# Social Mobility in the Long Run: A Temporal Analysis of China from 1300 to 1900*

Carol H. Shiue
University of Colorado, NBER, and CEPR

April 1, 2019

**Abstract**

Does inequality within the family play a significant role in explaining mobility patterns from one generation to the next? This paper exploits temporal changes in mobility over approximately 20 generations and six centuries to shed light on the sources of social mobility. Socioeconomic data on status and links at the individual level come from historical biographies of seven extended families (dynasties as based on the male surname) who lived in one region in China. The analysis documents a trend towards greater social mobility over time. Times of greater inequality between fathers, especially educational inequality, are times of lower social mobility. Moreover, geographic location strengthens the role of inequality for social mobility. Decomposing inequality into between versus within-dynasty components, however, shows that not all inequality is associated with persistence. While inequality between dynasties is conducive to persistence, inequality within the dynasty is associated with higher mobility, and this is true both upward and downward. Furthermore, among members of even closer kin in the dynasty, the positive relationship of inequality and mobility is stronger still. The results are robust to alternative measures of mobility, inequality, and definitions of status.

(JEL N35, D31, J62).

*Thanks to Miles Corak, Joseph Ferrie, Raquel Fernandez, Murat Iyigun, Wolfgang Keller, Naomi Lamoreaux, and Debraj Ray, as well as audiences at the CEPR Economic History Symposium, the Cliometrics Society Conference, the Conference on Deep Rooted Factors in Comparative Development at Brown, Iowa State, Kiel, Michigan, the NBER Summer Institute Development of the American Economy, Nebraska, Northwestern, NYU, the Paris School of Economics Conference on China and India, the Shanghai University of Finance and Economics (Shanghai), the Southwestern University of Finance and Economics (Chengdu), and the World Bank for useful comments. This project was supported by a grant from the National Institutes of Health (NICHD Award HD042731-01 "Intergenerational and Interfamily Economic Links"). T. Telford provided part of the data employed in this paper. Peiyuan Li provided excellent research assistance. Email: shiue@colorado.edu.

# 1 Introduction

Does inequality within the family play a significant role in explaining mobility patterns from one generation to the next? Across advanced countries today, more unequal societies tend to exhibit less intergenerational mobility.[1] The cross-sectional relationship suggests inequality inhibits mobility. However, our understanding of the determinants of social mobility is incomplete. While national policies have received much of the attention in the literature, less well-observed actions taken within the family may significantly redistribute wealth and resources among members, and thus impact on mobility in the next generation. For example, a shift towards more progressive government investments in education may increase mobility, but members of a family can also shift resources from richer to poorer members. It is well-known that the central purpose of many private social organizations and cooperatives, in the past and today, is to insure against risk and provide support to group members (Greif and Tabellini 2017), and that these informal groups, not only governments, provide important functions of redistribution. It is not known, however, whether greater inequality within the family is also consistent with less intergenerational mobility and whether this relationship changes over time.

This paper adopts a temporal perspective to better understand the inequality-mobility nexus. The long time period of study—approximately twenty generations over six centuries, the years 1300 to 1900—is useful for analyzing the relationship of inequality in the parent's generation with the mobility of the child, relative to his parents, for many cohorts.[2] The hypothesis is that if the members of the dynasty are altruistically linked, then inequality *within* the parental generation of the dynasty may translate into more mobility, even if the inequality-mobility relationship *between* dynastic groups is negative. By breaking down inequality into the between-dynasty and the within-dynasty components, we can better understand the role of progressive government policies versus redistribution within the family, respectively, in mediating overall mobility outcomes between parent and child.

The study uses a novel sample of families based on socioeconomic data at the individual level created from genealogical biographies of all the members of seven extended families (dynasties as based on the male surname) who lived in Tongcheng County, China.[3] There are approximately 40,000 individuals in the sample (of which there are around 10,000 married couples). The data allows one to ask whether there is a negative relationship between inequality and mobility in this sample overall, and if the same relationship holds also within members of the same dynasty. It also permits the further step of comparing the inequality and mobility relationship within even closer kin segments within the dynasty.

Typically, significant differences in population composition across countries preclude reliable conclusions

[1] Based on the inequality of children growing up in the 1970s for 12 countries, see Corak (2013). Chetty, Hendren, Kline, and Saez (2014) find a negative relationship between mobility and inequality across regions in the U.S. at roughly the same time. The former analysis yields one of the most well-known graphs in this literature, the "Great Gatsby Curve" (Krueger 2012).

[2] Temporal analysis using longitudinal data can yield mobility estimates that are more comparable and thus allow insights on the extent to which different societies offer equal opportunities (e.g., Corak 2013).

[3] In this paper, I mostly use the term dynasty. The larger extended family group is also referred to in the literature as clan, lineage, patrilines, or common descent group.

about the impact of inequality as a driver of intergenerational mobility.[4] Sharp variations in environment and cultural values, however, are likely to be limited in the Tongcheng sample of individuals because everyone resided in the same region. Nevertheless, the data reflect not only a wide range of high and low socioeconomic classes but also a substantial degree of variation in inequality, mobility, and other factors, by birth cohort, over time. If families reallocate resources among their members, we would expect to find a very different pattern between mobility and inequality within the dynasty compared to society as a whole. In addition, as these individuals resided in the county seat of Tongcheng and the more than 200 surrounding towns and villages, spatial differences in geographic locations can be observed. Exploiting both the temporal and geographical sources of heterogeneity produces a powerful lens with which to observe the factors that affect social mobility.

This paper presents several main results. First, inequality is an important determinant of social mobility. Times during which there is a high degree of inequality in the status of fathers are times of low mobility of their sons (or, high persistence). Educational inequality is even more strongly associated with low mobility than status inequality, suggesting that education is an important channel of mobility even in this period. When educational inequality is high, the son advantage for any given father advantage is almost twice of what it is during times of low educational inequality. Thus, inequality in society appears to lead to conditions of less mobility, consistent with explanations that emphasize how persistent inequality can come about through family transmission of wealth and ability in the presence of credit constraints, neighborhood segregation, as well as localized beliefs and aspirations (Piketty 2000, Genicot and Ray 2017).

Second, not all inequality is associated with status persistence. To examine the role of families and social groups for the inequality-mobility relationship, I decompose the evolution of overall inequality into the components due to between-dynasty and within-dynasty inequality. The results show that while inequality is negatively related to mobility for between-dynasty inequality, mobility is positively related to within-dynasty inequality. Furthermore, among members of even closer kin segments in the dynasty, the positive relationship of inequality and mobility is stronger still. Thus, when inequality within close kin in the father's generation is high, the mobility in the next generation—of the son relative to his father—tends to be high. This positive and significant relationship is robust to the inclusion of various other determinants of mobility, providing evidence that dynastic groups help to boost the upward mobility of poorer members while reducing the likelihood that sons of high-status fathers will stay in the same high-status position. While I do not observe the actual transfers, these results nevertheless provide a means to assess implied differences in resource pooling within informal networks.

Third, I show that one mechanism through which social mobility is enhanced is spatial mobility. Over time social mobility was increasing, consistent with anecdotal evidence on China of changing institutions from the Yuan (1271-1368) through the Ming (1368-1644) and most of the Qing (1644-1911). The analysis begins by documenting a strong positive association between the distribution of father and son status that carries over across generations. I show that persistence in status across generations is relatively high; a

[4] One example is ethnic composition; a higher measured mobility in one country compared to another does not necessarily imply that the former provides more equal opportunities if it is also ethnically more homogeneous than the other country.

prominent great-great-grandfather can confer a sizable boost to his offspring's status even five generations afterwards, indicating that social mobility is a multigenerational process. Sons from high-status families in the 19th century, however, can expect a much larger drop in status compared to the drop such sons experienced in the 16th century, and there is corresponding evidence that low-status children rise faster in the 19th century compared to earlier periods. The increase in social mobility over time thus applied to both the top as well as the bottom of the status distribution.[5] Utilizing more than 200 residential locations, I show that geographic inequality—in terms of distance to the region's capital—further strengthens the negative relationship between status inequality and mobility, and times in which individuals migrate from where they grew up are times of relatively high social mobility.

My analysis provides new information on the role of inequality in parental investments for social mobility. Becker and Tomes (1979) place central importance on parental investments in shaping the opportunities of their children. Status determines the parental resource budget constraint for these investments. Human capital in the parent generation is central in the framework of Becker and Tomes (1986), also because education affects the effectiveness of skill transmission in parent-child interactions (see Heckman and Mosso 2014). Inequality in status, education, and other dimensions generally affect not only the distribution of opportunities but also cultural values and aspirations of future generations (Greif and Tabellini 2017, Genicot and Ray 2017). At the same time, when migration to new locations is undertaken as a human capital investment, spatial mobility may affect social mobility as well (Schultz 1961, Ferrie 2005, and Abramitzky and Boustan 2017).[6] Consistent with these arguments, I find that that status inequality, and especially educational inequality is strongly associated with persistence. My work extends existing perspectives in two ways. First, I show that geographic inequality, measured in terms of distance from a key location such as the capital, is important in sharpening the influence of status inequality. Second, although cooperatives and private groups beyond the nuclear family are known to be important (Townsend 1994; Altonji, Hayashi, Kotlikoff 1992; Griliches 1979), this is one of the first studies of the impact of the larger extended family on the inequality-mobility relationship.

Second, this paper contributes to uncovering the primary determinants of social mobility. While there is influential cross-sectional work across contemporaneous OECD countries as well as U.S. regions (including Corak 2013 and Chetty, Hendren, Kline, and Saez 2014, respectively), the promise of temporal analysis of mobility has been well-recognized, and in the absence of explicitly linked panel data researchers have made progress employing grouped data with pseudo-links (Card, DiNardo, and Estes 2000, Aaronson and Mazumder 2008, Clark 2014, and Guell, Rodriguez-Mora, and Telmer 2015).[7] By analyzing explicit links over many generations for a given population, this paper sheds new light on the main correlates of mobility and alleviates some of the concerns about the comparability of cross-country results.[8] Similar in

[5] Also, a 10-percentage point father advantage gives the son a 6-percentage point advantage around 1600 but only a 4-percentage point advantage in the 19th century.

[6] See also Hoen, Markussen, and Roed (2018).

[7] Surveys include Black and Devereux (2011), Piketty (2000), and Solon (2014, 2018).

[8] A low coefficient in a regression of son on father outcomes may not provide a good indicator of how close that society comes to providing "equal opportunity" compared to another society with a high coefficient; see Björklund and Jantti (1997), Han and Mulligan (2001).

spirit to my temporal analysis is work based on linked records among family members for a single country (Abramitzky, Boustan, and Eriksson 2012; Collins and Wanamaker 2017; Bailey, Cole, Henderson, and Massey 2017; Long and Ferrie 2013).[9] By studying a longer time horizon—up to twenty generations and over a period of roughly six hundred years—my analysis captures a comparatively long period. This is important because often changes in the patterns of intergenerational mobility take a long time to materialize.[10]

Third, my research contributes to our understanding of the evolution of social mobility over long periods of time, specifically in China.[11] While we know more about the historical trends in mobility for industrialized nations, the historical trends for developing countries are still largely a black box. I find evidence for relatively low social mobility until the 17th century and higher mobility by the 19th century. This differs from some results in the literature, e.g., Clark and Cummins (2014), but is consistent with the lifting of occupational barriers, access to education, and the overall development of the Chinese economy. My analysis complements Ho's (1967) classic study, which saw increases in mobility into the elite classes. This paper expands beyond the elites by incorporating a much broader part of the population: the bulk of my sample (around 70%) consists of commoners who possessed none of the characteristics typically associated with status.

The remainder of the paper is as follows. Section 2 provides the necessary background on social mobility in the society analyzed in this paper and presents initial evidence showing social mobility increased in China during the sample period. Information on the data sample, with its distinctions of social status as well as supplementary information (education, geographic location), is given in Section 3. Any data that can provide a micro-database for the pre-industrial past is worthy of investigation, and genealogies are one such source. Given that genealogical information is relatively new to the study of social mobility, however, this section provides a discussion of the reliability, representativeness, and potential biases of the source, with additional details given in the Appendix. Section 4 presents social mobility estimates, establishing that mobility has increased over time, and shows that this finding is robust to measurement error and changes in the meaning of status over time. Section 5 presents the main results of the paper, showing that inequality–and in particular educational inequality–is a key driver of social mobility. I document that while inequality generally favors persistence, there is an important role for resource pooling among family members; importantly, inequality among closely-related kin favors social mobility. Finally, Section 6 offers a concluding discussion, while readers interested in further extensions and robustness checks may refer to the Appendix.

[9] A pioneering contribution is Ferrie (1996). See also Lindahl, Palme, Sandgren Massih, and Sjogren (2015) for results on Sweden, and Braun and Stuhler (2018) on Germany.

[10] For example, mobility for U.S. children entering the labor market today is roughly the same as for those born in the 1970s (Chetty, Hendren, Kline, Saez, and Turner 2014); see also Chetty, Grusky, Hell, Hendren, Manduca, and Narang (2016).

[11] See also Barone and Mocetti (2016) who analyze mobility in Florence, Italy, using pseudo-intergenerational links.

## 2 Social Mobility in China - Historical Background

This study covers the period from the late 13th to the late 19th century. My sample corresponds to three imperial dynasties of China, part of the Yuan (1271-1368), all of the Ming (1368-1644), and the major part of the Qing Dynasty (1644-1911). Total population is estimated to have grown from about 100 million in year 1280 to 400 million by 1900 (Cao 2000; Maddison 2007). China's turn towards increased contact with Western countries took place during the 19th century, with some early signs of industrialization, but growth in income per capita does not take place until the very end of the sample period.[12]

China's economy was largely based on agriculture, as was true for most of the world during this period. Government administration followed a hierarchical structure from the emperor and upper ministries in Beijing down to provinces, prefectures, and counties, the outlines of which was inherited from previous dynasties. Thus, the Yuan Dynasty's rulers administered over China within the social and political structures inherited from the previous Song Dynasty (960-1260), a period during which many see a peak in Chinese economic development. While the governing structure varied, especially from the Yuan to the late imperial Ming-Qing period, the effective power of government was organized around an autocratic central authority and lower administrative regions.[13] The state taxed lightly in international comparisons with other states. Moreover, the scope of markets for allocating scarce resources was more strongly limited by technology (e.g., transport technology, financial instruments) than by government regulation, in part because effective enforcement would have required more state resources than were in fact allocated. Recent research sees China's per-capita income at 95% of that of England in the year 1400, 77% by 1600, and 30% by 1800 (Broadberry, Guan, and Li 2017, Table 6).[14]

In this model of governance by bureaucratic-scholar officials, education and literacy was a cornerstone of entry into the political elite and a key aspect of the social contract between the state and local management. Consistent with low central taxation, the capacity of the state was limited. The provision of public goods was instead delegated to local (provincial and below) governments and the elites of local dynasties (Shiue 2004, 2005). In the social contract between the central goverment and the bureaucratic-scholar officials, education and literacy was a cornerstone of entry into the political elite.

Local leadership was held by the elites who exercised paternal authority over their extended families (local dynasties in mirror to the imperial dynasties). At the same time, the participation of local elites in the political decision-making body of the state was legitimized through the state run civil service examination system and its curriculum. There thus arose, in effect, a partnership between the central state and local dynastic elites, to mutual advantage. Particularly after the early Ming, when tournament-style national civil service examinations were held regularly, the arrangement strengthened and ensured the legitimacy of the central state on the one hand, while allowing local elites to earn high financial rewards and enjoy high status on the other hand. The result was a kind of equilibrium based on distinctively Chinese characteristics and different from, for example, European forms of social organization (Mokyr 2002; Greif

[12] See Keller, Li, and Shiue (2013).

[13] A survey of the literature in comparative economic development in history is Brandt, Ma, and Rawski (2017).

[14] Naturally, due to incomplete data these figures come with a sizable margin of error. On the debate on the timing of the Great Divergence between China and Western Europe, see Pomeranz (2001, 2011); also Shiue and Keller (2007).

and Tabellini 2017).[15]

**The Role of the State**

There were effectively two complementary sources of governance. One was determined by the central state level, and another was determined at the local dynastic level. In terms of the laws that were enforced at the state level, the descriptive evidence suggests there were increasingly fewer institutionalized and legal barriers over time. Hereditary aristocracies were eliminated already sometime over the Tang dynasty (670-906 AD), thus sweeping away an institutionalized artifact of social stratification. Beginning in the Song dynasty, some members of the political elite and high-ranking officials of the state were selected on the basis of formal examinations, although other channels were also used.

An important type of inherited status related to the way in which the state taxed labor. During the Yuan Dynasty, the state extracted services from the civilian population, possibly because a labor market for such services was not viable. To enforce the obligations of labor service, the state required households to be registered in segregated occupational and status groups — commoners, artisan, soldier, salt producer, miner, scholar, astrologer, and many other categories. The Ming Dynasty carried over the practice of compulsory occupational status registration from the Yuan at the beginning of the dynasty, but by the latter half the family histories of some high ranking officials who came to prominence in the 15th century reveal backgrounds in the artisan or other common status groups, which provides an indication that the hereditary nature of the categories had started to break down sometime during the Ming (Ho 1967).

While aspects of the older practice of hereditary occupations were still present, social mobility appear to have become more fluid as the Ming started to become laxer (or less able) to strictly enforce the labor obligations and resorted more to paying wages for desired services. The biographies of the successful exam candidates of the Ming period show that the family backgrounds of the highest level *jin-shi*[16] degree holders included also people from those special occupational status designations, such as the soldiers, army officers, horse breeders, medical officials, official cooks, and others (see Li Zhou Wang 1746).

Other important changes took place during the Ming Dynasty that held implications for inequality and mobility. Sumptuary laws, which prevented commoners from taking the civil examinations, were eliminated in the later part of the Ming and would have permitted more mobility. It would appear that by the start of the Qing in 1644, the only types of hereditary privileges and automatic status that remained belonged to the imperial dynasty, where the throne was passed from the emperor to one of his sons, and the leading families of the Eight-Banner system.[17] The Qing Dynasty also discontinued the Ming practice of family status registration, by which time there were no effective legal barriers to social mobility due to occupational status of the family (Ho 1967). Social status in China over time relied to a far greater degree on investments made by each generation on behalf of the next, increasing use of labor markets, and declining legal restrictions. These institutional changes would likely have increased mobility.

[15] See also Hajnal (1965), De Moor and van Zanden (2010), and Dennison and Ogilvie (2014).

[16] *Jin-shi* is the title of a man who passed the highest-level of China's civil service exam, the entrance way to top-level government jobs.

[17] The Eight-Banner system was an exclusive hereditary institution that dominated certain military and command functions of the Qing state, and banner families were eligible for special privileges.

As the formal and institutionalized barriers that prevented commoners from rising across social classes fell away, institutions governing the higher ranks of status also gradually changed over time. The new institutions, which involved relying more heavily on the civil service examinations to test a candidate's knowledge of classical education, were a central aspect of governance until the last years of the Qing dynasty (1644-1911). Individual merit earned through examinations and official titles coincided precisely with the Ming and Qing definition of social status, and moreover, they are the most important source of status in society at this time. Lower-level purchased degrees could be obtained by those who had some landed property, or wealth from commerce or trade. However, social status derived from passing the civil service examinations was the "ultimate source of power" Ho (1967, p. 51) from the beginning of the Ming onward. The ranking of degrees thus provides a close proxy to the status of the person holding the degree in question.

The *sheng-yuan* degree was the lowest of the recognized categories of government education, conferred upon those who had passed the district or prefectural degree threshold. The *sheng-yuan* who were scholastically more competent were awarded with the *gong-sheng*, "imperial student" title; above them in rank were the *ju-ren* (graduate of the provincial examinations), and above the *ju-ren* were the *jin-shi* (graduate of the national metropolitan examinations). There were no age requirements or limitations for advancement, but since the examinations required a high level of literacy and years of study, the earliest that one could attain the *jin-shi* degree would be in the early twenties, and it was not unheard of for a man in his fifties to still be a *sheng-yuan*. Not all *sheng-yuan* advanced to the next levels, and those who didn't may have given up and turned instead to working for officials in a secretarial capacity, or, helping to manage local affairs—settling disputes, organizing local public goods projects, improving welfare and security interests, or providing education in their community (Chang 1962). Over the Qing period, the ratio of the *sheng-yuan* relative to the higher level *jin-shi* likely rose, although there is limited information on the impact of such trends in the existing literature. In the empirical section, I examine what happens if assumptions about the rankings of status change over time.

For those who gained official appointments, the position permitted a substantial accumulation of wealth. Although the official salary was much more than the income of an average person, the official salaries were not extraordinarily high. In the Qing, a district magistrate would have received as official salary around 45 *taels* (silver dollars), and a governor would have earned 150 *taels* (based on Chang 1962). However, the official salary was only a small portion of the income of officials. The customary bonuses (sourced from local taxes, contributions, or gifts) of the district magistrate ranged from 1,000 to 1,800 *taels*, while a governor of a province might have received 12,000 *taels*. In addition, there were allowances to cover administrative expenses. These funds helped to pay for the officials' personal staff and other governing expenditures, but in a period when a common laborer earned 5-10 *taels* a year, these were still large amounts.

Most of the personal wealth of officials came from these extra sources of income, although there are no precise records of the actual figures of this extra income. In particular, the scholar-official was expected to use some of his income for charitable aims and public purposes — which effectively meant making sizeable financial contributions to schools, construction projects, and other public works. While officials

were not allowed to govern their own hometown regions, they would typically return to their hometown at retirement to exercise their considerable influence, prestige, and wealth in relation to the official who was in charge of governing. Even those who did not hold office, but were nonetheless educated, could provide valuable managerial services in these ways.

Throughout the Ming-Qing period, participation in the state exams and demonstration of literacy at the highest levels was the most respectable path towards high rank.[18] In 1371, national civil service examinations became more routine in nature, and over time, regional quotas for the various degrees also became less discretionary. Most of the observations in my sample are in the period after 1371. With this income obtained through their status, officials were able to uphold relatively high standards of living, contribute to local community projects, and make investments in landed property in their home localities. These returns were so attractive from a political status and economic point of view that wealthy merchants tried to invest in their sons' educations in order to climb the social ladder.

Anecdotal evidence of upward mobility can be found in biographies of high-level officials. Wu Chung-liang of Huizhou, who was awarded his civil service degree in 1593 and eventually became a magistrate of the Ming period, came from a blacksmith family. Other biographies show that a number of successful candidates were sons of weavers or merchants of the Yangzi Delta. A few cases delineate humble occupational beginnings: the grandfather of Shang Lo, a prime minister from the 15th century who ranked first in the national palace examinations in his cohort, earned his living through hunting and gathering firewood.

The question of the actual extent of mobility is open to empirical analysis, since the lack of institutional barriers does not necessarily imply that commoners were more likely to move up in status; or that upper status sons were in danger of losing the status of the family in which they were born. Given the fact that titles could not be inherited, downward mobility was a real possibility. The sons of a ranking official would have been part of the leisure class, and wealth and land could be inherited. But over time, the family's status would certainly have eroded if the descendants from later generations could not obtain the degrees and titles that would prevent sliding downward in the mobility ranks. Also here, anecdotal case studies have vividly illustrated the dramatic fall in household wealth among those descendants of famous officials who failed to earn any titles.

Although these are only a few examples, they each point to the decline in formal occupational barriers over time. Of course, both genetics as well as resources were still heritable, even if actual titles were not. The extent of mobility would have depended not only upon institutional barriers and their changes, but also incentives for investing in future generations' human capital, and change in those incentives over time in China.[19]

[18] During times of revenue shortage, the state also offered minor titles for sale. However, purchased degrees during the Qing did not entitle the title-holder to any of the high-ranking positions. My analysis below distinguishes purchased from non-purchased degrees.

[19] Shiue (2017) finds a strong negative relationship between child quality and quantity at times of high returns in the 17th and 18th century that faded away with the subsequent lower return to human capital in later years.

**Dynastic Rules**

The dynasty as an organization form of local society was present over the entire period of the analysis and played a key role in Chinese society.[20] Family or clan laws were written documents in genealogies and were generally guided by state laws.[21] The rituals of ancestral worship helped to unify the common ties and bonds of the group, but more than that, the written rules of behavior had the dual effect of increasing trust and social order within the group on the one hand, while on the other hand reducing costs of enforcement when interventions by the dynasty elders became necessary. While the emphasis was on moral persuasion, punishments for misbehavior— such as by flogging — were set out in these family laws as well. The rules were set out by custom and tend to share broad similarities across dynasties, although many differences in wording and specifics can also be identified.

Importantly, these family laws were not just empty ritual but an essential part of what made the dynastic organization a means of insurance and resource sharing. These laws and regulations applied to a wide set of concerns with respect to common property, awards, theft, poor relief, and social relationship within the clan as well as towards non-clan individuals (Liu 1959). Similar to how progressive government policies redistribute resources, family laws had the effect of creating a range of progressive dynastic policies that applied to members of the dynasty.

Given the lack of publically funded education at the time, the bulk of education was undertaken at the lineage level, through lineage schools or hired tutors. Dynasties that had more resources, in particular, often set aside common lands or dynastic funds for the education of the children of poorer kin as well as the fees and expenses associated with taking the examinations. Clans also customarily used their common funds and school lands for rewards and support of its students. According to Liu (1959):

> "The greatest emphasis is placed, however, upon the honoring of successful scholars and promising aspirants as well as assisting them in the consecutive stages of their career, as the public honors and prestige they gain are shared by the clan group".

The amounts varied according to clan wealth, but the range given to those who successfully earned degrees ranged from 2-10 *taels* for the first degree to 80 *taels* for the highest degree, which was a substantial sum.

In addition, most rules support the financial investment in education of its students while they studied. The genealogy of the Yeh dynasty states: "Our clan has long remained undistinguished, since few members ever become officials or degree holders. We should therefore give in order to help promising members study." One clan held semi-monthly essay contests at its ancestral hall to encourage study, while another set up rules to support examination fees; yet another clan decided to give priority of free schooling to the orphans and poorest members in the clan (Liu 1959).

Dynasties with school lands used rents to support students. One such clan stipulated that half of the rent would be used for teacher's salary and the other half to scholarships (split evenly between poor

[20] Dynasties and genealogies were prevalent in the 20th century until the communist party under Mao Zedong suppressed paternalistic sources of authority (Yang 1959a, 1959b).

[21] One clan rule reads (translation from Liu 1959): "Clan rules rely on the law of the state as their guide. The law, in turn, depends on clan rules to supplement it. In comparison, clan rules are more lenient than the law. But whoever violates clan rules will eventually find himself in violation of the law."

members and the descendants of the original donor of the land).[22] These strategies of upward mobility could potentially pay off if a member succeeded in the civil service examinations. Wealth would have meant better access to tutors, but kinship networks and connections among the dynasty in high status positions would have helped to solidify personal advantages for other dynasty members.

Anecdotal evidence paints a picture of how the rewards of high status was beneficial for the larger family entity. Chang (1962) documents officeholders giving up their share of the family inheritance to their brothers, or, in other cases, paying off debts. Although it is difficult to give exact magnitudes, the prestige of high income carried with it an expectation of charitable philanthropy to a larger community through financial contributions to schools, organizations, the welfare of the poor and widows. Indeed, the genealogy itself is seen by some historians and anthropologists as a document demarcating the members in the corporate group who are entitled to have access to dynasty lands, schools, and other resources (Watson 1982).

