NBER WORKING PAPER SERIES

INTERGENERATIONAL MOBILITY OF DAUGHTERS AND MARITAL SORTING: NEW EVIDENCE FROM IMPERIAL CHINA

Wolfgang Keller
Carol H. Shiue

Working Paper 31695
http://www.nber.org/papers/w31695

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2023

Thanks to Ran Abramitzky, to Hao Dong, our scheduled discussant, and to members of the seminar audience at the University of Hong Kong for comments. Thanks as well to Ted Telford who provided parts of the data. Financial support from NIH under grant NICHD Award HD042731-01 is gratefully acknowledged. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

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 They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2023 by Wolfgang Keller and Carol H. Shiue. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Intergenerational Mobility of Daughters and Marital Sorting: New Evidence from Imperial China
Wolfgang Keller and Carol H. Shiue
NBER Working Paper No. 31695
September 2023
JEL No. J62,N3

ABSTRACT

We study the role of marriage for women's intergenerational mobility during the Ming-Qing (1368-1911) period. Using status information based on the timing of marriage from family histories in Central China, already in the early 1500s it is the case that daughters from rich families attain higher status over their lifetime than daughters from poorer families. This intergenerational status persistence is partly due to marital sorting because daughters from high-status families tend to become the wives of sons who themselves come from rich families. Quantitatively, the correlation of 0.6 between the status of biological and in-law families means that marriage accounts for more than one third of total intergenerational status transmission, while not accounting for marriage overestimates mobility by more than 20 percent. Further underscoring the importance of marriage, typically the status of the in-law family plays a larger role for intergenerational status transmission than the child's biological grandparents. Over the period 1500 to 1900, the degree of marital sorting falls, as does intergenerational persistence. Lower investments in the marriage market to find a good match for a daughter go hand in hand with the fall in the returns to son education due to the decline of China's civil service examination.

Wolfgang Keller
Department of Economics
University of Colorado, Boulder
Boulder, CO 80309-0256
and NBER
Wolfgang.Keller@colorado.edu

Carol H. Shiue
Department of Economics
University of Colorado at Boulder
Boulder, CO 80309
and NBER
carol.shiue@colorado.edu

1 Introduction

From agricultural work over crafts production to the raising of children, women had economic functions since the earliest of times. And yet, the economic status of women has been difficult to characterize. One reason is lack of information on labor income, since even by the late 19th century, women's participation in formal labor markets was still low.[1] While one might equate the status of a woman with that of her married husband, this would make it impossible to distinguish the role of marital sorting from other forces of intergenerational status transmission. This paper studies the role of marriage for intergenerational mobility using original information on woman status for parts of Ming-Qing China (1368-1911) to separate the dual forces.

With more than 6,000 observations from seven genealogies of central China, we study the transmission of status from father to daughter using information on timing of the marriage. A key observation is that a woman who becomes first wife of her husband has outcompeted other women in the marriage market, including future wives of her husband. Thus, the expected contribution to the marriage of the first wife must be relatively high, and so is her status. Further, the role of marital sorting is examined by analyzing father-in-law status in conjunction with the status of the daughter's biological father. The coverage of the data, with a substantial number of observations before 1700, allows us to trace out trends in woman intergenerational mobility over time. Additional insights come from comparing the role of marital sorting for daughter mobility with its role for son mobility for the same clans.

The first finding is that whether a woman is first wife or a subsequent wife is indicative of her status. First wives typically have higher-status husbands than other wives, and there is stronger positive sorting between biological and in-law families for first-wife marriages than for other marriages. Starting in the early 15th century, daughters with a high-status father are more likely to be the first wife in their households, and the strength of the intergenerational correlation of status does not only depend on her father but also on the clan the daughter comes from.

Second, accounting for the role of marriage changes our understanding of the level, sources, and trends in the intergenerational mobility of daughters. In particular, with a correlation of around 0.6 between the statuses of biological and in-law families, marriage patterns are far from random, and analysis that abstracts from marital sorting would overestimate father-daughter mobility. Furthermore, the contribution of in-law family to daughter status is a sizable 40%, so that analysis based solely on biological family information might introduce error. In addition, the extent of marital sorting in the sample fell from the 17th to the 19th century. Thus, accounting for marital sorting is crucial for consistent comparisons of woman's intergenerational mobility across centuries.

Third, the implied level of mobility in the father-daughter relationship overall is comparable to that in the father-son relationship. Further, once one accounts for marital sorting the intergenerational mobility of daughters follows a similar trend as that of the son over the sample period as well.

[1] U.S. overall female labor force participation in 1890, for example, was 18.6%, while labor force participation of married women was lower (Olivetti 2013, Fernandez 2013).

The importance of marital sorting for intergenerational mobility is further highlighted by finding that a high-status father-in-law is typically more important for son status than either paternal or maternal grandfather.

One potential concern might be whether the status of a married woman changes over her life because of factors beyond her control. These are random, but we can examine the major cause of her change in status, which is whether the wife gives birth to a male heir. The above results are unchanged if one accounts for the number of sons. Moreover, marriage decisions are made based on what is expected from bride and groom at the time the marriage contract is signed, which is not affected by unforeseen events at that time.

Adding historical analysis of the intergenerational mobility of women to that of men has long been a focus in the study of social mobility.[2] Two challenges have been the comparatively low participation rate of women in formal labor markets and that women are often hard to trace in the data because they tend to change their names upon marriage. Recent advances include exploiting the status content of first names (Olivetti and Paserman 2015), harnessing marriage records (Craig, Eriksson, and Niemesh 2019, Espin-Sanchez, Gil-Guirado, and Vickers 2022), and taking a probabilistic, machine-learning approach to intergenerational linking (Bailey, Lin, Mohammed, Mohnen, Murray, Zhang, and Prettyman 2023, Buckles, Price, Ward, and Wilbert 2023). We complement this work by using an original female status measure that is implementable in economies where women do not participate in formal labor markets (including because they do not exist).[3] Our results provide evidence that in China, the intergenerational correlation of status between father and daughter was comparable to the intergenerational status correlation between father and son for the 17th century and earlier to the 19th century.

A broad literature emphasizes the role of marital sorting for women's economic status.[4] This paper extends existing work on historical patterns of marital sorting to earlier periods. In particular, the finding of falling marital matching in China from the 17th to the 19th century contrasts with flat 19th-20th century sorting levels in Quebec (Curtis 2022) and is more in line with falling marital sorting documented for the US (Buckles, Price, Ward, and Wilbert 2023, also for the 19th and 20th centuries.[5] The hypothesis that estimates of women's intergenerational mobility may be biased without accounting for marital sorting has been pursued for advanced economies during the 20th

[2]Chadwick and Solon (2002), Mazumder (2005), and Jäntti et al. (2006) present estimates on women's intergenerational mobility in advanced countries for the late 20th century.

[3]One may employ as the status of the wife her husband status (Craig, Eriksson, and Niemesh 2019, Dribe, Eriksson, and Scalone 2019, Bailey and Zin 2022, Buckles, Price, Ward, and Wilbert 2023), because in a world where market work is rare for married women, her economic status is largely determined by the status of her husband (Goldin 1983). Because we seek to separate marriage from other influences of women's intergenerational status transmission, we employ a measure of female status that is not derived from the status of her husband.

[4]Channels include complementarity or substitutability of traits in production (Becker 1973, 1974), cultural similarity (Kalmijn 1994), and sex ratios (Abramitzky, Delavande, and Vasconcelos 2011); Weiss (1997) provides a survey.

[5]See also Bailey and Zin (2022), Craig, Eriksson, and Niemesh (2019), and Olivetti, Paserman, Salisbury, and Weber (2022) on 19th and 20th century marital sorting in the United States. Keller and Shiue (2022) report on marital sorting in Ming-Qing China. Hamilton and Siow (2007) also present results on assortative mating for historical Quebec.

century (Chadwick and Solon 2002 and Ermisch, Francesconi, and Siedler 2006).[6] Buckles, Price, Ward, and Wilbert (2023) and Espin-Sanchez, Gil-Guirado, and Vickers (2022) present results for historical samples; the latter show that intergenerational mobility is lower once one accounts for marital sorting in 18th century Murcia (Spain). We show a similar result for a more broadly representative sample in China, and also document that the falling degree of marital sorting contributes to the observed increase in women's intergenerational mobility over time.

Recognizing that key features of an economy such as mobility, cultural assimilation, preferences, and trust rely on intergenerational transmission, a sizable literature has recently sought new sources for intergenerationally linked data. Progress has been made by the ex-post linking of cross-sectional data across generations (Ferrie 1996; surveyed by Abramitzky, Boustan, Eriksson, Feigenbaum, and Perez 2021). Because representativeness remains a concern (Bailey, Cole, Henderson, and 2020, Ward 2023), authors have pooled retrospective family surveys to estimate early 20th century US intergenerational mobility (Jacome, Kuziemko, and Naidu 2022). Other approaches are crowd-sourcing on genealogical websites to improve traditional linking methods (Price, Buckles, Van Leeuwen, and Riley 2021), synthetic genealogies that combine parish records with vital records (Curtis 2022), or, as in the present study, family genealogies.[7] The promise of any given approach will typically depend on the specifics of the sample as well as the particular questions asked. Key data issues in the present context are discussed in sections 3.2 and A.2.

The following section provides background on the institution of marriage as well as marital sorting in China during the sample period. Section 3 describes the data employed in this study. Estimation results can be found in section 4. We begin by studying female intergenerational mobility using data on father and daughter, before turning to the role of marital sorting by incorporating information on the status of the in-law family. Section 4 is completed by comparing the role of marital sorting for female and male intergenerational mobility. Section 5 summarizes key findings of this study and draws some conclusions. The Appendix provides additional details on the data as well as supplemental estimation results.

2 Marriage, Mobility, and Gender in the Ming-Qing: A Primer

Social scientists have made much progress on the subjects of marriage, inequality, and the specific role of women during the Ming and Qing dynasties (Watson and Ebrey 1991 is an introduction). For example, the conventional view in the literature holds that marriage was the ladder of success for women in late imperial China, that women stayed typically in the inner part of the Chinese house, and that marriage was primarily arranged by the parents.[8] At the same time, much of what we know is drawn from general sources such as the teachings of Confucius, the Qing code, or etiquette booklets. As far as actual behavior is concerned, historical sources impose tight limits on what is

[6] With positive marital sorting, mobility estimated from individual father-daughter data would be an overestimate since it ignores that daughters of rich biological fathers tend to marry sons that tend to have rich fathers as well.

