), \log(E_1))$-pairs). The proximity to the power-law is probably not important itself. We rely our arguments on the observation that the points $\big(E(\mu(y) \cdot P(\mu(y))), E(\mu(y))\big)$ obtained with the *recorded* matrices $P(\mu(y))$ exhibit a principally different pattern. The absence of the stable pattern found with a constant transition matrix signals that the observed transition matrices P depend on the current distribution μ_F of fathers' classes, $P = P(\mu_F)$. We also derive from Figure 4 that the dependence of the transition matrix on the distribution implies the amplification of fathers' inequality with respect to the case of the fixed transition matrix. Namely, under small inequality among fathers, the inequality among sons is smaller than in the case of the fixed transition matrix. On the contrary, large inequality among fathers indicates relatively large inequality among sons, whereas

the fixed transition matrix stronger reduces the differences between classes.
Figure 4 illustrates the idea of the choice of matrices "aggregation" of which leads us to the estimate of the steady state in the first epoch. Fixing a neighborhood of the point $(0.7, 0.7)$ we define the observed pairs of fathers and their sons, $(\mu_F(y), \mu_S(y))$ that are presumably close to the steady state. Admitting potential errors in the records of these pairs, we numerically determine a matrix T sufficiently accurately mapping each of the selected distribution of fathers to the distribution of their sons through our procedure called "aggregation". We verify that the rows of T^2 are rather similar and yet after the first iteration our vectors are (very) close to the (estimate of the) steady state. The entropy computed at the steady state is marked by the black cross in Figure 4. Interestingly, this black cross is closer to the point $(0.7, 0.7)$ than the points used to determine the matrix T.

5 Mobility measures

In this section, we switch to intergenerational mobility measurement, beginning with a brief introduction of existing approaches. Markov chain measures of intergenerational mobility could be categorized into two major classes, which correspond to *descriptive* and *normative* concepts of mobility.[1] From the descriptive perspective, one could think of mobility as the number of movements or the extent to which classes are changed from one generation to another. Within this *"movement"* approach, mobility indices can be built upon the examination of the transition matrix or by measuring changes in the distribution of classes. To illustrate the transition matrices measures, let's consider three elementary matrices:

$$P_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad P_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad P_3 = \begin{pmatrix} p_1 & p_2 \\ p_1 & p_2 \end{pmatrix}. \tag{6}$$

Intuitively, movement measures, constructed from a transition matrix, should assign maximum value to the matrix P_1, and minimum value for the P_2. In the general case, movement indices should capture shifts of the probability mass in the transition matrix that push the mass away from the main diagonal.[2]
The normative concept of mobility is predicated on a different motivation, which is close to the notion of equality of opportunity. From this perspective, a perfectly mobile society would be the one in which the probability of entering a particular social class for a child is independent of the class of his parent (Prais, 1955). Clearly, among the examples in (6), matrix P_3 represents this case of perfect mobility. Similar rows of the transition matrix P_3 postulate that children have an identical distribution of probabilities $(p_1; p_2)$ to end up at each class irrespective of the initial class of their parent. Measures of mobility in this case typically exploit the property of the convergence of the transition matrix to the steady state, which we discussed in Section 3. To recall, in the stationary Markov process as $t \to \infty$, transition matrix P^t converges to the limiting transition matrix P^∞, which consists of similar rows equal to the steady state distribution μ^∞. The classic example of mobility measures within this framework captures the speed of the convergence of a

[1] See Boudon (1973), Brumelle and Gerchak (1982), Dardanoni (1993), and Van de Gaer, Schokkaert, and Martinez (2001) for a broader review of Markov chain mobility measures.

[2] The most simple index, which arguably goes back to Matras (1961), quantifies the share of mobile individuals (whose class is different from their parent's class) in the total population. An alternative approach might be to count the number of crossed class boundaries in the transition process, known as Bartholomew's mobility index (Bartholomew, 1967). Distributional changes in measuring mobility might be captured by the elementary $L1$ norm distance between parental and child's class distributions from the Fields and Ok (1996) original paper, or by some more advanced versions of distributional change indices (Cowell, 1980, 1985; Fields & Ok, 1999; Mitra & Ok, 1998).

transition matrix to the origin-independent distribution at the steady state (the speed of class "mixing").[1] We call the mobility measures within this perspective the *"memory"* approach[2] because all the indices basically quantify how fast the memory of class-origin disappears in society. In this paper, we utilize both *movement* and *memory* approaches, and we employ some existing and propose new measures of intergenerational unified by a single theoretical framework alike of a well-known **mover-stayer model** (Frydman, 1984; Goodman, 1961; Spilerman, 1972).

5.1 Mover-stayer model

Consider a simple model where each individual first chooses to automatically retain the class of his father with the probability $\mathfrak{p}$ and otherwise make a decision of the class choice (which includes his father's class) μ^{∞} from the limiting transition matrix P^{∞} with the probability $(1-\mathfrak{p})$. Then the conditional probability to move from a state i to a state j is

$$\mathcal{P}_{ij} = \begin{cases} \mathfrak{p} + (1-\mathfrak{p})\mu_j^{\infty}, & \text{if } i = j \\ (1-\mathfrak{p})\mu_j^{\infty}, & \text{otherwise} \end{cases} \tag{7}$$

given $\mathfrak{p} \in [0,1]$ and $\mu^{\infty} \in [0,1]^n$. Model (7) employs both memory and movement mobility approaches. The opportunities are equal for all classes at the steady state. They are represented by the lottery μ^{∞} available for each individual. Nevertheless, this lottery is taken with the probability of $1-\mathfrak{p}$ that exhibits openness to changes. On the contrary, the probability $\mathfrak{p}$ represents proneness to the father's social class and contributes to the probability of immobility, expressed as $\mathfrak{p} + (1-\mathfrak{p})\mu_i^{\infty}$ for each i.
In the following subsections, we introduce a set of intergenerational mobility measures within "memory" and "movement" theoretical mobility frameworks such that they coincide with $\mathfrak{p}$ in the case of model (7). These measures will be discussed in the framework of both HMC and NHMC assumptions.

5.2 Memory approach

Our first candidate for measuring the speed of vanishing of societal memory is the estimate of a coefficient $\mathfrak{p}$ from equation (7) per se. In (7), the coefficient $\mathfrak{p}$ reflects the willingness of the individual to maintain the same social class as his father. Therefore, intuitively, the greater is $\mathfrak{p}$, the larger is the social memory. To estimate $\mathfrak{p}$ we propose employing an auxiliary measure $\mathbf{\Delta M^t}$ which is related to movements of individuals:

$$\mathbf{\Delta M^t} = \left|(1 - \sum_i \mu^{\infty} P_{ii}) - (1 - \sum_i \mu_i^{\mathbf{t-1}} P_{ii})\right|, \quad t \geq 1. \tag{8}$$

Here, the first term $(1 - \sum_i \mu^{\infty} P_{ii})$ defines the probability of changing social class at the steady state under the same transition matrix P. The second term $(1 - \sum_i \mu_i^{\mathbf{t-1}} P_{ii})$ represents the

[1] The largest branch of such measures is based on the eigenvalues of the transition matrix (Conlisk & Sommers, 1979; Shorrocks, 1978; Sommers & Conlisk, 1979; Theil, 1972). The common indices include one minus modulus of the second largest eigenvalue (λ_2), one minus modulus of the geometric mean of eigenvalues of the transition matrix, etc. Another example of measuring the speed of the transition matrix convergence is the Dobrushin (d') contraction coefficient (Dobrushin, 1956), which we introduced in Section 4. A different approach in this literature considers the mean first passage time of a transition matrix (Kemeny, Snell, et al., 1969). Conlisk (1990) proposed to measure immobility as a number of generations that must pass before the descendants of one randomly chosen individual will achieve the class of another randomly chosen individual.

[2] We follow Blume, Cholli, Durlauf, and Lukina (2024) in this.

expected probability of changing the class for each son from $\mu^{\mathbf{t}}$ compared to his father from $\mu^{\mathbf{t-1}}$. Therefore, $\mathbf{\Delta M^t}$ measures the average discrepancy between the probability of changing the class under the matrix P for the tth generation of sons and for the descendants from the steady state. Obviously, as $t \to \infty$, $\mathbf{\Delta M^t} \to 0$. It's important to notice that the second term in (8) for $t = 1$ is equivalent to the historical mobility measure introduced in Section 5.3.

Calculations described in the Eq. (21) of Appendix D, justify that $\mathbf{\Delta M^t}/\mathbf{\Delta M^\infty} \sim \mathfrak{p}^t$ in the model (7). Therefore, $\log \mathbf{\Delta M}^t$ decays linearly in t with the rate $\log \mathfrak{p}$ under the transition matrix $\mathcal{P}$ from the model (7). For an *arbitrary* transition matrix P taken from the real data, we can still regress $\log \mathbf{\Delta M^t}$ on t and determine the regression coefficient β which gives us the estimate $\hat{\mathfrak{p}}$ of $\mathfrak{p}$. This estimate involves the counterfactual evolution toward the steady state governed by the matrix $\mathbf{\Delta M}$.