While there is little controversy that dynasties were an important underpinning of society, with the notable exception of de la Croix, Doepke, and Mokyr (2018) little is known at present about the economic impact of dynasties, but on the whole different dynasties would have been likely to face similar choices and constraints, conditional on resources. Below, the observable heterogeneity in dynasty characteristics is incorporated in the empirical analysis.

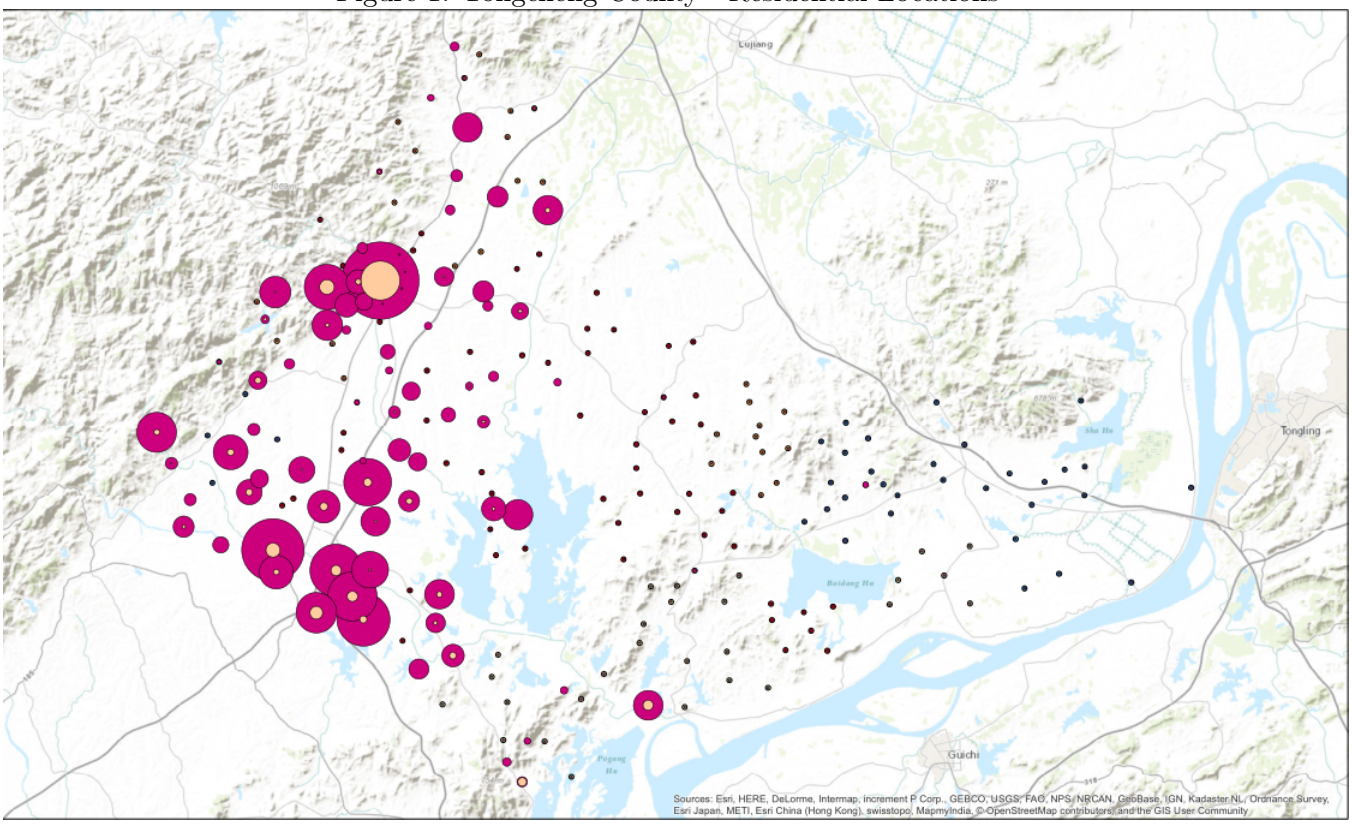
## 3 Characteristics of the Data and Summary Statistics

The data of this paper comes from genealogies of individuals and households who lived in Tongcheng County of Anhui Province. Tongcheng County is approximately 30 miles by 60 miles, about 150 miles from Nanjing, the early Ming Dynasty capital, and 650 miles from Beijing, the later Ming and Qing capital. Figure 1 shows a map of Tongcheng county. The county is located just north of the Yangzi River, on the lower right in Figure 1, about 300 miles inland from the coast of the East China Sea. Shown as well are the location of villages and towns in where the members of the seven Tongcheng dynasties resided. The lighter color shading indicates the number of men having acquired a substantial amount of human capital, an important aspect of status. The central town of the county, Tongcheng city, is the largest point, located in the upper left of Figure 1. Tongcheng county had about 1.3 million inhabitants in the year 1790 (Beattie 1979).[23] For comparison, this was similar to the combined population of the newly independent states of Massachusetts, Connecticut, and New York at this time.[24] The county was representative of the more developed and densely settled regions of China, one of the many thriving economic regions near the lower Yangzi. The region was mainly a rice-producing area where the wealthiest families were typically landowning gentry. Over the Ming and Qing Dynasties, the region gained some fame for having produced a number of the highest officials of the empire.

The dataset is created from genealogies of seven dynasties of Tongcheng County. The uses and reasons

[22] According to a survey of the awards of the School and Ritual Land in genealogies by Liu (1959).
[23] Beattie (1979) is a good source on the history of Tongcheng.
[24] Figure including slaves; https://www.census.gov/library/publications/1793/dec/number-of-persons.html

Figure 1: Tongcheng County - Residential Locations




**Notes**: The frequency of observations for a village, 1300-1900, is proportional to the point's radius. Pink (darker) is the total observations while tan (lighter) reflects the number of educated.

for why the genealogies were kept suggest that it would have been in the interests of members to have an accurate record of their members. The tradition of compiling genealogies began around the time of the Song Dynasty (960-1279) (Zhao 1997). Although today, many of these genealogies are in libraries, originally the genealogy was a valuable private record, stored for safekeeping in the hometown of the family — in the home of an elder or in the ancestral halls — they were never meant for public exhibition. Genealogies could be compiled and updated by the literate members of the dynasty, but government genealogical bureaus also assisted with these records (Pan 1929, Telford et al. 1983, Zhao 2001). It originated in part as a critical aspect of ancestral worship, explaining why some types of information such as burial locations of the deceased are carefully recorded in genealogies. The genealogy also defined the membership of the dynasty, family rules of conduct, and played an important role in acknowledging the individuals who were entitled to the lands, the schools and other resources of the dynasty. Thus, genealogies had economic and political functions above and beyond ancestral worship.[25]

Typically, genealogies start with the progenitor of the dynasty from which all following dynasty members descend. The members were related by birth or by marriage, where not only women, but men sometimes married into a dynasty and adopted the surname of their spouse's family.[26] According to the genealogical principles of compilation, all male members were eligible for inclusion. In the Tongcheng genealogies, the dynasties' progenitor is recorded usually in the 14th century, with the earliest date being the year 1298. These genealogies cover typically 18 consecutive generations, with a maximum of 21. The latest death recorded in the data set is 1925.[27]

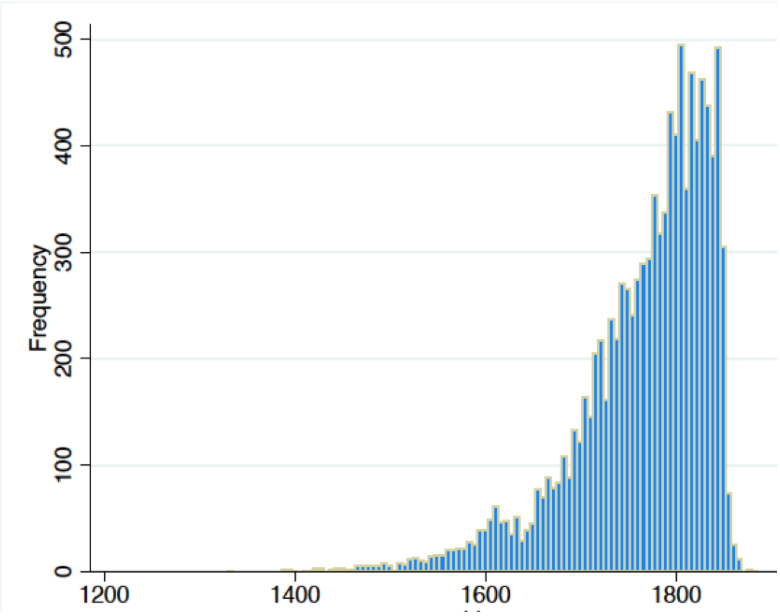
Generally, the coverage of genealogies at the turn to and into the 20th century becomes patchy. While my sample covers part of the Yuan and the Ming dynasties the large majority of observations are for the Qing. Figure 2 shows a histogram of the data over time for the population of all seven local dynasties in the Tongcheng genealogies by birth year. Note the transition from Ming to Qing around 1644 results in a dip in population. The starting date depends on the birth date of the progenitor, but all seven dynasties are present throughout the sample period (Figure A.1 in the Appendix shows the relative size of each of the dynasties by sub-period). Genealogies differ from census data and other administrative data that records information at a certain point in time. In the ideal case, the information in the census gives a complete representation of all strata of the population. To study intergenerational mobility with census information, an individual must be observed in one census as a child living with his or her parents, and then in a second census when the child is old enough to have his or her own status level. Thus, records from different cross-sectional observations have to be then linked, which can be difficult and therefore accuracy and sample attrition are critical concerns (Bailey, Cole, Henderson, and Massey 2017). Chinese genealogies, in contrast, are conceptually similar to pedigree charts. Identifying the same individual in different cross-sections is unnecessary. One advantage of these data is that genealogical biographies are summaries that record the highest lifetime position and other accomplishments for each individual. In this sense, the data immediately captures intergenerational outcomes. On the other hand, the retrospective

[25] See Liu (1978), Telford (1986) and Telford et al. (1983) for surveys of the content and scope of Chinese genealogies.

[26] On the other hand, there is the omission or undercounting of infants and children who died early, criminals, and daughters.

[27] The Tongcheng genealogies are by no means unique in the length of the period covered; Fei and Liu (1982), for example, examine ten dynasties over the period of 1400 to 1900.

Figure 2: Entries in the Tongcheng Genealogies by Birth Year




**Notes**: Figure shows number of observations by birth year. Note the dynastic change around 1644.

and non-administrative nature of the compilation may lead to data problems. The following discusses these issues.

### 3.1 Data in Genealogies

The remainder of this section explains how the textual biographies in the genealogies were translated into measures of socioeconomic rank, and of the data thus created. Below, I show that while each dynasty on its own may not be represent the average, the aggregation of the seven dynasties together is very close to being representative of the population.

Historical demographers have classified genealogies into three general types: branch, lineage, and clan genealogies.[28] The "branch" genealogies lack generational depth and are therefore not as useful for mobility studies. They tend to include only a small number of people living in one location, and although there are many branch genealogies in existence, these give an incomplete picture of the larger lineage.[29] "Lineage" genealogies have very significant generational depth, and are multi-volume records that give records of sometimes thousands of individuals, including vital statistics and short biographies, of multiple patrilines in the lineage who resided over nearby villages within a single county. "Clan" genealogies are heavily edited lineage genealogies with tens or hundreds of volumes, and include thousands of individuals, sometimes across multiple provinces. While large in geographical scope, a drawback of these clan genealogies is that they are heavily edited and tend to omit the level of individual detail that would be essential for studying mobility.

[28]The following classification is based on the survey in Telford (1986).

[29]These branch records, often handwritten, may be the primary source material for the lineage genealogies.

The Tongcheng genealogies were chosen because they are among those mid-sized genealogies that include people in a county and contain a certain level of demographic and socioeconomic detail with a relatively high quality of compilation (Telford 1995). With 40,000 individuals including wives and children, the number of people is sizable although not as large as some genealogies covering populations in an entire province. There is information on about 9,800 individual men in the sample. As mentioned previously, Tongcheng's population in the year 1790 was around 1.3 million, and I have roughly a 1.5% sample of the population.[30] The biographical entry includes information on the major stages of each man's life, evidence of wealth, various aspects of elevated status (such as being the main guest at the county banquet), temporary functions (such as village head), and specific actions (such as large donations or setting up ancestral estates). There is also some information on a person's occupation, his education, and whether he was a government official (and at which rank). The entry also lists the man's wife (or wives) as well as each couples' children.

### 3.2 Status Coding

Genealogical entries summarize the most noteworthy and highest achievements attained over a person's lifetime. The kinds of achievements mentioned can be classified into twenty-three categories from status 0 (no status) to status 22 (highest status). These status categories are drawn from the meticulous work by Telford (1986, 1992) for the Ming, and are in turn based on generally agreed upon rankings of education and status in Chang (1955), Eberhard (1962), and Ho (1962) on Qing society. Descriptions of the twenty-three status categories and the frequency with which they occur are given in Table 1 (columns 2 to 5). It is worth noting that the biographical entries of the genealogies are undirected in nature and do not provide all dimensions of information for all individuals. Thus, we do not know if, for example, a minor civil official (status 13 in Table 1, column 2) ever set up large donations (status 8 in Table 1, column 2).

Given the fact that the highest and most noteworthy achievements in a person's lifetime are recorded in the biography, however, if we can rank these noteworthy achievements then we can obtain a reasonably accurate ranking of status across all individuals. Thus, if there is nothing other than vital statistics in the individual's biography, and this person had no titles, degrees, or noteworthy evidence of wealth, then he is coded with status 0 (see Table 1). Likewise, if we observe an individual married more than once but had no other evidence of status higher than having married more than once, then that person would be given a status 2 (in Table 1, column 2).

Table 1 indicates that this data covers a relatively large fraction (71%) of low-status persons (status level for the lowest category was coded 0). This shows that genealogies were not only or primarily records of the lives of the elites. At the same time, it is evident that status gradations are captured in more detail towards the high end of the status distribution, if only because among the top classes status differences are relatively easily identified (for example, by particular titles or positions). Therefore, I aggregate the 23 status levels to 6 levels in the baseline analysis for two reasons. First, to avoid excessive detail in

[30] I observe about 3,600 men that would be alive in the year 1790 in my sample. These men had more than 4,200 wives, and the data records more than 7,500 sons and 4,100 daughters, for a total of just under 20,000 persons. My main sample are N = 8,893 observations on sons linked to their fathers and grandfathers.

some of the sections of the status distribution; and second, because the status definition is hybrid in nature in the sense that it combines elements that are unambiguously rankable (such as the level of an individual's government position), with information that is more difficult to rank (such as how a wealthy merchant compares with an individual who purchased a minor office, or how a person who was a village head compares to someone who married more than once). Over the 6 status categories, however, the rankings are unambiguous. The 23 and the 6 status levels are both shown together with their frequency in Table 1, in columns 1 and 2.

### 3.3 Converting Status to Rank

Several points are worth highlighting. First, because there is no monetary measure of status, the typical log income of son on log income of father specification employed in intergenerational mobility studies cannot be adopted.[31] Following Dahl and DeLeire (2008) and Chetty, Hendren, Kline, and Saez (2014) I instead employ the position of rank percentile of father and son in their respective status distributions. Second, since the goal is to obtain the highest lifetime status, it is not necessarily crucial that the biographical entries do not cover all dimensions for all individuals as long as we know the maximum. For example, if someone was a *jin-shi* (status levels 20 to 22) information on whether he also was village head at some point is unnecessary because being village head could not possibly raise his social status.

Third, there may be some uncertainty about a person's lifetime status. For one, status is observed only up to a relatively small number of discrete categories (namely, 23 in the raw data and 6 in my baseline status definition). While information at a more granular level would be desirable, discrete status levels are conceptually similar to studies with income data when individuals have the same income (in particular, zero income). Related to this, the sons of ranking officials are given a small raise in rank even if there are no other attainments. This is a historically accurate description of this society, but may raise concerns. In Section 4.2 (Table 8), I show results from recoding the status of the sons of officials to zero in order to examine the robustness of this and other results to errors in coding.

As in Chetty, Hendren, Kline, and Saez (2014), I assign all individuals of a given status the mid-point of the percentile rank in the status distribution. For example, when in a particular birth cohort 70% of all men have status level zero (the lowest status level), each one of them is assigned the percentile rank of 0.35. Finally, there may still be uncertainty about a person's relative status, based on the status category to which he was assigned. This can be thought of as measurement error. It is different from measurement error in the canonical intergenerational mobility study which results because income data shows earnings at the stages of the lifecycle, when the data was collected rather than complete lifetime outcomes (Solon 1999). Here, the question is rather, for example, whether a person with a purchased office indeed has a one level lower status than a student of the Imperial Academy (levels 11 and 12, respectively, see Table 1). It is also unclear whether status levels were constant over the relatively long sample period. To address these issues, below I examine a range of alternative status definitions. Furthermore, I will explore the influence of various degrees (and directions) of measurement error for the results (Section 4).

[31] As noted above, data on remuneration in Chang (1962) and other sources is limited to particular official positions; there is no income information for most of the social strata, and there is no systematic information at the individual level.

Table 1: Social Status - Description and Distribution

| (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|
| Status | Status-23 | N | % of Sample | Description | Educated |
| 0 | 0 | 6,320 | 71.08 | No title, degree, and evidence of wealth | 0 |
| 1 | 1 | 35 | 0.39 | Honorary or posthumous title; village head; main guest at the county banquet | 0 |
| 1 | 2 | 741 | 8.33 | Multiple wives in consecutive marriage (two or more not living at the same time) | 0 |
| 1 | 3 | 824 | 9.27 | Father a *sheng-yuan*, minor official, or official student; or evidence of wealth, | 0 |
| | | | | *jian-sheng*, expectant official | |
| 1 | 4 | 20 | 0.22 | Grandfather a *ju-ren*, *gong-sheng*, *jin-shi*, or official | 0 |
| 1 | 5 | 31 | 0.35 | Father a *ju-ren*, *gong-sheng*, *jin-shi*, or official | 0 |
| 1 | 6 | 145 | 1.63 | Educated, scholar, no degrees or office; editor of genealogy; | 1 |
| | | | | refused office, or prepared but did not pass exam | |
| 2 | 7 | 79 | 0.89 | Concubinage (i.e. polygyny, two or more wives or concubines at the same time) | 0 |
| 2 | 8 | 11 | 0.12 | Substantial evidence of wealth and property; set up ancestral estates, | 0 |
| | | | | large donations, philantrophy; wealthy farmer, landowner, or merchant | |
| 2 | 9 | 163 | 1.83 | Official student | 1 |
| 2 | 10 | 1 | 0.01 | Military *sheng-yuan*, minor military office | 0 |
| 3 | 11 | 133 | 1.50 | Purchased *jian-sheng* and/or purchased office | 0 |
| 3 | 12 | 93 | 1.05 | Student of the Imperial Academy (non-purchased) | 1 |
| 3 | 13 | 48 | 0.54 | Civil *sheng-yuan*; minor civil office | 1 |
| 3 | 14 | 95 | 1.07 | Expectant official; no degrees | 0 |
| 3 | 15 | 4 | 0.04 | Expectant official one of the lower degrees | 1 |
| 3 | 16 | 23 | 0.26 | Military *ju-ren*, *jin-shi*; major military office | 1 |
| 4 | 17 | 38 | 0.43 | Civil official with no degree, minor degree, or purchased degree | 0 |
| 4 | 18 | 23 | 0.26 | *ju-ren*, *gong-sheng*, with no office | 1 |
| 4 | 19 | 47 | 0.53 | *ju-ren*, *gong-sheng*; with expectant office | 1 |
| 4 | 20 | 0 | 0.00 | *jin-shi*, no office | 1 |
| 5 | 21 | 11 | 0.21 | *jin-shi* with official provincial post or expectant official | 1 |
| 5 | 22 | 7 | 0.08 | *jin-shi* with top-level position in Imperial bureaucracy | 1 |
| | | | | (Hanlin Academy, Grand Secretariat, Five Boards, Prime Minister) | 1 |

**Notes**: Table gives frequency of socioeconomic status of son (N = 8,892). Sample is all married males linked over three generations (son, father, and grandfather). Baseline status definition with six levels in first column; twenty-three status levels, based on Telford (1986, 1992), Chang (1955), Ho (1962), and Eberhard (1962), given in second column.

In addition to information on lifetime socio-economic status, the data provides relatively complete information on birth year and month for each man, together with his dynasty, segment of dynasty, and generation in the dynasty. Within each dynasty, there are branches of the family tree, called segments. Everyone belonged to both a dynasty and a segment within the dynasty; among the seven dynasties, I observe 39 segments. Furthermore, there is information on each married man's wife (or wives), children, and other identifiers. Recorded by couple in the raw data, the links between generations can be established with virtual certainty because male children re-appear in the genealogy as married men. The status levels of a father and son pair are directly available from the created data. Every person is listed as a son of a specific man who obtained a certain lifetime status. Later in the genealogy the son—with his name, dynasty, generation, and vitals—is listed again as a married man with his own lifetime status.[32] Since the genealogy contains information on male dynasty members as children irrespective of whether they grow up to be married, I am able also to assess the role for mobility of males who drop out of the sample. Among the reasons for dropping out include the consequences of early death, outmigration, and, quantitatively most important, men who stay single throughout their lives; Sections 4 and 5 below analyze these factors in more detail.

### 3.4 Summary Statistics and Validity

Summary statistics on status and basic demographic characteristics for about 8,900 father-son pairs are given in Table 2.[33] The birth year of the first father in the data set is 1298, as noted above, while the latest birth year of a son is in the year 1885. The earliest-born mother in the Tongcheng genealogies is recorded for the year 1300, only two years different from the earliest-born father. Table 2 lists summary statistics for the men's birth and death month; the average for both is close to 6.5, which is what one might expect if the timing of births and deaths are random events and there is no artificial age heaping.[34] Table 2 also shows a total of three brothers in the average household. The number of wives averages 1.17, reflecting the fact that the great majority of men married only once. Multiple marriages typically took place sequentially after the first wife died, and polygyny was rare (seen only in a very small fraction of households).

Recall that the Tongcheng genealogies are privately-compiled records of specific social groups, namely histories of local dynasties. It is natural, thus, to be concerned about representativeness and external validity of the data. Furthermore, it is important to ask about the reliability of the information, as well as to examine the presence or absence of various biases, including recall, success, and survivor biases. I turn to these issues now.

Dates given in genealogies are not in numerical format but in terms of traditional calendrical symbols, which are difficult to falsify (Zhao 1997). This increases the reliability of the recorded dates, which are key to computing longevity from vital statistics. Longevity as calculated from the vital statistics on birth year

[32]Vital statistics (birth year and month, death year and month), birth order, dynasty generation and status uniquely identify 99.4% of all men in the data.

[33]There are 9,787 individual men in the Tongcheng genealogies for whom I have information on father status; lagging by one generation to identify grandfather status brings the number of observations down to just under 8,900.

[34]Information according to Chinese lunar months has been concorded to solar months following Shiue (2002).

Table 2: Summary Statistics

| | | N | Average | Std. Dev. |
|---|---|---|---|---|
| Father | Status | 8,893 | 2.20 | 4.17 |
| | Year of Birth | 8,893 | 1732.04 | 70.85 |
| | Month of Birth | 8,893 | 6.92 | 3.49 |
| | Year of Death | 8,658 | 1787.44 | 70.59 |
| | Month of Death | 8,658 | 6.47 | 3.43 |
| Mother | Year of Birth | 8,893 | 1735.63 | 70.99 |
| | Year of Death | 8,254 | 1789.47 | 71.62 |
| Son | Status | 8,893 | 1.63 | 3.64 |
| | Year of Birth | 8,893 | 1763.90 | 71.30 |
| | No. of Brothers | 8.893 | 3.28 | 1.62 |
| | No. of Wives | 8,891 | 1.17 | 0.45 |

**Notes:** Status levels 0 to 22, as shown in Table 1, column 2. Number of Brothers includes son himself.

and death year can therefore be used to provide a check on the information on status in the genealogies. If status is correlated with income and wealth, then a person's longevity should be correlated with status because since higher income and wealth should provide better living conditions.

This is confirmed in Figure 3, which plots the longevity of adult married men against their status: low-status men lived typically about 56 years, compared to high-status men who typically reached 63 years.[35] The finding that high status is associated with higher life expectancy provides a straightforward validation of the status definitions.[36] Age specific mortality rates provides another point of comparison. Population figures at the regional level are typically based on gazetteers, which are local histories about a certain place.[37] In addition, there are official accounts for subsets of the population, such as the Qing population registers, which are the product of the Eight Banner registration system.[38] Telford (1990) compares demographic patterns in the Tongcheng genealogical data and the Banner populations for 1774 to 1873, when the latter starts to become available. He finds a very similar variation in the probability of dying for different age categories across the two sources (see Telford 1990, Figure 2).[39]

Next, we can compare the fraction of people in different status classes in my sample with other available estimates in the literature based on larger populations. These published estimates are themselves based on rough estimates, since there does not exist accurate population head-counts at a regular frequency for

[35]The average of age at death is in the high fifties, which is higher than average life expectancy at birth in China during this period because the baseline sample excludes both males who died at young age and those who remained single. Section 4.2 below presents mobility results that include single men.

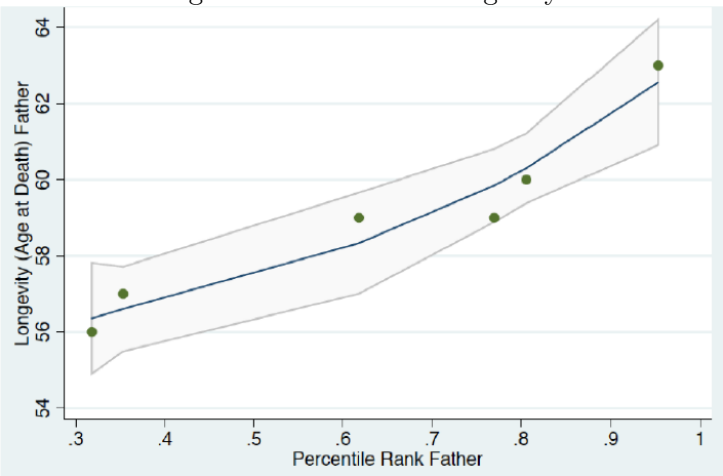
[36]These are statements about averages; the person with the highest lifespan in the sample, 91 years, had status zero.

[37]Three county-level gazetteers about Tongcheng cover the period under analysis: Tongcheng xian zhi (1490), Tongcheng xian zhi (1696), Tongcheng xuxiu xian zhi (1827).

[38]These data are available for areas in China's northeast, in today's Liaoning and Heilongjiang Provinces, these lands were organized under the Imperial Household Agency and the Jilin Military Yamen, an office in the General Office of the Eight Banner Command. See https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/265. For the imperial household dynasty, there are observations going back to the seventeenth century (Lee et al. 1993).

[39]Campbell and Lee (2002) compare data from genealogies of Liaoning to the household registers. They find evidence of higher mortality rates in the genealogies compared to what was reported in the registers. This contradicts the idea that mortality in underreported in the genealogical sample. It also not what one would expect if more privileged and educated men would be more likely included in the genealogies.