[7] Shiue (2016) is a recent survey on the nature and content of Chinese genealogies.

[8] See Mann (1991, p.204), Ebrey (2023), and Shan (2013), respectively.

known, especially on lower status individuals and women.[9] While this paper seeks to contribute some new evidence from a sample population of Tongcheng, the following section sketches some generally agreed-upon features for Ming-Qing China.

Typically in a Chinese marriage the husband would gain the productive value of the wife, and in exchange the wife would receive a claim to the privileges of the husband's ancestral rites, as well as a stake in the economic fortunes of her husband, even if he died (Watson and Ebrey 1991). Marriage was the union of two extended families (clans), not of individuals, and as such the parents, grandparents, and even the father's brothers and sisters determined who a younger relative should marry (Shan 2013). The families of women and men could gain status through marriage, which was an occasion to adjust their economic position (Shan 2013), and would require often substantial expenditures.

Children were typically betrothed at the age of 8 to 10 years, and even younger for high-status families. In general, marriage was an occasion in which the wife changed her residence. Not only would the wife move from her father's family to her husband's family (patrilineal), but she would also move from her father's home to her husband's home (patrilocal). A man had usually a single female partner, his married wife, during his lifetime. The couple might fulfill their marital goals, in particular a certain number and gender of children, or not. Producing a male heir to ensure that the family line lives on was one of the most important roles of the wife. If the wife does not have (male) children, or the wife dies, as a rule the family line was broken, unless an adoption of a male heir could be arranged.

The exception to this is that the husband remarries and has a male child with a second or later wife. The ability to remarry depended largely on the status of the husband–given the considerable marriage expenses, most men could not afford to marry a second time.[10] From the point of view of the second wife, remarriage came with certain risks of alienating offspring (from the first wife) and losing one's authority over the coming generation (Mann 1991). A stepmother had to deal with such things as the resentment of her stepchildren who would consider her an outsider, and the insubordination of her daughter-in-law who they were supposed to manage (Mann 1991). The status of a concubine was subordinate to that of the wife in the household, because the former had no role in ancestral rites.

On the other hand, if the husband died early, his widow ranked in the household under the male heir if there was one, and she could be a chaste widow. If there was no male heir, the woman's position in the household was precarious, leading often to her remarriage even though this was generally frowned upon in this society (Mann 1991). Women who would remarry due to their

[9] Anecdotally, Ebrey (2023) notes that already by the Song dynasty (960-1279) there were Chinese widows who ran inns, midwives delivering babies, pious women who spent their days chanting sutras, nuns who called on such women to explain Buddhist doctrines, girls who learned to read with their brothers, famers' daughters who made money by weaving mats, childless widows who accused their nephews of seizing their property, and women who drew from their dowries to help their husband's sisters marry well.

[10] In addition, whether a man could find additional polygynous wives or concubines depended on his status as well, though in general polygamy (multiple wives living at the same time) was rare.

husband's death tended to do so with a man of lower status (Telford 1992).

As a rule families had a choice of whether to invest in the marriage of their daughters or in the education of their sons. Well connected in-laws or an educated wife could provide benefits for the next generation because of access to well-connected networks of knowledge, and an educated mother would have been the first teacher to the young males of the next generation. In that sense, marriage was a ladder to success for women. Because status could not be passed on directly from parents to their children, the economic success of the next generation depended to a greater extent on the investments parents made, and with that, it depended on the match made at the time of marriage. Marriages, and the associated investments through searching for a good marriage, were one of the channels through which wealth was passed on to the next generation.

The question of whether marriage perpetuated intergenerational inequality would depend on the specific patterns of marital sorting. If marriage is mostly between likes it raises cross-household inequality, while if differences attract each other marriage is a force towards equality. While one might expect that families want their daughters to always to "marry up", there were reasons that men might want to "marry down"– because such brides were thought to be harder workers and more easily satisfied with their situation (Ahern 1974, Freedman 1979).[11]

The patterns of marital sorting depended also on the availability of spouses, the age of bride and groom, and other time-varying factors (Mann 1991, Telford 1992). For example, women of the 17th century Jiangnan regions increased their cachet as wives by becoming increasingly educated, triggering a rising incidence of (non-arranged) "companionate marriages", and this can be traced to commercial growth and urbanization (Ko 1994). There are also examples of girls during the 18th century who despite being in an arranged marriage with a scholar husband would stay in their parents household and learn alongside their older brother from their mother until puberty mandated sex segregation (Mann 1997).

Two additional trends of the 17th and 18th century marriage market are worth noting (Watson and Ebrey 1991). First, unbalanced sex ratios meant that there were fewer women than men, leading to a relatively high excess of single men. Under these demographic circumstances when there are many fewer women than men, brides could reasonably expect to marry up. Second, declining likelihood of success in the civil service examinations meant that there was a smaller pool of elite scholarly families who were traditionally the most well-off among all of society. There was greater occupational fluidity in the mid-Qing (Mann 1991). Thus, daughters of elite families were likely to marry down because of the declining trends in opportunities for the educated elite.

Meanwhile the costly expenses of investments in education for sons competed with the costs of dowry in marriage. As the returns to education declined, however, fewer families would have found

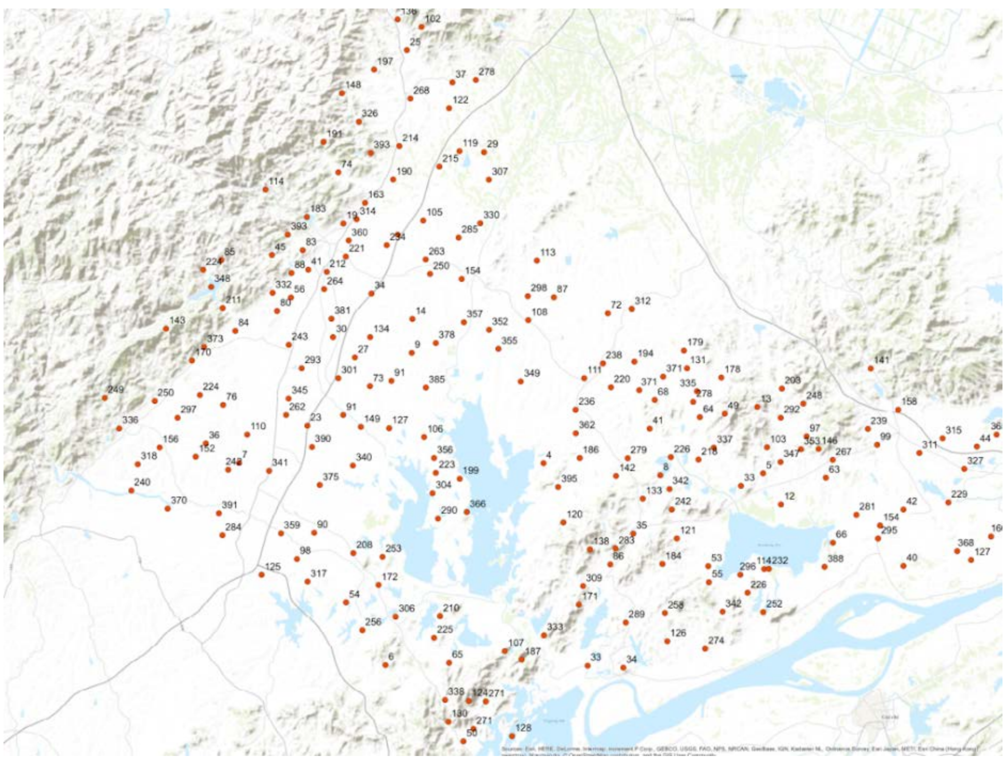
[11] Often, parents would seek a so-called "matching doors" marriage in which families sought matches between couples of comparable social backgrounds. While gifts were presented both from bride's family to groom's (akin to a dowry) and from groom's family to bride's (akin to a bride price), in no period was China a dowry society comparable for example to India (Watson and Ebrey 1991). It was important that gifts were not too one-sided lest one gives the impression that a spouse is "purchased" (Mann 1997).

this a productivity channel of investment. Since positive assortative matching typically requires considerable search costs, declining likelihood of success in officialdom would likely have decreased the returns to high marriage search costs or dowry. Thus, the flip side of the decline in the return to education is a decline in the return to marital sorting (Watson and Ebrey 1991). One would expect a decline in marriage matching as the returns to male education fell.

3 Data

This study examines intergenerational mobility using genealogical data for seven male clans in Tongcheng county, part of Anhui province for parts of the Ming (1368-1644) and Qing dynasty (1644-1911).[12] Genealogies are a classic source of socio-economic data for China. Estimates put the number of existing genealogies into the tens of thousands (Wang 2008), and dozens of them have been employed for Tongcheng by earlier researchers such as Beattie (1979). Figure 1 shows a map of Tongcheng county that indicates towns and villages in the sample.

Figure 1: Tongcheng County and the Location of Sample Villages



Notes: Location of towns and villages in Tongcheng county, in Anhui province, with basic topography and roads.