Our second introduced memory measure involves the notion of the directional derivative of the transition matrix P. We are making an important observation here: several well-known memory measures from the Markov chain literature, e.g., d' from equation (4) and the second largest eigenvalue λ_2 of the transition matrix P is based on the fact that certain characteristics of the transition matrix under the homogeneous Markov chain are multiplied by a constant less than one when the transition matrix is raised to the power of t. If the second largest eigenvalue is the measure of interest, then this constant is λ_2 itself, which characterizes the evolution of the transition matrix multiplied by the second largest eigenvector over t. Other measures, such as $\hat{\mathfrak{p}}$ defined above, exhibit multiplication by a constant as an asymptotic property when the Markov chain approaches its steady state. We emphasize here that all these *measures can be described by the ratio of the norm of the directional derivative to the norm of the vector that determines the direction*. This feature can be interpreted as the rate of convergence to the steady state.

Since the multiplication by a matrix P defines a linear mapping, its directional derivative in the direction of a vector x is simply the product $x \cdot P$, independent of the vector μ where the derivative is computed. Thus,

$$\frac{\partial(\mu \cdot P)}{\partial x}\Big|_{\mu} \stackrel{\text{def}}{=} \lim_{t \to 0} \frac{(\mu + tx) \cdot P - \mu \cdot P}{t} = x \cdot P \quad \text{and} \quad \texttt{measure}(P, t, x) = \frac{\|x \cdot P^t\|}{\|x\|}. \tag{9}$$

Since the multiplication is applied to distributions, vectors μ with the sum of the coordinate equaled to 1, it follows that the coordinates of x are summed to 0, $\sum_i x_i = 0$.

A perturbation of the current distribution μ^0 is given by the exchange between two coordinates: a coordinate i is increased by some number, whereas a coordinate j is decreased by the same number. This defines the direction $x = e^i - e^j$. Averaging (9) with this x over all pairs (i, j), produces our second proposed measure of societal memory

$$\bar{d}'(P, t) = \frac{2}{n(n-1)} \sum_{i>j} \frac{\|e^i \cdot P^t - e^j \cdot P^t\|_{TV}}{\|e^i - e^j\|_{TV}} = \frac{1}{n(n-1)} \sum_{i>j} \sum_{k=1}^{n} |P^t_{ik} - P^t_{jk}|, \tag{10}$$

where e^i is an indicator vector, and P^t_{ik} denotes the i^{th} row of matrix P^t. This measure captures the evolution of an average distance between the rows of the transition matrix P over t. As $t \to \infty$, the rows of the transition matrix P become similar to each other, which transforms P^t into the limiting transition matrix P^∞, and drives $\bar{d}' \to 0$. If we focus on the single direction among $e^i - e^j$, the largest directional derivative yields the measure d' formulated in (4). The approach based on the directional derivatives formulated in (10) conceptually unifies the measures d' and $\bar{d}'$. Both indices vanish over time t, but the dynamics of the average versus maximum row distance might not necessarily be similar. Leveraging the average impact of all matrix rows, the

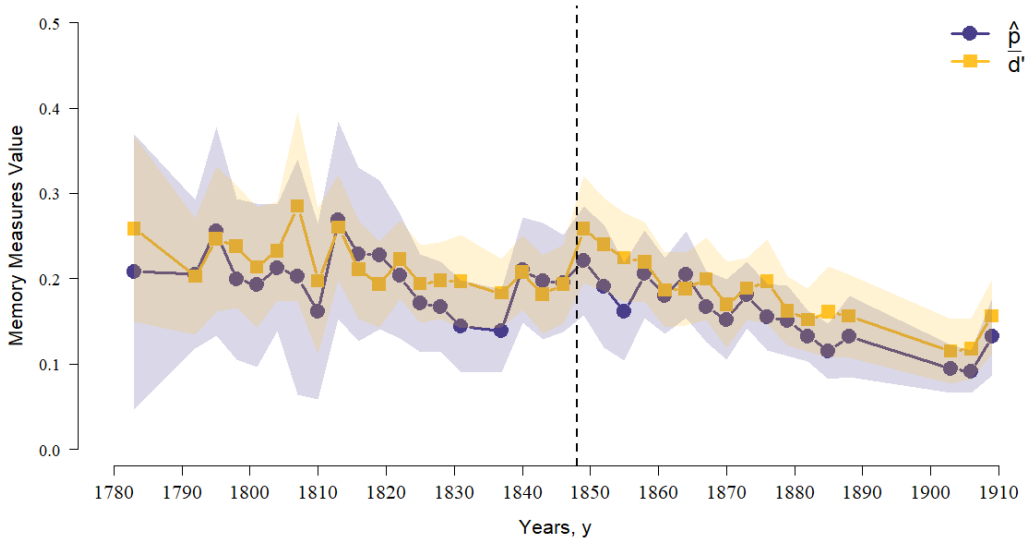
measure $\bar{d}'$ better corresponds to the measure $\hat{\mathfrak{p}}$, than the d' statistics which describes the "worst-case" distance between matrix rows.
Since $\bar{d}' = d' = \hat{\mathfrak{p}} = \mathfrak{p}$ in the model (7), the measures $\bar{d}'$ and d' may be interpreted as the probability of the decision to follow fathers' social class (which is different from the probability to stay at the same social class, see (7). However, the estimation of $\bar{d}'$, d', and $\hat{\mathfrak{p}}$ for an arbitrary transition matrix P from the real data can result in any level of discrepancy between those indices. Importantly, it can be noted that measures $\bar{d}'$ and d' characterize only the properties of the transition matrix P, whereas $\hat{\mathfrak{p}}$ also takes into account the initial distribution of father statuses μ^0. The similarity/dissimilarity of the dynamics of $\bar{d}'$ and d' versus $\hat{\mathfrak{p}}$ might provide suggestive evidence of the relative importance of the transition matrix versus the initial condition in the Markov chain converging process.
Formally, the mobility measure λ_2, the second largest eigenvalue of the transition matrix, may also be obtained by using the directional derivative approach that relies on equation (9) if x is the corresponding eigenvector (Halmos, 1947). In the case of the model (7), $\lambda_2 = \mathfrak{p}$, and the coordinates of the second largest eigenvector sum up to 0.[1] The direction $x = \mu - \nu$ between arbitrary lotteries μ and ν is characterized by the property $x_1 + \ldots + x_n = 0$ that holds for the eigenvectors corresponding to λ_2 of the matrix defined by (7). However this property is not necessarily to be held for a general case of any transition matrix estimated from the data. As a result, we prefer $\bar{d}'$ and d' over λ_2.
We note that equation (9) defines a class of indices within a broader set introduced in (Brumelle & Gerchak, 1982). Measures $\bar{d}'$, d', and λ_2 from this class are united by the fact that they are equal to $\mathfrak{p}$ from the model (7). In Appendix E we recall additional memory measures defined by (9) that employ the directional derivative of the transition matrix toward the steady state.

5.2.1 Dynamics of societal memory in the Late Qing

Figure 5 shows trends in the societal memory measures $\bar{d}'$ and $\hat{\mathfrak{p}}$ estimated from the cross-section CMGPD-LN data. The observed dynamics indicate the societal memory decay in the late Qing Dynasty China. The heterogeneity of lotteries for sons from different initial social classes, measured by $\bar{d}'$, gradually shrinks. Eventually, the understanding of $\bar{d}'$ as a derivative implies that individuals' social choices became less susceptible to external perturbations. The similarity in trends between $\bar{d}'$ and $\hat{\mathfrak{p}}$ is suggestive of the dominant role of the transition matrix convergence in the memory decay process, as opposed to the role of the cross-section distributions. The results from the panel six-generations data exhibit similar patterns (Figure F1 in Appendix F). Figure F2 in Appendix F shows the comparison of our memory measures in the cross-section data to the well-known d' and λ_2 from the Markov chain literature. It can be observed that in the CMGPD-LN dataset, all four memory measures reveal similar dynamics of the gradual societal memory decay. We emphasize that we exploit all known measures, d', λ_2, $\bar{d}'$, that are defined by (9) and coincide with $\mathfrak{p}$ in the case of (7). These measures are *not* expected to exhibit similar dynamics in the general case. Since they do so for the CMGPD-LN dataset, we associate them with $\mathfrak{p}$, interpreted as the stickiness to the father's class. Thus, the consideration of all these measures has enabled us to extract and trace the dynamics of the stickiness to the father's social class during the final period of the Qing dynasty.

[1] Due to its special symmetric structure, the matrix in (7) has the eigenvalue of 1 with multiplicity 1 and corresponding eigenvector $(1, \ldots, 1)$, and the eigenvalue p with multiplicity $n-1$ and eigenvectors orthogonal to $(1, \ldots, 1)$. The latter yields that the sum of the coordinates is 0.