Figure 3: Status and Longevity




**Notes**: Figure shows median percentile rank and longevity for all men for whom there is information on birth and death year; 90% confidence interval shaded.

China as a whole, but it may nevertheless be informative to see if the genealogical data of Tongcheng is within a reasonable range of the estimates in the historical literature.

Much of the available estimates focuses on the upper classes as a percent of the population. While there is no consensus on who should be considered to belong to the upper classes, there is wide agreement that education and success in the civil service examinations were important components. Chang (1955) takes the view that *sheng-yuan* holders and above were in the upper class, and estimates that they were in the top 2% of the total population in the later half Qing period. In my analysis, the part of the population corresponding to Chang's (1955) definition are status levels 13 and above (Table 1); they account for 3.3% of the sample, which is quite comparable. Fei (1946) presents another, wider estimate of the upper classes, which he placed at 20%. In my analysis, status levels 2 to 22 correspond to Fei's definition of high status—and the share of these groups in my sample is 20.2%. Both these comparisons suggest that the Tongcheng genealogies are fairly representative of China's population as a whole with respect to the size of top status groups as well as the relative size of higher versus commoner classes.

Apart from status, one can check if the percentage of successful examinees in the Tongcheng genealogies broadly lines up with national averages. The most systematic evidence on education in China during Ming-Qing is related to the civil service examinations. In particular, the number of *sheng-yuan*, the individuals that passed the initial state examination, was about 500,000 in the year 1700 (Elman 2000), or roughly 0.3% of the population. In the Tongcheng sample, about 0.76% of the men around the year 1700 were licentiates. Accounting for women, children, and elderly indicates that the fraction of licentiates in Tongcheng was similar, or perhaps somewhat lower than in China as a whole.

In summary, overall comparisons of percentages of degree holders, upper class categories, and mortality rates suggest that the status distribution in the Tongcheng sample is not very different from what one

Table 3: Statistics by Dynasty

| | N | Status | | Distance to Tongcheng city |
|---|---|---|---|---|
| | | Average | Std. Dev. | Average (km) |
| Chen | 291 | 0.43 | 1.65 | 61.31 |
| Ma | 627 | 6.95 | 6.35 | 22.98 |
| Wang | 4,682 | 1.50 | 3.10 | 77.87 |
| Ye | 1,607 | 4.00 | 5.44 | 48.28 |
| Yin | 604 | 1.15 | 3.04 | 69.40 |
| Zhao | 769 | 0.91 | 2.31 | 125.28 |
| Zhou | 314 | 0.72 | 2.25 | 54.44 |
| All | 8,893 | 2.20 | 4.17 | 70.99 |

**Notes**: Status of the father, based on twenty-three levels (Table 1, column 2). The last column is the distance of each man's residential location to the county capital (see Figure 1).

might expect from a randomly drawn sample. Genealogies were compiled for ancestral rituals, so there would not be an obvious incentive to systematically create false entries, but nevertheless, one may be concerned that achievements themselves were falsified. To check for this, I compare the list of people who claimed to have obtained the *jin-shi* degree against other known lists of degree holders from Tongcheng County (Fang 2010; Cao 2016; Wang 2017). There were over 51,000 *jin-shi* degree holders from the Yuan, Ming, and Qing that are part of the records of the state. Information on top degree holders can be cross-checked for accuracy by referring to known lists of *jin-shi* degree holders from the Chinese state, which give the name, the date on which someone received his degree, and his hometown. I have verified that the information on the 18 *jin-shi* in my Tongcheng sample is in line with the information of these official lists. This provides additional evidence that supports the reliability and accuracy of the genealogical records.

### 3.4.1 Other Checks on Internal Consistency of the Data and External Validity

Another useful aspect for assessing the data are differences across dynasties. Table 3 presents statistics separately for each of the seven local dynasties, showing substantial differences between the dynasties, not only in size (with the Wang being most populous), but also in terms of social status.

Specifically, the Ma and the Ye dynasties tend to have relatively high status. Table 3 also provides information of the geographic distance for each dynasty member to the capital of the county, Tongcheng city. There are significant differences between the dynasty in the typical distance of their residence from the main city. Dynasties whose members have relatively high status tend to be located relatively close to the capital of the region. In particular, 51% of the members of the Ma dynasty live in Tongcheng city, compared to only 0.8% of the Zhao dynasty members. Geographic location is an important aspect of status that I will return to in section 5 below. This stark variation across dynasties can be employed to shed light on the size of several forms of selection and biases in the data that might affect my mobility estimates. For example, I ask whether times in which the members of a particular dynasty have relatively high status are also times when this dynasty accounts for a relatively high share in the sample—the answer

Table 4: Intergenerational Transitions between Status

| | | Father Status | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | Total |
| Son Status | 0 | 86.74 | 66.47 | 0 | 0 | 0 | 0 | 71.07 |
| | 1 | 11.39 | 23.95 | 66.33 | 61.43 | 28.12 | 10.71 | 20.20 |
| | 2 | 1.04 | 2.87 | 13.42 | 9.89 | 17.97 | 7.14 | 2.86 |
| | 3 | 0.75 | 5.16 | 15.70 | 21.88 | 35.16 | 46.63 | 4.45 |
| | 4 | 0.07 | 1.17 | 4.05 | 5.83 | 17.19 | 28.57 | 1.21 |
| | 5 | 0 | 0.37 | 0.51 | 0.97 | 1.56 | 7.14 | 0.21 |
| | Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

**Notes:** Transition probabilities (%) between six status levels; columns sum to 100.

is No. This and other analyses are presented in the Appendix, section A.

In summary, there is no evidence that major biases exist in this Tongcheng sample of seven genealogies. While it is impossible to generalize since genealogies come in a variety of qualities, systematic checks of internal consistency and external validation of the Tongcheng data suggests that while measurement error is present the records appear to be fundamentally sound.[40] Furthermore, information in the sample is consistent with what we know and expect based on other sources for larger parts of China. To a significant extent this is because the sample consists of seven genealogies that each describe rather different local dynasties. While each dynasty on its own could hardly be expected to yield a representative description of society in China, the sample of seven dynasties appears to be large and diverse enough so that critical differences to what we know about China as a whole are absent.

# 4 Patterns of Social Mobility

## 4.1 Up and Down Transitions in Status

Having evaluated the representativeness and accuracy of the genealogical sample for Tongcheng, and provided evidence supporting the fundamental soundness of the data, this section describes mobility patterns in the sample. After characterizing the joint distribution of parent and child outcomes for the entire sample period, this section will examine the extent to which mobility has changed over time.

Transition matrices give a first perspective on mobility. Table 4 shows mobility in terms of the six discrete status levels defined above (from the minimum level 0 to the maximum level 5). There is some clear evidence of status persistence. First of all, for low-status families (Father Status 0), almost 9 out of 10 of their sons retain that bottom level of status. In contrast, for sons coming from top-level *jin-shi* families (Father status 5), the chance of retaining *jin-shi* status is 7%, about 35 times the population share of *jin-shi*, and one in three of the *jin-shi* sons retains the high status levels four or five (top 2% of the population). Turning to mobility, let us first consider upward mobility, which has unambiguously positive welfare implications. Table 4 shows that the probability that a son rises to a certain level of status

[40] Section 4 discusses measurement error. See Appendix A for additional analyses.

Table 5: Transition Matrix for Early Half of Sample

| | | Father Status | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | Total |
| Son Status | 0 | 91.61 | 66.90 | 0 | 0 | 0 | 0 | 71.28 |
| | 1 | 6.22 | 22.07 | 62.32 | 60.99 | 28.89 | 0 | 17.87 |
| | 2 | 1.31 | 4.37 | 16.20 | 9.89 | 20.00 | 18.18 | 3.99 |
| | 3 | 0.75 | 5.75 | 16.90 | 24.45 | 37.78 | 36.36 | 5.55 |
| | 4 | 0.11 | 0.80 | 3.87 | 4.12 | 11.11 | 27.27 | 1.11 |
| | 5 | 0 | 0.11 | 0.70 | 0.55 | 2.22 | 18.18 | 0.20 |
| | Total | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

**Notes:** Transition probabilities (%) between six status levels before the year 1780; columns sum to 100.

is typically increasing in his father's status, as one would expect. For example, the chance of reaching status level 4 is 0.07% for a son of a status-0 father, and 1.17%, 4.05%, and 5.83% for a son of a status-1, status-2, and status-3 father, respectively. Similarly, the probability to become a top-level *jin-shi* (status 5) is 1.56% for a son of a status-4 father, about 1% for the son of a status-3 father, 0.5% in the case of a status-2 father, 0.4% for a status-1 father, and the chance for a son from a status-0 family to become top-level *jin-shi* is zero in this data. The last column in Table 4 gives the distribution of status in the son generation.

If the chance of moving to the very top is typically increasing in father status, the chance of dropping to the bottom of the status distribution is generally decreasing in father status. For example, the chance of falling to status level 1 is about 11% for a son of a status-5 father, but 28%, 61%, and 66% for a son of a status-4, status-3, and status-2 father, respectively. These results are interesting, and it is plausible that the probability that the son has high (or, correspondingly, low) status is declining in the distance to this status from his father's status.

It is also interesting to compare the transition matrices over time. Tables 5 and 6 show them for the earlier and the later half of the sample. We see that in the earlier period more than 91% of sons from lowest-status families stay in the bottom class of the distribution, in contrast to the later period when this share has fallen to 82%. Similarly, notice that in the earlier period 18% of the sons of top-level *jin-shi* are *jin-shi* themselves, in contrast to the later period when none of the sons of *jin-shi* fathers can hold on to their father's status. This is despite the fact that the bottom and top status shares are roughly the same in the early and later part of the sample, at 71% and 0.2%, respectively (see the Total columns in Tables 5 and 6).

More generally, there is some evidence that mobility has increased from these results. For example, according to the mobility measure $M$ proposed by Shorrocks (1978), $M = \frac{n-trace\,P}{n-1}$, where $P$ is the square transition matrix and $n$ its dimension, $M_{early} = 0.83$ while $M_{late} = 0.87$, consistent with mobility being higher in the later period. Transition matrices based on status levels depend on the size of each of the classes. In contrast, I now turn to transitions based on the status rank distribution, see Table 7. I aggregate the lower three quintiles because this is roughly the size of the status-0 group. In terms of upward mobility, there is an 8.8% probability that a son who is born to a father in the bottom 60%

Table 6: Transition Matrix for Later Half of Sample

| | | Father Status | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 0 | 1 | 2 | 3 | 4 | 5 | Total |
| Son Status | 0 | 82.22 | 66.11 | 0.00 | 0.00 | 0.00 | 0.00 | 70.85 |
| | 1 | 16.19 | 25.57 | 76.58 | 62.06 | 26.32 | 17.65 | 22.51 |
| | 2 | 0.79 | 1.59 | 6.31 | 9.88 | 13.16 | 0.00 | 1.73 |
| | 3 | 0.76 | 4.66 | 12.61 | 18.18 | 28.95 | 52.94 | 3.36 |
| | 4 | 0.03 | 1.49 | 4.50 | 8.30 | 31.58 | 29.41 | 1.32 |
| | 5 | 0.00 | 0.59 | 0.00 | 1.58 | 0.00 | 0.00 | 0.22 |
| | Total | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

**Notes:** Transition probabilities (%) between six status levels in the year 1780 and after; columns sum to 100.

Table 7: Percentile Rank Transition Matrix

| | | Rank Father | | | |
|---|---|---|---|---|---|
| | | [0,60] | [60,80] | [80,100] | Total |
| Rank Son | [0,60] | 3,904<br>73.18 | 1,019<br>57.28 | 412<br>23.16 | 5,335<br>60.00 |
| | [60,80] | 963<br>18.05 | 404<br>22.71 | 412<br>23.16 | 1,779<br>20.00 |
| | [80,100] | 468<br>8.77 | 356<br>20.01 | 955<br>53.68 | 1,779<br>20.00 |
| | Total | 5,335<br>100.00 | 1,779<br>100.00 | 1,779<br>100.00 | 8,893<br>100.00 |

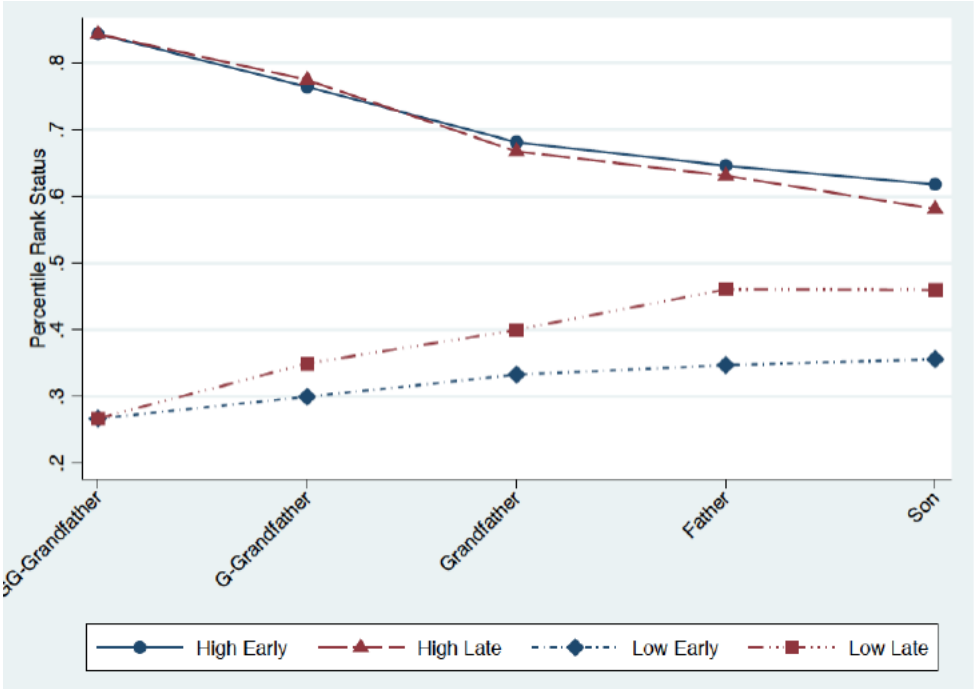
**Notes:** N and Transition probabilities (%) between the lower three, the fourth, and the fifth quintile.

of the rank distribution is able to moves into the top [80,100] quintile (Table 7). The probability more than doubles for sons of fathers in the fourth [60,80] quintile, and the probability is equal to 54% for sons of fathers who are in the top [80,100] quintile of the rank distribution (Table 7). It is also noteworthy that 57% of the sons of fathers in the fourth quintile fall to the bottom sixty percent of the status rank distribution. In contrast, the share of sons from top-quintile fathers falling to the lowest 60% is only 23%. This finding is, interestingly, consistent with contemporary evidence of a greater advantage to children at the very top of the distribution for a range of contemporaneous economies including the Nordic countries and Canada (Corak 2013).

My analysis employs several measures of social mobility to address different question of interest.[41] Exploiting the long-run nature of the data, I now turn from father-son pairs to long-run social mobility over five generations. Using the percentile rank of each individual in the status distribution, the following analysis provides additional evidence for an increase in mobility over time. Figure 4 shows results for the first versus the second half of sample. The two lines on top describe downward mobility conditional on the great-great-grandfather being in the top 30% of the status distribution. In the first half of the sample, a son with such a high-status great-great-grandfather could expect to be at rank 62 of the percentile status distribution, whereas in the second half the expected rank of the son would be 58. Thus, there is a

[41] See Fields and Ok (1999) for a discussion of some key elements of the theory of mobility measurement.

Figure 4: Five-Generations Mobility is Increasing Over Time




**Notes**: Figure shows average percentile status rank for early vs. late period (son's birth year before versus including and after 1786), for five generations. GG-Grandfather: great-great-grandfather, G-Grandfather: great-grandfather.

higher level of downward mobility in the second half of the sample (based on point estimates). Moreover, with five generations it becomes clear that the increase in downward mobility—the difference between the two downward trending lines--does not materialize in the first or second generation; instead, it is more pronounced in subsequent generations.

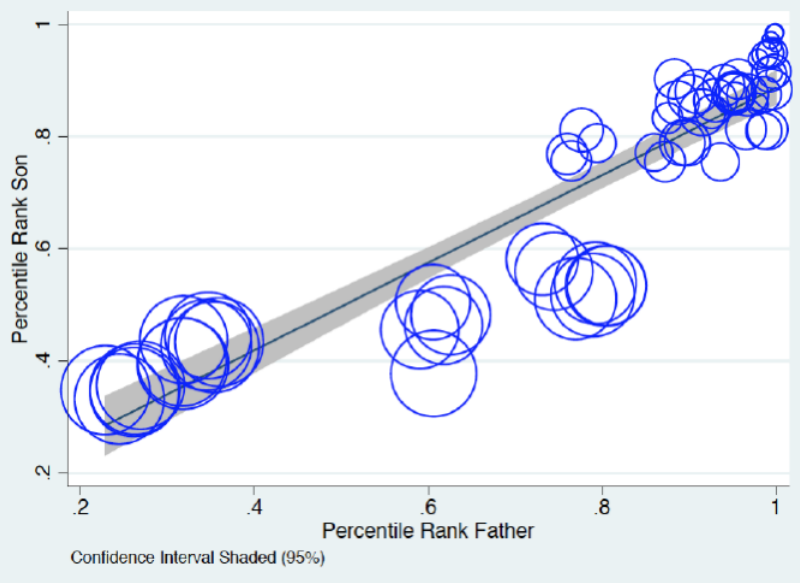
The two lines at the bottom of Figure 4 show evidence of changes in upward mobility. While in the first half of the sample the son of a low-status great-great-grandfather could expect to be at rank 36 of the percentile distribution, in the second half of the sample period the son could expect to be at rank 46, a full ten percentile ranks higher. Furthermore, the figure indicates that the increase in upward mobility is strongest for the second to fourth generation; for the fifth generation, in contrast, there is little difference in upward mobility in the early versus the late part of the sample. Thus, there is some evidence for an increase in mobility over time from five-generation-linked observations.

The following analysis returns to the relationship between father status and son status. Based on the six status levels shown in Table 1, I obtain the percentile status rank of each son in the son's birth cohort, defined as a 25-year time window (such as years 1575 to 1600).[42] As before, I assign all individuals of a given status the mid-point of the percentile rank in the status distribution; here the status of any individual is measured as his percentile rank for his birth cohort. Similarly, I calculate the percentile rank of all fathers in the sons' birth cohort using information on father status. By conditioning on the status distribution of each birth cohort, changes in average status between birth cohorts, for example, will not affect the results. Birth-cohort specific status distributions are helpful given my focus on the role of inequality for social mobility over time. As will be discussed in section 5 below), however, the findings are robust to assigning status based on the status distribution for the full six hundred years as well as explicitly accounting for changes in average status. Figure 5 shows the relationship between average rank of father and average rank of son, based on all father-son pairs. For each of the six status levels there are 12 circles, one for each birth cohort. The figure makes clear that son status is positively related to father status, on average. The slope of the line is less than one, implying that on average, father status does not fully determine son status. For example, the lower left in Figure 5 shows different cohorts of fathers and sons without status (level 0): across the twelve birth cohorts, fathers without status have an average percentile rank of about 0.3, while their sons have on average a rank of 0.39. The fact that on average the son's rank is higher than that of the father's at this percentile rank reflects regression to the mean. The $R^2$ of the relationship is 0.84; notice that few status group-cohort combinations are located outside the confidence interval of the fitted line.

Importantly, there is a very substantial number of observations in the lower range of socio-economic status, in contrast to the pioneering work by Ho (1962) which focused on the top-most percentile ranks—status level 16 and above according to the 23 category classification in this paper, which make up only 1.77% of my sample—in Figure 5 this corresponds to the upper-right corner. The slope of the average rank-rank relationship is relatively steep in the top two percentiles, roughly one third steeper than between the 80th and 90th percentile. Thus, while Ho (1962) concluded relying on national lists of *jin-shi* that China was a highly mobile society during the Ming-Qing, Figure 5 indicates that mobility at the very top was likely

[42]For data availability reasons I treat father-son pairs before 1575 as one birth cohort, and similarly pairs after 1825.

Figure 5: Relationship between Average Rank Father and Average Rank Son




**Notes**: Relationship between average percentile rank father and average percentile rank son for six status levels and twelve birth cohorts. Size of marker is proportional to number of observations.

lower than at status levels more typical for the population.

### 4.2 Relative Mobility in a Regression Framework

The relationship between rank of the father and rank of the son just shown depicts relative mobility, in the sense that we see the difference in status rank of sons that come from high-status versus lower-status families. In a regression framework with individual-level data, the relationship shown in Figure 5 takes the form

$$R_S(i) = \alpha + \beta R_F(i) + \epsilon(i), \qquad (1)$$

where $R_S(i)$ and $R_F(i)$ are the percentile rank of the son and the father, respectively, in pair $i$, = 1,..., 8,893. The mean-zero error term $\epsilon(i)$ captures all other influences affecting pair $i$. A regression coefficient $\beta$ close to one indicates low mobility because the son's rank in the status distribution is strongly determined by father rank, whereas there is high mobility when $\beta$ close to zero. I estimate an OLS coefficient of $\beta$ = 0.53 (see Table 8, column (1)). This means that a father with a 10 percentage points rank advantage over another father can expect that his son has a 5.3 percentage point rank advantage over the other son. Put differently, the advantage in the father generation is roughly cut in half in the generation of their sons. How does this estimate compare with existing figures? As noted in the Introduction, it is difficult to compare such mobility estimates across economies to draw inferences about equality of opportunity, because in key ways the economies could be quite different. Moreover, many existing studies employ a

Table 8: Alternative Status Definitions and Samples

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Baseline | Single Men | Multiple Wives | Father Official or Wealth |
| $\beta$ | 0.528 | 0.377 | 0.625 | 0.306 |
| $(s.e.)$ | (0.009) | (0.007) | (0.008) | (0.012) |
| $E\left[R_S(i)\,\middle|\,R_F(i)=0.25\right]$ | 0.418 | 0.459 | 0.418 | 0.467 |
| $E\left[R_S(i)\,\middle|\,R_F(i)=0.80\right]$ | 0.620 | 0.595 | 0.622 | 0.550 |
| $R^2$ of Average Relation | 0.841 | 0.704 | 0.835 | 0.552 |

**Notes**: Estimation of $\beta$ (equation (1)) by OLS; $E\left[R_S(i)\,|R_F(i)=0.25\right]$ is the expectation of percentile rank of sons with fathers at percentile rank $< 0.5$. $E\left[R_S(i)\,|R_F(i)=0.80\right]$ is the expectation of percentile rank of sons with fathers at percentile rank $> 0.6$. N = 8,893 except in column (2) where N = 14,905.

log-log specification based on earnings or study mobility in terms of education, not status mobility using the percentile rank approach of equation (1). Keeping these issues in mind, Braun and Stuhler (2018) obtain an estimate of $\beta = 0.56$ for educational mobility in Germany in the late 19th and 20th century, Lindahl, Palme, Sandgren Massih, and Sjogren (2015) find $\beta = 0.33$ for educational mobility in Sweden since the late 19th century, while Clark and Cummins (2014) estimate intergenerational elasticities based on rare surnames around 0.7 to 0.8 (I will return to this below). For contemporaneous economies, Chetty, Hendren, Kline, and Saez (2014) estimate a value of $\beta = 0.34$ for late 20th century birth cohorts in the U.S., while Corak (2013) shows a range of about 0.2 to 0.5 for countries such as Denmark and Italy, respectively. My estimate of $\beta = 0.53$ is thus among the higher estimates, indicating relatively low social mobility. However, I will show below that over time since the year 1300 social mobility has increased.

### 4.3 Robustness to Alternative Status Definitions and Samples

Section 3 of the paper used a variety of comparisons and found the Tongcheng sample was fundamentally representative of the broader population. Nevertheless, there may still be concerns that the status variable has been coded with error, since there is no consensus on exactly how certain information (such as being a wealthy landowner) maps into socio-economic status. Furthermore, status definitions may have changed over the course of the sample period. Therefore, this subsection explores how the findings presented thus far would change as I change the definitions of status, or, as I employ different samples. Recall that in the baseline, the coefficient $\beta$ in the rank-rank specification, equation (1), is 0.53, with a robust standard error of 0.009 (first row of Table 8, column (1)). This rank regression estimate is complemented by several other measures of mobility (rows 2 to 4 of Table 8, column (1)). First, a simple measure of upward mobility is the expected rank of a son whose father is at rank 0.25 of the distribution.[43] This is equal to 0.42: thus, on average, such sons are expected to rise by 17 percentage points in the rank distribution. Second, as a simple measure of downward mobility, I consider the expected rank of a son whose father is at the 80th rank percentile.[44] This turns out to be equal to 0.62, implying an expected 18 percentage points decline

[43] Calculated as the expected rank of a son whose father is in the lower half of the status distribution (percentile rank $<$ 0.50).

[44] This is calculated as the expected rank of a son whose father has percentile rank 60 and higher.

in status relative to the father (row 3 in column 1, Table 8). Of course, in addition to these one may consider other measures of mobility; I do so in section 5 to show that my findings are robust to measuring mobility differently than is done here. The last row in Table 8 gives the $R^2$ of the relationship between the averages of rank father and rank son, which is shown in Figure 5 above. Recall also that the baseline sample consists of all married males, that is, individuals who show up in the family tree both as a son and then as a married husband.

In a society where illegitimacy rates were very low, mobility from parents to children would require, first and foremost, the examination of the married population. I now expand the sample to include individuals who appear in the genealogy as sons but not as married husbands. While marriage was nearly universal among women, the fraction of single men is generally estimated to have been substantially higher.[45] This can be explained by at least two reasons. First, female mortality rates were higher than those for males, and second, men tended to remarry more frequently than women did upon the death of a spouse. In a stratified society, this implied it was not uncommon for wealthier and more successful men to marry multiple times while lower status men stayed single. For the analysis of single men, I focus on men who reached at least fifteen years of age and where we know the date of death. The presence of a death date record increases the likelihood that the individual did not out-migrate.[46] Further, because there is no status information on unmarried men, their status has to be estimated. I assign these individuals to the lowest status level, a plausible assumption. At the same time, this approach (weakly) underestimates the status of single men and should therefore be seen as a bound on how social mobility would change if single men were included.

The relative mobility estimate $\beta$ declines from 0.53 in the baseline to 0.38 (column (2), row 1). Further, in terms of upward mobility sons of fathers in the lower half of the status distribution can expect to rise to rank 0.46, compared to 0.42 in the baseline (row 2). Also, a son of a father at the 80th percentile can expect to drop to the 60th, in contrast to 62nd percentile rank in the baseline (row 3). In part, mobility is higher with single men because father-son pairs are added to the sample in a way that tends to increase downward mobility. In fact, some single sons to whom I assign status level zero are likely to have had status above level zero. Further, the extent to which this was the case may have be correlated with status of the father.[47] To understand how important this might be for estimating mobility, instead of status level zero I assign status level 1 to a randomly selected subset of the single men, plus a boost in status proportional to the status of their fathers.[48] With this approach the relative mobility estimate would be $\beta = 0.48$, not too different from the baseline estimate of 0.53.[49] Thus, under plausible assumptions the inclusion of single men with estimated status levels would yield a moderately higher social mobility.