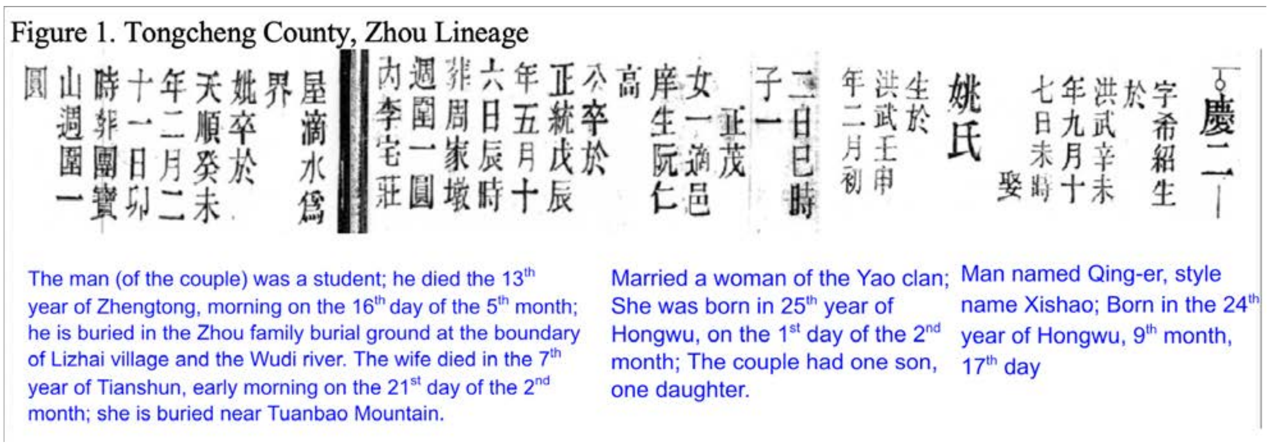
Figure 1 shows some of the historical towns and villages of Tongcheng county, which is located north of Anqing, the capital of Anhui province. The Yangzi river is shown in the south-eastern part of the Figure. Tongcheng county had approximately 1.5 million inhabitants in the year 1790

[12] Chinese clans are also referred to as lineages or common descent groups in the literature.

(Beattie 1979).

Because documenting intergenerational links is a key purpose, genealogies are particularly useful for studying intergenerational family responses. The genealogies employed in this paper are essentially richly annotated pedigree charts, though for practical reasons the information is organized by married couple, which also underlines the central importance of marriage to these families. We also know which of the men married, and which did not. Among the men who married, we know how many wives they had, and, through examining their death dates, we can determine if the wives were alive at the same time or whether it is likely that the marriages happened sequentially. The importance of descent is underlined by the that the genealogies list the children that were born to each wife separately. Figure 2 presents part of one of the genealogies employed in this paper as an example.

Figure 2: Zhou Clan Genealogy Example



Notes: Part of the source data, see text for additional information.

The following section describes the information in the sample on female and male status, before turning to the representativeness and potential biases in the sample.

3.1 Status Information

3.1.1 Female Status

The status of women during the sample period cannot be based on labor income since most women were not formally employed, and neither were they eligible to participate in China's civil service exam. Instead, this study employs information on the timing of marriage. In particular, we know whether a given woman was the first wife of her husband, in the sense of being the female who was married first to the husband in a temporal sense.[13] Table 1 summarizes this information. About

[13]We have information on the temporal ordering of all male-female partnerships in the household if there is more than one. This includes partnerships with concubines (a relationship that does not involve marriage). Information on the actual date of marriage is rare in Chinese genealogies.

thirteen percent of all women are the first wife of their husband. The status variable of such women takes the value of one, while for other women its value is zero.

Table 1: Female Status as First Wife

Role in Household	N	Percent
First Wife	813	13.2
Not First Wife	5,365	86.8
Total	6,178	100.0

Notes: Shown is distribution of first wives versus not first wives in the sample. Group of Not First Wife encompasses four female roles in household, of which that of single wife is the largest (Table A.5).

A relatively high status for first wives is consistent with the economic analysis of marriage (Becker 1973, Weiss 1997, and Choo and Siow 2006). Individuals and their parents in society have many potential partners, which creates competition between families for their marriage partners. In the case of Imperial China, as noted above, marriage partners are chosen primarily by the two families rather than by bride and groom themselves. Marriage is a voluntary assignment of males to females, and different assignments create different marriage outcomes.[14]

Why do bride and groom marry? We assume that the objective is to maximize expected marital output. An important part of output is a certain quantity of children, including at least one male child because given the patrilineal structure that child is central for continuing the family line. Also the quality of the children would matter, in particular skills, because that affects material welfare of the family in the next generation.

Turning to the contribution of each partner to the marriage, in the simplest case each female is endowed with a single characteristic, f, which affects positively marital output. For concreteness, let the female's contribution be her ability to be an effective teacher. As noted above, the mother would often be the first teacher of her children–non-private schooling was rare during the sample period. Assume further that each male is endowed also with a single characteristic, m, and that also has a positive effect on marital output; suppose that the male's contribution is money.

A particular assignment of brides to grooms is a marriage equilibrium if (i) no married person would rather be single and (ii) no two (married or unmarried) persons would prefer to form a new union. Marriage patterns in this setting depend on the relationship between the contributions of females and males to expected marital output. If the relationship between the female contribution f and the male contribution m to marital output is positive, the equilibrium assignment will be such that high-f females will tend to marry high-m males (see Weiss 1997). This follows because if m and f contributions are subject to diminishing returns by themselves and the two characteristics are positively related, maximizing marital output requires to pair men and women with identical values

[14] This is true whether explicit price mechanisms (such as dowry or bride price) are observed or not. As noted in section 2, one-sided transactions from one family to the other –such as dowry–were not common in China.

of m and f (the special case of perfect positive assortative matching). Thus, one expects that rich grooms will tend to marry brides who have high teaching abilities.

In the marriage market, thus, rich groom families will tend to search for high-f brides that match their son's high level of money, m. At the same time, there is no perfect information about each individual's f and m levels, because marriage commitments are made when brides and grooms are young (see section 2). In forming expectations about marital output, a rich groom family will concentrate its search for a bride in the part of the marriage market that attracts families with brides that have high abilities, summarized by f, whereas families with less fortunate sons and daughters will primarily seach in another part of the marriage market.

Under these circumstances, a woman who is chosen as first wife by her husband can be expected to be endowed with a relatively high level of ability, f. First, given the observed choice of the husband, the first wife has outcompeted other women who are in the marriage market. This implies that the first wife's contribution to expected marital output must be relatively high. Other females that the first wife dominates include female partners that her husband would pick later.[15] In particular, the expected contribution of the first wife to marital output must be higher than the wife her husband marries as his second wife in case the first wife dies; such remarriage is empirically the norm for multiple lifetime female partners (rather than multiple married wives simultaneously [polygamy] or concubines).[16] Marriage commitments between families are made based on what is expected at age 10 of the future spouses or earlier, and unforeseen events such as early death (or infertility) do not affect these choices.

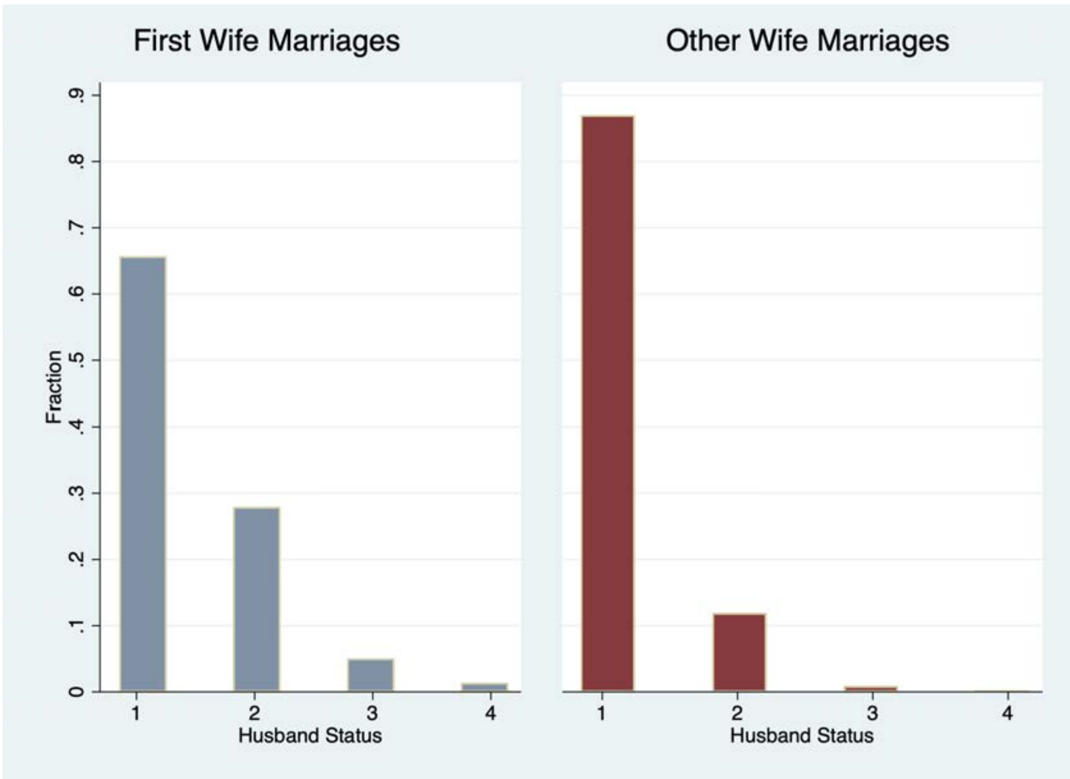
Furthermore, a first wife's contribution to marital output is also larger, in expectation, than that of a woman who remains the single lifetime wife of her husband. A key reason for this is that marriage expenditures were high so that most families could afford to marry only once during their lifetime (see section 2), leading to self-selection of families to search in a particular part of the marriage market. A relatively poor family with a bride of average ability would primarily search in the part of the marriage market where the expectation is a single lifetime relationship; in contrast, a family searching in the high end of the marriage market knows that there would be a chance that the husband might have multiple female partners during his lifetime. Search for a suitable partner thus means that rich groom families select relatively early brides with high expected abilities, and equilibrium in the marriage market is such that on average women who are first wives have relatively high status. There is abundant historical evidence for a relatively high status of the first wife. In particular, because a woman who becomes second wife of her husband has to potentially deal with the resentment of her stepchildren (see section 2), women endowed with the highest ability levels f will not want to take up this role; as a consequence, the status of a second wife must be lower than the status of the first wife. Additional evidence based on the present sample is provided in the following.

[15] Later partners of the husband are usually alive when the first wife is picked; the average age difference is 4-5 years.

[16] Only 2% of all wives in the sample live in a polygamous household; see also Table A.5.

First, consider the densities of husband status for those women who are first wives versus those who are not (Figure 3 compares left side versus right side, respectively).

Figure 3: First Wive Marriages and Husband Status



Notes: Husband status increasing from level 1 to level 4; Kolmogorov-Smirnov test rejects equality of distribution ($p < 0.01$).