Figure 5: Dynamics of memory measures: cross-section data



Notes: This figure plots the evolution of $\bar{d}'$ and $\hat{\mathfrak{p}}$ memory measures in the cross-section data for the observed fathers and sons ($t = 1$). Vertical dashed line depict a threshold between epochs $\mathcal{E}_1$ and $\mathcal{E}_2$. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

5.2.2 Class heterogeneity

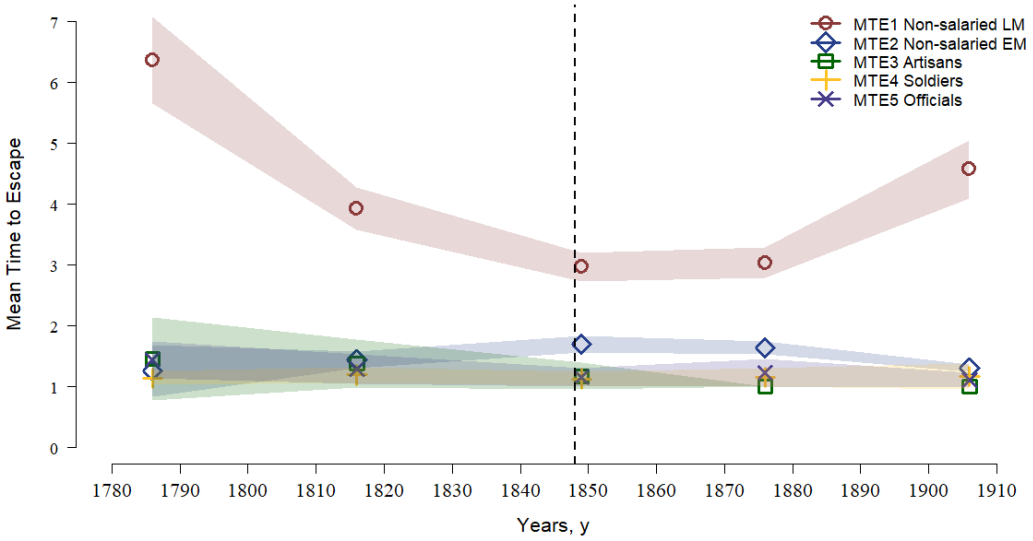
The memory framework allows us to study heterogeneity in memory decay for different initial social classes. To do this, we consider mean exit times to escape from a given state (Kemeny et al., 1969):

$$MTE(P,t) = (1 - P_{ii}) \cdot \sum_{t=1}^{\infty} t \cdot P_{ii}^{t-1} \tag{11}$$

This measure shows the expected number of generations it would take for an individual's descendants to exit each particular social class: the larger this expected value, the less mobile are individuals in this class.

Figures 6 and 7 exhibit the evolution of the mean time to exit from each initial social class, estimated from the six-generations panel and from the cross-section CMGPD-LN data, respectively. The results suggest that the poorest class of late-married non-salaried was the least mobile compared to all other classes in late Qing dynasty China. The dynamics from the six-generations panel data suggest that in the 1783s once an individual reached the poorest class, it took, on average, around six generations for their descendants to upgrade their social position. In the 1850's the stickiness of the poorest class went down: for the descendants of parents in 1850 it would take twice faster (3 generations, on average) to quit from the late-married non-salaried class. Notably, after the 1850s, the mobility of the poorest class went down, and the mean time to exit from this class is 4.5 generations. During the entire period under consideration, it never took more than two generations to quit any other social class. Further, from Figures 6 and 7b, we can see that the early married non-salaried class before 1850 was approximately the same mobile as any salaried class. However, in the epoch $\mathcal{E}_2$ it became less mobile, and the mean time to exit from the early married non-salaried class became significantly higher compared to salaried classes up to 1909. At the end of the Qing dynasty, this class returns to the same high level of mobility as in the 1783s-1790s.

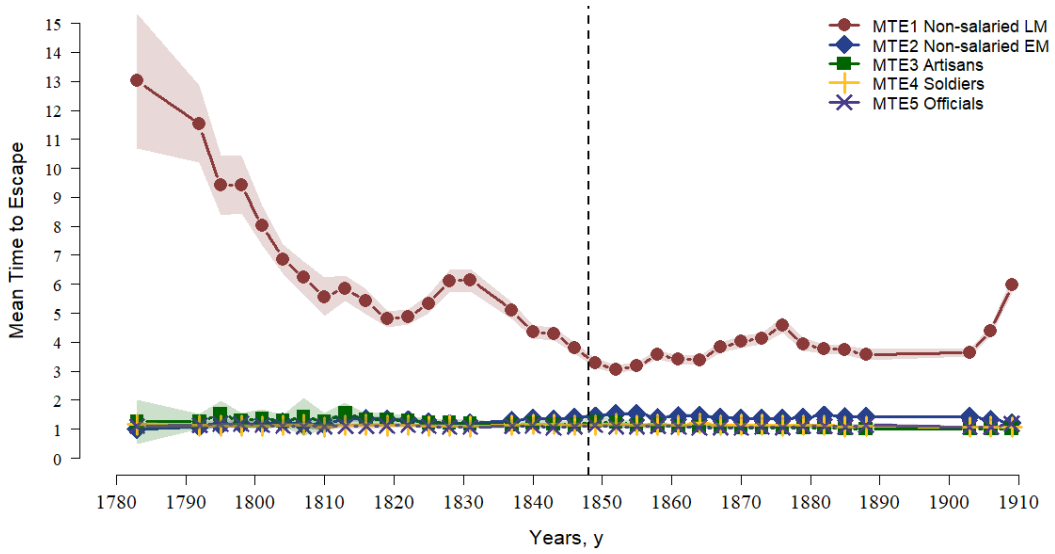
Figure 6: Mean time to exit from each class: six-generations panel data.



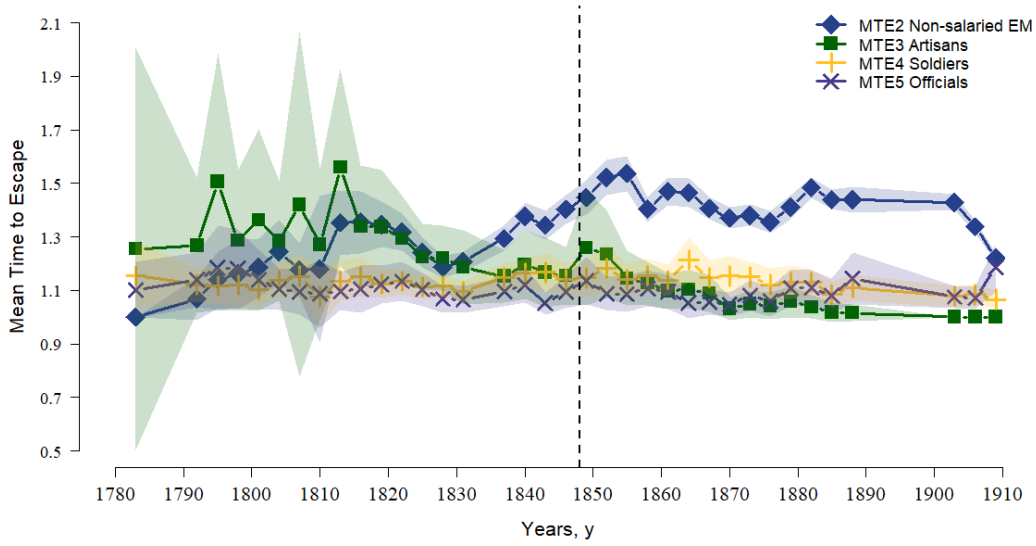
Notes: This figure plots the evolution of the MTE memory measure from eq. (11) for each social class in late Qing dynasty China estimated from the six-generations panel built from the CMGPD-LN. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

Figure 7: Mean time to exit from each class: cross-section data.

(a) All classes.



(b) Except Non-salaried LM.



Notes: This figure plots the evolution of the MTE memory measure from eq. (11) for each social class in late Qing dynasty China estimated from the cross-section data built from the CMGPD-LN. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

5.3 Movement approach

Considering mobility as "movement", we are able to distinguish between the notions of *exchange* and *structural* mobility, which have been widely studied in the sociology literature (McClendon, 1977; Sobel, 1983; Sobel, Hout, & Duncan, 1985). This distinction is important because changes in distribution structure might happen due to exogenous shocks, which might be conflated with the internal factors shaping the transmission of social status across generations. For example, changes in factors of labor supply, labor demand, or changes in the demographic structure might substantially affect structural mobility, but they seem to be unrelated to the true "inherited" part of mobility. For example, industrialization leads to a sizable reduction in the class of farmers and agricultural workers in many countries. This, naturally, yields a lower share of children whose parents had been farmers become farmers as well. However, this rise in mobility is purely driven by structural changes. Therefore, confounding trends of structural and exchange mobility might lead one to misleading conclusions (Jarvis & Song, 2017).

Our approach extends the well-known Broom and Jones (1969) and Featherman, Jones, and Hauser (1975) *circulation* and structural mobility indices from the sociology literature. The main distinction is that we consider the full dynamics of the Markov process up to the steady state and not only the first (observed) step of the process. We begin by defining historical mobility in a similar way to the simplest movement index from Matras (1961). In this setting, in accordance with our mover-stayer model introduced in Section 5.1, we would only care about the distinction between the probability of following versus the probability of not following parental social class. Therefore, we define **historical mobility** as the expected fraction of dynasties whose position in generation $t+1$ is different from the position in generation t:

$$HM(P, \mu^t, t) = 1 - \sum_i \mu_i^t \cdot P_{ii}. \tag{12}$$

It can be noted that HM can be decomposed into upward and downward mobility. To do so, equation (12) can be estimated separately for cases where individuals upgrade their position relative to their father's status, defining upward historical mobility, and where they had lower status compared to their fathers, defining downward mobility.