I also explore the influence of particular elements of a person's biography on the definition of status. One assumption used in status definition is that men who are sequentially married more than once would have

[45] Telford (1986) finds that the proportion of unmarried men in Tongcheng during the Ming was above 20%.

[46] In the event single men out-migrated from Tongcheng, it is possible that news of their marriage did not eventually get recorded. However, migration over longer distances was rare; for example, Telford (1986) reports a figure of 1.9%.

[47] Consistent with this there is evidence that the fraction of sons that would not marry is declining with status during the Ming (Telford 1986, Table 5.1).

[48] I assume that sons born in January have a status of one plus a 10% boost depending on their fathers' 23-level status.

[49] The estimate of $\beta = 0.48$ is not shown in Table 8 for space reasons.

Table 9: Status Measured with Error

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Small Error | Large Error | Same Error Father | Error Father and | Error Father and |
| | Father | Father | and Son | Son orthogonal | Son opposite |
| $\beta$ | 0.528 | 0.421 | 0.553 | 0.506 | 0.486 |
| $E\left[R_S(i)\,\middle|\,R_F(i)=0.25\right]$ | 0.418 | 0.430 | 0.412 | 0.421 | 0.425 |
| $E\left[R_S(i)\,\middle|\,R_F(i)=0.80\right]$ | 0.618 | 0.605 | 0.629 | 0.617 | 0.610 |
| $R^2$ of Average Relation | 0.841 | 0.815 | 0.869 | 0.844 | 0.838 |

**Notes**: Estimation of $\beta$ (equation (1)) by OLS; $E\left[R_S(i)\,|R_F(i)=0.25\right]$ is the expectation of percentile rank of sons with fathers at percentile rank $< 0.5$. $E\left[R_S(i)\,|R_F(i)=0.80\right]$ is the expectation of percentile rank of sons with fathers at percentile rank $> 0.6$. N = 8,893.

attained a certain level of status.[50] While this is supported by historical accounts that suggest a man who could afford to marry more than once would have likely had been somewhat better off, an alternative is to drop this category. If we were to instead code these men no differently from someone in the lowest status group (status level 0), the relative mobility estimate $\beta$ changes to 0.63, while upward and downward mobility remain largely unchanged (column (3)). This suggests that the way in which this particular group was coded has only a limited influence on the findings. Another item that is among the biographical information is whether a man's father was a provincial-level government official (*sheng-yuan*), an expectant official (*jian-sheng*), and whether the father had evidence of wealth.

One might be concerned with assigning status to the son for the sole reason that his father had some socio-economic status. In the context of this society, it would be somewhat implausible, from a historical standpoint, to view the sons of officials no differently from commoners. Consistent with that, men whose father was such a *sheng-yuan* or who had evidence of wealth live typically three years longer than men without status.[51] This is the reason the status coding gives the sons of officials a slight bump up in status, yet still ensures they are much lower than those who did attain actual degrees. As a matter of computation, however, if instead I treat these as men without status (level zero), mobility is estimated to be higher because the re-coding of status increases the fall in status of sons whose father was a *sheng-yuan* or had evidence of wealth (see column (4)). Given the importance of the definition of status, I will return to this question when discussing the influence on mobility of changes of status definitions over time below (section 5).

The preceding analysis has examined alternative status definitions that are deterministic in nature. In the following, I generalize this approach by treating true status as a variable that is observed with error. Depending on the size of the error, a person's status may be very similar, or it could also be quite different, from what is given in the Tongcheng genealogies. While this approach does not 'undo' whatever measurement error is present in the data, it allows to see, both qualitatively and quantitatively, the influence of measurement error for mobility. My starting point is a mean-zero error in father status with a

[50] Level 2 in the status classification with 23 levels, see Table 1.

[51] Furthermore, the hypothesis that being the son of a high-status father provides status in itself finds support in the fact that the official *jin-shi* lists of the Chinese states mentioned the name of the immediate kin of the *jin-shi* (including father, mother, wives, brothers, and sons).

standard deviation of 0.01. Results are shown in Table 9. As one would expect, the effect of this relatively small error on the mobility estimate is small (Table 9, column (1)). When the size of the error is drastically increased to a mean-zero error with standard deviation of 0.5, the relative mobility estimate falls from 0.53 to 0.42 (column (2)). Note that father status is the independent variable in the rank regression. Therefore, the change in β towards zero is in line with classical measurement error that biases the regression coefficient towards zero. In comparison, changes in upward and downward mobility estimates, which are not based on regressions, are relatively small. This is one of the reasons why I examine multiple measures of mobility throughout the paper.

The following three specifications add errors to both the father and son status data. I begin with the same measurement error for father and son (column (3)), which can be thought of as perfect correlation in terms of the mismeasurement of status. Generally, doing this leads to more persistence than before. The coefficient in the rank regression is now 0.55, up from 0.53 in the baseline, upward mobility is slightly lower and downward mobility is somewhat lower as well. Persistence is higher because a term is introduced that is the same for father and son in a given pair.

If, in contrast, the error in father status is orthogonal to the error in son status, mobility is estimated slightly higher than in the baseline analysis (compare columns (4) and (1)). Finally, column (5) shows results when the error in father and son status is negatively related with a correlation of -1; the result is a higher level of mobility. For example, β in the rank regression falls from 0.53 to 0.49. Generally, positive correlation in measurement error is more plausible than negative correlation for a given father-son pair. As seen above, positively correlated measurement error will tend to give lower mobility estimates (column (3)). To the extent that in the data there is positively correlated measurement error in a given father-son pair, the β estimate would be biased towards persistence, as in column (3), and the true β would be lower than the baseline estimate of 0.53. At the same time, comparing columns (1) and (3) of Table 9 suggests that, quantitatively, the difference is likely to be small.

### 4.4 Social Mobility as a Multi-Generational Process

In line with most research on social mobility, I have so far mostly examined the role of father status for the son. However, in some models, such as Solon (2014), generations before the father generation may influence son mobility. This could be due to direct contacts, as may be the case between grandfather and son, or due to indirect effects, such as from accumulated resources and values in earlier generations. In this section, I provide evidence on the roles of the paternal grandfather, great-grandfather, and great-great-grandfather status for social mobility. Instead of the father-son pair, the unit of observation now is the five-generation quintuplet, from great-great-grandfather to son. The following are simple extensions of the rank regression, equation (1):

$$R_S(i) = \alpha + \beta_1 R_F(i) + \beta_g R_g(i) + \epsilon(i), \qquad (2)$$

where $g$, $g = \{GF, GGF, GGGF\}$ indicates grandfather, great-grandfather, and great-great-grandfather, respectively, and $R_{GF}(i)$, for example, is the percentile rank of the grandfather in quintuplet $i$. With

Table 10: Social Mobility over Five Generations

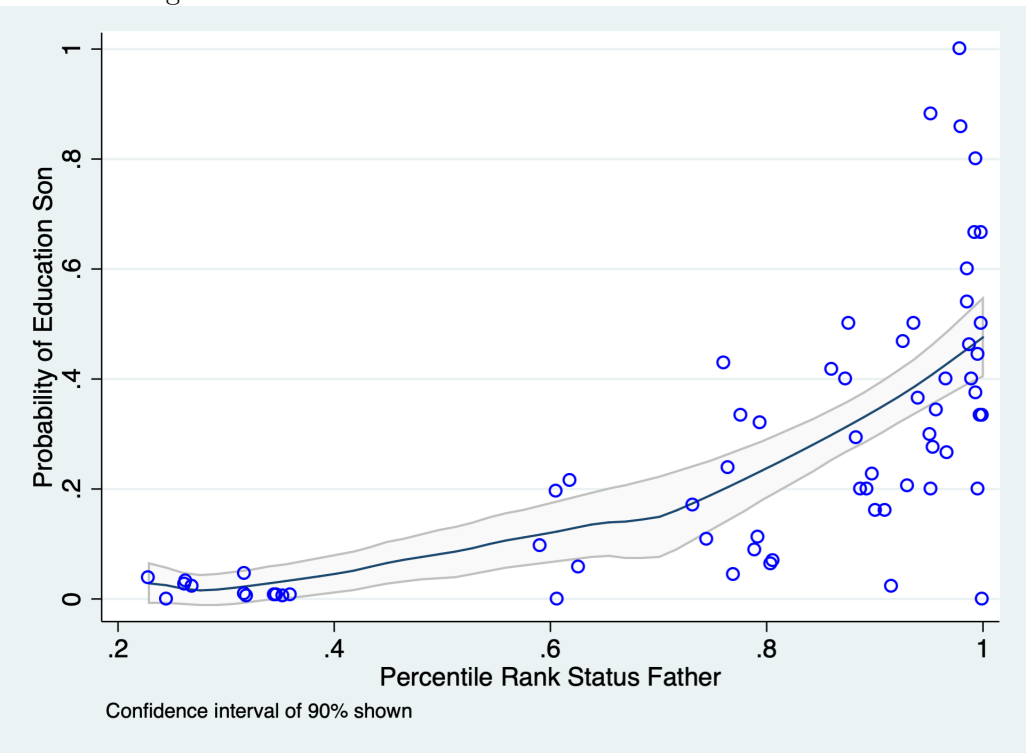
| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Rank Father | 0.528** (0.009) | 0.533** (0.009) | 0.484** (0.013) | 0.488** (0.011) | 0.498** (0.010) |
| Rank Grandfather | | | 0.072** (0.013) | | |
| Rank Great-Grandfather | | | | 0.087** (0.011) | |
| Rank Great-Great-Grandfather | | | | | 0.078** (0.009) |
| $R^2$ | 0.311 | 0.322 | 0.326 | 0.329 | 0.329 |
| N | 8,893 | 7,328 | 7,328 | 7,328 | 7,328 |

**Notes**: Dependent variable is percentile rank status of son. Estimation by OLS. Robust standard errors in parentheses. ** means significant at a 1% level.

roughly thirty years between each two generations, a five-generation linked quintuplet covers about 150 years of calendar time. Similar to how lagging in a time series regression reduces the number of observations, the linking of five generations results in a fall in the number of observations. I begin by comparing the regression results for the father-son sample and the five-generation sample. Results in Table 10 shows there is little difference in the relative mobility estimates, see columns (1) and (2). This indicates that the results are not strongly affected by whether there are five generations or not. The first set of results for equation (2) are for Rank Grandfather, shown in column (3). The coefficient for Rank Father falls somewhat while the coefficient on Rank Grandfather is about $\beta_{GF} = 0.07$. According to these results, the status advantage that sons from higher class families have is not exclusively due to his father, but rather to a mix of the rank of father and grandfather. Moving to additional generations, the coefficient of great-grandfather rank is positive at 9% (column (4)), while the great-great-grandfather variable enters with a coefficient of 8% (column (5)). Notice that the size of the coefficient for the earlier generation does not monotonically fall as the generational distance to the son increases. This suggests that the results capture more than the direct contact between these persons.

Quantitatively, the results of column (5) imply that a 10-percentage point advantage in father status translates on average into an advantage of the son between 5.1 percentage points and 5.8 percentage points; the former is the case when the great-great-grandfather has no status, the latter when he has top status. This means that having a high-status great-great-grandfather increases the status advantage of the privileged son by a sizable 14%. The sum of the coefficients of father plus great-grandfather, or of father plus great-great-grandfather (columns (4) and (5)) is somewhat above the estimate of $\beta_1$ when only father status is included (column (2)). This indicates that if a family is high status for multiple generations the advantage to the son is higher than if only the father had the high status. Furthermore, due to the positive correlation of status across generations, including all five generations simultaneously does not lead to important additional insights. Note though that the father rank coefficient remains quite similar to its value in columns (3) to (5). This indicates that the additional mobility-relevant information coming from earlier generations is largely shared, and moreover, it does not reduce the role of father status.

Figure 6: Education as an Intermediate Outcome of Status



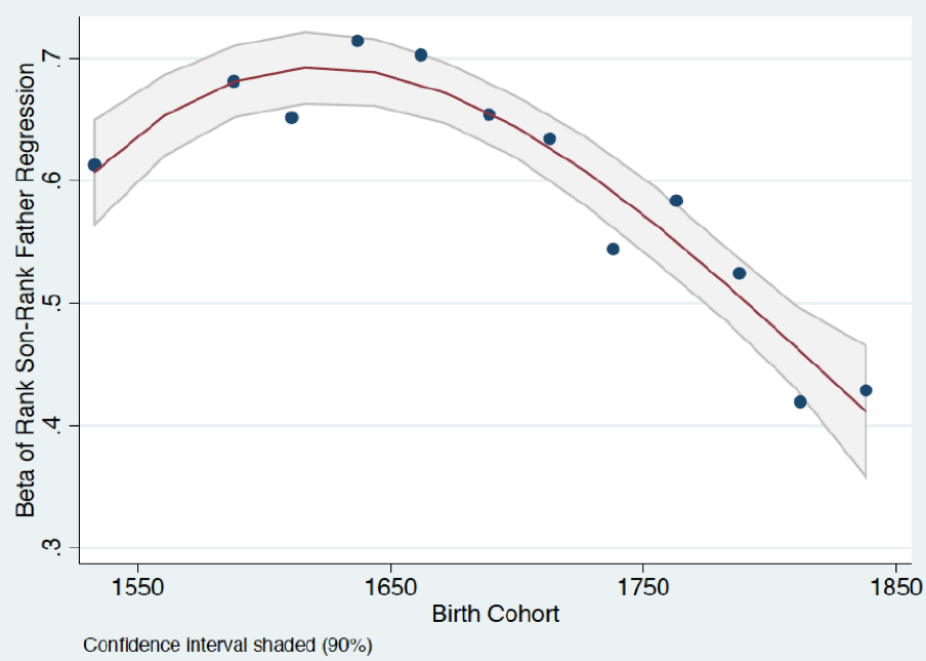

### 4.5 Education as an Intermediate Outcome of Status

The genealogy gives information on acquired formal education as evidenced by participation in the civil service entrance exams. These exams required significant investments, which were made by about 6.5% of the men in the sample. In the absence of widespread public schooling, these investments were largely undertaken privately in a bid to be awarded government office and high status. Education, therefore, is an intermediate outcome of son status, and participation in the civil service exams may be considered an indicator for having undertaken investments in education. Figure 6 indicates that coming from a high-status family gives a substantial advantage to the son's chances of accumulating human capital. The measure of education includes those who merely studied for the exam but failed to take any degrees (status level 6 in the 23 category assignment). In particular, the probability to become educated for a son that is born to a lowest-status family is close to zero, while it is around 50% for a son of a *jin-shi*. Furthermore, the probability that the son is educated is 14% if the father is in the top two quintiles, compared to 1% when the father is in the lower half of the status distribution. The relationship appears to be stronger at the high end of the status distribution.

Overall, higher status of the father translates not only into higher status of the son but also into more education, an intermediate outcome of status. While it is possible to separate men who passed from those who did not, as well as distinguish different exam levels, doing so does not change the qualitative findings below.[52]

[52] See Figure A.10 in the Appendix.

Figure 7: Social Mobility Over Time




**Notes**: Shown are $\beta_c$ coefficients from OLS regressions of percentile rank son on percentile rank father (equation (1)). Ten twenty-five year birth cohorts (1575 to 1825), plus birth cohort before 1575 and after 1825. The horizontal axis gives median birth year of son in cohort. Earliest birth year is 1330.

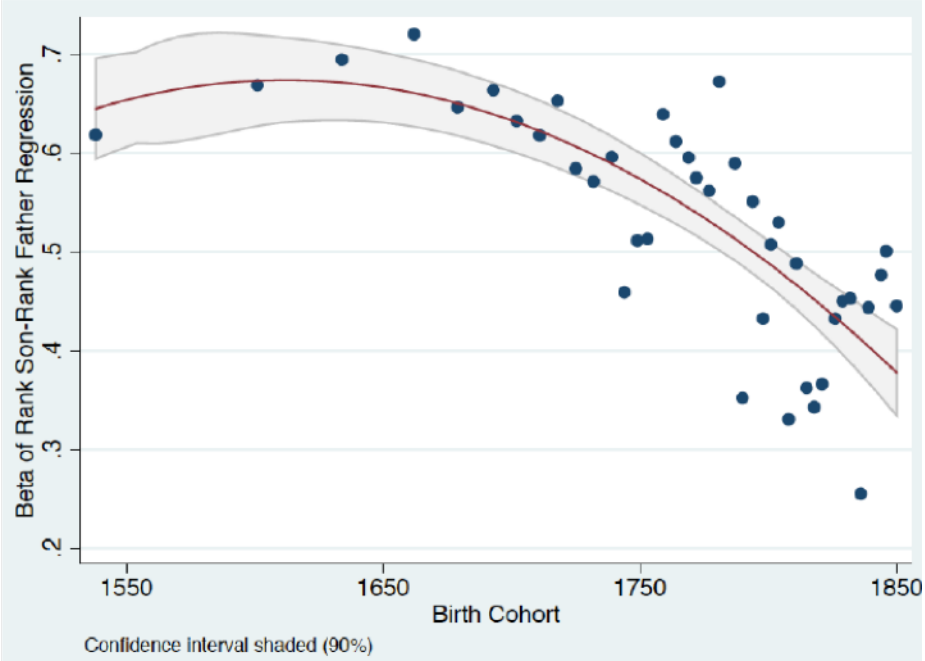
# 5 Temporal Variation in Social Mobility

## 5.1 Main Findings

This section examines changes in social mobility over time. I begin by analyzing mobility for ten 25-year birth cohorts, $c$, between 1575 to 1825, plus two additional cohorts. One of the two additional cohorts span the open period before 1575, while the other covers the period after 1825. In each of these twelve cohorts, the rank of a son is relative to other sons in his birth cohort. Similarly, the rank of each son's father is relative to all fathers in this birth cohort. Let $\beta_c$ be the coefficient from the regression of percentile rank son, $R_S(i)$), on percentile rank father, $R_F(i)$, for all pairs $i$ belonging to birth cohort $c$, $c = 1, ..., 12$.

The evolution of mobility in terms of the coefficient in the rank son on rank father regression is shown in Figure 7. This analysis employs data on all father-son pairs over the six-hundred years long sample period even though the first marker on the horizontal axis gives the year 1550. This is due to the fact that the first $\beta_c$ estimate is given for 1533, the median birth year of all sons in the first birth cohort; however the earliest birth year for the first cohort is in the year 1330. The figure shows that until 1700, $\beta_c$, is typically between 0.6 and 0.7, but by the early 19th century, it has fallen to around 0.45. Thus, social mobility increased over time. With an uneven number of father-son pairs over time, some cohorts have more observations than others. One might be concerned that the variation in the number of father-son pairs in each birth cohort will affect the resulting trends. To address this concern, the following shows

Figure 8: Mobility Over Time with a Fixed Number of Observations per Birth Cohort



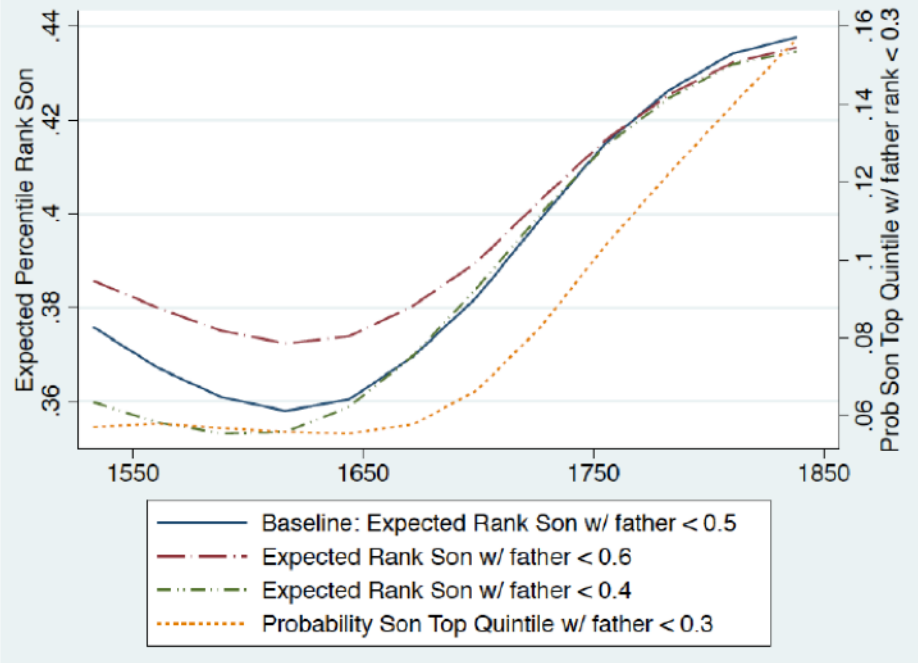

**Notes**: Shown are $\beta_c$ coefficients from OLS regressions of percentile rank son on percentile rank father (equation (1)). Forty birth cohorts with equal numbers of observations each. The horizontal axis gives median birth year of son in cohort. Earliest birth year is 1330.

mobility estimates based on a fixed number of father-son pairs in a given birth cohort. With this approach, the distribution of the data towards the end of the sample translates into more birth cohorts towards the end of the sample. Figure 8 shows that keeping a fixed number of observations for each birth cohort leads to the same qualitative pattern of increasing mobility that is seen for 25-year birth cohorts. Quantitatively, the increase in mobility is now somewhat stronger than before. In particular, $\beta_c$ falls from around 0.64 in the year 1500 to below 0.4 by 1850. This indicates that this finding is robust to defining birth cohorts in alternative ways.

The findings of Figures 7 and 8 for relative mobility is supported by the evidence on upward and downward mobility. The evolution of upward mobility, using the baseline measure and various additional measures, is shown in Figure 9. An increase in upward mobility is found in all cases. In particular, the probability that the son reaches the top quintile of the status distribution when his father has rank 0.3 or lower almost triples between 1550 and 1840 (less than 6% to more than 15%). Thus, upward mobility increased substantially during the sample period. A parallel analysis of alternative measures of downward mobility shows that also in this dimension there has been a sizable increase in mobility over time; see Figures A.3 and A.6 in the Appendix.[53] Finally, instead of status distributions that are specific to each cohort, I have estimated mobility when each son is ranked in the status distribution relative to sons over the entire sample period of six hundred years. The findings are quite similar to those with birth-cohort specific

[53]Results in the case of forty birth cohorts of equal size are shown in Figures A.4 and A.5 for upward and downward mobility, respectively.

Figure 9: Upward Mobility Over Time - Alternative Measures




**Notes**: Shown are alternative measures of upward mobility for twelve birth cohorts. Earliest birth year is 1330. Locally smoothed series. (1) Baseline: expected rank of son given his father has rank below 0.5. (2) Expected rank of son given his father has rank below 0.6. (3) Expected rank of son given his father has rank below 0.4. (4) Probability that son has rank 0.8 or above given that his father has rank below 0.3.

Table 11: Mobility Changes Over Time - Robustness

| | | (1) | (2) | (3) | (4) |
|---|---|---|---|---|---|
| | | Single Men | Father Official | Measurement Error | Lower Rank |
| | | | or Wealth | in Status | Officials > 1780 |
| Mobility | Upward | 0.941 | 0.926 | 0.983 | 0.976 |
| | Downward | 0.962 | 0.907 | 0.985 | 0.968 |
| | Relative | 0.922 | 0.820 | 0.968 | 0.966 |
| | Average | 0.942 | 0.884 | 0.979 | 0.970 |

**Notes**: Shown is the correlation for three mobility measures with the baseline specification for four alternative specifications, as described in the text; twelve birth cohorts.

status ranks, which indicates that the increase in mobility over the sample period is not driven by ranking father and son status in their specific birth cohort (see Figure A.7 in the Appendix). It also suggests that who was considered high status, and the relative definitions of status, did not change much over late imperial China. This may be because the state's institutions and the civil service examinations persisted throughout the period, even as emperors and reign names changed.

The trends towards higher social mobility is in line with the historical evidence on China over the sample period (recall Section 2 above), but it is at odds with Clark's (2014) result that the intergenerational mobility coefficient is time-invariant and in the range of 0.7 to 0.8 (p.212). There could be a number of reasons for relatively high $\beta_c$ coefficients, one of which is group effects.[54] Mobility between groups in society tends to be low compared to individual-level mobility: here, for example, the $\beta$ estimate of the mobility between dynastic groups is above 0.8. One interpretation could be that time-invariant intergenerational mobility estimates, as in Clark (2014), may be the result of group differences that change little over time. In the present case, this effect plays less of a role because between-group differences have come down over time as I show below in Section 6.

## 5.2 Temporal Variation in Social Mobility: Robustness

This section examines changes in mobility over time for alternative samples, status definitions, and mobility measures. Table 11 reports the strength of the correlation between baseline results on relative, upward, and downward mobility and corresponding results for four alternatives. These are (1) a broader sample to include single men, (2) a redefinition of status so that sons of fathers with status—father was a *sheng-yuan*, *jian-sheng*, or had wealth—do not receive status and are demoted to status level 0, and (3) substantial measurement error in father status. The fourth alternative is a downgrade in the status of minor officials and their sons in the second half of the sample, in line with some accounts of this period (column (4)). Table 11 correlates the results from the redefined mobility measures across 25-year windows with the baseline specification of relative mobility, upward mobility, and downward mobility (Figures 7, 9, and A.3). The results indicate that that the changes in social mobility found in the baseline above are obtained for the alternative specifications as well. The correlation with the baseline numbers is never lower than 0.82, and

[54] See Solon (2018) for a broader discussion.

the average is 0.94. Overall, the results show that social mobility increased over time, and the finding is robust to redefinitions of those status categories in the coding that might be most concerning.

# 6 Explaining Mobility Differences

The previous section showed that mobility—relative mobility, upward mobility, and downward mobility—increased over time, and the trends are robust to redefinitions of the benchmark status categories. This section examines the main drivers of this increase in social mobility to show that inequality plays a key role in mobility, but the sign may be positive or negative. I begin by examining several important dimensions of individual inequality: the location where a person lives, the socioeconomic status of an individual's father, and especially inequality in the individual's father's level of education. Next, I turn to group effects and provide graphical evidence on the importance of dynasty-level inequality for social mobility, both across groups as well as within groups. I also present new evidence on the role of migration for social mobility. The section concludes with results presented from multiple regressions to obtain independent variation and comparable coefficient magnitudes.

## 6.1 Inequality in Status

Parental investments are central to intergenerational mobility (Becker and Tomes 1979), and status inequality determines which parents can make these investments and which cannot. Solon (2004) modifies the Becker and Tomes framework, showing that a shift to more progressive government investment in human capital can increase intergenerational mobility and at the same time reduce cross-sectional inequality. Investments in human capital within dynastic families — such as through lineage schools — would have the same progressive effects, but the impact would be felt only within the group.