Figure 3 illustrates two important points. First, the husband status distribution for first wife marriages is shifted to higher levels compared to that of non-first-wife marriages. While about 66% of all first-wives have husbands with the lowest status level, 87% of all non-first wives have a husband with that level. Conversely, the fraction of first wives with husbands of status 3 or 4 is more than five times as high as the corresponding fraction of non-first wives. Figure 3 provides empirical support that women who are chosen as first wives have higher status than other women.

Second, there is no perfect positive assortative marriage matching on husband and wife status. This could be for several reasons, including that there is more than a single dimension describing husband and wife's contribution to marital output, especially if some of these interact negatively in the production of marital output. Irrespective of the reason, Figure 3 suggests that assuming that the status of a married wife is equal to the status of her husband may generate substantial and potentially non-classical measurement error.

Additional support for a woman status measure based on the timing of marriage is that the degree of positive sorting between parent and in-law families is higher for first-wife marriages than for other marriages (0.66 versus 0.56, respectively, based on the status of biological father and father-in-law). The higher level of sorting for first-wife marriages is consistent with incurring higher marriage search costs. This is what one would expect if the marriage is relatively important, and consistent with this, the status of the woman in her new family will tend to be relatively high.

3.1.2 Male Status

Status of men in this study is based on up to 30 descriptors in the genealogy that characterize his status in society, including any official positions (and at what level), levels of education, wealth, donations, and honors. We aggregate this information to form four groups with increasing status, which are shown in Table 2. This grouping is employed for all five male status variables in the present study, that is, that of the husband, the husband's father, the husband's grandfather both on the husband's father and on the husband's mother side, as well as the status of the father of the wife.

Table 2: Status Classification for Men

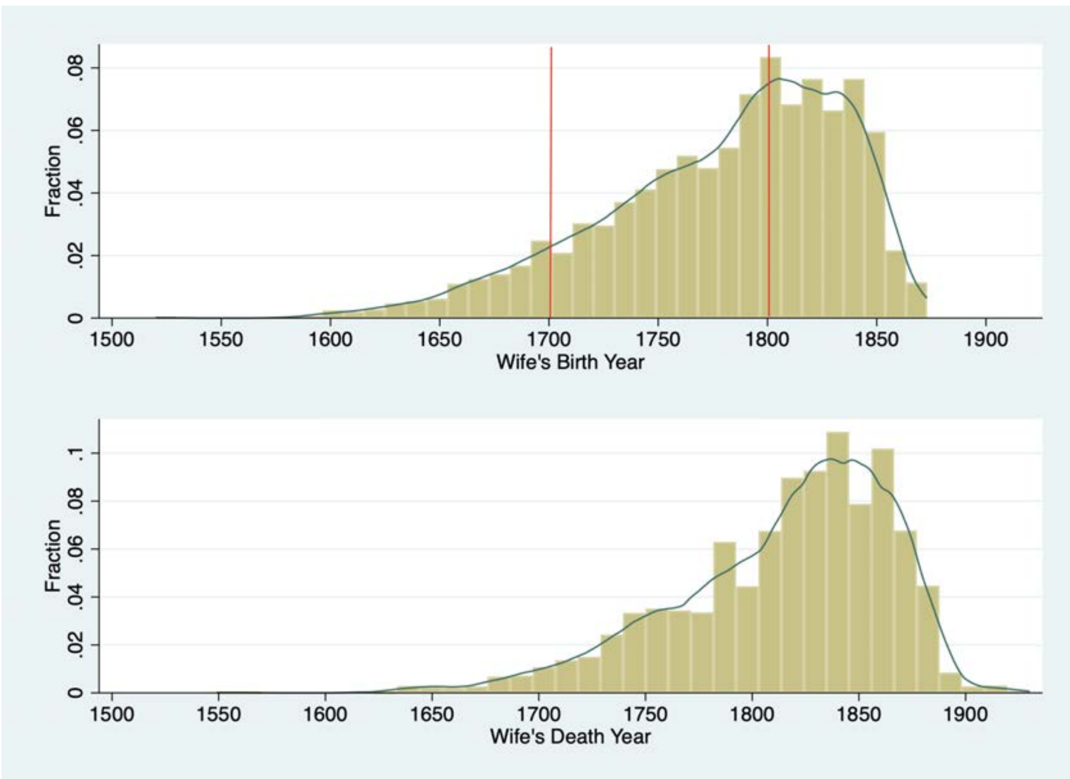
Group	Descriptors	N (%)
1	No title, degree, and evidence of wealth; Honorary title; village head Moderate wealth and status of 1st and 2nd degree family member	5,202 (84.2)
2	Educated, scholar, no degrees or office; editor; prepared but did not pass exam; Wealthy farmer, landowner, or merchant; set up lineage estates, large donations, Official Student; Student of the Imperial Academy; Military or civil *shengyuan*; major military office	866 (14.0)
3	Civil *juren* (graduates of the provincial civil service exam)	88 (1.4)
4	Civil *jinshi* (graduates of the national civil service exam)	22 (0.4)
All		6,178 (100.0)

Notes: Shown is status classification for males; classification based on Telford (1986), Chang (1955, 1962), Ho (1967), and Eberhard (1962). More details are given in Table A.6. Distribution of status in final column is for sons.

Table 2 shows that 84.2% of all men belong to the lowest status group. That is consistent with other evidence for China during this period. Furthermore, we see that graduates of the national civil service exam (*jinshi*) rank in the top 0.5% of the status distribution.

In the regression analysis below, status is employed as percentile rank (following Dahl and DeLeire 2008), which has the advantage over the status categories on the left in Table 2 that it accounts for the size of the group, among other advantages. For example, with 84.2% of men belonging to the lowest status class, each of them would be assigned 0.421 (the midpoint between 0 and 0.842). Because the distribution of status changes over time, we compute each man and each woman's percentile rank relative to the subperiod-specific distribution. Three subperiods are distinguished by birth year, namely (i) before 1700, (ii) between 1700 and 1800, and (iii) 1800 and later. The breakdown of the observations by cohort are shown in Figure 4.

Figure 4: Distribution of Sample Across Cohorts



Notes: Shown is distribution of observations by wife's birth year and death year, as well as by cohort; N = 601 in cohort Before 1700, N = 2,848 in 18th century, N = 2,696 in 19th century; earliest birth year 1520, latest year of death 1930.

Overall, the slow evolution of the sample population parallels what we know about the overall population dynamics in Tongcheng (see Figure A.1). With just over 600 observations before the year 1700, the sample size for the first cohort is also large enough to allow for a quantitative analysis. The decline in the number of observations towards the 20th century is partly explained by a general decline in the focus on genealogical traditions with the end of the Qing dynasty (1911) and the founding of the People's Republic of China (1949).[17]

[17]Recent years have seen a revival in the use of genealogies in China.

3.2 Representativeness

Genealogical information might in fact be inaccurate. In particular, the information is self-reported, and there are no penalties for misrepresentation. At the same time, there is only sparse official data available during the sample period, and it does not appear to be as reliable as the clan records (see Figure A.1). This is not too surprising–holding regularly broadly representative population counts (censuses) was not a practice in many countries during the earlier parts of the sample period.

On the upside, beyond ancestral worship Chinese genealogies served a number of important economic functions for the clans themselves that would require accurate information. First of all, property rights turn on information recorded in genealogies, because they establish and sustain village settlement rights for specific clans. Second, genealogies are critical as a means of defense, including war, because by determining who is member of the clan and who is not it defines allegiances, rights, and responsibilities in times of conflict (both versus other clans and versus the government). Third, genealogies provide information on taxation and public goods provision. On the one hand, the state delegates to local clans the right to tax as well as the responsibility to fund public works such as irrigation. On the other hand, a clan's genealogy would specify assessments (essentially taxes) on their members to found and maintain common clan property.

The following analysis describes key features of the sample and how it compares to other information we have on China during this period. We begin with the full sample of seven clans before turning to the part of the data which allows to study the role of marital sorting for woman's intergenerational mobility.

Tongcheng Sample of Seven Clans The sample was created using a targeted approach by considering more than three dozen genealogies from Tongcheng county and selecting a subset with the goal of generating a broadly representative sample. Seven clans have been chosen, with the following last names: Chen, Ma, Wang, Ye, Yin, Zhao, and Zhou. Conditional on choosing the genealogy of a given clan, all entries of the genealogy become part of our sample. Given these clans' characteristics and size, we have about 20% of the sample belonging to the upper class as defined by Fei (1946), and the fraction he gives nationally for that class is also 20%. Also, the fraction of local and provincial civil service examination graduates in the sample is comparable to Chang's (1955) figure of 2%. Thus, groups with different level of status account for similar shares in the sample and in various analyses of China overall.

The seven clans yield a broadly representative sample in part because the clans are quite different from each other. First, the clans differ in size, see Table 3. The largest clan is the Wang clan (48%), while the smallest is the Chen clan (3%). Given that genealogies require skill and resources to produce, positive selection in the sense that richer clans report more members is a natural concern. This can be evaluated by relating clan size to the distribution of clan status, shown in the lower part of Table 3.

Table 3: Summary Statistics Across Male Clans

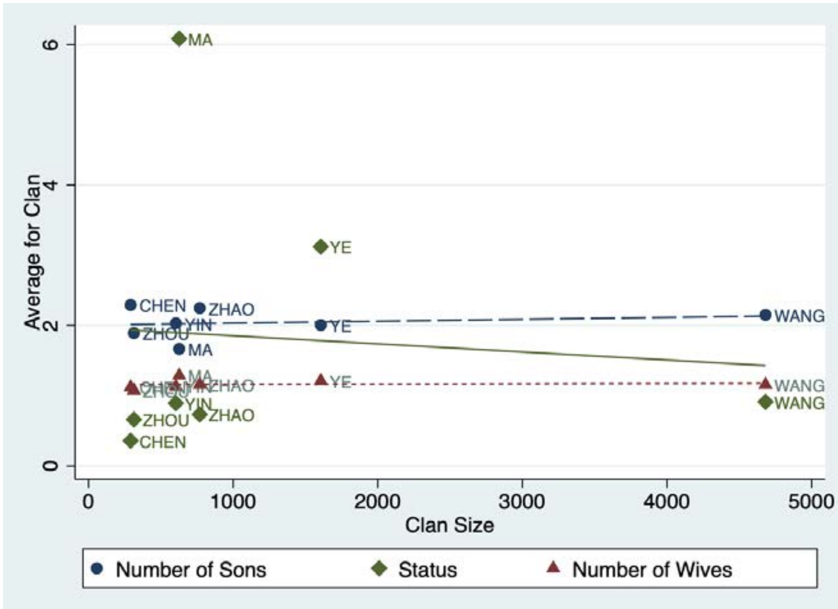
	Chen	Ma	Wang	Ye	Yin	Zhao	Zhou	All
N	353	917	5,798	2,147	758	1,028	376	11,377
Status								
1	348 (98.6)	412 (44.9)	5,448 (94.0)	1,602 (74.6)	714 (94.2)	981 (95.4)	357 (95.0)	
2	5 (1.4)	429 (46.8)	349 (6.0)	442 (20.6)	43 (5.7)	47 (4.6)	19 (5.0)	
3		60 (6.5)	1 (< .1)	86 (4.0)	1 (0.1)			
4		16 (1.7)		17 (0.8)				

Notes: Given is son status by clan for N = 11,377 couples. Percentage of clan total given in parentheses.