We assume that historical mobility consists of two parts: exchange and structural mobility. **Structural mobility** is driven by changes in marginal distributions, and is defined as a part of movement that changes the distribution of dynasties over classes:

$$SM(P, \mu^t, t) = \|\mu^t - \mu^t \cdot P\|_{TV} = \frac{1}{2} \cdot \sum_i |\mu_i^t - \mu_i^{t+1}|. \tag{13}$$

This index mimics a simple index of distributional change from Fields and Ok (1996). In defining exchange mobility, we follow a classical definition from Berger and Snell (1957) and assume **exchange mobility** is a part of movement that does not change the distributional structure at every step of the transition t:

$$EM(P, \mu^t, t) = HM(P, \mu^t, t) - SM(P, \mu^t, t). \tag{14}$$

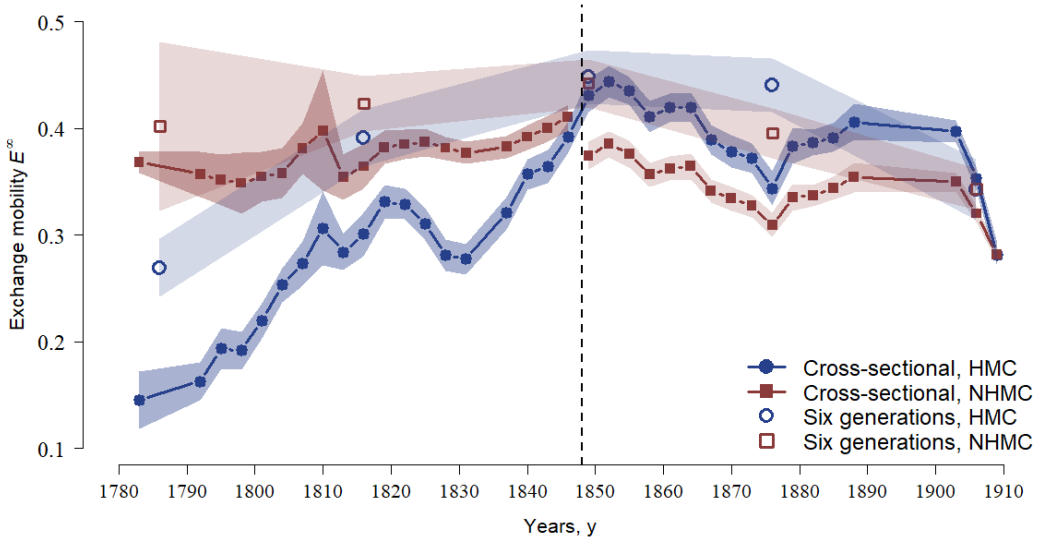
Structural mobility vanishes over time because as $t \to \infty$ marginal distributions converge to the steady-state distribution μ^∞. At the steady state, the historical mobility is equal to the exchange mobility and is computed as:

$$EM^\infty(P, \mu^\infty) = HM^\infty(P, \mu^\infty) = 1 - \sum_i \mu_i^\infty \cdot P_{ii}. \tag{15}$$

Exchange mobility at the steady state EM^{∞} is our preferred measure of intergenerational mobility within the "movement" concept because it captures the number of movements in the Markov chain net of the effect of changes in the structure of cross-section distributions. Namely, it measures the expected fraction of dynasties whose position changes after μ^t stabilized at μ^{∞}. Since (15) includes μ^{∞}, we can estimate EM^{∞} both within homogeneous and nonhomogeneous frameworks. In HMC, to estimate $\hat{\mu}^{\infty}$ we will raize matrix P to the power of t, until $P^t \to P^{\infty}$. In NHMC, instead of μ^{∞} we will plug in the exogenous estimate of $\hat{\mu}^{\infty\mathcal{E}_1}$ and $\hat{\mu}^{\infty\mathcal{E}_2}$ into (15) for the first and second epochs, as described in Section 4.

5.3.1 Dynamics of exchange and structural mobility in the Late Qing

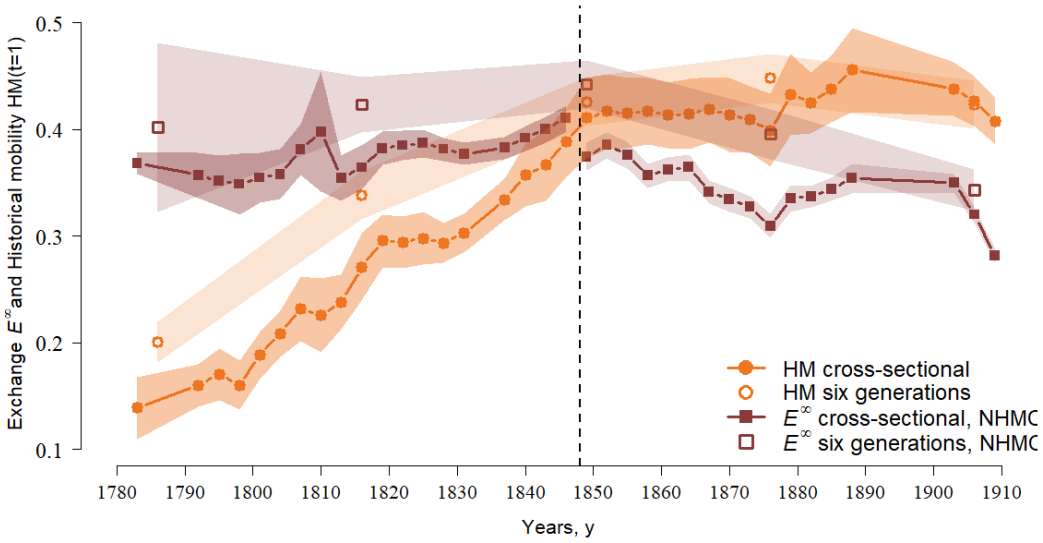
Figure 8: Exchange mobility at the steady state.



Notes: This figure plots the evolution of exchange mobility at the steady state EM^{∞} under on-homogeneous Markov chain (HMC) versus nonhomogeneous Markov chain (NHMC) assumptions. Vertical dashed line depict a threshold between epochs $\mathcal{E}_1$ and $\mathcal{E}_2$. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

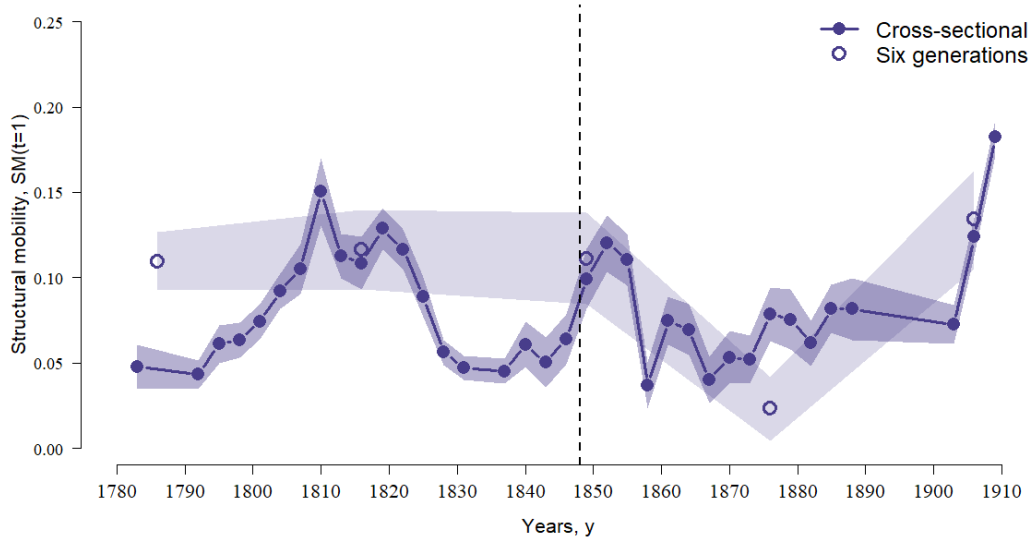
Figure 8 shows the dynamics of the exchange mobility at the steady state EM^{∞} from eq. (15) estimated both from the cross-section and panel six-generations data. Under HMC assumption, net of the effect of changes in distribution structure, exchange mobility at the steady state indicates that the probability of not following parental class in late Qing dynasty China grew from around 25% in the 1783s to 45% in the 1850s, and then fell to 35% in the 1900s. NHMC results demonstrate a more stable trend: the probability of not following parental class at the steady state was around 40-45% in the 1783s-1850s, and then fell to 35% in the 1900s. Figure 9 shows the trend of historical mobility from eq. (12) versus our preferred measure of exchange mobility at the steady state under the nonhomogeneous Markov chain assumption. We would clearly make a different conclusion about the mobility pattern if we only rely on the probability to move for the father-son pairs observed in the data.
Figure 10 asses trends in the structural mobility SM from eq. (13) over time (y). We can see that structural mobility was stable in the first epoch $\mathcal{E}_1$ up to the 1850s ($\approx 11\%$), then dropped to only 2% in the 1870s, and then steadily raised back to 13% in the 1900s. It could also be noticed from the evolution of cross-section distributions reported in Figure 1 that structural changes were

Figure 9: Exchange mobility at the steady state versus historical mobility.