The following analysis exploits the temporal variation in mobility documented in the previous section. The results are given for the case of forty different birth cohorts over more than five centuries (the earliest birth year is 1330). Recall that the $\beta_c$ coefficients from the rank regressions, cohort-by-cohort, are in the range of 0.25 to 0.72, and that the general trend is towards lower values (higher mobility) over time (Figure 8). These cohort-specific $\beta_c$ estimates are now related to inequality, measured by the Theil (1967) index, in father percentile rank status in cohort $c$.[55] The index is defined as

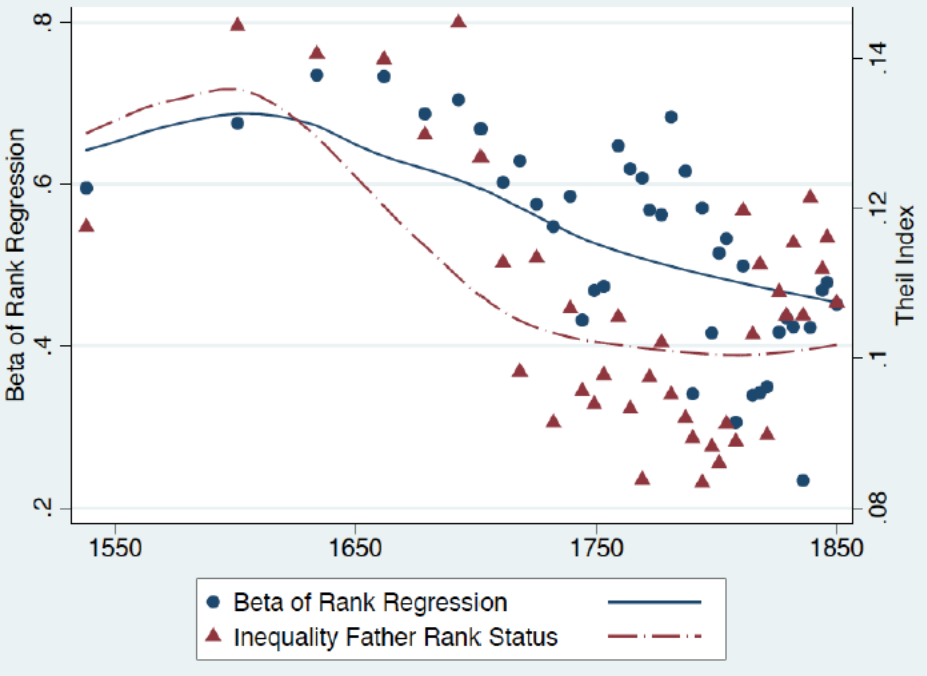
$$T_c = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{R_F(i)}{\mu_c} ln \left( \frac{R_F(i)}{\mu_c} \right), \qquad (3)$$

where $R_F(i)$ is the percentile rank status of the father in father-son pair $i$ belonging to cohort $c$, $\mu_c$ is the average percentile father rank in cohort $c$, and $N_c$ is the number of fathers in cohort $c$.

Father status is based on the baseline definition with six different levels, see Table 1, first column. The

[55]The Theil index is a well-known member of the single parameter Generalized Entropy Class indices. See Bourguignon (1979) and Shorrocks (1980) for discussions of inequality measures. My results are similar when I use other measures, including the Gini index, as shown in the Appendix.

Figure 10: Status Inequality and Relative Mobility




**Notes**: Shown is $\beta_c$ and status inequality in the father generation for forty birth cohorts with equal number of observations; first birth year is 1330.

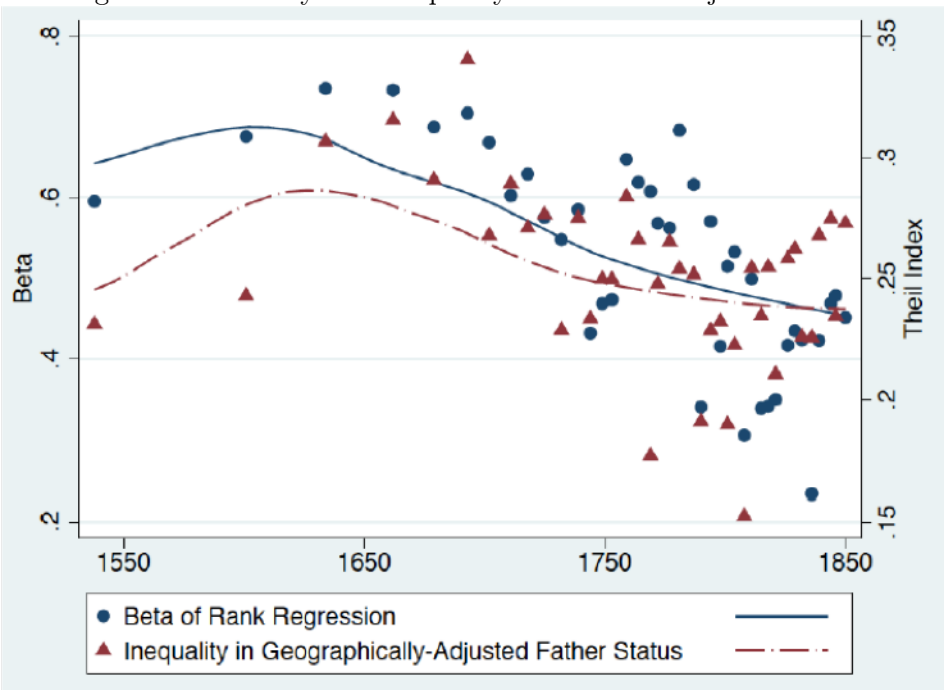
circles (fitted with a solid line) in Figure 10 plot the relative intergeneration mobility of sons in the different cohorts; for each of these cohorts there is a triangle (fitted with a dash-dot line) that gives the inequality in status observed in the father's generation. Figure 10 shows that times of relatively high inequality in the father generation are associated with a high $\beta_c$, or, low mobility. Before the year 1700, $\beta_c$ tends to be above 0.6, and these are the birth cohorts for which the Theil index is relatively high (around 0.12). In contrast, by the year 1850 $\beta_c$ is lower, around 0.45, while status inequality has fallen to around 0.10. Thus, status inequality is found to be negatively correlated with social mobility (correlation with $\beta_c$ of 0.43).[56] In addition, there is also a significant (and in fact stronger) correlation of status inequality with upward and downward mobility (correlations of -0.79 and 0.79, respectively). Specifically, while a son from a poor family (25th percentile father) can expect to rise under the high-inequality conditions around 1600 (Theil = 0.13) only to the 35th rank, in 1800 such a son can expect to rise to the 44th rank as inequality is lower.

## 6.2 The Impact of Geographic Location on Social Mobility

The results thus far provide evidence, based on long-run temporal variation, that inequality is detrimental to social mobility. A plausible explanation is that times of relatively high inequality give high status families a relatively better chance to use their position and resources to help placing their sons into high socioeconomic positions in the next generation. It is also interesting that inequality is correlated similarly

[56]The correlation is significant at standard levels, with a 95% confidence interval of [0.14, 0.65]. Confidence intervals for the correlations shown in Figures 10 to 18 are shown in Table A.4 in the Appendix.

Figure 11: Mobility and Inequality in Location-Adjusted Status

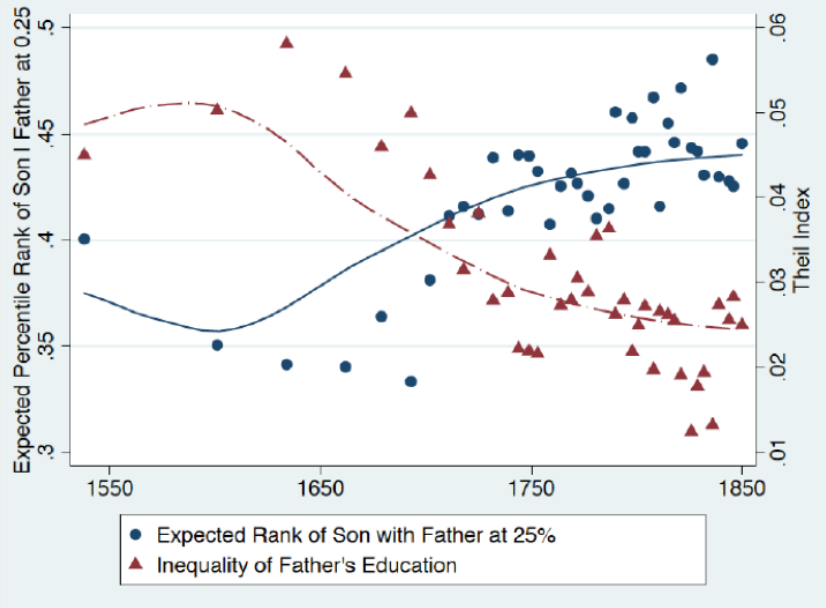



**Notes**: Shown is $\beta_c$ and location-adjusted status inequality in the father generation. Horizontal axis shows median birth year in birth cohort; first birth year is 1330.

strongly with both upward and downward mobility. This suggests that during times of high inequality high-status families are able to avoid a rapid regression to the mean as much as poor families are unable to promote their sons into high-level status positions. In addition to inequality in status, geographical location may matter too. A high-status father could be instrumental in advancing the career of his son—however, if the family lives far away from the most important offices and decision makers, the influence of the father may be limited even if he has high status. Because location is endogenous—e.g., high-status individuals are likely to move to an important location—the analysis is not meant to estimate causal effects. The most important location in Tongcheng County was the capital, Tongcheng City. I have employed GIS methods to obtain for each individual the distance, as the crow flies, from his location to Tongcheng City. To examine the possible influence of geographic location on social mobility, I define geographically-adjusted father status as the percentile rank divided by distance to Tongcheng City, raised to some power $\delta$: $\frac{R_F}{dist^{\delta}}$, where I choose $\delta = 0.2$ based on empirical fit; alternative values of $\delta$ give quantitively similar results (see Figure A.8 in the Appendix).

Figure 11 shows that inequality in geographically-adjusted status is also positively correlated with the rank regression coefficient. While the adjusted status inequality is not higher in 1550 versus 1850, in contrast to the unadjusted status equality measure, the correlation between inequality and is now stronger (0.58, compared to 0.43). This is consistent with geographic location playing a role for mobility.

Figure 12: Inequality in Education and Upward Mobility




**Notes**: Expected percentile rank of sons with fathers of percentile rank of less than 0.5 and Theil index of educational inequality in the father generation. Horizontal axis shows median birth year in birth cohort; first birth year is 1330.

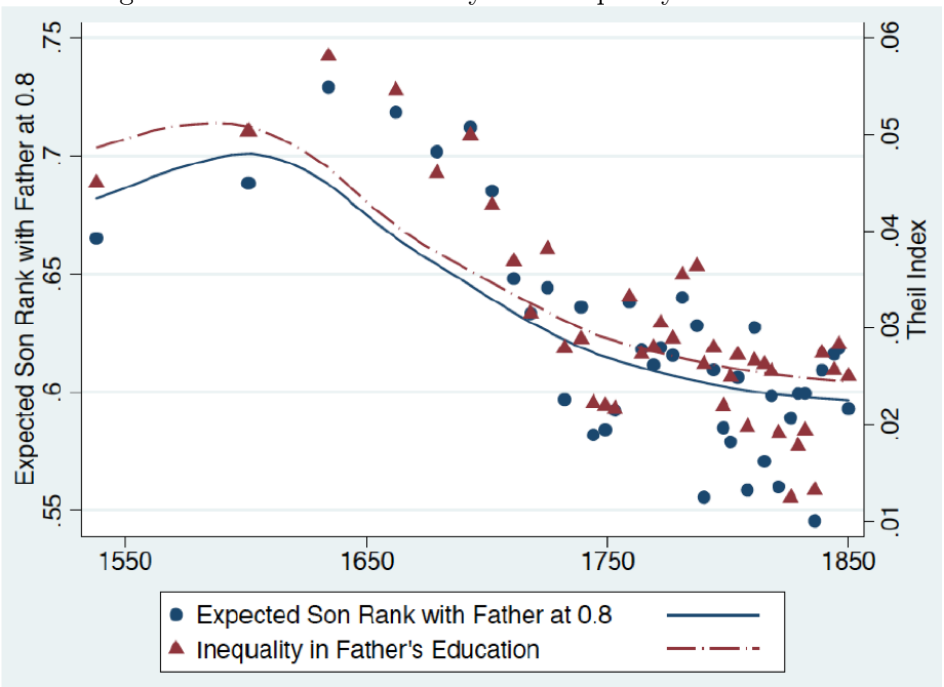
### 6.3 Educational Inequality

This section turns to the relationship between educational inequality and mobility in my sample. In models of intergenerational mobility, inequality in parental human capital often plays a key role (Becker and Tomes 1986). Figure 6 showed that education is an important intermediate outcome of status. Human capital is generally seen as one of the most important investments parents can make into their children, but it was of key importance in this context as well. Given the absence of public education during the sample period, such investments came primarily from the family, local dynastic schools, and to a much lesser extent from state support. Thus, I examine whether inequality in terms of father's education matters. Figure 12 shows the relationship between upward mobility and educational inequality. Using the Theil index, educational inequality is defined as

$$T_c^e = \frac{1}{N_c} \sum_{i=1}^{N_c} \frac{E_F(i)}{\mu_c^e} ln \left( \frac{E_F(i)}{\mu_c^e} \right), \qquad (4)$$

where $E_F(i)$ is an indicator variable for education of the father in pair $i$ belonging to cohort $c$, and $\mu_c^e$ is the average father education level in cohort $c$. One is added to all values of the indicator. There is a negative relationship between educational inequality and upward mobility. At times during which education is unequally distributed among fathers of each cohort of sons, upward mobility of the sons is low. The correlation, at -0.91, is extraordinarily strong. Indeed, the temporal correlation between mobility and educational inequality is stronger than between mobility and status inequality (discussed above). This finding is powerful evidence that many channels through which parents seek to improve the prospects of

Figure 13: Downward Mobility and Inequality in Education




**Notes**: Shown is expected percentile rank of sons with fathers of percentile rank of more than 0.6 and Theil index of educational inequality in the father generation. Horizontal axis shows median birth year in birth cohort; first birth year is 1330.

their children appear to not be operative when education is unequally distributed.

How about educational inequality and downward mobility? Given the evidence so far one might think that downward mobility of sons relative to their fathers is low when educational inequality in the father's generation is high. The hypothesis is that unequally distribution human capital allows high-status families to slow down the regression to the mean of their offspring. Figure 13 shows the expected rank of a son, over time, whose father has percentile rank 0.80.[57] For all birth cohorts, the son's expected rank is below 0.80 on average, which reflects regression to the mean. However, during periods of highest educational inequality, around 1600, sons fell relatively few ranks in status on average; in a later period, around 1850, when educational inequality among fathers was lowest, sons could expect to fall more than twice the earlier drop (fall of 9 versus 21 percentile ranks, respectively). The correlation of educational inequality with downward mobility is just as high as for upward mobility, demonstrating that elites were not shielded from downward mobility. The correlation of educational inequality with relative mobility ($\beta_c$) is also strong, at 0.83. Additional results in the Appendix confirm the finding that inequality in father's education is a major barrier to social mobility is robust to different definitions of the cohort (see Figure A.9 in the Appendix), different measure of education (see Appendix, Figure A.10), and employing other measures of inequality, such as the Gini index (see Appendix, Figure A.11).

[57] Computed as the expected rank of a son whose father has rank between 0.60 and 1.

### 6.4 Inequality, Mobility, and the Role of the Family

In sum, the previous results showed there is strong evidence that educational inequality is a major barrier to social mobility. I now turn to the analysis of the role of inequality at the group level. Recall that the Tongcheng genealogies cover seven dynasties with different characteristics (Table 3). The first step is to consider the extent to which status inequality is due to between-dynasty versus within-dynasty inequality. Next, I ask whether this has changed over time during the sample period. If the Yuan and Ming state required families to be registered in occupational and status categories, but these rules were not enforced by the late Ming and Qing periods, then we can expect that over time inequality between families declined.

Let $m^D_{F,c}$ denote the average percentile rank status of all fathers of dynasty $D$ in father-son pairs $i$ belonging to cohort $c$. Between-dynasty inequality in cohort $c$ is defined analogously to equation (3), with individual father percentile rank being replaced by the group (dynasty) mean, $m^D_{F,c}$:

$$T^B_c = \frac{1}{N_c}\sum_{i=1}^{N_c} \frac{m^D_{F,c}}{\mu_c} ln\left(\frac{m^D_{F,c}}{\mu_c}\right). \qquad (5)$$

The between-dynasty component is thus the inequality that would be observed if status levels were identical within each dynasty. The within-dynasty inequality is obtained by subtracting the between-dynasty component from total inequality:

$$T^W_c = T_c - T^B_c. \qquad (6)$$

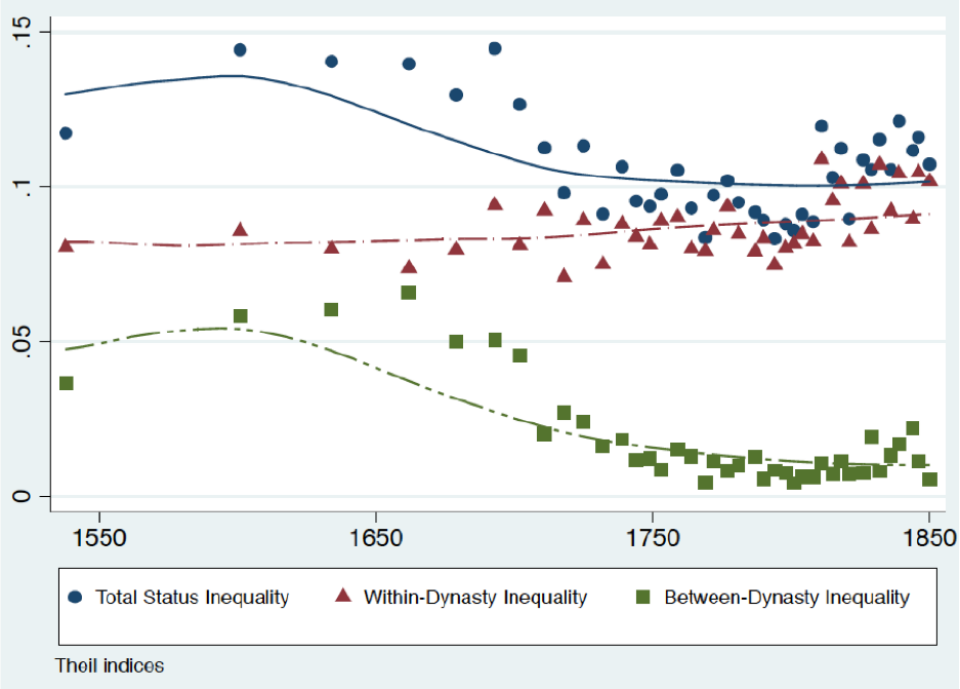
Figure 14 shows the decomposition in between- and within-dynasty inequality.

Overall inequality in father status fell over time. This pattern is shown again with the top line in Figure 14 (circles). The figure shows that within-dynasty inequality (plotted with triangles) accounts for the large majority of all inequality (84%). Not only is within-dynasty inequality the major part of total inequality, it has also risen somewhat over time. One reason for this might be the lifting of occupational barriers over time (recall section 2), to the extent that that led to more diverse occupations and status of members within the average dynasty. The evolution of overall inequality parallels the evolution of between-dynasty inequality, which has fallen over time. In the youngest birth cohort, about 1850, between-dynasty inequality accounts for only 5% of total inequality. In other words, cross-dynasty differences in status are much reduced from the 13th to 19th centuries. It is interesting to see that within- and between-dynasty inequality have different trends, and it suggests that they play a quite different role for social mobility.

#### Kin Segments within the Dynasty

Before examining the relationship of these components of inequality with social mobility, it is useful to decompose even further the inequality at the dynastic level into the segment level. Recall that each of the seven dynasties has a number of branches, called segments; all in all there are 39 of these segments. Let $m^S_{F,c}$ be the average percentile rank status of all fathers of segment $S$ in father-son pairs $i$ belonging to cohort $c$. Between-segment inequality, $T^{B,S}_c$, is defined analogously to between-dynasty inequality

Figure 14: Decomposition into Between- and Within-Dynasty Inequality




**Notes**: Total, within-, and between dynasty inequality in forty birth cohorts; year on horizontal axis is median birth year of cohort. Earliest birth year is 1330.

(equation (5)) as

$$T_c^{B,S} = \frac{1}{N_c}\sum_{i=1}^{N_c} \frac{m_{F,c}^S}{\mu_c} ln\left(\frac{m_{F,c}^S}{\mu_c}\right). \qquad (7)$$

Within-segment inequality, $T_c^{W,S}$, is given by

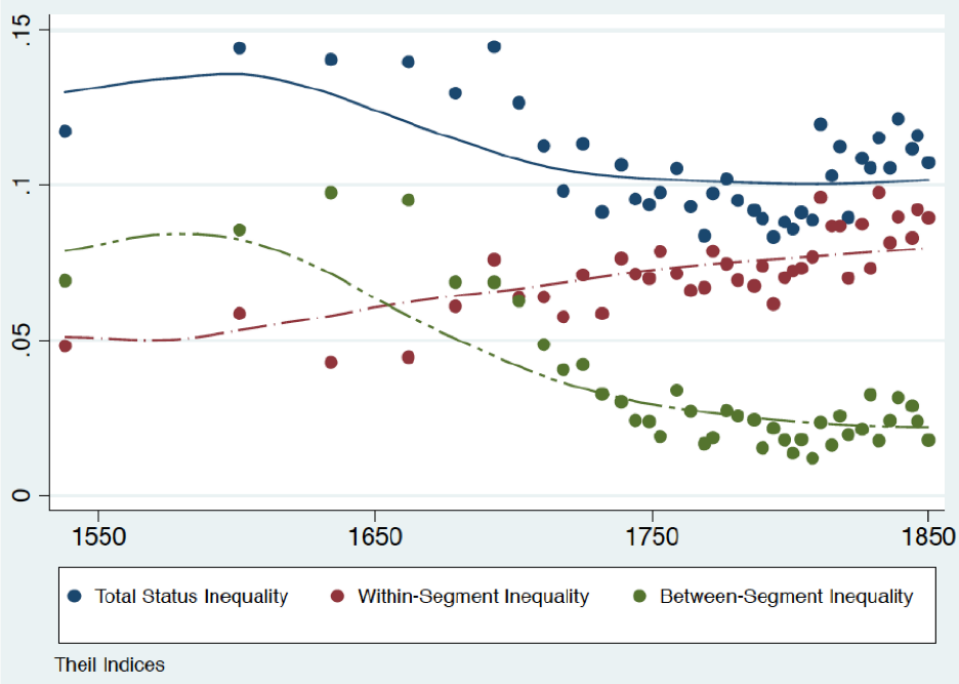
$$T_c^{W,S} = T_c - T_c^{B,S}. \qquad (8)$$

With just under 8,900 observations and forty birth cohorts in the sample, this means that on average in a given birth cohort a dynasty segment has about 6 members. By comparison, there are an average of 32 members of the same dynasty who are in the same birth cohort. It is highly plausible that direct personal ties between the segment members are stronger than between all dynasty members present in a given birth cohort, especially since segments tend to be closer related family members and are also more likely to be residing in closer proximity.

The decomposition of status at the segment level attributes a higher share to within inequality, about 70% on average, see Figure 15; note there are a greater number of groups (39 segments instead of 7 dynasties). The evolution of within- versus between inequality components are now qualitatively similar to those at the dynasty level above. The over time increase in within-inequality, however, is more pronounced at the segment level compared to the dynasty level.

The changing patterns of inequality within the same dynastic groups that lived in the same region suggest

Figure 15: Decomposition into Between- versus Within-Group Inequality at the Segment Level




**Notes**: Total, within-, and between-group inequality in forty birth cohorts at the segment level. Horizontal axis gives median birth year of cohort; birth year of first son is 1330.

that these changes occurred despite shared genetics and environment, since segments are made up of people in the dynasty that tend to live in the same village and are close-in relatives within the larger extended family.

### 6.4.1 Mobility and Inequality

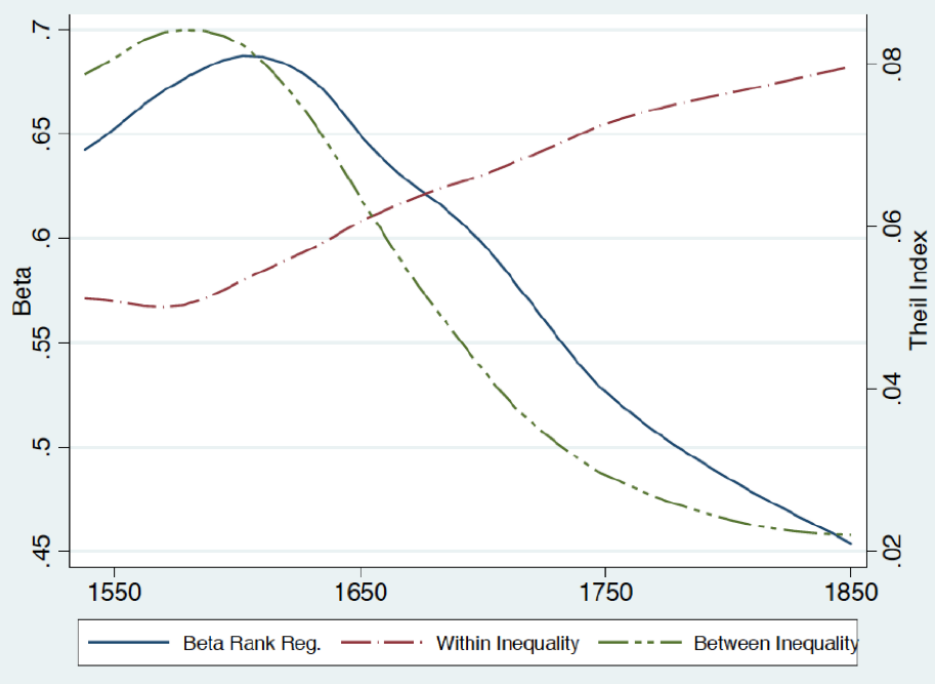
Over time, inequality within the group starts to make up a large and increasing share of total societal inequality. After decomposing inequality into these components, I now turn to the relationship of within-group and between-group inequality with mobility. Figure 16 presents the first result.

Figure 16 shows relative mobility over 12 cohorts, $\beta_c$, is positively correlated with between-group inequality (correlation of 0.68). Times of high between-group inequality (in the father generation) are times of persistence (rank mobility of son and his father). Second, Figure 16 shows that within-group inequality is negatively correlated with $\beta_c$ (correlation of -0.64). Hence, times of high within-group inequality are times of high mobility. Thus, not all inequality is a barrier to mobility. Instead, there is a striking contrast between within- versus between-group inequality, with the former being associated with higher mobility.

### 6.4.2 Implications

Despite the increase in within-dynasty inequality over time, mobility rose. Compared to an earlier period when occupational barriers were higher—in effect implying higher between-dynasty inequality—in the later

Figure 16: Social Mobility and Within- versus Between-Group Inequality




**Notes**: Shown are $\beta_c$, between-, and within-group inequality across forty birth cohorts; locally smoothed series. Horizontal axis shows median birth year in birth cohort; first birth year is 1330.

period lower occupational barriers permitted increasing heterogeneity, and inequality, within dynasties, as people within the extended family started to be able to work in different occupations. Even though total inequality fell, the shift from the Yuan-Ming society where inequality was primarily between dynasties, to the Ming-Qing society where the inequality was primarily within dynasties was consequential. Specifically, it had further positive consequences for mobility. This is because Chinese families followed clan rules that were equality promoting. Thus, inequality within the Chinese dynasty tended to raise mobility in the next generation. High inequality within-groups can be consistent with high mobility as long as the group is committed to resource sharing and investment within the dynasty.