Table 3 confirms that the clans differ in terms of status. There are two clans who have members of the highest status level, the Ma and the Ye clans. In the other clans, the fraction of clan members that have the lowest status is 94% or higher. Moreover, for the Ma clan, the majority of clan men a status above the lowest level. If rich men were overrepresented in the same, one would expect that the Ma clan accounts for a large fraction of the sample overall–and yet there are three clans with higher numbers in the sample. This corresponds to the principle of Chinese genealogies that all men irrespective of status are included.

Table 3 indicates that there is little evidence that clan size simply reflects a clan's resources or that high-status clans systematically report more new members than lower-status clans. The result is confirmed in Figure 5 which shows a flat relationship between clan size and average status. In addition, Figure 5 considers other indicators of resources. In particular, wealthier men could afford multiple wives and maintain larger families (Harrell 1985). The figure shows the relationship between clan size and the number of wives is essentially flat. Furthermore, fertility as measured by the number of male children is not strongly related to clan size either, see Figure 5.

Figure 5: Clan Size versus Status, Fertility, and Female Partners per Man



Notes: Figure shows three relationships, between average (i) number of sons, (ii) husband status, and (iii) number of females per man, each with the number of clan members across seven male clans.

Section A.2 examines these issues further by considering a range of forms of selection and biases, including recall bias, progenitor bias, and survivor bias, finding little evidence that the sample would not be broadly representative of the Tongcheng area.

Information on Status of Female's Father To study daughter's intergenerational mobility, this study employs information on the status of her father. This information is available only for N = 6,178 observations. The restricted sample differs to some extent from the rest of the sample, for example, the average percentile rank of women in the restricted sample is 0.505, compared to 0.494 in the remainder (see Table A.3). Consistent with a higher status of first wives, this difference is in part due to the status of fathers of first wives being recorded at a higher rate than the status of fathers of non first-wives. We have re-estimated key specifications with a re-weighted sample that accounts for the impact of this sample restriction, finding that results are similar; see Table A.4 for these results.

4 Intergenerational Mobility of Women

4.1 Alternative Measures of Daughter Status

This section shows results on the intergenerational relationship between the status of the daughter and the status of her father. We begin by employing the following OLS regression

$$daug_{i,t}^{fm} = \beta_0 + \beta_1^b father_{i,t}^{fm} + \mu_t + \nu_f + \nu_m + \varepsilon_{i,t}^{fm}, \quad (1)$$

where $daug_{i,t}^{fm}$ is the daugther's percentile status rank in father-daughter pair i with birth year t, of clan f and marrying a man of clan m. Two alternative measures of daughter status are employed. In the first, her status is measured by the status of the husband she marries, while in the second, daughter status is based on her status in the marriage as first wife. A lower value of β_1^b indicates more intergenerational mobility, in the sense that if father status were low it would not greatly determine the status that the daughter achieves during her lifetime; in the extreme case of β_1^b equal to zero, each generation of daughters starts with a clean slate. Equation (1) includes a separate fixed effect for each of the birth years of the daughters, μ_t, to account for shocks and secular trends over the sample period (earliest birth year is 1413, latest birth year is 1885). Further, equation (1) includes fixed effects for each of the 135 female and 7 male clans, ν_f and ν_m, respectively. Conditional on these variables, $\varepsilon_{i,t}^{fm}$ is assumed to be a mean-zero error term.

On the left side of Table 4, results are shown with female status equal to first wife as the dependent variable. The positive coefficient shows that daughters who have a higher-status father have a higher chance to be the first wife in their husbands households (Panel A, column (1)). Accounting for birth year effects changes the coefficient little, while including clan fixed effects lowers the β_1^b estimate somewhat (Part A, columns (2) and (3), respectively). These clan fixed effects eliminate the influence of between-clan differences from the analysis of father-daughter intergenerational mobility. Finding higher mobility using within-clan differences indicates that clans are a force towards persistence.[18]

Part B of Table 4 shows results by cohort. Father status is positively correlated with first-wife status in all cohorts. With clan fixed effects, point estimates across cohorts differ somewhat more but the β_1^b estimates are not significantly different (Panel B, columns (1), (2), and (3)). These results provide evidence that father status is positively correlated with the status that his daughter has in her marriage starting in the 15th century.

On the right side of Table 4, analogous results are shown for employing as measure of female status the status of her husband. Then, the coefficient β_1^b is estimated at 0.5, see column (4) of Panel A. It means that a ten percentage point higher father status translates into a five percentage point higher daughter status. Accounting for shocks and long-run trends by including a fixed effect for each daughter birth year in the sample does not greatly change the estimate of β_1^b, as column (5) of Panel A indicates, but clan fixed effects lowers the estimate of β_1^b from 0.5 to 0.35, see Panel A column (6).

Thus, higher father status is associated with higher daughter status whether measured by first wife or by husband status. While either measure of daughter status yields a positive intergenerational coefficient, their levels differ. Specifically, a given percentage point higher father status raises daugther status more if one uses her husband status as measure of her status (0.13 versus 0.5, columns (1) and (4), Panel A, respectively). One reason for that may be that husband and father status are defined using the same four status groups, in contrast to first wife which captures some aspects of status that may not have a counterpart in the male context. [19]

[18]The result is analogous to the role of race in studies of US intergenerational mobility, e.g. Ward (2023), Jacome et al. (2022). It also parallels findings for father-son mobility during Ming-Qing China in Shiue (2019).

[19]For example, high female status might be achievable also for a wife who is the only lifetime wife of her husband,

Table 4: Female Intergenerational Mobility: Alternative Measures of Female Status

	Wife Status = First Wife			Wife Status = Husband Status		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Intergenerational Correlation						
Father Status	0.130** (0.015)	0.127** (0.016)	0.103** (0.017)	0.499** (0.017)	0.497** (0.018)	0.355** (0.015)
Panel B: Intergenerational Correlation Over Time						
Father Status <1700	0.130** (0.020)	0.102** (0.037)	0.073+ (0.039)	0.498** (0.017)	0.478** (0.042)	0.343** (0.041)
Father Status 18th c	0.132** (0.016)	0.148** (0.022)	0.126** (0.023)	0.500** (0.018)	0.507** (0.025)	0.363** (0.025)
Father Status 19th c	0.128** (0.016)	0.107** (0.027)	0.083** (0.029)	0.498** (0.018)	0.493** (0.031)	0.350** (0.031)
Birth Year FE		Y	Y		Y	Y
Male Clan FE			Y			Y
Female Clan FE			Y			Y
N	6,178	6,145	6,101	6,178	6,145	6,101

Notes: Dependent variable female status, defined on top of column. Estimation of equation (1). All status variables employ percentile ranks; FE stands for fixed effects; robust standard errors reported in parentheses; **/*/+ indicates significance at 1%/5%/10% level.

In sum, there is a positive intergenerational correlation between father and daughter status in this sample that holds irrespective of whether we employ a direct of derived measure of daughter status (first-wife and husband status, respectively). Quantitatively, the two measures do not predict the same role for intergenerational transmission, which is presumably due to measurement error created by the imperfect match of female and male status categories. At the same time, estimates for either measure change similarly as one adds different sets of fixed effects, and they also vary in a similar way for different cohorts. Given existing work that has successfully employed husband's status as the wife's measure of status, one can view Table 4 as validating first-wife status as an original alternative measure of female status.

The following section reexamines these findings by accounting for the role of the daughter's father-in-law.

4.2 The Role of Marital Sorting for Intergenerational Mobility of the Daughter

In this section we report results from extending the analysis to account for the status of the daughter's father in law. The estimation equation is now given by

$$daug_{i,t}^{fm} = \beta_0 + \beta_1^b father_{i,t}^{fm} + \beta_2^l father\, in\, law_{i,t}^{fm} + \mu_t + \nu_f + \nu_m + \varepsilon_{i,t}^{fm}, \tag{2}$$

where the dependent variable is measured as the daughter's status as first wife. Table 5 shows the results.

What intergenerational coefficient is estimated when daughter status is related to the status that she obtains in some sense through marriage, the status of her father in law? Results are shown in column (2) of Table 5, Panel A. The father-in-law coefficient β_2^l is estimated to be positive, indicating that first-wife status likelihood of the daughter is increasing in father-in-law status. Including biological father and father-in-law status simultaneously in the regression, both β_1^b and β_2^l are estimated to be positive, see Panel A, column (3). This is an important result because it provides new evidence that biological and in-law fathers have independent roles for the daughter's status. Furthermore, one way to gauge their relative importance for daughter status is based on the coefficient sizes β_1^b and β_2^l; along these lines, the share of the biological father's role is 60% ($= 0.09/(0.09+0.06)$). With the remaining 40% accruing to the father-in-law, this means that marriage plays a significant role for the intergenerational mobility of daughters.[20]

Furthermore, marriage is not just important for better understanding the sources of female intergenerational mobility, it also affects what the estimated level of mobility is. For according to column (3) a ten percentage points higher status of both biological father and father-in-law raises the daughter's status by 1.5 percentage points ($= 0.09 \times 0.1 + 0.06 \times 0.1$). If one abstracts from

especially if she can fulfill the families' fertility goals of the families. Conversely, more than 40% of first wives have a commoner father. Note that different means does not explain why the β_1^b coefficients are different in columns (1) and (4), as both are percentile ranks with a mean of 0.5 by construction.