Notes: This figure plots the evolution of exchange mobility at the steady state EM^{∞} under the nonhomogeneous Markov chain (NHMC) assumption versus the historical mobility $HM(t=1)$. Vertical dashed line depict a threshold between epochs $\mathcal{E}_1$ and $\mathcal{E}_2$. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

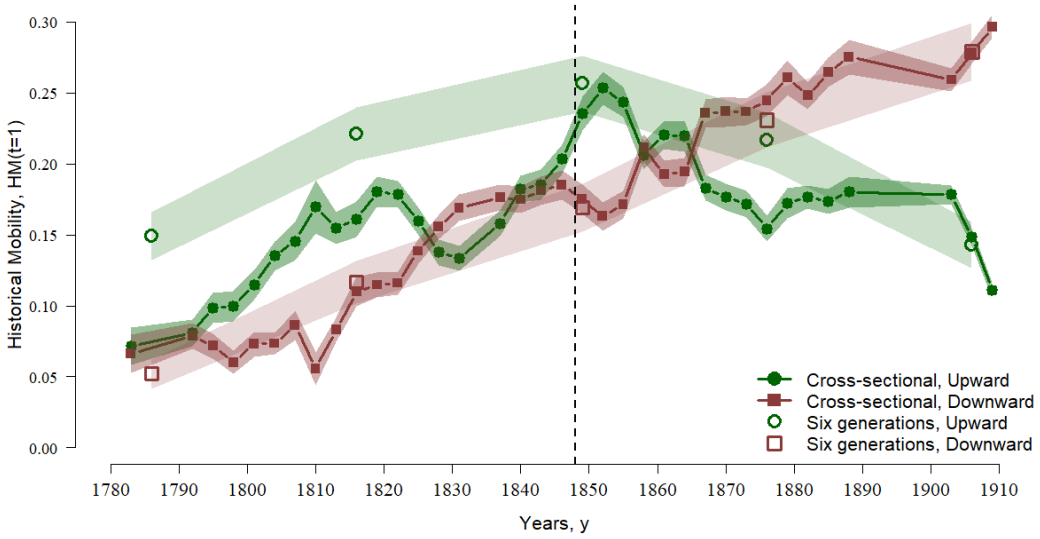
Figure 10: Structural mobility.



Notes: This figure plots the evolution of structural mobility $SM(t=1)$. Vertical dashed line depict a threshold between epochs $\mathcal{E}_1$ and $\mathcal{E}_2$. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

mostly driven by the dynamics of early- and late-married nonsalaried classes, because non-salaried accounts for around 98% of the total population over the whole period under consideration. We can see from Figure 1 that the late-married non-salaried class (the poorest class) dropped from 92% in generation $t=0$ in our six-generations panel data to only 61% in

Figure 11: Upward and downward historical mobility.



Notes: This figure plots the evolution of upward and downward historical mobility $HM(t=1)$. Vertical dashed line depict a threshold between epochs $\mathcal{E}_1$ and $\mathcal{E}_2$. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

generations $t=3$ and $t=4$ and then razed to 78% in generation $t=4$[1].
Figure 11 shows the evolution of upward (green color) and downward (red color) historical mobility from eq. (12). We can see that upward historical mobility dominated downward mobility in the late Qing dynasty before the 1870s, and afterward the observed pattern reversed.

5.4 Average expected outcomes

To get a better understanding of trends in intergenerational mobility in late Qing dynasty China, we introduce an additional measure aimed at capturing the evolution of the average "levels" of social status. Mobility may be high or low, but it's important to understand around which social classes, on average, all the existing "movements" happen within the movement concept of mobility. In turn, within the memory concept, it's also important to understand not only how fast the lotteries of descendants from different classes equalize but also to assess how good or bad the equilibrium lottery μ^{∞} is in terms of the prediction of the expected average social class. To answer these questions, we introduce a measure called average expected outcome.
For every initial group i and a fixed transition matrix, we can calculate **expected outcome** of the child whose parent belongs to group i:

$$EO_i(P,o) = \sum_{j=1}^{n} p_{ij} o_j, \tag{16}$$

where p_{ij} are transition probabilities and o_j is a score assigned to class j. This definition is similar to the *ex-ante* expected evaluation of the income prospects of the children at the time of their birth, which was formulation as the opportunity index in (Van de Gaer et al., 2001, pp. 527–528).

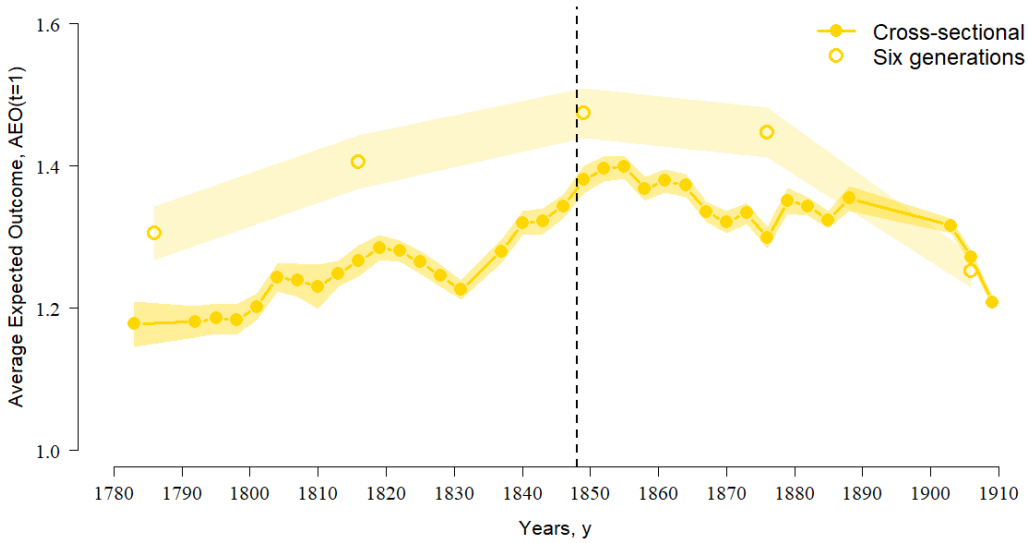
[1]To recall, generations t in the six-generations panel corresponds to the cross-section data years y in accordance with the Table 2.

One can then introduce the **average expected outcome** in society at every step of the Markov process, which is defined as EO_i weighted with fractions of the population in each class i in period t:

$$AEO^t(P, o, \mu^t, t) = \sum_{i=1}^{n} \mu_i^t \sum_{j=1}^{n} p_{ij} o_j. \tag{17}$$

Figure 12 shows the dynamics of average expected outcomes computed from the CMGPD-LN

Figure 12: Average expected outcomes.



Notes: This figure plots the evolution of average expected outcomes $AEO(t = 1)$. Vertical dashed line depict a threshold between epochs $\mathcal{E}_1$ and $\mathcal{E}_2$. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

data. In line with the dynamics of upward and downward mobility, we find that average expected outcomes grew before the 1870s and then declined up to the 1900s.

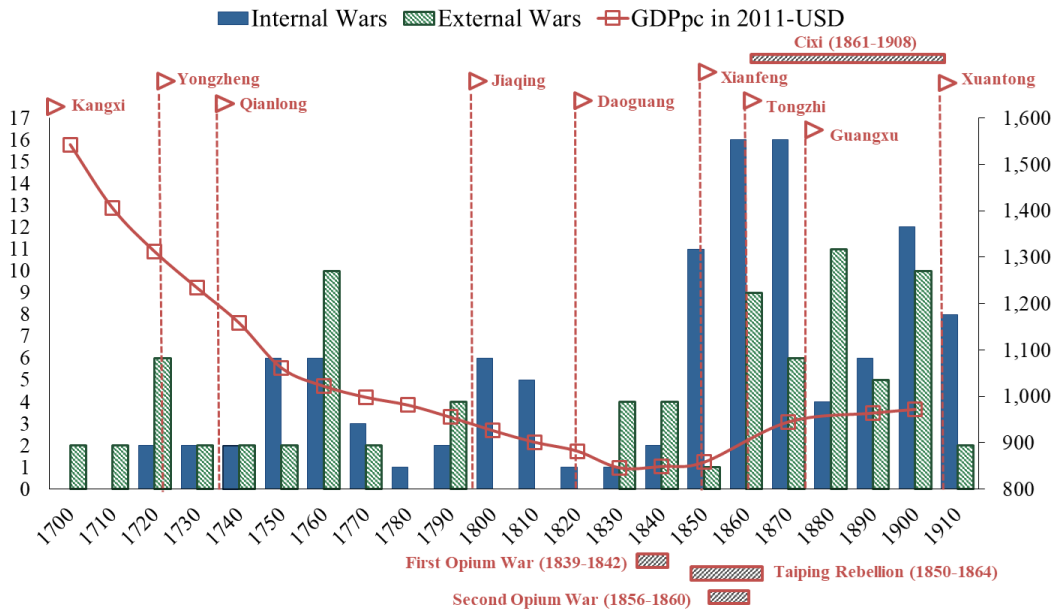
6 Back to history

We highlight that GDP per capita is negatively correlated with the average expected outcome as well as with the (counterfactual) exchange mobility EM^{∞} computed under the assumption of HMC (compare Figure 13, 8, and 12). The decrease in GDP per capita until the mid-18th century aligned with the general population growth that approached a plateau by 1840. The increase in the exchange mobility (EM^{∞}, HMC) observed by the mid-18th century, likely indicates the gradual improvement of opportunities.

Our findings on a new mobility regime emerging after 1850 naturally link to the beginning of the Taiping rebellion in 1850 and the beginning of the Second Opium War in 1856. An interesting manifestation of this change is the emergence of a sudden drop in the structural mobility around 1880. This occurred during a short period of partially effective modernization in military production, transportation, and education. It seems that under the foreign aggression and domestic response with reforms and unrest, the demand for occupations stabilized.

In the last decades of the Qing dynasty, the entropy inequality measures for fathers and sons diverge, and so there is no basis for extrapolating to a steady state (see Figure 4 and the discussion around it).