In addition, when inequality within the group is relatively high there must be some group members of relatively high status and others of relatively low status. If resource transfers from the relatively high-status to low-status families within the group enable a child from a poor family to rise, this contributes to high status mobility (since the child's father is poor). Thus, inequality within a close-knit group increases mobility in a dynamic adjustment that occurs as inequalities within the group are smoothed out in the following generation.

Intuitively, close-knit groups are relatively effective when it comes to cooperation and resource-sharing. However, when all members of a close-knit group have the exact same resources they are less able to exploit their within-group comparative advantage because the benefits of resource sharing are in that case not as large. Moreover, given that 70% of the sample are persons without status, times of low inequality could

also typically be times where all group members have low status. Under such low-inequality conditions there are no resources to transfer from one group member to another, the group cannot make investments in some of their children, and consequently mobility is low. The finding that within-group inequality is associated with higher levels of mobility is consistent with the importance of transfers and investments of resources at the group level.

Aspects of the way society was organized point also point to the pooling of resources within the group as a likely important factor. Since total inequality of society is comprised of between group inequality and within group inequality, we can hypothesize that the forces that bring about high persistence operate primarily at the inter-dynastic level. The bi-variate correlation between mobility and between-inequality is substantial and in fact stronger than the correlation between mobility and total status inequality (0.8 versus 0.43). That between-group inequality is a stronger predictor of mobility than individual-level mobility suggests that dynasties made investments to advance the careers of the dynasty's sons but excluded those outside their own dynasty, and there were no incentives or institutions in place to enforce the sharing of resources across dynasties. If true, this suggests inter-dynastic cooperation was weak and helps to explain why, from an institution perspective, the provisions of local public goods in China was weak throughout its imperial history.

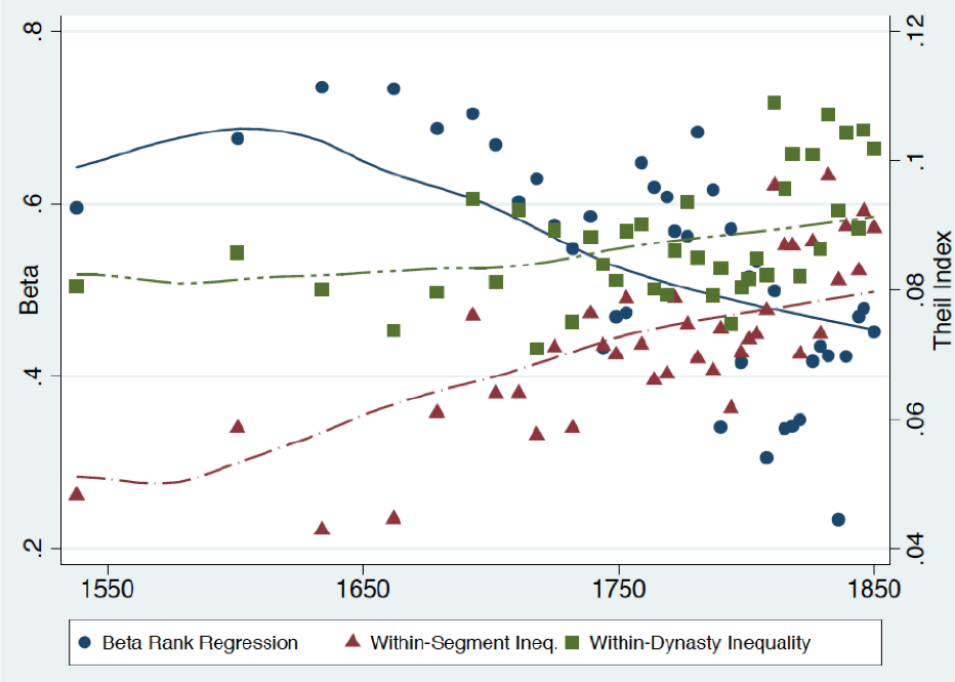
The previous figures showed that high inequality in a given cohort between, say, Ma segments and Chen segments is associated with low mobility, while high inequality within-groups is associated with high mobility. In addition, within group inequality (both within dynasties and within segments of dynasties) increased over time. We can further explore the hypothesis by asking whether close-knit groups are relatively effective in resource sharing by comparing the inequality-mobility relationship for members of the dynasty (the extended family) to the segment, which is an even tighter kin group within the dynasty.

Figure 17 shows the relationship between mobility and within-group inequality calculated at two different levels: the dynasty level versus the segment level. Recall that with seven dynasties and thirty-nine segments, one can think of the former as the relatively large and the latter as the relatively small group.

Within-group inequality for either large or small group is positively related to mobility (that is, the correlation with $\beta_c$ is negative). However, the correlation of small-group inequality with mobility is stronger than for large-group inequality: $corr(\beta_c, T_c^{W,S}) = -0.64$, while $corr(\beta_c, T_c^{W}) = -0.40$. Thus, inequality in a relatively small group is more conducive to mobility than inequality in a larger group. This result is intuitive from a resource-pooling perspective. First, a relatively small group will be more tightly knit to members of their own than to members of a larger group. Second, free-rider and monitoring problems are likely to be more limited in a smaller group, so that any investments into the groups' children subsequent to the resource transfer within the group are more effective.

Overall, it appears that the results are consistent with the role of inequality for mobility that depend on whether economic agents are incentivized by individual interests versus group interests.

Figure 17: Within-Group Inequality and Mobility - Small vs. Large Group




**Notes**: Shown are $\beta_c$, within-dynasty inequality and within-segment inequality across forty birth cohorts. Horizontal axis shows median birth year in birth cohort; first birth year is 1330.

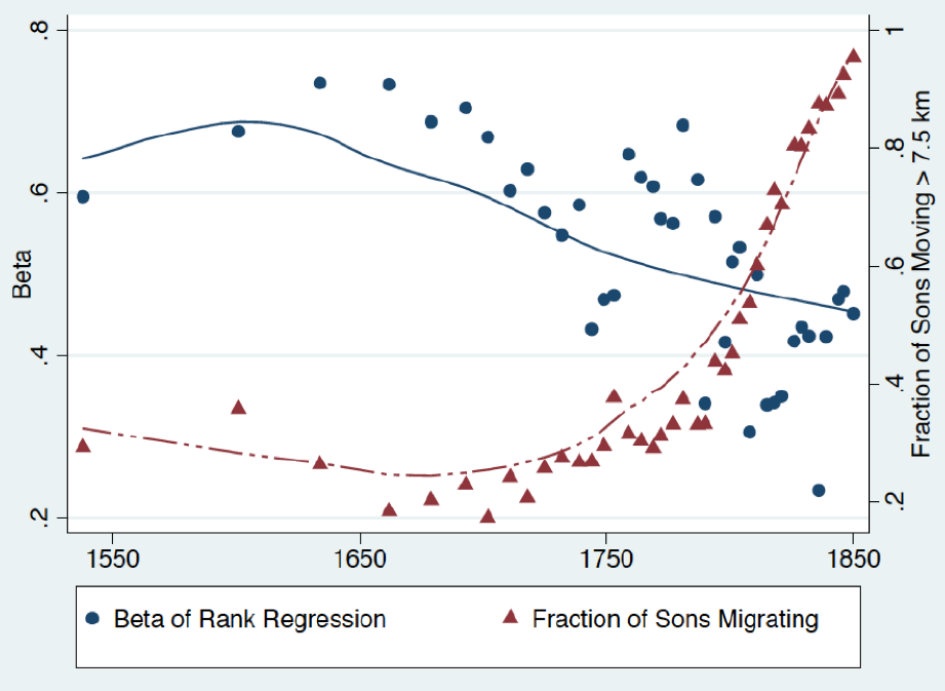
## 6.5 Spatial Mobility and Social Mobility

Ferrie (2005) shows that from the late 19th to the 20th century in the U.S. both internal migration and intergenerational occupational mobility declined, in line with the hypothesis that spatial mobility is one of the mechanisms of social mobility. The following turns to evidence on the relationship between spatial and social mobility in China, see Figure 18. The figure shows that broadly speaking, times in which a relatively high fraction of sons moves a substantial distance away from where their fathers lived are times of higher social mobility. In particular, in the late 18th and early 19th centuries the fraction of migrants (defined as moving away more than 15 *li*, approximately 7.5 kilometer) is above 50% while mobility is at its highest ($\beta_c$ below 0.5). There are some reasons for caution because the migration variable is constructed from limited burial location information. Also, the figure indicates that migration is not the only channel affecting mobility rates; for example, by the late 17th century migration rates are still falling while social mobility is increasing. Nevertheless, these results provide some new evidence that migration is an important way how social mobility changes.

## 6.6 Inequality and Mobility - Multi-variate Extensions

The following extends the analysis of the relationship between inequality and mobility to a multi-variate regression framework. The results cannot be interpreted as causal because all independent variables are likely endogenous. Nevertheless, the framework may allow us to understand better the independent role

Figure 18: Spatial and Social Mobility




**Notes**: Shown are $\beta_c$ and the fraction of sons migrating for forty birth cohorts, defined as moving away from the father's location by more than 15 *li* (7.5 kilometers), as identified by burial location. Horizontal axis shows median birth year in birth cohort; first birth year is 1330.

of various factors that may be correlated. I will also examine whether the relationship between inequality and social mobility is affected by other factors that act as omitted variables in the bi-variate analysis. Table 12 shows OLS results for relative mobility,

$$\beta_c = \pi_0 + \pi_1 X_c + \varepsilon_c, \qquad (4),$$

where $X_c$ is a vector of correlates of the estimated birth-specific $\beta_c$. Reported are standardized coefficients once variables are transformed to have mean zero and standard deviation of one (also called beta coefficients), with robust t-statistics in parentheses. The first variable introduced on the right hand side is father status inequality, $T_c$, defined in equation (3). The result in column (1) indicates that status inequality is significantly associated with persistence, confirming the patterns in Figure 10. With a t-statistic of close to 3.5, the correlation is significant at standard levels. Results for inequality in location-adjusted father status are shown in column (2). Geographic location sharpens the role of status inequality for mobility, with the $R^2$ almost twice of what is obtained in column (1). When both location-adjusted and unadjusted status inequality variables are included, the former is more powerful in accounting for variation in relative mobility (column (3)).

Next, Table 12 confirms the strong relationship between educational inequality, $T_c^e$, and mobility (column (4)). Fathers' human capital differences, a single variable, account for more than two thirds of all the variation in relative mobility. The analysis also shows that educational inequality is a more powerful

Table 12: Inequality and Mobility - Multi-Variate Analysis

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Father Status Inequality | 0.429<br>(3.449) | | 0.065<br>(0.346) | | | | |
| Geogr. Adj. Father Status Inequality | | 0.575<br>(4.432) | 0.530<br>(2.853) | | 0.157<br>(1.232) | | |
| Educational Inequality | | | | 0827<br>(10.924) | 0.739<br>(7.771) | | 0.998<br>(4.597) |
| Within Group Inequality | | | | | | −0.324<br>(−2.462) | −0.216<br>(−2.128) |
| Between Group Inequality | | | | | | 0.465<br>(3.420) | −0.351<br>(−1.644) |
| $R^2$ | 0.184 | 0.331 | 0.333 | 0.684 | 0.701 | 0.524 | 0.723 |

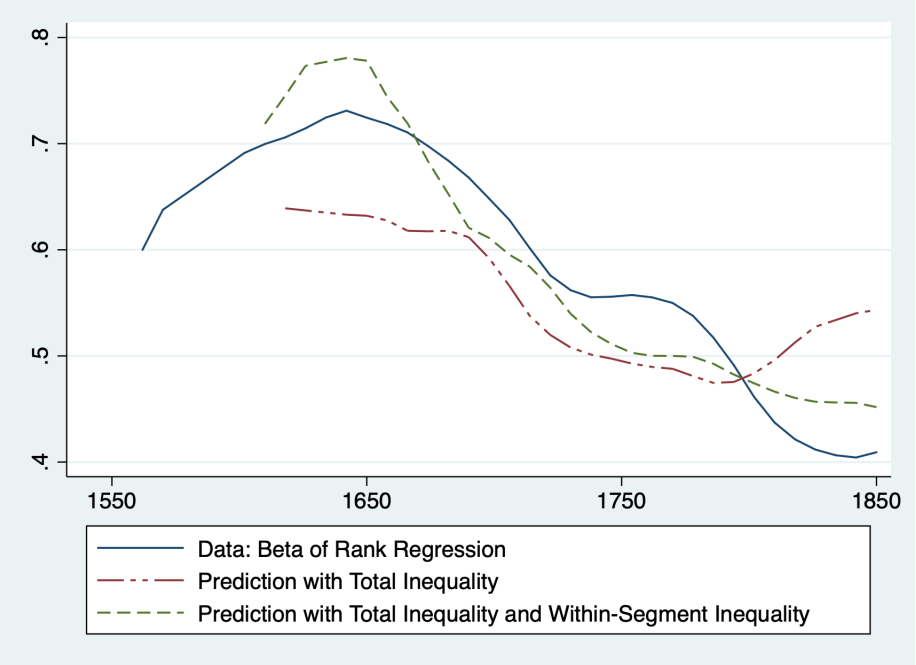
**Notes**: Dependent variable is $\beta_c$ for N = 40 birth cohorts. Estimation by OLS; standardized coefficients reported, with robust t-statistics in parentheses. Inequality is Theil index; within- vs. between- decomposition at the level of the segment.

predictor of social mobility than status inequality, even if location-adjusted (column (5)). At the same time, the strength of different variables in a regression setting will depend on the relative extent of measurement error, and if education is measured with less error than status the former will tend to dominate in the regression.

The specification in column (6) introduces within- and between-group inequality as separate variables ($T_c^{W,S}$and $T_c^{B,S}$, respectively). Doing so yields several interesting results. First, decomposing inequality in its within- and between components increases the variation in mobility that can be accounted for to about three-fold. This is important because if mobility would only reflect changes in between-dynasty inequality, the $R^2$ would not be higher than in column (1) (recall that within-group inequality is the remainder). Second, and related to the previous point, both within-inequality and between-inequality are significantly correlated with mobility, with opposite signs, in the multi-variate analysis. Third, with standardized coefficients of -0.32 and 0.47, respectively, within- and between-group inequality are quantitatively comparable in their importance for mobility. Specifically, an interquartile increase in between-group inequality is associated with a five percentage point *increase* in β, while an interquartile increase in within-group inequality is associated with a five percentage point *decrease* in β. Thus, a given change in both within- and between-group inequality leaves social mobility approximately unchanged.

Another way to quantify the importance of within-group inequality for the change in mobility is presented in Figure 19. It shows the change in $\beta_c$ across forty birth cohorts together with two predictions of this change: first, using only total father status inequality, $T_c$, and second, using both total inequality and within-segment inequality ($T_c$ and $T_c^{W,S}$, respectively). Notice that the (smoothed) change in $\beta_c$ from 1600 to 1850 is roughly 29 percentage points (approximately from $\beta_c = 0.7$ to $\beta_c = 0.41$). Employing total inequality only, predicted mobility changes from around 0.64 to 0.55, or by 9 percentage points, whereas employing both total and within-segment inequality leads to a decline of predicted mobility by about 26 percentage points betwen 1600 and 1850 (from about 0.72 to 0.46). Thus, changes in within-group inequality are crucial to quantitatively account for the observed change in mobility. Furthermore, without

Figure 19: Accounting for Mobility Changes: The Importance of Within-Group Inequality




**Notes**: Figure shows smoothed $\beta_c$ for N = 40 birth cohorts, together with two alternative predictions, one with total inequality ($T_c$) and the second with total inequality and within-segment inequality ($T_c^{W,S}$). Predictions are fitted values from OLS regressions.

factoring in changes in within-group inequality it is impossible to explain the further increase in mobility after the year 1800.

The final specification of Table 12 introduces within- and between-group status inequality together with educational inequality. The between-inequality coefficient changes signs (and weakens) while the size of the education coefficient increases; this is likely due to the positive correlation between the two variables. Within-group inequality continues to be significantly associated with higher mobility, as before. Judging by the coefficients, educational inequality is the most powerful predictor of relative mobility, a force towards persistence, while the role of within-group inequality is roughly one-fourth as large, but is associated with higher mobility. Analogous results for upward and downward mobility are shown in the Appendix, Tables A.2 and A3, respectively. They show that educational inequality is strongly negatively correlated with upward and strongly positively correlated with downward mobility. The single variable accounts for close to 85 percent of the variation in upward and downward mobility differences, up from 68 percent of relative mobility differences. Furthermore, between-group inequality is a stronger predictor of upward- and downward mobility than within-group inequality, in contrast to relative mobility where their coefficient sizes are comparable. The difference in findings is due to the fact that upward and downward mobility is calculated for particular points in the status distribution, in contrast to β which is identified from the entire status distribution.[58]

[58]Recall that the measure of upward (downward) mobility is the expected percentile rank of the son if the father has percentile rank below 0.5 (above 0.6). The regression analysis of upward and downward mobility in the Appendix also finds

One might be concerned that not only inequality but other variables—family size or average education, for example—may affect these results. The following set of results examines the role of average status, as well as average education levels and average longevity (age at death), for the results. Recall that father status is in terms of percentile rank for a given birth cohort, so that the average is 0.5 by construction. However, if one calculates the status average in a given birth cohort using the six status levels, zero to five (see Table 1), changes in average status might account for some of the variation in mobility. In addition, other factors such as average grandfather status or family size might also help to account for mobility differences. These questions are addressed in the following Tables 13 and 14.

Table 13 contains a number of additional results that examine the importance of inequality versus other potential determinants of mobility. I first show the relationship of these potential determinants of mobility by themselves before adding various measures of inequality that were discussed in the text. Status is measured here as the average of levels 0 to 5, see Table 1. Column (1) shows that $\beta_c$, the regression estimate of relative mobility, is higher at times when the average status of fathers is higher. A similar result is obtained for average levels of education (column (2)), and, albeit weaker, for longevity, see column (3). These results are broadly speaking consistent with the hypothesis that higher levels of resources, independent of their distribution, lead to lower social mobility.

Next, I turn to the grandfather generation. Column (4) shows that inequality in grandfather status is associated with less mobility, similar to inequality in father status. In contrast, geographic dispersion, measured as the average of distance from Tongcheng village in a given birth cohort, is a force towards mobility (column (5)), and so is a relatively large family size (column (6)). The latter result is consistent with recent findings that a low number of children raised the probability of success for one of them in the civil service entrance exams (Shiue 2017).

The following set of results in columns (7) to (12) add to these the within- and between-group inequality measures discussed in the text. Recall that introduced by themselves, within-inequality is associated with more mobility (negative sign) while between-inequality is a force towards persistence (positive sign). First, we see that including these inequality measures increases the coefficient of average status while the sign of the between-inequality variable turns negative, presumably because the two measures are positively correlated. The within-inequality coefficient continues to be negative, as before. Results are similar for average education levels (except that the between-inequality coefficient is not significant, column (8)). Average status level appears to be a stronger predictor of mobility compared to average education. Upon inclusion of the inequality measures, the size of the longevity and grandfather inequality coefficients shrink but remain significant at standard levels (column (9) and (10), respectively), whereas geographic dispersion and family size are not significantly correlated with mobility anymore (columns (11) and (12), respectively). Interestingly, between-group inequality turns insignificant when included together with grandfather inequality (column (10)), in contrast to within-group inequality which has consistently a negative sign (higher mobility).

The set of results shown in Table 14 introduce educational inequality instead of within- and between-

that status inequality remains a significant predictor of mobility next to educational inequality.

Table 13: Inequality and Mobility - Other Influences I

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Status Mean* | 0.66 (7.83) | | | | | | 1.25 (4.28) | | | | | |
| *Education Mean* | | 0.77 (6.67) | | | | | | 0.89 (2.69) | | | | |
| *Longevity Mean* | | | 0.43 (2.57) | | | | | | 0.27 (2.87) | | | |
| *Grandfather Inequality* | | | | 0.68 (7.71) | | | | | | 0.49 (3.44) | | |
| *Geogr. Dispersion* | | | | | −0.38 (−2.78) | | | | | | 0.10 (0.60) | |
| *Family Size* | | | | | | −0.42 (−3.85) | | | | | | −0.05 (−0.46) |
| *Within Inequality* | | | | | | | −0.87 (−4.92) | −0.25 (−2.32) | −0.26 (−2.07) | −0.42 (−3.48) | −0.31 (−2.21) | −0.30 (−2.26) |
| *Between Inequality* | | | | | | | −1.02 (−2.88) | −0.31 (−1.05) | 0.45 (3.77) | 0.00 (0.02) | 0.54 (2.68) | 0.45 (3.27) |
| $R^2$ | 0.44 | 0.59 | 0.19 | 0.46 | 0.14 | 0.18 | 0.69 | 0.63 | 0.59 | 0.61 | 0.53 | 0.53 |

**Notes**: Dependent variable is $\beta_c$; estimation by OLS. Family size is number of brothers. Standardized coefficients shown, robust t-statistics in parentheses. N = 40.

Table 14: Inequality and Mobility - Other Influences II

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| *Status Mean* | −0.13 (−0.83) | | | | | |
| *Education Mean* | | −0.73 (−2.34) | | | | |
| *Longevity Mean* | | | 0.16 (2.95) | | | |
| *Grandfather Inequality* | | | | 0.07 (0.48) | | |
| *Geogr. Dispersion* | | | | | 0.15 (1.29) | |
| *Family Size* | | | | | | −0.03 (−0.39) |
| *Educational Inequality* | 0.94 (5.55) | 1.34 (4.42) | 0.77 (11.26) | 0.77 (5.28) | 0.92 (9.57) | 0.81 (9.91) |
| $R^2$ | 0.69 | 0.71 | 0.71 | 0.69 | 0.70 | 0.69 |

**Notes**: Dependent variable is $\beta_c$; estimation by OLS. Family size is number of brothers. Standardized coefficients shown, robust t-statistics in parentheses. N = 40.

group inequality in the specifications. Recall that in a bivariate analysis educational inequality enters with a coefficient of 0.83 (Table 12). Adding educational inequality together with average status turns the latter coefficient to close to zero (column (1)). The regression with average and inequality in education has a positive coefficient on inequality and a negative coefficient on the average. Note that this is the opposite of its bivariate coefficient (see column (2), Table 14), and moreover the educational inequality coefficient is now 1.5. This suggests that while both inequality and average education are positively correlated with $\beta_c$ and with each other, inequality is the more powerful predictor, so the sign on average education is flipped. We see in column (3) that average longevity continues to be significant and positive once educational inequality is added. This is consistent with health having an independent influence on social mobility, although quantitatively, the role of educational inequality is about five time as large (recall that the coefficients are standardized). In contrast, there is no significant independent correlation of grandfather inequality, geographic dispersion, and family size anymore once inequality in education is included (columns (4), (5), and (6)).

Summarizing, the results of Tables 13 and 14 indicate that inequality is a robust and in fact the most important predictor of social mobility. In particular, there is little evidence that differences in status or education levels matter once educational inequality has been accounted for. I find similar results in the case of twelve birth cohorts, see Table A.5 in the Appendix.

# 7 Conclusions

Two premises underlie this paper. First, temporal variation in characteristics for a given society is useful to learn about determinants of social mobility. Second, the case of China, with its emphasis on kinship-based relationships in the family is a good way to learn about ways in which individuals versus groups shape social mobility. These premises led naturally to employing genealogies as the empirical source of this study. This study shows that they are a useful source, but genealogies are not the only way to study long-run social mobility. Whether the relationships found in this paper hold true in other times and places can only be determined by future research and socio-economic data on broad sets of the population for many past generations based on other sources.

I document a strong correlation between the joint distribution of father and son outcomes, in line with existing results in the literature. I also find that mobility is a multi-generational process, and that inequality, especially in education, is a major barrier to social mobility, a result that permeates much recent research and public discussion. The context of this paper helps to cut through to additional perspectives because cultural differences or ethnic composition cannot play much of a role in the current findings. Changes in social mobility over time were significant, and to a large extent we are able to see these changes only because the horizon of this study is long enough so that such changes can be observed.

Over time, mobility increased while overall inequality decreased. This paper shows that changing government institutions resulted in a shift in inequality, reducing between-dynasty inequality while raising within-dynasty inequality. The resulting changes could not have been predicted at the time, but the shift from the Yuan-Ming society where inequality was primarily between dynasties, to the Ming-Qing society where the inequality was primarily within dynasties, meant that most of the societal inequality was encountered within the family by the Qing period. Despite the increase in within-dynasty inequality over time, mobility rose. This can be understood in the context of Chinese clan rules, which customarily promoted investment in education and resource sharing. Thus, both the decline in between-group inequality and the rise in inequality within-groups contributed to increasing mobility over time.

For many countries and societies, especially in the past, extended families were an important economic unit. But the question is relevant more generally for any social organizational group. A key finding in the paper is that group-level interactions are crucial for inequality to be a force towards mobility. By showing that inequality within the group is associated with higher mobility I provide evidence that resource-shifting within a group (though outside the nuclear family) may be important for social mobility. Once we are able to identify the self-professed boundary of the "within" group, it becomes clear they behave differently towards each other than *vis-a-vis* non-related group members, and so much so that the relationship between inequality and mobility is no longer always negative. These results provide an empirical means to assess the importance of the role of informal networks in resource pooling, investment, and other economic behavior.

## References


[1] Aaronson, Daniel, and Bashkar Mazumder. 2008. "Intergenerational Economic Mobility in the United States, 1940 to 2000." *Journal of Human Resources* XLIII(1), pp. 139-172.

[2] Abramitzky, Ran and Leah Platt Boustan. 2017. "Immigration in American Economic History." *Journal of Economic Literature* 55(4), pp. 1311-45.

[3] Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. 2012. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review*, 102 (5): 1832-56.

[4] Altonji Joseph, Fumio Hayashi, and Laurence J. Kotlikoff. 1992. "Is the Extended Family Altruistically Linked? Direct Tests Using Micro Data." *The American Economic Review*, 82:5 1177-1198.

[5] Bailey, Martha, Connor Cole, Morgan Henderson, and Catherine Massey. 2017. "How Well Do Automated Linking Methods Perform in Historical Samples? Evidence from New Ground Truth." NBER Working Paper 24019.

[6] Barone, G., and S. Mocetti. 2016. "Intergenerational mobility in the very long run: Florence 1427-2011. Banca D'Italia working paper # 1060, April.

[7] Beattie, Hillary. 1979. *Land and Lineage in China: A Study of T'ung-ch'eng County, Anhwei, in the Ming and Ch'ing Dynasties.* New York: Cambridge University Press.

[8] Becker, Gary S., and Nigel Tomes. 1979. "An equilibrium theory of the distribution of income and intergenerational mobility." *Journal of Political Economy* Vol. 87(6), pp. 1153-89.

[9] Becker Gary S., Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics* 4(3): S1–39

[10] Bjorklund, A., and M. Jantti. 1997. "Intergenerational income mobility in Sweden compared to the United States." *American Economic Review* 87: 1009-1018.

[11] Black, Sandra, and Paul Devereux. 2010. "Recent Developments in Intergenerational Mobility." In *Handbook of Labor Economics*, Volume 4B, ed. Ashenfelter O., Card D., 1487–1541. Amsterdam: Elsevier.

[12] Bleakley, Hoyt and Joseph Ferrie. 2016. "Shocking Behavior: Random Wealth in Antebellum Georgia and Human Capital Across Generations." *The Quarterly Journal of Economics*, 131(3): 1455-1495.