[20] Variation in the status variables is similar; therefore, using standardized variables with mean zero and standard deviation of 1 yields a similar share for the biological father (56%).

Table 5: Female Intergenerational Mobility: The Role of Marriage

	(1)	(2)	(3)
Panel A: Intergenerational Correlation			
Father Status	0.127** (0.016)		0.090** (0.018)
Father-in-Law Status		0.100** (0.012)	0.060** (0.014)
Panel B: Intergenerational Correlation Over Time			
Father Status <1700	0.102** (0.037)		0.005 (0.047)
Father Status 18th c	0.148** (0.022)		0.112** (0.024)
Father Status 19th c	0.107** (0.027)		0.088** (0.030)
Father-in-Law Status <1700		0.153** (0.031)	0.150** (0.039)
Father-in-Law Status 18th c		0.107** (0.017)	0.059** (0.018)
Father-in-Law Status 19th c		0.070** (0.022)	0.032 (0.024)
N	6,145	6,145	6,145

Notes: Dependent variable female status, measured as first wife indicator. Estimation of equation (2). Birth year fixed effects included. Status variables employed in percentile ranks specific to the three subperiods shown; robust standard errors reported in parentheses; **/*/+ indicates significance at 1%/5%/10% level.

the role of the father-in-law, in contrast, a ten percentage point higher father status raises daughter status by only 1.27 percentage points (column (1), Panel A). Thus, accounting for marriage lowers the intergenerational correlation for daughters by 18% (= $(1.5 - 1.27)/1.27$).

The results also provide evidence on the strength of sorting between biological and in-law families. If the true model of daughter status is equation (2) but instead equation (1) is estimated, then $\hat{\beta}_1^b = \beta_1^b + \rho\beta_2^l$, where ρ is the correlation between the status of biological father and father in-law. Thus, the correlation between daughter status and father status that one estimates with equation (1) in column (1) is the sum of the direct effect of the biological father and the indirect effect that is equal to the degree of assortative mating times the direct effect from the father-in-law, from column (3); plugging in the estimates, one obtains $0.127 = 0.090 + \rho\, 0.06$,or $\rho = 0.62$.

This correlation of 62% between status of the biological father and status of the father in law is not too different from the correlation of 72% estimated for the status of father and mother of brides in 18th century Murcia (Espin-Sanchez, Gil-Guirado, and Vickers 2022). One reason for the higher estimated assortative mating for 18th century Spain might be that the Murcia sample focuses on individuals of relatively high status. Also in our sample, the degree of marital sorting tends to be higher at the top end of the status distribution.[21] As another reference point, the degree of status sorting of 0.62 is similar to what one obtains for educational sorting in China during the 20th century, in particular for cohorts born around the year 1970 (Dong and Xie 2023). Our results indicate that there is a long history of positive assortative matching in China that goes back to at least the early 1500s.

Panel B of Table 5 extends this analysis by allowing for differences by cohort. One finding is that the point estimate of β_2^l falls over time when only father-in-law status is included as independent variable (column (2)). These results would suggest that female intergenerational mobility has increased over time from the 17th to the 19th century.

The inclusion of both biological father and father in-law variables yields the by-cohort results shown in Panel B, column (3). The results indicate that the role of father-in-law for daughter status was diminished over time while the role of her biological father was rising. Thus, it can be challenging to estimate changes in female mobility by observing only her biological father *or* her father-in-law status. Parallel to the decline in the role of father-in-law for daughter status is a lower level of marital sorting between biological and in-law families. The correlation between status of biological father and father in-law, ρ, is 0.65 before the year 1700, falling to 0.61 in the 18th century, and to 0.59 during the 19th century (based on columns (1) and (3), Panel B).

If one accounts for both fathers by adding the point estimates of β_1^b and β_2^l cohort by cohort, the intergenerational correlation can be higher by as much as 50% relative to the estimate based only on biological father.[22] In addition, employing the sum of β_1^b and β_2^l estimates one concludes that

[21] For example, if we drop the top 5 percent of the status distribution (based on biological father status), the estimated degree of assortative mating falls to 0.60, from 0.62 for the full sample.

[22] For the first cohort, $(15.5 - 10.2)/10.2 = 0.52$.

daughter mobility is highest in the 19th century, whereas based on biological father status alone this is not the case.[23] These findings are confirmed by specifications that include female and male clan fixed effects, and the extent to which estimates that account for marital sorting by including father-in-law tend to be even larger when clan fixed effects are included, see Table A.7.

Overall, accounting for the role of marriage matters for the sources, level, and trends in woman intergenerational mobility. The following analysis presents results on the intergenerational mobility of sons with which our mobility results for daughters can be compared.

4.3 Comparison to Intergenerational Mobility of Sons

4.3.1 Father-Son Mobility

The estimation equation is analogous to the framework for women above. The following OLS regression is estimated

$$son_{i,t}^{mf} = \beta_0 + \beta_1^b father_{i,t}^{mf} + \mu_t + \nu_f + \nu_m + \varepsilon_{i,t}^{mf}, \quad (3)$$

where $son_{i,t}^{mf}$ is the percentile status rank of the son of father-son pair i with birth year t who belongs to clan m and marries a woman of clan f. Table 6 shows the results in columns (1), (2), and (3) on the left. We begin by estimating equation (3) to obtain one coefficient β_1^b for the entire sample period, before turning to the estimation of cohort-specific coefficients, as before (Panel A and Panel B, respectively).

The coefficient β_1^b is estimated at just under 0.48 in column (1), Panel A, indicating that the advantage of a higher father status is roughly halved in one generation. Notice that this is similar to what is obtained when employing husband status as the measure of daughter status (Table 4). Accounting for temporal effects related to birth year does not greatly change this result, while the inclusion of clan fixed effects lowers the estimate of β_1^b to about 0.36 (columns (2) and (3), respectively, Panel A).

Panel B shows results for each of the three cohorts. The β_1^b estimates do not vary strongly across cohorts, which is similar to the case of female intergenerational mobility above. At the same time, based on point estimates there is some evidence for a lower coefficient in the father-son regression over time, in particular in column (2). Thus, in comparison with female intergenerational mobility the evidence for rising male intergenerational mobility is stronger.

[23] $(\beta_1^b + \beta_2^l) = 0.088 + 0.032 = 0.12$ for the third cohort, compared to 0.155 for the first; in contrast, based on β_1^b alone, point estimates are 0.107 for the third and 0.102 for the first cohort, see Panel B, columns (3) and (1), respectively.

Table 6: Marriage and Intergenerational Mobility of Husbands

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Intergenerational Correlation						
Father	0.479**	0.476**	0.359**	0.478**		0.355**
	(0.014)	(0.014)	(0.015)	(0.014)		(0.017)
Father In Law					0.497**	0.280**
					(0.018)	(0.020)
Panel B: Intergenerational Correlation by Cohort						
Father Status <1700	0.487**	0.498**	0.355**	0.533**		0.427**
	(0.020)	(0.032)	(0.033)	(0.032)		(0.043)
Father Status 18th c	0.481**	0.492**	0.376**	0.491**		0.373**
	(0.015)	(0.018)	(0.018)	(0.020)		(0.023)
Father Status 19th c	0.473**	0.445**	0.335**	0.437**		0.306**
	(0.015)	(0.021)	(0.015)	(0.026)		(0.029)
Father-in-Law Status <1700					0.478**	0.204**
					(0.042)	(0.051)
Father-in-Law Status 18th c					0.507**	0.280**
					(0.025)	(0.027)
Father-in-Law Status 19th c					0.493**	0.307**
					(0.031)	(0.035)
Birth Year FE		Y	Y	Y	Y	Y
Male Clan FE			Y			
Female Clan FE			Y			
N	6,178	6,146	6,104	6,145	6,145	6,145

Notes: Dependent variable is son status. Estimation of equation (3). All status variables employ percentile ranks based on 4 groups; robust standard errors reported in parentheses; **/*/+ indicates significance at 1%/5%/10% level.

4.3.2 Marriage and Men's Intergenerational Mobility

This section provides evidence on the role of the father-in-law for the son's intergenerational mobility. The estimation equation is now given by

$$son_{i,t}^{mf} = \beta_0 + \beta_1^b father_{i,t}^{mf} + \beta_2^l father\, in\, law_{i,t}^{mf} + \mu_t + \nu_m + \nu_f + \varepsilon_{i,t}^{mf}, \tag{4}$$

where the dependent variable is the son's percentile rank. Results are shown in Table 6, on the right side. First, consider the estimate of β_2^l on father- in-law in column (5) of Panel A; at 0.497, the point estimate is somewhat larger even than the estimate of β_1^b for the son's biological father, shown in column (4). This finding speaks to the great importance of the father-in-law, and the son's marriage to bride, for son status.

With the inclusion of both father and father-in-law status variables, the respective coefficients β_1^b and β_2^l are estimated at 0.355 and 0.280, see column (6), Panel A. While in a direct comparison the point estimate for the biological father is thus larger than that for the father-in-law, the latter is still somewhat more important for son mobility than he is for daughter mobility.[24]

Over time, the estimate of β_1^b for the father contribution tends to decrease while the estimate of β_2^l for the father-in-law contribution increases, see column (6), Panel B. This contrasts with the roles of biological father and father-in-law for the mobility of daughters, where the relative contribution of the biological father rises over time (Table 5). What might explain this result?