Figure 13: Economic and political environment in the late Qing dinasty China.



Notes: Internal and external wars characterize number of ongoing conflicts over a period of plus-minus five years around the year on the x-axis. Data: GDP per capita is taken from Maddison Project Database 2020 (https://www.rug.nl/ggdc/), number of conflicts is computed based on Orlandi et al. (2023).

These changes in dynamics followed the Hundred Days' Reform in 1898, the foreign intervention associated with the Boxer Rebellion (1900), and the New Policies (initiated in 1901), which reformed the military, administration, and education. A distinct occurs around 1900 with the abolition of the Imperial examination system for officeholding. However, the trend of upward mobility shifted from ascent to decline in the mid-18th century, making it unlikely that changes to the rules governing entry into civil service would immediately impact the quality of the elite, especially on a positive side. The reign of the Dowager Empress Ci Xi from 1875 to 1901 is often regarded as a period of bad governance and long run decay of the Imperial regime and may be a more natural explanation of the unstable dynamics we observe. We emphasize that changes in the social status dynamics, measured through entropy and exchange mobility, showed a rapid deviation from the steady state in the 1900s, as if the distribution of social classes had been driven out of a region of equilibrium stability. We cannot know whether this was movement towards some new equilibrium; the Qing dynasty was overthrown instead.

7 Conclusion

In this paper, we study long-run trends in intergenerational mobility over the last 130 years of the Qing dynasty China. We build our analysis employing Markov chains for modeling the transmission of social status from parent to child, which is a common tradition in the sociology literature. Our approach consolidates existing Markov chain mobility measures as well as new measures derived from the mover-stayer theoretical framework. We consider two main conceptual approaches for measuring mobility: mobility as movement and immobility as memory. We see the technical contribution of our paper in offering measures that unify both concepts.
We find strong evidence that the mobility process of the late Qing qualitatively changed in the

period of the Opium Wars and Taiping Rebellion. The first epoch of 1783-1850 exhibits converging mobility dynamics, while the second epoch of 1850-1911 reveals the instability during the last years of the Qing.

Within the memory approach, we find that societal memory decays over time, and for the descendants of the latest cohorts around 1909 it took substantially faster to reach class independence (equality of opportunity) compared to the early cohorts in 1783. We also find that the poorest social class of late-married non-salaried was substantially less mobile than all other classes, and for the sons whose fathers came from this class, it might take from three to six generations to upgrade their social position.

Considering mobility as movement, we document low levels of structural mobility. The share of movements that changed the distribution of dynasties over classes was around 11% in the first epoch, then it dropped in the 1870s to 2% and caught up to 13% in the 1900s. Exchange mobility at the equilibrium, estimated under the nonhomogeneous Markov chain assumption, governs the probability of changing social class without changing social structure was around 40-45% before the 1850s, and then fell to 35% up to the 1900s. We also find that upward mobility dominated downward mobility in the late Qing dynasty before the 1870s, and afterward, the observed pattern reversed. The quality of the average lottery of class choice, measured by average expected outcomes confirms the trend in directional mobility: descendants faced improving lotteries of higher expected classes before the 1870s, and afterward up to the 1900s the quality of lotteries declined presuming lower expected class .

In terms of future research, we make two suggestions. Methodologically, we show the power of nonhomogeneous Markov chain methods to give novel insights into mobility dynamics. These tools naturally require multi-generation data and so suggest the importance of developing such methods. We also think that our distinctions between mobility and lack of memory of parental status and mobility as movement away from parents can matter in terms of characterizing mobility dynamics.

Substantively, we think a natural next step involves integration of the stylized mobility facts we identify into interpretations of the causes of the 1911 Chinese Revolution in particular and revolutions in general. The findings that the crises of the 1850 shifted the Chinese mobility regime in a way that produced decreasing upward mobility exhibits the interplay of contingency and structure. Contingency involves the particular crises that occurred while structure involves the emergence of a new regime in which upward mobility deteriorates. Links between the absence of upward mobility and revolutions are a standard idea in the historiographical literature and our work highlights their empirical salience for China.

References

Bartholomew, D. J. (1967). *Stochastic models for social processes.* London: Wiley.

Berger, J., & Snell, J. L. (1957). On the concept of equal exchange. *Behavioral Science*, *2*(2), 111–118.

Beshers, J. M., & Reiter, S. (1963). Social status and social change. *Behavioral Science*, *8*(1), 1–13.

Blume, L., Cholli, N. A., Durlauf, S. N., & Lukina, A. (2024). *Immobility as memory: Some new approaches to characterizing intergenerational persistence via markov chains* (Tech. Rep.). National Bureau of Economic Research.

Boudon, R. (1973). *Mathematical structures of social mobility* (Vol. 2). Jossey-Bass.

Broom, L., & Jones, F. L. (1969). Career mobility in three societies: Australia, italy, and the united states. *American Sociological Review*, 650–658.

Brumelle, S., & Gerchak, Y. (1982). The structure of indices of social mobility and inheritance. *Journal of Mathematical Sociology*, *8*(2), 251–264.

Chen, B., Campbell, C., & Dong, H. (2018). Interethnic marriage in northeast china, 1866–1913. *Demographic Research*, *38*, 929–966.

Chen, S. (2017). *State-sponsored inequality: The banner system and social stratification in northeast china*. Stanford University Press.

Chen, S., Campbell, C. D., & Lee, J. (2014). Categorical inequality and gender difference: Marriage and remarriage in northeast china, 1749-1913. In *Similarity in difference: Marriage in europe and asia, 1700-1900.*

Conlisk, J. (1976). Interactive markov chains. *Journal of Mathematical Sociology*, *4*(2), 157–185.

Conlisk, J. (1990). Monotone mobility matrices. *Journal of Mathematical Sociology*, *15*(3-4), 173–191.

Conlisk, J., & Sommers, P. (1979). Eigenvector status proxies in markov chain mobility models. *Sociological Methods & Research*, *8*(2), 159–178.

Cowell, F. A. (1980). Generalized entropy and the measurement of distributional change. *European Economic Review*, *13*(1), 147–159.

Cowell, F. A. (1985). Measures of distributional change: An axiomatic approach. *The Review of Economic Studies*, *52*(1), 135–151.

Dardanoni, V. (1993). Measuring social mobility. *Journal of Economic Theory*, *61*(2), 372–394.

Ding, Y., Guo, S., Lee, J., & Campbell, C. (2004). *Liaodong yimin zhong de qiren shehui [banner society and the settlement of liaodong]*. Shanghai: Shanghai shehui kexue chubanshe.

Dobrushin, R. L. (1956). Central limit theorem for nonstationary markov chains. i. *Theory of Probability & Its Applications*, *1*(1), 65–80.

Duncan, O. D. (1966a). Methodological issues in the analysis of social mobility. In *Social structure & mobility in economic dovelopment* (pp. 51–97). Routledge.

Duncan, O. D. (1966b). Occupation trends and patterns of net mobility in the united states. *Demography*, *3*, 1–18.

Featherman, D. L., Jones, F. L., & Hauser, R. M. (1975). Assumptions of social mobility research in the us: The case of occupational status. *Social science research*, *4*(4), 329–360.

Fields, G. S., & Ok, E. A. (1996). The meaning and measurement of income mobility. *Journal of Economic Theory*, *71*(2), 349–377.

Fields, G. S., & Ok, E. A. (1999). Measuring movement of incomes. *Economica*, *66*(264), 455–471.

Frydman, H. (1984). Maximum likelihood estimation in the mover-stayer model. *Journal of the American Statistical Association*, *79*(387), 632–638.

Goodman, L. A. (1961). Statistical methods for the mover-stayer model. *Journal of the American Statistical Association*, *56*(296), 841–868.

Halmos, P. R. (1947). *Finite dimensional vector spaces* (No. 7). Princeton University Press.

Jarvis, B. F., & Song, X. (2017). Rising intragenerational occupational mobility in the united states, 1969 to 2011. *American sociological review*, *82*(3), 568–599.

Jiang, Q., & Kung, J. K.-s. (2021). Social mobility in late imperial china: reconsidering the "ladder of success" hypothesis. *Modern China*, *47*(5), 628–661.

Keller, W., & Shiue, C. H. (2023). *Intergenerational mobility of daughters and marital sorting: New evidence from imperial china* (Tech. Rep.). National Bureau of Economic Research.

Kemeny, J. G., Snell, J. L., et al. (1969). *Finite markov chains* (Vol. 26). van Nostrand Princeton, NJ.

Lee, J. Z., & Campbell, C. (2014). China multi-generational panel dataset, liaoning (cmgpd-ln), 1749–1909. icpsr27063-v10. *Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Available at: https://doi. org/10.3886/ICPSR27063. v10*.

Mare, R., & Song, X. (2014). Social mobility in multiple generations. *UCLA CCPR Population Working Papers*.

Matras, J. (1961). Differential fertility, intergenerational occupational mobility, and change in the occupational distribution: Some elementary interrelationships. *Population Studies*, *15*(2), 187–197.

Matras, J. (1967). Social mobility and social structure: Some insights from the linear model. *American Sociological Review*, 608–614.

McClendon, M. J. (1977). Structural and exchange components of vertical mobility. *American Sociological Review*, 56–74.

Mitra, T., & Ok, E. A. (1998). The measurement of income mobility: A partial ordering approach. *Economic Theory*, *12*, 77–102.