[13] Bourguignon, François. 1979. "Decomposable Income Inequality Measures." *Econometrica,* 47(4), pp. 901–20.

[14] Bourguignon, François, and Christian Morrisson. 2002. "Inequality Among World Citizens: 1820-1992." *American Economic Review*, 92 (4): 727-744.

[15] Brandt, Loren, Debin Ma, and Thomas G. Rawski. 2014. "From Divergence to Convergence: Reevaluating the History behind China's Economic Boom." *Journal of Economic Literature*, 52 (1): 45-123

[16] Braun, Sebastian, and Jan Stuhler. 2018. "The Transmission of Inequality Across Multiple Generations: Testing Recent Theories with Evidence from Germany." *Economic Journal.*

[17] Cao, Guangling, 2016. "Ming Dai Anqing Fu Keju Rencai Yanjiu" (Research on Civil Service Examination Candidates from Anqing Fu in the Ming Dai). Ph.D. Dissertation, Anhui Normal University.

[18] Cao Shuji. 2000. Zhongguo Renko Shi Mingqing Shiqi (A Demographic History of China, the Qing Period), vol. 5. Shanghai: Fudan University Press.

[19] Card, D., J. DiNardo, and E. Estes. 2000. "The More Things Change: Immigrants and the Children of Immigrants in the 1940s, the 1970s, and the 1990s. In *Issues in the Economics of Immigration*, edited by G. Borjas, University of Chicago Press for the NBER.

[20] Chang, Chung-li. 1955. *The Chinese Gentry.* Seattle: University of Washington Press.

[21] Chang, Chung-li. 1962. *The Income of the Chinese Gentry.* Seattle: University of Washington Press.

[22] Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, Jimmy Narang. 2017. "The fading American dream: Trends in absolute income mobility since 1940." *Science*, 356 (6336): 398-406.

[23] Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States." *Quarterly Journal of Economics* 129, pp. 2633-2679.

[24] Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, Nicholas Turner. 2014. "Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility." *American Economic Review*, vol. 104(5), pages 141-47, May.

[25] Clark, Greg. 2014. *The Son Also Rises: Surnames and the History of Social Mobility.* Princeton University Press.

[26] Clark, Greg, and Neil Cummins. 2015. Intergenerational Wealth Mobility in England, 1858-2012: Surnames and Social Mobility." *Economic Journal.*

[27] Collins, William, and Marianne Wannamaker. 2015. "Up from Slavery", paper presented at University of Colorado, Boulder.

[28] Corak, Miles. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *Journal of Economic Perspectives*, 27 (3): 79-102.

[29] Dahl, Molly and Thomas DeLeire. 2008. "The association between children's earnings and father's lifetime earnings:estimates using administrative data." IRP Discussion Paper 1324-08.

[30] Daqing luli (The Great Qing Code), 1740 edition.

[31] Daqing huidian shili (1899 edition) Ch. 370.

[32] De Moor, Tine and Jan Luiten Van Zanden. 2010. "Girl Power: the European marriage pattern and labour markets in the North Sea region in the late medieval and early modern period." *The Economic History Review*, 63(1).

[33] Durand, John D. 1977. "Historical Estimates of World Population: An Evaluation." *Population and Development Review* 3(3): 253-296.

[34] Eberhard, Wolfram. 1962. *Social Mobility in Traditional China.* Leiden: E.J. Brill.

[35] Elman, Benjamin. 2000. *A Cultural History of Civil Examinations in Late Imperial China.* Univ. of California Press.

[36] Fang Xu-ling. 2010. "Qingdai Tongcheng Jinshi Shuliangkao" (A study of the number of successful jinshi candidates from Tongcheng during the Qing Period) , *Journal of Anqing Teacher's College*, 29(5).

[37] Fei, H.T. 1946. "Peasantry and gentry: an interpretation of Chinese social structure and its change. *American Journal of Sociology.* 52(1).

[38] Fei, John C. H. , and Ts'ui-jung Liu. 1982. "The Growth and Decline of China Family Clans", *Journal of Interdisciplinary History* XII:3 (Winter): 375-408.

[39] Ferrie, J. P. 1996. "A New Sample of Males Linked from the 1850 Public Use Micro Sample of the Federal Census of Population to the 1860 Federal Census Manuscript Schedules." *Historical Methods* 29(4): 141-156.

[40] Ferrie, J. P. 2005. "The End of American Exceptionalism? Mobility in the United States since 1850.".*Journal of Economic Perspectives* Vol.19(3): 199-215.

[41] Fields, G. S. & Ok, E. A. 1999. "The measurement of income mobility: An introduction to the literature" [Electronic version]. In J. Silber (Ed.) *Handbook on income inequality measurement* (pp. 557-596). Norwell, MA: Kluwer Academic Publishers.

[42] Genicot, Garance and Debraj Ray, 2017. "Aspirations and Inequality." *Econometrica*, 85(2): 489-519.

[43] Greif, Avner, and Guido Tabellini, 2017. "The clan and the corporation: Sustaining cooperation in China and Europe." *Journal of Comparative Economics* 2017, 1–35.

[44] Griliches, Zvi. 1979. "Sibling Models and data in Economics: Beginnings of a Survey." *Journal of Political Economy*, 87(5) Part 2: Education and Income Distribution, S37-S64.

[45] Hajnal, John. 1965. "European marriage patterns in perspective." In *Population in History*, ed. D.V. Glass and D. E. C. Eversley. London.

[46] Han, S., & Mulligan, C. 2001. Human Capital, Heterogeneity and Estimated Degrees of Intergenerational Mobility. *The Economic Journal*, 111(470), 207-243.

[47] Heckman, James, and Stefano Mosso. 2014. "The Economics of Human Development and Social Mobility". Working Paper, University of Chicago, May.

[48] Ho, Ping-ti. 1967. *The Ladder of Success in Imperial China: Aspects of Social Mobility (1368-1911).* New York: Columbia University Press.

[49] Hoen, M., S. Markussen, and K. Roed. 2018. "Immigration and Social Mobility". IZA Discussion Paper No. 11904, October.

[50] Keller, Wolfgang, Ben Li, and Carol H. Shiue. 2011. "China's Foreign Trade: Perspectives from the past 150 Years." *The World Economy*, 34(6).

[51] Krueger, A. B. 2012. "The Rise and Consequences of Inequality in the United States", speech at Center for American Progress, January 12. https://cdn.americanprogress.org/wp-content/uploads/events/2012/01/pdf/krueger.pdf

[52] Lee, C.I., and G. Solon. 2009. "Trends in Intergenerational Income Mobility." *Review of Economics and Statistics.*

[53] Li, Zhou Wang. Guochao like timing beilu chuji . 1746 edition.

[54] Lindahl, M., Palme, M., Sandgren-Massih, S., & Sjögren, A. 2014. A Test of the Becker-Tomes Model of Human Capital Transmission Using Microdata on Four Generations. *Journal of Human Capital*, 8(1).

[55] Lindahl, M., M. Palme, S. Sandgren Massih, and A. Sjögren. 2015. Long-Term Intergenerational Persistence of Human Capital: An Empirical Analysis of Four Generations. Journal of Human Resources, Volume 50, Number 1, pp. 1-33

[56] Liu Hui-Chen Wang. 1959. *The Traditional Chinese Clan Rules.* New York: J.J. Augustin Incorporated Publisher.

[57] Liu Ts'ui-jung. 1978. "Chinese genealogies as a source for the study of historical demography", In *Studies and essays in commemoration of the golden jubilee of Academia Sinica.* Taipei: Academia Sinica.

[58] Long, Jason, and Joseph Ferrie. 2013. "Intergenerational Occupational Mobility in Great Britain and the United States since 1850." *American Economic Review*, 103(4): 1109-37.

[59] Pan, Guangdan. 1929. "Zhongguo jiapu xue lueshi" (A historical outline of Chinese genealogical studies), Dongfang Zazhi (The Eastern Miscellany). 26(1): 107-120.

[60] Piketty, T. 2000. "Theories of Persistent Inequality and Intergenerational Mobility." *Handbook of Income Distribution*, eds. A.B. Atkinson and F. Bourguignon.

[61] Pomeranz, Kenneth. 2001. *The Great Divergence: China, Europe, and the Making of the Modern World Economy.* Princeton University Press.

[62] Pomeranz, Kenneth. 2011. "Ten Years After: Responses and Reconsiderations." *Historically Speaking*, 12(4).

[63] Schultz, T. W. 1961. "Investment in Human Capital." *American Economic Review* Vol. 51(1), pp. 1-17.

[64] Shiue, Carol H. 2002. "Transport Costs and the Geography of Arbitrage in Eighteenth Century China." *American Economic Review*, 92(5): 1406-1419.

[65] Shiue, Carol H. 2004. "Local Granaries and Central Government Disaster Relief: Moral Hazard and Intergovernmental Finance in Eighteenth and Nineteenth Century China." *Journal of Economics History*, 64(1): 101-125.

[66] Shiue, Carol H. 2005. "The Political Economy of Famine Relief in China, 1740-1820." *Journal of Interdisciplinary History*, 36(1): 33-35.

[67] Shiue, Carol H. 2016. "A Culture of Kinship: Chinese Genealogies as a Source for Research in Demographic Economics." *Journal of Demographic Economics*, December 2016, 82(4), pp. 459-482.

[68] Shiue, Carol H. 2017. "Human Capital and Fertility in Chinese Clans Before Modern Growth." *Journal of Economic Growth.*

[69] Shiue, Carol H., and W. Keller. 2007. "Markets in China and Europe on the Eve of the Industrialization." *American Economic Review*, 97(4):1189-1216.

[70] Shorrocks, A.F. 1978. "The Measurement of Mobility." *Econometrica*, 46(5): 1013-1024.

[71] Shorrocks, A.F. 1980. "The Class of Additively Decomposable Inequality Measures." Econometrica, 48(3): 613-625.

[72] Solon, Gary. 1999. "Intergenerational Mobility in the Labor Market", Chapter 29 in Orley C. Ashenfelter and David Card (eds.), *Handbook of Labor Economics,* vol. 3A, Amsterdam, North-Holland, 1999.

[73] Solon, G., 2004. A model of intergenerational mobility variation over time and place. In: Corak, M. (Ed.), *Generational Income Mobility in North America and Europe.* Cambridge University Press, Cambridge, pp. 38–47 (Chapter 2).

[74] Solon, Gary. 2014. "Theoretical Models of Inequality Transmission across Multiple Generations", *Research in Social Stratification and Mobility* 35, March, pp. 13-18.

[75] Solon, Gary. 2018. "What do we know so far about multigenerational mobility?" *Economic Journal* Vol. 128, pages F340-F352.

[76] Telford, Ted A., et al. 1983. An Annotated Bibliography of Chinese Genealogies on Microfilm at the Genealogical Society of Utah. Taipei: Cheng Wen.

[77] Telford, Ted A. 1986. "A survey of social demographic data in Chinese genealogies." *Late Imperial China* 7: 118-48.

[78] Telford, Ted A. 1992. "Covariates of Men's Age at First Marriage: The Historical Demography of Chinese Lineages." *Population Studies*, 46(1): 19-35.

[79] Telford, Ted A. 1995. "Fertility and Population Growth in the Lineages of Tongcheng County, 1520-1661." *Chinese Historical Microdemography*, ed. S. Harrell. Berkeley: University of California Press.

[80] Theil, H. (1967). *Economics and Information Theory*. Amsterdam: North Holland

[81] Townsend, R. 1994. "Risk and Insurance in Village India." *Econometrica*, 62(3), 539-591.

[82] Wang, Jing. 2017. "Qing dai Tongcheng jinshi yanjiu" (Research on the jinshi from Tongcheng), Ph.D. Dissertation. Zhejiang University.

[83] Watson, James. 1982. "Chinese kinship reconsidered: Anthropological perspectives on historical research." *China Quarterly* 92:589-622.

[84] Yang, C.K. 1959a. *A Chinese village in early communist transition*. Cambridge, MA: Technology Press, MIT.

[85] Yang, C.K. 1959b. *The Chinese family in the communist revolution*. Cambridge, MA: Technology Press, MIT.

[86] Zhao, Zongwei. 1997. "Long-Term Mortality Patterns in Chinese History: Evidence from a Recorded Clan Population." *Population Studies*, 51(2): 117-127.

# A For Online Publication - External Validity and Potential Biases

## A.1 Representativeness - Top Status and Other Status Levels

One way to gauge the representativeness of the Tongcheng sample is to look at the number of top status individuals as a percentage of the population. Given the genealogy is a written document, if literate individuals only recorded information about themselves their and immediate kin, the percentage of top status people in the genealogy should be very high. Alternatively, if genealogies recorded extended family who were not of high status—rules of ritual say that all adult male members are eligible, regardless of education or status—the percentage should typically be low.

In his classic study based on national lists of *jin-shi*, which are extremely reliable, Ho (1962) reports that during the Qing in Anhui there were 41 *jin-shi* per one million population, or, 0.0041 percent. Anhui is very close to the national average. There were regional variations. The province of Anhui, it should be

noted, was below the provincial average in terms of *jin-shi* per capita in Qing China (Ho 1962, p. 228). In my sample of Tongcheng individuals compiled from genealogies, there were a total of 14 *jin-shi* during the Qing, which comes to 0.045 percent of the population in the data.[59] Thus, there are about ten times more *jin-shi* in the Tongcheng sample than in Qing Anhui overall. The reason for this is that Tongcheng was an important urban center in Anhui that had a well-known reputation for producing high-status individuals during the Qing (Beattie 1979). At the same time, *jin-shi* were rare, and some parts of Anhui province did not produce a single *jin-shi* over centuries of time. Furthermore, Tongcheng was not among the areas of China where top status individuals were most prevalent. Some areas had a number of *jin-shi* that was higher by an order of magnitude compared to Tongcheng.[60] Therefore, while the number of men with the highest status level in Tongcheng was higher than in the local surrounding area, Tongcheng was a fairly typical region. It was noteworthy at a local, perhaps provincial level, but it was not an unusual place in China.

Moreover, it is important to note that the variation in *jin-shi* across dynasties in the Tongcheng sample dwarfs differences in the population, for example across provinces. At the top of the list, the Ma dynasty had 9 *jin-shi* relative to 627 men, a ratio of 1.4%, whereas other Tongcheng dynasties do not have a single *jin-shi*. Put simply, the potential sample selection in the genealogy, as a genre, is likely to be minor in comparison to the rather large and pronounced differences that we see in the achievements of different dynasties.

This analysis of the representativeness of top status in the sample can be supplemented by examining the representation of other status levels. There may be a tendency to exaggerate the status of people in the genealogy, or to drop the poorest segments of the dynasties—in both cases there would be many more officials or educated people in the genealogical sample than in the society at large. While there exists no generally agreed-upon status classification for China, another comparison to other sources is that in the Liaoning Eight Banner sample 98% of males had "No Status" while 2% were "Officials".[61] This compares to about 71% of men having "No Status" in the Tongcheng sample, while about 1.4% have an official position. The relatively high fraction of "Officials" in the Eight Banner population might be related to the fact that it was a less densely populated area in the North of China. In addition, according to estimates from Ho (1962), during the Qing Dynasty, the number of Bannermen who were awarded *jin-shi* status as a percentage of the population ranked first among all provincial regions of China.

[59]There are 8,291 married men during the Qing in the sample. Telford (1986) finds that the proportion of unmarried men in Tongcheng during a somewhat earlier period of the Ming was above 20%. I assume that 20% of all men did not marry , and that the Qing population was composed of below-age-of-marry/men/women to one-third each. This gives a scaling factor of 3.75: 14 *jin-shi*/(8,291 x 3.75) = 0.045 percent. If there are 20% of men not marrying, and there is universal marriage of women, then there must be 20% fewer daughters than sons, and 20% fewer women than men.

[60]Zhejiang and Jiangsu were among the provinces with high densities of *jin-shi*. Ho (1962) reports that a single prefecture in China could have as many as 1,004 *jin-shi* during all years of the the Qing (Ho 1962, p. 247). With typically seven counties to a prefecture, this means that there could be as many as 1,004/7 = 143 *jin-shi* per county during the Qing. With 14 *jin-shi* in Tongcheng county, The county was not exceptional.

[61]Source: Author's computations from the China Multigenerational Dataset, Liaoning1749-1909, http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/27063.

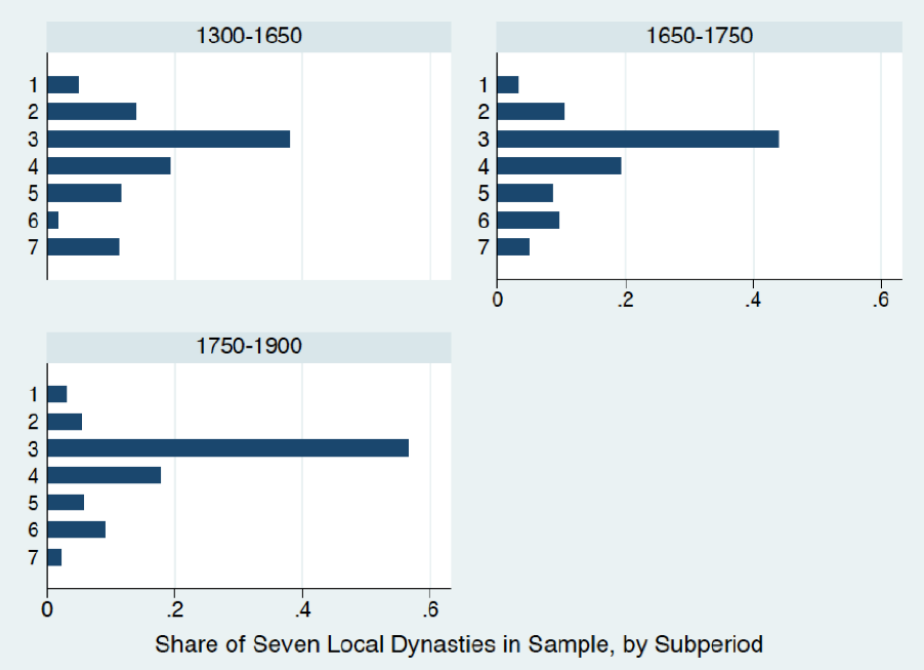
## A.2 Selection, Recall, and Survivor Bias

Overall, the figures from the previous subsection tend to confirm the conclusion in the paper that the Tongcheng sample is broadly representative the distribution of socioeconomic status in China as a whole. Variation across dynasties is useful for examining the factors that might affect whether a particular person, or observation, is included in the genealogy. As noted earlier, there were a number of different reasons why the genealogical tradition emerged. However, one concern is that genealogies might often begin with a particularly noteworthy man, who then becomes the progenitor of the dynasty. Part of his noteworthiness might come from a high level of education, which is one of the most important signs of status and one of the most consistently reported characteristics of noteworthy persons. Alternatively, perhaps later generations were more likely to select a noteworthy progenitor. In either case, the implication would be a trend of declining status over time.

In the Tongcheng genealogical sample, there are three dynasties whose records begin with an educated progenitor: the Chen (progenitor born in 1298), the Wang (1358), and the Ma (1408). However, the status of these three progenitors was not more than what might be considered an intermediate level, not the highest level (*jin-shi*). For the other four Tongcheng dynasties, the highest levels of status are typically found nine generations after the inception of the dynasty. Thus, the status patterns in the data are not simply driven by particularly noteworthy individuals that started the dynasty records as progenitors. Figure A.1 shows the size of the seven dynasties over time. Relative size does change somewhat, in particular the Wang and Zhao dynasties grow while the Ma and Zhou dynasties shrink, although all dynasties have a sizable presence in virtually all periods. Selection would also arise if genealogical records contain many more entries of success compared to failure. If a genealogy were more apt to record success, we might expect that it would do this on multiple measures. One way to check for this would be to see if dynasty size (more male births as an indicator of success) and average status (higher status as another type of entry of success) are correlated. The correlation between the first two columns in Table 3 is virtually zero, providing evidence that there was is no strong relationship between dynasty size and its average status. Thus, there is no evidence that on average, more successful dynasties have included more entries in their genealogies. We can also check if this kind of effect might be present over time to see if it is true that periods during which a dynasty is successful are also those when relatively many dynasty members are recorded. Breaking down the overall sample period of 1300 to 1900 into twelve subperiods that typically have 25-year windows, as in the text, I find that the correlation between status and the number of dynasty members in that time window is close to zero (and negative, -0.10).

Furthermore, because the updating of the genealogy was retrospective one might believe that it is in times right after a dynasty had been relatively successful when a relatively high number of dynasty members would be recorded. This might happen if the appearance of a high-status individual (success) correlates with a better memory of all the family members who were related to this locally famous individual, compared to periods when no one in the dynasty was particularly successful or famous. However, there is no positive relationship between the number of dynasty members recorded and the average dynasty status either (the correlation is insignificant at -0.09).

Figure A.1: Relative Size of Dynasties Over Time




**Notes**: Dynasty names are as follows: 1 is Chen, 2 is Ma, 3 is Wang, 4 is Ye, 5 is Yin, 6 is Zhao, and 7 is Zhou.

Additionally, it is possible to check whether high-achievement dynasties tend to be overrepresented towards the end of the sample, as would be the case if there was survivor bias. If so, one would expect that dynasties with average status account for a relatively large share of the post-1800 observations in the Tongcheng data. Across dynasties, however, I do not find a strong relationship between average status and the share of post-1800 observations (insignificant correlation of 0.07).

Overall, the previous subsection argued that there is little evidence in the data that suggests selection and other biases existed over a range of the most potentially concerning issues. Discussion of additional aspects of the Tongcheng genealogical data can be found in Shiue (2016).

## A.3 Alternative Status Definitions and Samples

This section examines the robustness of the mobility results for alternative samples and status definitions; see Table A.1 for the results. Note that the classification with six status levels has comparatively many distinctions in the top percentiles (see Figure 5), which comes from the fact that there is more discriminating information on these top status individuals. How important are the top status individuals in driving the results? The answer is not much. When I drop the top and bottom 2.5 % to focus on the central 95 percent of father-son pairs, the results turn out to be not very different (Table A.1, column (2)). In contrast, results would be quite different if one simply drops observations when father status is zero. Relative mobility then would be estimated at $\beta = 0.68$, compared to 0.53 (not shown). Next, I employ the percentile rank approach with the finer status gradations available in the biographies (23 status groups, as shown in Table 1). The relative mobility estimate is now 0.57. Perhaps more importantly, the relationship

Table A.1: Alternative Status Definitions and Samples

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| | Baseline | Central 95% | Rank Status | Six Discrete | Purchased and Exam- |
| | | | with 23 Levels | Status Levels | Passed Same Status |
| $\beta$ | 0.528 | 0.506 | 0.570 | 0.552 | 0.528 |
| $E\left[R_S(i)\,\middle|\,R_F(i)=0.25\right]$ | 0.418 | 0.417 | 0.416 | n/a | 0.419 |
| $E\left[R_S(i)\,\middle|\,R_F(i)=0.80\right]$ | 0.620 | 0.610 | 0.626 | n/a | 0.619 |
| $R^2$ of Average Relation | 0.841 | 0.806 | 0.712 | 0.977 | 0.839 |

Notes: Estimation of β (equation (1)) by OLS; $E\left[R_S(i)\,|R_F(i)=0.25\right]$ is the expectation of percentile rank of sons with fathers at percentile rank $< 0.5$. $E\left[R_S(i)\,|R_F(i)=0.80\right]$ is the expectation of percentile rank of sons with fathers at percentile rank $> 0.6$. N = 8,893 except in column (2) where N = 8,760.

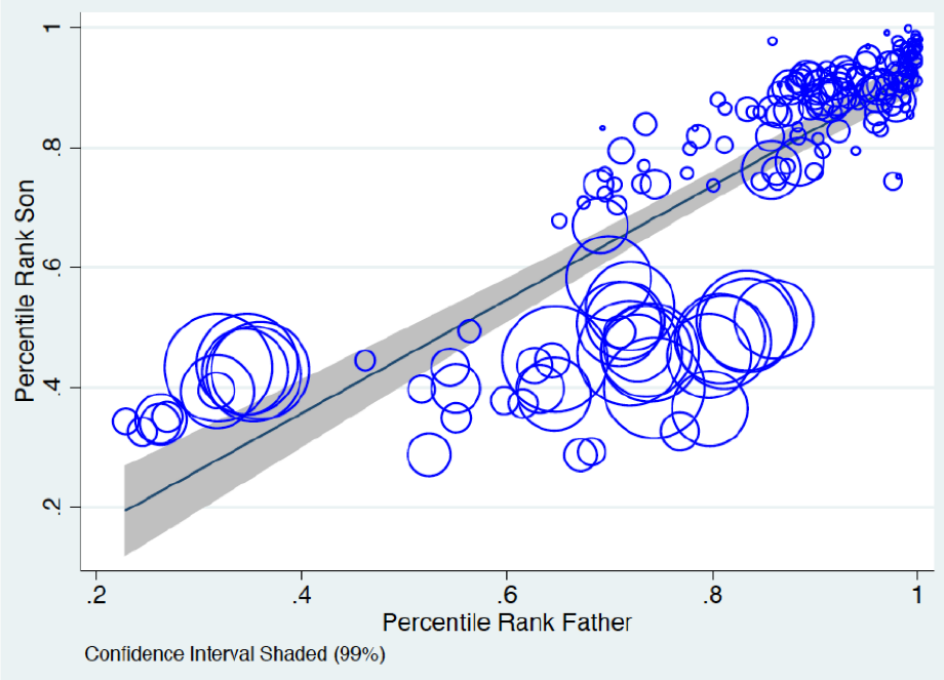
between average rank father and average rank son is more non-linear than with the fewer status groups in the baseline specification, and hence the $R^2$ is lower, see Figure A.2. The figure provides evidence that employing twenty-three categories may be too many. Instead of ranks in the status distribution, one may employ a smaller number of discrete status levels, and six discrete status levels in the baseline are used (values 0 to 5). This makes specific assumptions on how far one status level is from another; that is, the step from level zero to one is exactly as large as the step from level three to four, for example. Employing this I find that it would lead to a somewhat higher level of persistence ($\beta = 0.55$), see column (4) in Table A.1.

Finally, one of the status levels coded in the data refers to those who purchased a degree as a way to get a minor official position. As discussed in the text (section 2), status in China was often obtained with a degree by passing the official government entrance exam. People who were not able to pass the exam, but were sufficiently wealthy could buy their way in. If one were to assume that a purchased degree or office would give the holder a comparable status—perhaps because the purchase proves that the man has the resources to do so—this does not change the estimated mobility substantially (see Table A.1, column (5)). One reason for this is that the re-coding of status affects only about 1.5% of the sample.