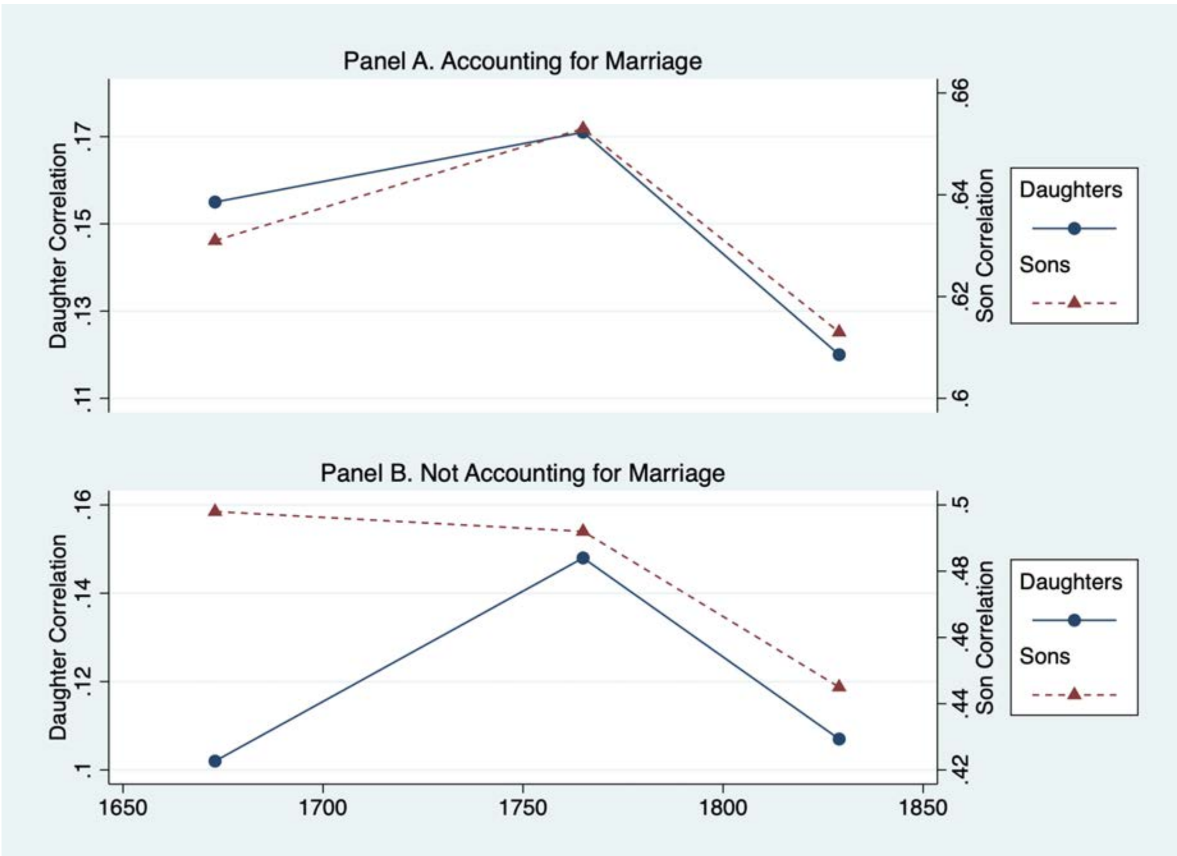
Given that our analysis of female and male intergenerational mobility is based on the same couples, the biological fathers from the point of view of the sons are the same persons as the father-in-laws from the point of view of the daugthers. This might suggest that the finding is due to the biological fathers of the seven clan sons' in the sample being less accomplished over time. However, changes over time in the status for the seven male clans are comparable to the changes in the status of the fathers of the in-marrying women. Furthermore, if between-clan differences were important for the result that for women the importance of the biological father is rising while for men it is falling over time, we would obtain different results once we focus on within-clan differences by including clan fixed effects. Yet, results on the changing importance of father-in-law for men versus women turn out to be similar when clan fixed effects are included, see Table A.7. Thus, the rising importance of marriage for son mobility at the same time when the importance of marriage for daughter mobility is falling is not due to differential status trends across clans.

Figure 6 summarizes several of the major findings of this study. Panel A shows trends in intergenerational mobility for daughters and sons based on the correlation of the status of the child with their biological father (i.e., coefficient β_1^b),while Panel shows these trends based on biological father plus father-in-law (i.e., $\beta_1^b + \beta_2^l$). First, note that mobility levels are different once marital sorting is taken into account, as indicated by the different vertical scales. Accounting for marital sorting

[24]In the case of sons, 0.28/(0.28 + 0.355) = 44%, in contrast to 40% for the in-law father in the case of daughters; see Tables 6 and 5, respectively.

yields a higher level of status persistence. For example, the intergenerational correlation for sons in the 18th century is 0.65 once the status of the father-in-law is taken into account, compared to 0.49 without.

Figure 6: Trends in Intergenerational Mobility: The Role of Marital Sorting



Notes: Shown in Panel A is $\beta_1^b + \beta_2^l$, the sum of biological father and father in-law point estimates, while in Panel B the point estimate of biological father, β_1^b, is presented. Horizontal axis is median birth year in cohort. Left axis gives father-daughter status correlation, right axis father-son status correlation.

Second, assessing trends in intergenerational mobility without accounting for marital sorting yields often wrong results. In particular, without taking into account the role of marital sorting, the evidence suggests that son mobility rises from the 17th to the 18th century (Panel B), when it fact it falls (Panel A). Similarly, without marital sorting one would conclude that daughter mobility if anything declines from the 17th to the 19th century (Panel B), when in fact daughter mobility tends to rise from the 17th to the 19th (Panel A). Third, once the role of marital sorting is taken into account, trends in male and female intergenerational mobility are quite similar. There is a fall in mobility from the early centuries to the 18th century before mobility rises towards the 19th century (Panel A). This is what one might have expected given that marriage creates ties between daughters and sons that are implicitly broken when the role of the in-laws is not accounted for.

4.3.3 Multigenerational Mobility of Sons and Marriage

This section turns to the role of marriage in the context of multigenerational mobility, defined as intergenerational relationships beyond that of parent and child. We estimate versions of the following OLS regression

$$son_{i,t}^{mf} = \beta_0 + \beta_1^b father_{i,t}^{mf} + \beta_2^l father\,in\,law_{i,t}^{mf} + \beta_3^b gff_{i,t}^{mf} + \beta_4^b gfm_{i,t}^{mf} + \mu_t + \nu_f + \nu_m + \varepsilon_{i,t}^{mf}, \quad (5)$$

where the newly added terms are the status levels of the son's grandfathers on father and mother side ($gff_{i,t}^{mf}$ and $gfm_{i,t}^{mf}$,respectively). Table 6 shows the results. For data availability reasons the focus will be on the mobility of the son. Adding another generational lag means that the sample size shrinks because it is now conditional on observing the triplet son-father-grandfather, and for the earliest two generations in the sample grandfathers are not observed. We begin by showing that the sample change associated with the additional generational lag does not have a major impact on our results, see column (1) of Table 7.

Table 7: Marriage and Multigenerational Mobility

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Father	0.344**	0.357**	0.291**	0.239**	0.314**	0.278**	0.231**
	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Father-in-Law	0.266**		0.221**	0.165**	0.220**	0.196**	0.152**
	(0.022)		(0.022)	(0.022)	(0.023)	(0.023)	(0.023)
Grandfather (f)		0.171**	0.125**	0.100**		0.107**	0.089**
		(0.014)	(0.014)	(0.013)		(0.014)	(0.014)
Grandfather (m)					0.110**	0.074**	0.049**
					(0.018)	(0.019)	(0.018)
Birth Year FE	Y	Y	Y	Y	Y	Y	Y
Male Clan FE				Y			Y
Female Clan FE				Y			Y
N	4,973	4,973	4,973	4,932	4,973	4,973	4,932

Notes: Dependent variable is son percentile rank status. Estimation of equation (5). Grandfather (f) is the father's father, Grandfather (m) is the mother's father. All status variables employ percentile ranks based on 4 groups; robust standard errors reported in parentheses; **/*/+ indicates significance at 1%/5%/10% level.

If the father in-law status variable is replaced by the son's grandfather on the male side–his father's father–, results shown in column (2) are obtained. Consistent with other research (e.g. Solon 2014, Braun and Stuhler 2018, and Shiue 2019), the positive coefficient on grandfather (m) indicates that generations before that of the parent matter. However, notice that the point estimate for grandfather (m) is smaller than that on father-in-law status. When both grandfather (m) and father-in-law are included, both variables have a positive coefficient but that on father-in-law is significantly larger (p-value of 0.1%; column (3)). Also in the specification with clan fixed effects, the coefficient on father-in-law is significantly larger than that of (male) grandfather (column (4)). We conclude that according to these the results the role of the in-marrying family is larger than the role of higher-order ancestors.

The remaining specifications provide evidence for mobility also on the role of the maternal grandfather. Notice that the point estimate is 0.11, which compares to a point estimate of 0.17 for the paternal grandfather (columns (5) and (2), respectively). When both grandfather status variables are included, both turn out to have positive coefficients (column (6)). Furthermore, the relative size of the maternal grandfather coefficient is about 70% (= 0.074/0.107). This suggests that the role of the husband's mother's father relative to his father's father is comparable to the role of his father-in-law relative to his father (from column (6), 0.196/0.278 = 0.71). This indicates that there is a certain continuity in the importance of marriage for mobility across generations. Finally, results with clan fixed effects are qualitatively similar, see column (7).

5 Summary and Conclusions

With more than 6,000 observations from seven genealogies of central China, we study the transmission of status from father to daughter using original information on the status of women. We have provided evidence that a woman who becomes first wife of her husband has higher status than other women, because there is relatively strong positive sorting between biological and in-law families for first-wife marriages, and moreover, the husbands of first wives have relatively high status.

First, we find that starting in the early 15th century, daughters who have a high-status father are more likely to be the first wife in their households, and the strength of the intergenerational correlation of status does not only depend on her father but also on the clan the daughter comes from. Second, accounting for the role of marriage changes our understanding of the level, sources, and trends in the intergenerational mobility of daughters. In particular, with a correlation of around 0.6 between the statuses of biological and in-law families, marriage patterns are far from random, and analysis that abstracts from marital sorting would considerably overestimate father-daughter mobility. Furthermore, the contribution of in-law family to daughter status is with 40% sizable, and analysis based solely on biological family information might introduce error. In addition, the extent of marital sorting in the sample fell from the 17th to the 19th century. Thus, accounting for marital sorting is crucial for consistent comparisons of woman's intergenerational mobility across centuries.

Third, overall the implied level of mobility in the father-daughter relationship is comparable to that in the father-son relationship. Further, once one accounts for marital sorting the intergenerational mobility of daughters follows a similar trend as that of the son over the sample period. The importance of marital sorting for intergenerational mobility is further highlighted by finding that a high-status father-in-law is typically more important for son status than either paternal or maternal grandfather.