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

Pullum, T. W. (1975). *Measuring occupational inheritance*. Amsterdam: Elsevier.

Shiue, C. H. (2016). *Social mobility in the long run: An analysis with five linked generations in china, 1300-1900* (Tech. Rep.). Working paper.

Shorrocks, A. F. (1978). The measurement of mobility. *Econometrica: Journal of the Econometric Society*, 1013–1024.

Sobel, M. E. (1983). Structural mobility, circulation mobility and the analysis of occupational mobility: a conceptual mismatch. *American Sociological Review*, 721–727.

Sobel, M. E., Hout, M., & Duncan, O. D. (1985). Exchange, structure, and symmetry in occupational mobility. *American Journal of Sociology*, *91*(2), 359–372.

Sommers, P. M., & Conlisk, J. (1979). Eigenvalue immobility measures for markov chains. *Journal of Mathematical Sociology*, *6*(2), 253–276.

Song, X. (2021). Multigenerational social mobility: A demographic approach. *Sociological Methodology*, *51*(1), 1–43.

Spilerman, S. (1972). Extensions of the mover-stayer model. *American Journal of Sociology*, *78*(3), 599–626.

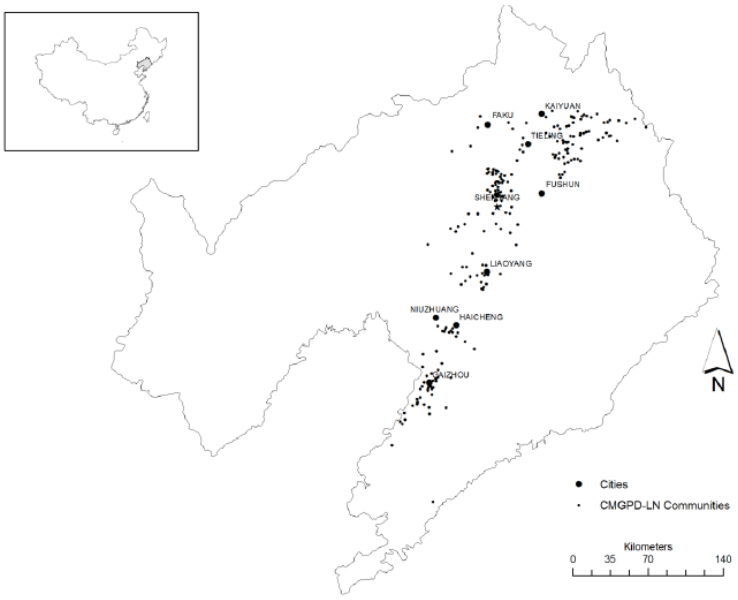
Theil, H. (1972). *Statistical decomposition analysis*. Amsterdam: North-Holland.

Van de Gaer, D., Schokkaert, E., & Martinez, M. (2001). Three meanings of intergenerational mobility. *Economica*, *68*(272), 519–538.

Appendix

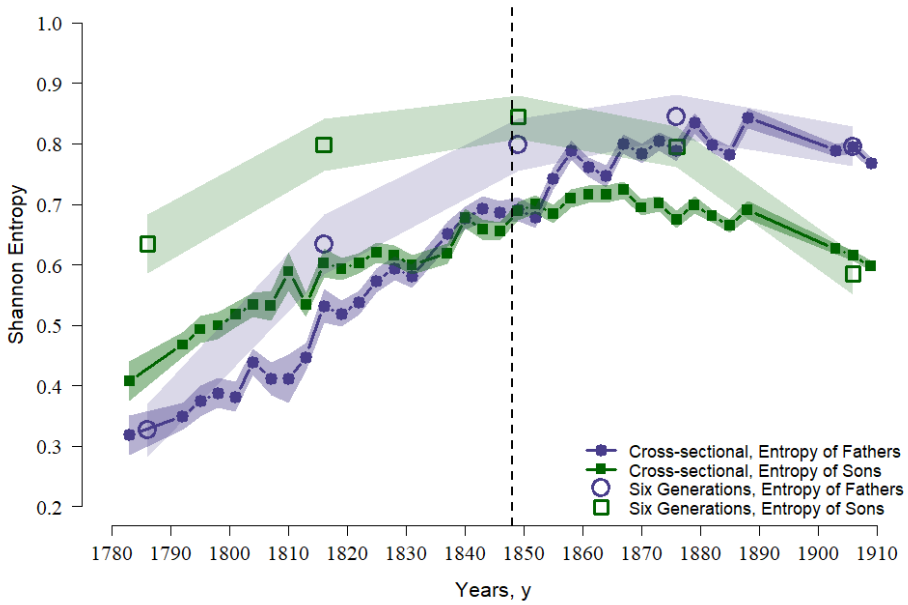
A CMGPD-LN: detailed

Figure A1: CMGPD-LN communities.



B Shannon entropy: all ages

Figure B1: Dynamics of the Shannon entropy for the entire dataset



Notes: This figure plots the dynamics of the entropy of fathers $E(\mu_F)$ and the entropy of sons $E(\mu_S$ separately for the entire cross-section and six-generations data without any age restrictions. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

C Aggregation of transition matrices

We aggregate the set of matrices $P(y)$ parameterized by integers y from a given set Y to a new matrix T. The aggregation also relies on vectors $\mu(y)$ and $\nu(y)$ such that $\nu(y) = \mu(y) \cdot P(y)$, $y \in Y$. In our application, $\mu(y) = \mu_F(y)$ and $\nu(y) = \mu_S(y)$ represent the distributions of fathers and sons prescribed to years $y \in Y$ in the cross-section data. The aggregation is performed as the regularization problem of the type

$$\mathfrak{L} = \sum_y \sum_j \left(\sum_i T_{ij}\mu_i(y) - \nu_j(y) \right)^2 + \lambda \sum_y \sum_i \sum_j (T_{ij} - P_{ij}(y))^2 \longrightarrow \min \tag{18}$$

subject to the constrain $T_{ij} \geq 0$. We stress that on one hand, our purpose is to find the matrix T that maps available vectors $\mu(y)$ as close to $\nu(y)$, $y \in Y$, as possible. However, the optimization (18) with $\lambda = 0$ is ill posed because the infinite number of matrices satisfies the equation $\mu \cdot M = \nu$. On the other hand, as λ grows, the matrix T that solves (18) becomes closer to the mean of the given matrices $P(y)$, $y \in Y$. We utilize $\lambda = 0.1$ in the paper to secure the exclusion of an extraneous matrix T ($\lambda = 0$) still maintaining the goal of the proximity between $\mu(y) \cdot M$ and $\nu(y)$. As soon as this goal is attained with a small λ, the solution of (18) delivers a kind of smoothing of the matrices $P(y)$. A further mixing with the mean, corresponding to the increase in λ, affects the matrix T only insignificantly. Thus a wide range of λ leads to quantitatively similar results.
The solution of the unconstrained optimization problem (18) is given by the matrix equation

$$\left(\left(\frac{1}{N} \sum_y \mu_i(y)\mu_j(y) \right) + \lambda\delta_{ij} \right)_{i,j=1}^{n} \cdot \begin{pmatrix} T_{1j'} \\ T_{2j'} \\ \vdots \\ T_{nj'} \end{pmatrix} = \begin{pmatrix} \frac{1}{N}\sum_y \left(\mu_1(y)\nu_{j'}(y) + \lambda P_{1j'}(y)\right) \\ \frac{1}{N}\sum_y \left(\mu_2(y)\nu_{j'}(y) + \lambda P_{2j'}(y)\right) \\ \vdots \\ \frac{1}{N}\sum_y \left(\mu_n(y)\nu_{j'}(y) + \lambda P_{nj'}(y)\right) \end{pmatrix}$$

$\delta_{ij} = 1$ if $i = j$ and 0 otherwise. If the constrain is violated, a corner solution is chosen.

D Memory measures in the Mover-Stayer model

Measures $\bar{d}'$ and d' for $t = 1$. Equation (9) with $t = 1$ involves the differences between two rows of the transition matrix: $\sum_k |P_{ik} - P_{jk}|$. In the case of (7), $P = \mathcal{P}$, $\mathcal{P}_{ik} - \mathcal{P}_{jk} = \mathfrak{p}$ if $k = i$ and $k = j$, but $\mathcal{P}_{ik} - \mathcal{P}_{jk} = 0$ otherwise. Therefore, $\sum_k |\mathcal{P}_{ik} - \mathcal{P}_{jk}| = 2\mathfrak{p}$ for any i and j. This yields that

$$\bar{d}'(\mathcal{P}, 1) = \frac{1}{n(n-1)} \sum_{i>j} \sum_{k=1}^{n} |\mathcal{P}^t_{ik} - \mathcal{P}^t_{jk}| = \frac{1}{n(n-1)} \sum_{i>j} \sum_{k=1}^{n} 2\mathfrak{p} = \mathfrak{p}$$

and

$$d'(\mathcal{P}, 1) = \max_{i,j} \sum_{k=1}^{n} |\mathcal{P}^t_{ik} - \mathcal{P}^t_{jk}| = \mathfrak{p}$$

too.

Second eigenvalue λ_2. We note that $\lambda_2(\mathcal{P}, t) = \mathfrak{p}^t$ in our version of the mover-stayer model, since $\mathfrak{p}$ is the second largest eigenvalue of the matrix $\mathcal{P}$ given by Equation (7).