# B Additional Results on Mobility over Time

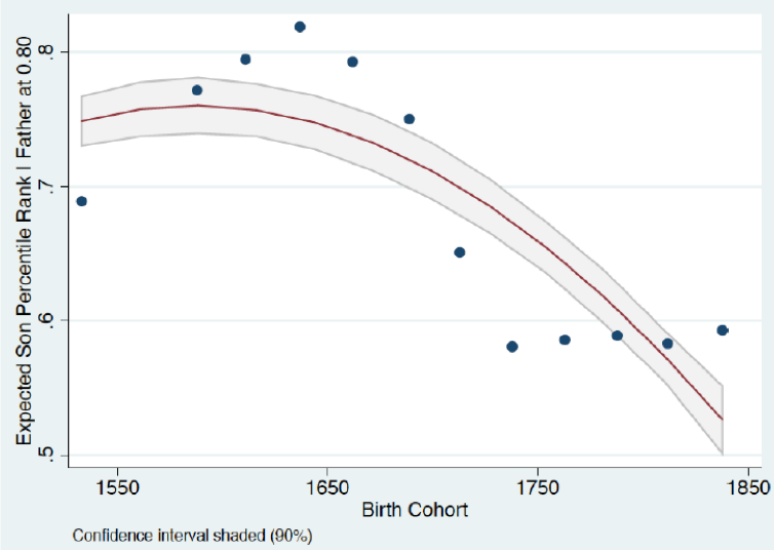
Figure A.3 shows the evolution of the expected rank of sons whose father was at the 80th percentile of the status distribution. Around 1600, such a son could expect to be at the 75th percentile, whereas by the early 19th century the son could expect to be closer to the 55th percentile, a 27 percent drop in expected rank for these sons from high-status families. Figure A.4 presents evidence for a higher level of social mobility towards the end of the sample period in the case of birth cohorts of equal size, similar to what was shown for the case of birth cohorts with an equal number of years (see main text). Figure A.5 shows the results for downward mobility in the case of birth cohorts with equal numbers of observations. The expected rank of a son from a family at rank 0.8 falls from by 14 percentile ranks between years 1550 and 1850, from about 0.73 to 0.59. The results the measures of upward and downward mobility confirm

Figure A.2: Average Father Rank vs. Average Son Rank with 23 Status Levels



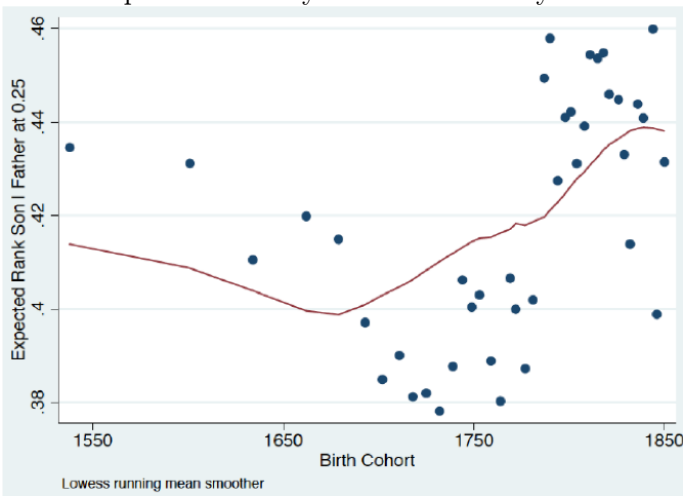

**Notes**: Relationship between average percentile rank father and average percentile rank son for 23 status levels and twelve birth cohorts. Size of marker is proportional to number of observations.

Figure A.3: Downward Mobility over Time




**Notes**: Shown is the expected percentile rank of sons with father percentile rank status $>$ 0.6 for twelve birth cohorts. Horizontal axis gives median birth year in cohort; earliest birth year is 1330.

Figure A.4: Upward Mobility over Time - Forty Birth Cohorts




**Notes**: Shown is expected percentile rank of son with fathers in the lower half of rank distribution for forty birth cohorts. Horizontal axis gives median birth year in cohort; earliest birth year is 1330.

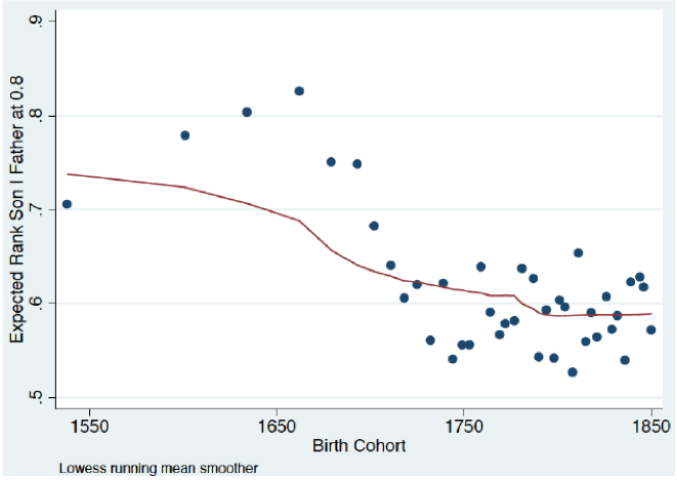
the finding of increasing mobility in terms of rank regression coefficient $\beta_c$. The following Figure A.6 complements the analysis in the text for alternative measures of absolute upward mobility over time by showing results on the evolution of alternative measures of downward mobility. The baseline results for downward mobility are shown in Figure A.6 with a solid line. Downward mobility increased over time: sons of fathers with rank above 0.6 could expect a rank of just under 0.7 around 1600 whereas by 1800 such sons could only expect a rank of about 0.6. Results are qualitatively similar for sons of fathers with rank above 0.8 (dash-dot) and above 0.5 (dash-dot-dot). Figure A.6 presents also evidence on downward mobility between sets of quintiles. In particular, the probability that a son from a family in the top two quintiles (above 0.6) ends up in the lower two quintiles (below 0.4) is about 20% in 1600 but more than 50% by the beginning of the 19th century. The above results indicate that trend towards higher downward mobility noted in the text is robust to alternative measures.

Further, instead of status distributions that are specific to each cohort, I also estimated changes in mobility when each son is ranked in the status distribution relative to sons over the entire sample period of six hundred years. The result, shown in Figure A.7, is quite similar to the findings with birth-cohort specific status ranks: $\beta$ is estimated around 0.65 until about the year 1700, before the parameter estimate $\beta$ drops to around 0.4 by the early 19th century. The results in Figure A.7 indicate that the increase in mobility over the sample period is not driven by the way in which the father and son were ranked in their status relative to others in their specific birth cohort.

# C Drivers of Mobility: Additional Results

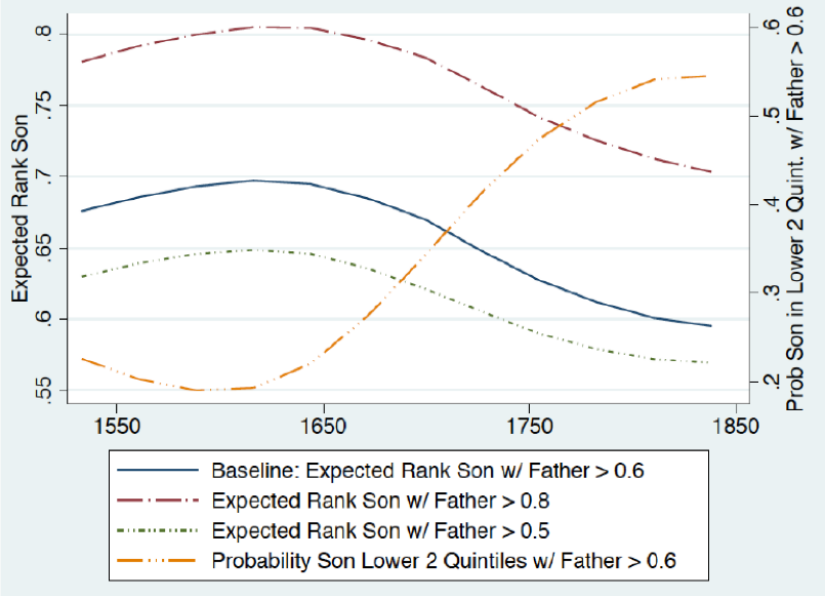
The definition of geographically-adjusted father status in the text employs a distance decay of $\delta = 0.2$. The following Figure A.8 presents results for two other values of $\delta$, namely $\delta = 0.1$ and $\delta = 0.05$. The

Figure A.5: Downward Mobility over Time - Forty Birth Cohorts



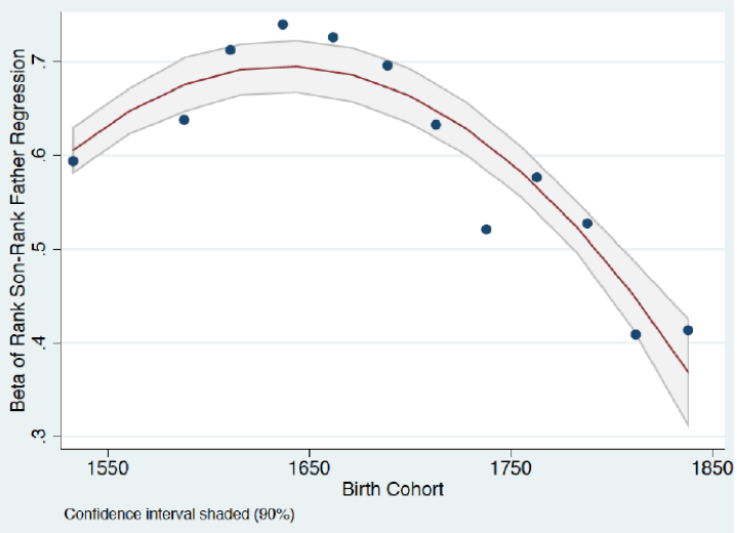

**Notes**: Shown is expected percentile rank of son with fathers of percentile rank 0.6 and higher for forty birth cohorts. Horizontal axis gives median birth year in cohort; earliest birth year is 1330.

Figure A.6: Downward Mobility over Time - Alternative Measures



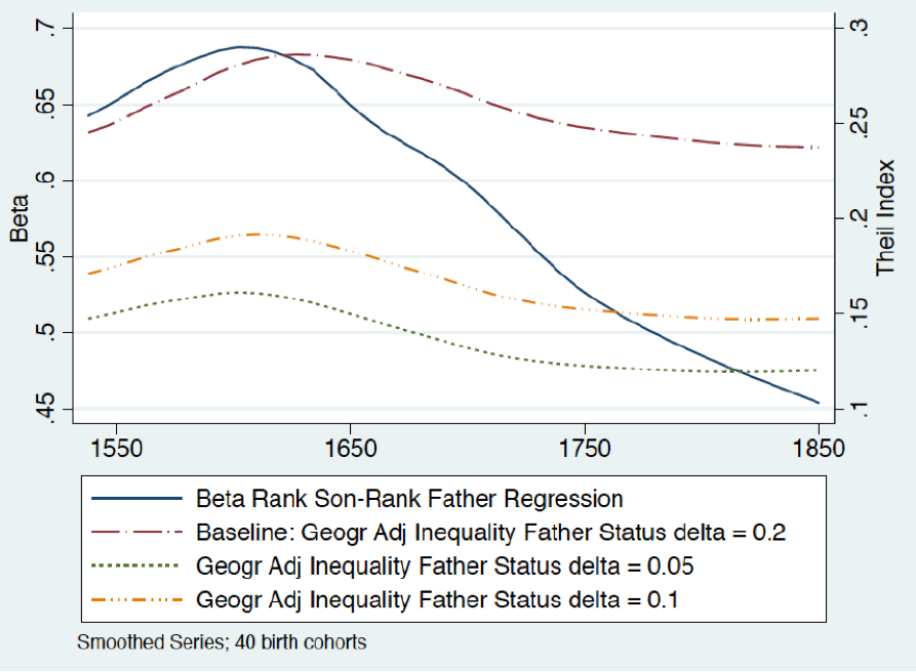

**Notes**: Shown is expected percentile rank of son with fathers of percentile rank greater than 0.6, greater than 0.8, and greater than 0.5, as well as the probability that a son is the lower two quintiles of the status distribution given the father has percentile rank above 0.6. Locally smoothed series across birth cohorts. Horizontal axis gives median birth year in cohort; earliest birth year is 1330.

Figure A.7: Mobility over Time with Global Status Distribution



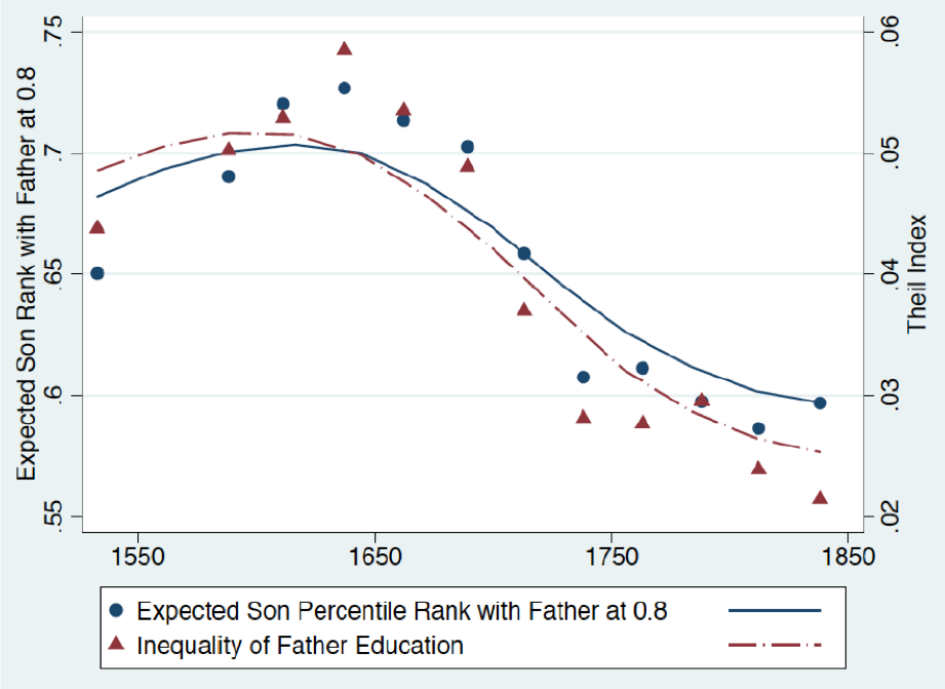

**Notes**: Shown is β (equation (1)) for twelve birth cohorts based on constant distributions of father and son status for the entire sample period; the first birth year is 1330.

Figure A.8: Location and Status Inequality - Alternative Assumptions




**Notes**: Shown is locally smoothed β (equation (1)) and alternative measures of geographically-adjusted father status inequality for forty birth cohorts; the first son birth year is 1330.

Figure A.9: Educational Inequality and Downward Mobility over Time - Twenty-Five Year Birth Cohorts



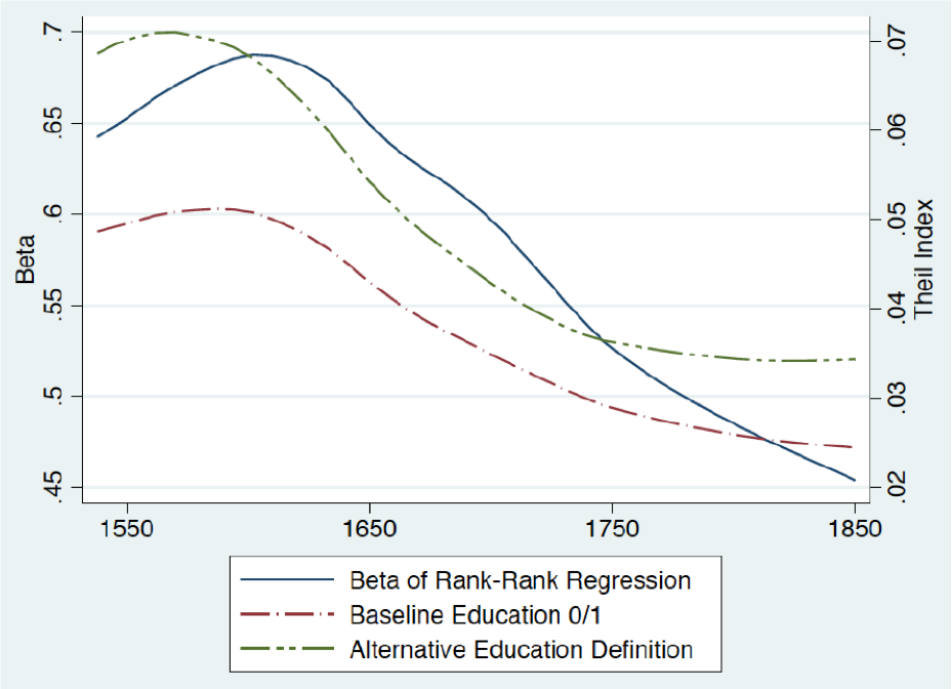

**Notes**: Shown is the expected rank of sons with fathers at percentile rank status $> 0.6$ and the Theil index of father education for twelve birth cohorts. Horizontal axis shows median birth year in cohort; birth year of first son is 1330.

relation of these alternative geographically-adjusted father status inequality measures with mobility are similar as in the baseline.

Next, I present the relationship between downward mobility and educational inequality for the case of twelve birth cohorts, to examine the robustness of the results with respect to the definition of birth cohort; see Figure A.9. Figure A.9 indicates that there is a strong relationship between the two variables irrespective of how birth cohorts are defined.

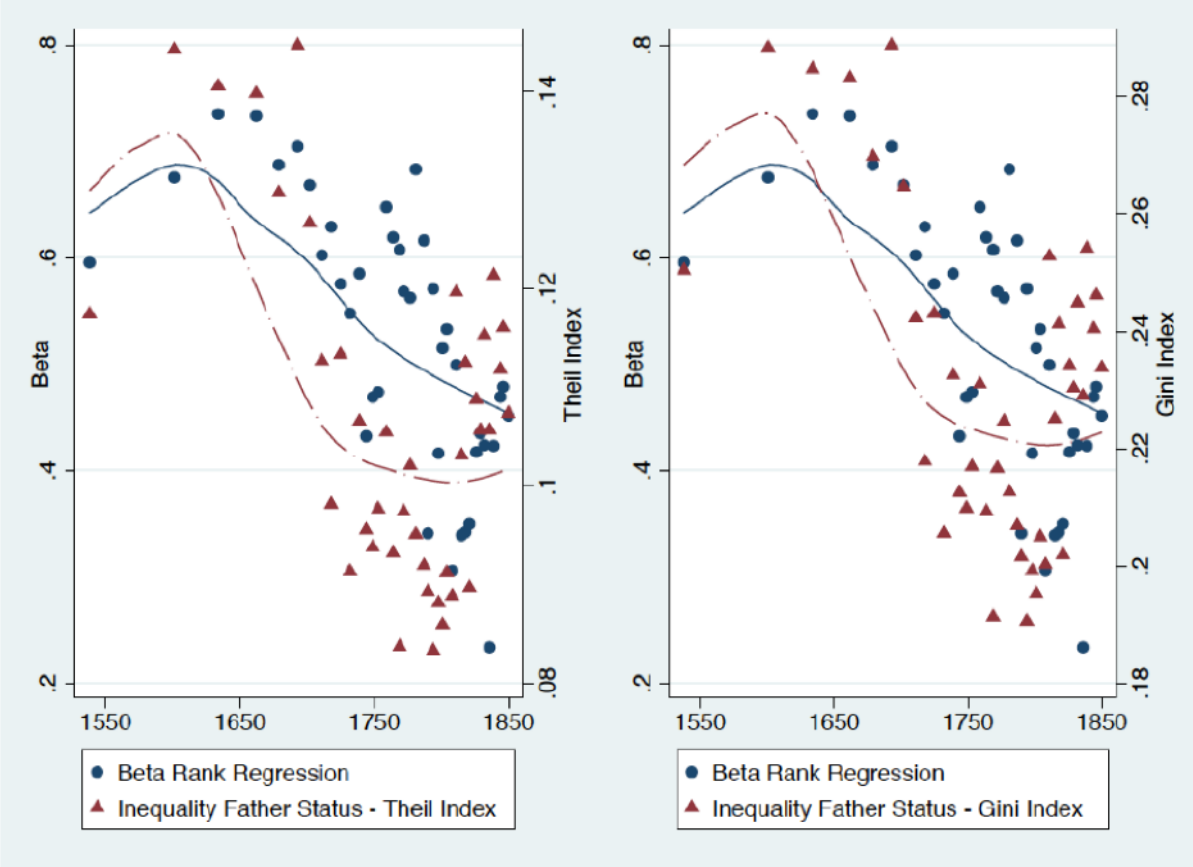
The main analysis in the text provides evidence that educational inequality in the father generation is strongly and negatively correlated with social mobility. The measure of education in this baseline is an indicator variable of whether a person has prepared to take the civil entrance exam at any level, see Table 1, last column. Distinguishing those that passed at a high levels from those that only passed at a lower level (or prepared but did not pass) does not change the strong relationship between educational inequality and mobility, as Figure A.10 shows. To examine the robustness of the results to different definitions of inequality, I compare the results in the benchmark, which used the Theil index, with the Gini index. The following figures compare the relationship of mobility and inequality using two well-known indices of inequality, the Theil index and the Gini index. As Figure A.11 indicates, the correlation between $\beta_c$ and inequality is virtually the same employing the Theil and the Gini index. Furthermore, results for the Gini and the Theil indices are similar in the case of upward and downward mobility over time as well (not shown). I have also explored other well-known measures of inequality, including the standard deviation of

Figure A.10: Mobility and Alternative Measures of Inequality in Father Education




**Notes**: Shown are locally smoothed series of $\beta$ (equation (1)) and Theil indices for two alternative definitions of education for forty birth cohorts. Horizontal axis shows median birth year in cohort, birth year of first son is 1330. Baseline education definition is indicator variable (Table 1, last column). Alternative definition has education variable taking the value of 1 for preparation and passing of lower exams (status levels 6, 9, 12, and 13) and a value of 2 for passed examinations at a higher level (status levels 15, 16, and above 17), see Table 1.

Figure A.11: Alternative Measures of Inequality - Theil versus Gini Index




**Notes**: Shown is $\beta$ (equation (1)) and inequality in father percentile rank status for forty birth cohorts; Theil index on the left, Gini index on the right. Horizontal axis shows median birth year in cohort; birth year of the first son is 1330.

Table A.2: Upward Mobility and Inequality

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Father Status Inequality | −0.802<br>(−8.337 | | −0.622<br>(−5.173) | | | | |
| *Geogr. Adj. Father Status Inequality* | | −0.690<br>(−4.741) | −0.263<br>(−4.741) | | −0.254<br>(−2.736) | | |
| Educational Inequality | | | | −0.914<br>(−12.477) | −0.771<br>(−9.820) | | −0.527<br>(−4.131) |
| Within Group Inequality | | | | | | −0.112<br>(−1.216) | −0.169<br>(−2.114) |
| Between Group Inequality | | | | | | −0.991<br>(−9.831) | −0.559<br>(−4.735) |
| $R^2$ | 0.644 | 0.476 | 0.680 | 0.836 | 0.879 | 0.844 | 0.900 |

**Notes**: Dependent variable is upward mobility, defined as the expected percentile rank of son if the father has rank $< 0.5$ for N = 40 birth cohorts. Estimation by OLS; standardized coefficients reported, with robust t-statistics in parentheses. Inequality is Theil index; within- vs. between- decomposition at the level of the segment.

Table A.3: Inequality and Downward Mobility over Time

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Father Status Inequality | 0.801<br>(9.337) | | 0.626<br>(5.259) | | | | |
| Geogr. Adj. Father Status Inequality | | 0.684<br>(5.518) | 0.255<br>(2.064) | | 0.239<br>(3.077) | | |
| Educational Inequality | | | | 0.921<br>(14.872) | 0.786<br>(10.111) | | 0.606<br>(4.093) |
| Within Group Inequality | | | | | | 0.125<br>(1.479) | 0.191<br>(3.061) |
| Between Group Inequality | | | | | | 0.991<br>(11.606) | 0.496<br>(3.539) |
| $R^2$ | 0.641 | 0.468 | 0.676 | 0.849 | 0.888 | 0.830 | 0.904 |

**Notes**: Dependent variable is downward mobility, defined as the expected percentile rank of son if the father has rank $> 0.6$ for N = 40 birth cohorts. Estimation by OLS; standardized coefficients reported, with robust t-statistics in parentheses. Inequality is Theil index; within- vs. between- decomposition at the level of the segment.

logs, the coefficient of variation, and the relative deviation from the mean, with similar results as when the Theil index is employed. Therefore, I conclude that the finding that inequality is negatively correlated with inequality is robust to alternative measures of inequality.

## C.1 Additional Multi-variate Results on Inequality and Mobility

The following set of regressions provide supplemental multi-variate regression results to the analysis of relative mobility in the text (Table 12). The first set of results examine the relationship of upward mobility and inequality, see Table A.2. The set of results in Table A.3 concerns the relationship between downward mobility and inequality. These results show that the relationship of inequality on the one and downward and upward mobility on the other hand is broadly similar to that of relative mobility and inequality. Table A.4 reports confidence intervals for a number of figures in the text; the correlations are significant at

Table A.4: Bi-variate Correlations and Confidence Intervals

| Figure | Mobility | *Inequality/ Spatial Mobility* | *Correlation* [95% *CI*] |
|---|---|---|---|
| 10 | $\beta$ : relative | Percentile Rank Status | 0.43 [0.14, 0.65] |
| 11 | $\beta$ : relative | Geogr. Adj. Percentile Rank Status | 0.58 [0.32, 0.75] |
| 12 | Upward | Education | −0.91 [−0.84, −0.95] |
| 13 | Downward | Education | 0.92 [0.85, 0.96] |
| 15 | $\beta$ : relative | Within-segment status | −0.64 [−0.41, −0.79] |
| 15 | $\beta$ : relative | Between-segment status | 0.68 [0.47, 0.82] |
| 17 | $\beta$ : relative | Within-segment status | −0.64 [−0.41, −0.79] |
| 17 | $\beta$ : relative | Within-dynasty status | −0.40 [−0.10, −0.63] |
| 18 | $\beta$ : relative | Spatial Mobility | −0.69 [−0.49, −0.83] |

**Notes:** Table shows bi-variate correlations and confidence intervals of variables shown in the figures given in column 1.

standard levels. Finally, Table A.5 shows results for a number of other potential influences of mobility in the case of twenty five-year cohorts. They broadly confirm the findings for the case of forty birth cohorts in Tables 13 and 14 in the text.

Table A.5: Inequality and Other Influences - Twelve Birth Cohorts

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Status Mean* | 0.87 (6.64) | | | | | | −0.01 (−0.02) | | | | | |
| *Education Mean* | | 0.89 (5.65) | | | | | | −0.33 (−0.85) | | | | |
| *Longevity Mean* | | | 0.08 (0.23) | | | | | | 0.17 (2.40) | | | |
| *Grandfather Inequality* | | | | 0.83 (4.11) | | | | | | 0.18 (1.97) | | |
| *Geogr. Dispersion* | | | | | −0.52 (−1.88) | | | | | | 0.27 (3.16) | |
| *Family Size* | | | | | | −0.61 (−2.68) | | | | | | 0.08 (0.64) |
| *Educational Inequality* | | | | | | | 0.94 (2.85) | 1.26 (2.80) | 0.95 (9.24) | 0.78 (5.60) | 1.12 (9.91) | 0.99 (6.11) |
| $R^2$ | 0.75 | 0.80 | 0.01 | 0.68 | 0.27 | 0.37 | 0.87 | 0.87 | 0.90 | 0.88 | 0.90 | 0.87 |

**Notes**: Dependent variable is $\beta_c$; estimation by OLS. Standardized coefficients shown, robust t-statistics in parentheses. N = 12.