$\Delta\mathbf{M}$ The probability to keep the class at the current distribution $\mu^{\mathbf{t-1}}$ is $\sum_i \mu_i^{\mathbf{t-1}}\mathcal{P}_{ii}$. Then the probability to move is

$$\Delta\mathbf{M}_{\text{part}}(\mathcal{P}, \mu^{\mathbf{t-1}}, t) = 1 - \sum_i \mu_i^{\mathbf{t-1}}\mathcal{P}_{ii},$$

which is the part of $\Delta\mathbf{M}$ defined in (8). Using the definition $\mathcal{P}_{ii} = \mathfrak{p} + (1-\mathfrak{p})\mu_i^\infty$ we have that $\sum_i \mu_i^{t-1}\mathcal{P}_{ii} = \mathfrak{p}\sum_i \mu_i^{\mathbf{t-1}} + (1-\mathfrak{p})\sum_i \mu_i^{\mathbf{t-1}}\mu_i^\infty = \mathfrak{p} + (1-\mathfrak{p})\sum_i \mu_i^{t-1}\mu_i^\infty$ and

$$\Delta\mathbf{M}_{\text{part}}(\mathcal{P}, \mu^{\mathbf{t-1}}, t) = (1-\mathfrak{p})(1 - \sum_i \mu_i^{\mathbf{t-1}}\mu_i^\infty). \tag{19}$$

Let's prove by induction that

$$\mu^{\mathbf{t}} = \mathfrak{p}^t\mu^{\mathbf{0}} + (1-\mathfrak{p}^t)\mu^\infty. \tag{20}$$

The base of induction, $\mu^{\mathbf{1}} = p\mu^{\mathbf{0}} + (1-p)\mu^\infty$, is equivalent to the definition. Further, given $\mu^{\mathbf{t-1}} = \mathfrak{p}^{t-1}\mu^{\mathbf{0}} + (1-\mathfrak{p}^{t-1})\mu^\infty$ and $\mu^{\mathbf{t-1}} = \mathfrak{p}\mu^{\mathbf{t-2}} + (1-\mathfrak{p})\mu^\infty$, we have

$$\mu^{\mathbf{t}} = \mathfrak{p}(\mathfrak{p}^{t-1}\mu^{\mathbf{0}} + (1-\mathfrak{p}^{t-1})\mu^\infty) + (1-\mathfrak{p})\mu^\infty = \mathfrak{p}^t\mu^{\mathbf{0}} + (\mathfrak{p}-\mathfrak{p}^t)\mu^\infty + (1-\mathfrak{p})\mu^\infty$$

ending up with (20). Substituting (20) to (19), we get

$$\Delta\mathbf{M}_{\text{part}}(\mathcal{P}, \mu^{\mathbf{t-1}}, t) = (1-\mathfrak{p})\Big(1 - \sum_i \big(\mathfrak{p}^{t-1}(\mu^{\mathbf{0}} - \mu^\infty) + \mu^\infty\big)\mu^\infty\Big)$$

As $t \to \infty$ the terms with $\mathfrak{p}^{t-1}$ disappear, and

$$\Delta\mathbf{M}(\mathcal{P}, \mu^{\mathbf{t-1}}, t) = (1-\mathfrak{p})\mathfrak{p}^{t-1}\sum_i(\mu_i^{\mathbf{0}} - \mu_i^\infty). \tag{21}$$

Thus, $\Delta\mathbf{M}(\mathcal{P}, \mu^{\mathbf{t-1}}, t)$ as a function of t behaves as Cp^t, where $C = (1-\mathfrak{p})/\mathfrak{p} \cdot \sum_i(\mu_i^{\mathbf{0}} - \mu_i^\infty)$.

E Additional memory measure

Measure involving the rate of convergence to the steady state. The direction $x = \mu^\infty - \mu^0$ from the current lottery μ^0 toward μ^∞ allows one to define the measure

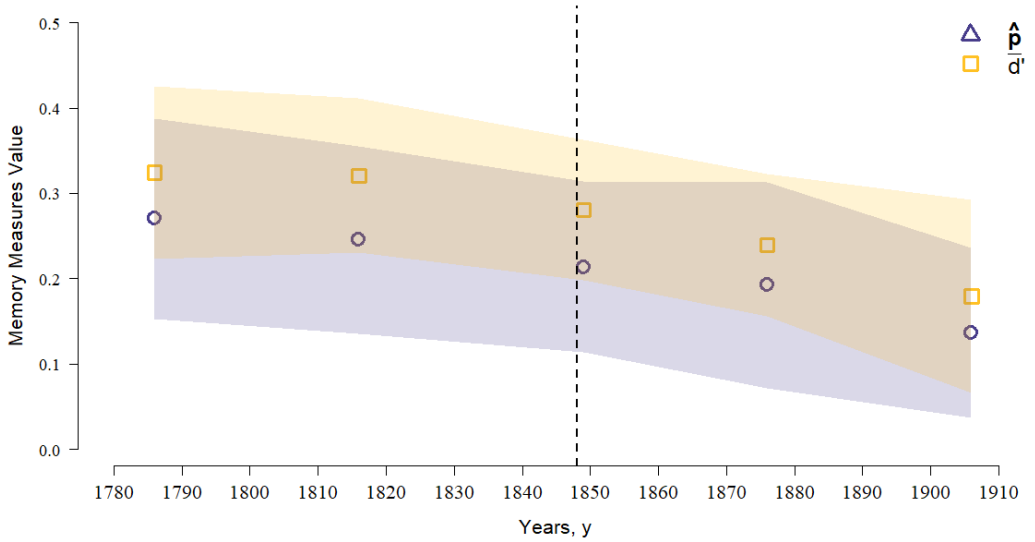
$$D_\infty(P, t) = \frac{\|(\mu^0 - \mu^\infty)\cdot P^t\|}{\|\mu^0 - \mu^\infty\|} \tag{22}$$

in line with (9). The measure D_∞, which quantifies the diminishing role of the initial lottery μ^0 as the number of iterations as t increases, can be included in the variety of memory measures. However, D_∞ *has nothing in common with the proneness to follow fathers' decision* measured by the above measures. In the computation, we limit ourselves by $D_\infty(P, 1)$, which is also well defined under the assumption of the nonhomogeneous Markov chain. In the latter case, μ^∞ is estimated with this nonhomogeneous Markov chain. We emphasize that D_∞ itself measures only the speed of the convergence to the steady state avoiding the question of the proximity of the current lottery to the steady state. For example, the decrease of D_∞ in 1860–1909 accompanied the approaching of the status distribution to the (considered counterfactual) steady state of epoch $\mathcal{E}_2$, whereas the steady approaching of the lottery distribution to the steady state in epoch $\mathcal{E}_1$ is characterized by multidirectional changes in D_∞. Thus the convergence rate and the distance to the steady state must be considered together.

We note that the measure (22) involves not only the transition matrix and the initial lottery, but also the steady state. Clearly, such a dependence can be incorporated into a measure in various ways.

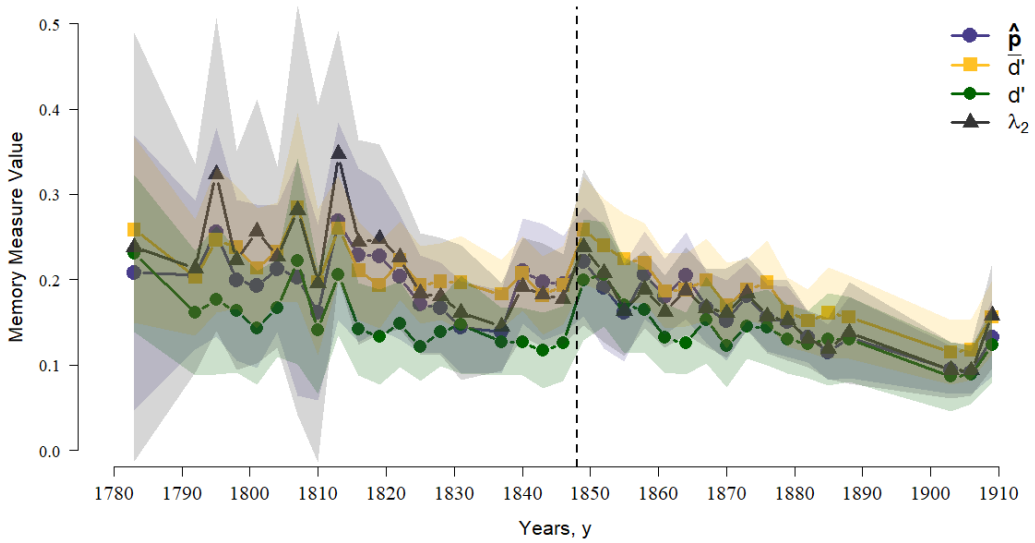
F Additional results on memory measures

Figure F1: Dynamics of memory measures: six-generations panel data.



Notes: This figure plots the evolution of memory measures $\bar{d}'$ and $\hat{\mathbf{p}}$ in the six-generations data. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).

Figure F2: Comparison to the conventional memory measures.



Notes: This figure plots the evolution of conventional Markov chain literature memory measures d' and λ_2 versus our memory measures $\bar{d}'$ and $\hat{\mathbf{p}}$ in the cross-section data for the observed fathers and sons ($t = 1$). Vertical dashed line depict a threshold between epochs $\mathcal{E}_1$ and $\mathcal{E}_2$. Shaded areas represent 95% bootstrap CIs (based on 1000 bootstrap samples).