References

[1] Abramitzky, R., L. Boustan, K. Eriksson, J. Feigenbaum, and S. Perez (2021), "Automated Linking of Historical Data", Journal of Economic Literature Vol. 59(3): 865-918.

[2] Abramitzky, Ran, Adeline Delavande, and Luis Vasconcelos (2011), "Marrying Up: The Role of Sex Ratio in Assortative Matching", American Economic Journal: Applied Economics Vol. 3(3), pp. 124-157.

[3] Ahern, Emily M. 1974. "Affines and the Rituals of Kinship." In Religion and Ritual in' Chinese Society , ed. Arthur P. Wolf. Stanford: Stanford University Press.

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

[5] Bailey, Martha J., and Peter Z. Lin (2022), "Marital Matching and Women's Intergenerational Mobility in the Late 19th and Early 20th Century U.S.", UCLA working paper, November.

[6] Bailey, M., Lin, P. Z., Mohammed, A. R. S., Mohnen, P., Murray, J., Zhang, M., & Prettyman, A. (2023), "The Creation of LIFE-M: The Longitudinal, Intergenerational Family Electronic Micro-Database Project", working paper, June.

[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, G. S. (1974). A Theory of Marriage: Part II. Journal of Political Economy, 82(2, Part 2), S11–S26.

[9] Becker, G. S. (1973). A Theory of Marriage: Part I. Journal of Political Economy, 81(4), 813–846.

[10] Braun, Sebastian Till, and Jan Stuhler (2018), "The Transmission of Inequality Across Multiple Generations: Testing Recent Theories with Evidence from Germany", *Economic Journal* 128, pp. 576-611.

[11] Buckles, Kasey, Joseph Price, Zach Ward, and Haley Wilbert (2023), "Family Trees and Falling Apples: Intergenerational Mobility Estimates from U.S. Genealogy Data", paper presented at the NBER Summer Institute, July 2023.

[12] Bull, Hans Henrik (2005), "Deciding Whom to Marry in a Rural Two-Class Society: Social Homogamy and Constraints in the Marriage Market in Rendalen, Norway, 1750–1900", International Review of Social History Vol. 50, Supplement 13, pp. 43-63.

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

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

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

[16] Chadwick, Laura, and Gary Solon (2002), "Intergenerational Income Mobility Among Daughters", American Economic Review 92(1), pp. 335-344.

[17] 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", The Quarterly Journal of Economics 129 (4): 1553-1623.

[18] Choo, Eugene, and Aloysius Siow (2006), "Who is Marrying Whom and Why", Journal of Political Economy Vol. 114, No. 1, pp. 175-201.

[19] Craig, Jaqueline, Katherine Eriksson, and Gregory T. Niemesh (2019), "Marriage and the Intergenerational Mobility of Women: Evidence from Marriage Certificates 1850-1910", working paper.

[20] Curtis, Matthew (2022), "The her in inheritance: how marriage matching has always mattered, Quebec 1800-1970", ECARES working paper, September.

[21] Dahl, Molly, and Thomas DeLeire (2008), "The Association between Children's Earnings and Fathers' Lifetime Earnings: Estimates Using Administrative Data", University of Wisconsin-Madison, August.

[22] Dong, Hao, and Yu Xie (2023), "Trends in Educational Assortative Marriage in China Over the Past Century", *Demography* Vol. 60(1), pp. 123-145.

[23] Dribe, Martin, Bjorn Eriksson and Francesco Scalone (2019), "Migration, marriage and social mobility: Women in Sweden 1880-1900", Explorations in Economic History 71(C), pp. 93-111.

[24] Dribe, Martin, and Christer Lundh (2005), "Finding the Right Partner: Rural Homogamy in Nineteenth-Century Sweden", International Review of Social History Vol. 50, Supplement 13, pp. 149-177.

[25] Ebrey, Patricia Buckley (2023), "Women in Traditional China", Asia Society, New York, NY, https://asiasociety.org/education/women-traditional-china, accessed Aug 17, 2023

[26] Ebrey, Patricia Buckley (1991), "Shifts in Marriage Finance from the Sixth to the Thirteenth Century." In Watson, Rubie S. and Patricia Buckley Ebrey, Eds. *Marriage and Inequality in Chinese Society.* Los Angeles, CA: University of California Press.

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

[28] Espin-Sanchez, Jose-Antonio, Salvador Gil-Guirado, and Chris Vickers (2022), "La "Doña" è Mobile: The Role of Women in Social Mobility in a Pre-Modern Economy", Journal of Economic History 82(1): 1-41; doi: 10.1017/S0022050722000018

[29] 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).

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

[31] Fernández, Raquel (2013), "Cultural Change as Learning: The Evolution of Female Labor Force Participation over a Century", American Economic Review 103: 472-500.

[32] 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.

[33] Freedman, Maurice (1979), The Study of Chinese Society, ed. G. William Skinner. Stanford: Stanford University Press.

[34] Goldin, Claudia (1983), "The changing economic role of women: A quantitative approach", The Journal of Interdisciplinary History 13 (4), 707–733.

[35] Hamilton, Gillian, and Aloysius Siow (2007), "Class, Gender and Marriage," Review of Economic Dynamics, vol. 10(4), pages 549-575, October.

[36] Harrell, Stevan (1987), "On the Holes in Chinese Genealogies", Late Imperial China, Volume 8, Number 2, pp. 53-79

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

[38] Jácome, Elisa, Ilyana Kuziemko, and Suresh Naidu (2022), "Mobility for All: Representative Intergenerational Mobility Estimates over the 20th Century", NBER # 29289, September.

[39] Jäntti, Markus, Bernt Bratsberg, Knut Roed, Oddbjorn Raaum, Robin Naylor, Eva Osterbacka, Anders Bjorklund, and Tor Eriksson (2006), "A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States", IZA Discussion Paper # 1938, January.

[40] Kalmijn, Matthijs (1994), "Assortative Mating by Cultural and Economic Occupational Status", American Journal of Sociology, Vol. 100, No. 2 (Sep., 1994), pp. 422- 452.

[41] Keller, Wolfgang, and Carol H. Shiue (2022), "Marriage Matching over Five Centuries in China", NBER # 30625.

[42] Ko, Dorothy (1994), *Teachers of the Inner Chambers: Women and Culture in Seventeenth-century China.* Stanford University Press.

[43] Mann, Susan. (1997), *Precious Records: Women in China's Long Eighteenth Century.* Stanford University Press.

[44] Mann, Susan (1991), "Grooming a Daughter for Marriage: Brides and Wives in the Mid-Qing Period", in Watson and Ebrey (eds.), *Marriage and Inequality in Chinese Society.* Los Angeles, CA: University of California Press.

[45] Mare, Robert (1991), "Five decades of educational assortative mating", American Sociological Review 56: 15-32.

[46] Mazumder, Bashkar (2005), "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data", Review of Economics and Statistics Vol. 87(2), pp. 235-255.

[47] Olivetti, Claudia (2013), "The Female Labor Force and Long-run Development: The American Experience in Comparative Perspective", November 2013.

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

[49] Olivetti, Claudia, and M. Daniele Paserman, Laura Salisbury, and E. Anna Weber (2022), "Who married, (to) whom, and where? Trends in marriage in the United States, 1850-1940", working paper, Boston University, October.

[50] Price, Joseph, Kasey Buckles, Jacob Van Leeuwen, and Issac Riley (2021), "Combining family history and machine learning to link historical records: The Census Tree data set", Explorations in Economic History 80, https://doi.org/10.1016/j.eeh.2021.101391

[51] Shan, Chun (2013), "Marriage law and Confucian Ethics in the Qing dynasty", Frontiers of Law in China, Vol. 8, Issue 4, December.

[52] Shiue, Carol H. (2019), "Social Mobility in the Long Run: A Temporal Analysis of China from 1300 to 1900", CEPR Discussion Paper # 13589.

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

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

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

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

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

[58] Ward, Zach (2023), "Intergenerational Mobility in American History: Accounting for Race and Measurement Error", January; forthcoming, American Economic Review.

[59] Watson, Rubie S. (1981). "Class Differences and Affinal Relations in South China." *Man* 16: 593-615.

[60] Watson, Rubie (1985), *Inequality among Brothers, Class and Kinship in South China*. Cambridge University Press.

[61] Watson, Rubie S. and Patricia Buckley Ebrey, Eds. (1991), *Marriage and Inequality in Chinese Society.* Los Angeles, CA: University of California Press.

[62] Weiss, Yoram (1997), "The formation and dissolution of families: Why marry? Who marries whom? And what happens upon divorce", Chapter 3 in *Handbook of Population and Family Economics*, Volume 1A, (eds.) M. Rosenzweig and O. Stark, Elsevier Publishers, pp. 81-123.

[63] Wolf, Arthur P. and Chieh-Shan Huang (1980), *Marriage and Adoption in China, 1845-1945.* Stanford: Stanford University Press.

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

Appendix

A Data

A.1 Sample Construction and Characteristics

As noted in the text, the sample was created with a targeted approach that selected a subset of genealogies from Tongcheng with the explicit goal of obtaining a broadly representative sample. Seven clans have been chosen for inclusion into the sample, with the following last names: Chen, Ma, Wang, Ye, Yin, Zhao, and Zhou. Conditional on choosing the genealogy of a given clan, all entries of the genealogy become part of the sample. The seven clans' genealogies are also relatively complete in terms of recording vital information. It is well-known that Chinese genealogies vary in the degree of completeness. Among the 75 genealogies from Anhui provinces surveyed by Telford (1986), for example, there is vital information on both husband and wife in 76% of the cases. In the present estimation sample, there is vital information for 79.5% of the husbands and 79.1% of the wives.

Figure A.1 shows data for Tongcheng county's overall population that is available from official sources ("registered" population from tax registers). Official data is only available for 21 years. Given this sparsity, as well as virtually unchanged levels for more than a century during a period for which genealogies indicate steady population growth suggests that official population data for this period in Tongcheng is, at a minimum, incomplete.

The limited availability of reliable official data for this period means that one cannot establish the representativeness of the sample by making comparisons to official data for all of Tongcheng. Instead, we rely on broader comparisons wherever data is available, at the national or regional level, and for subsets of our sample period.

A.2 Sample Representativeness

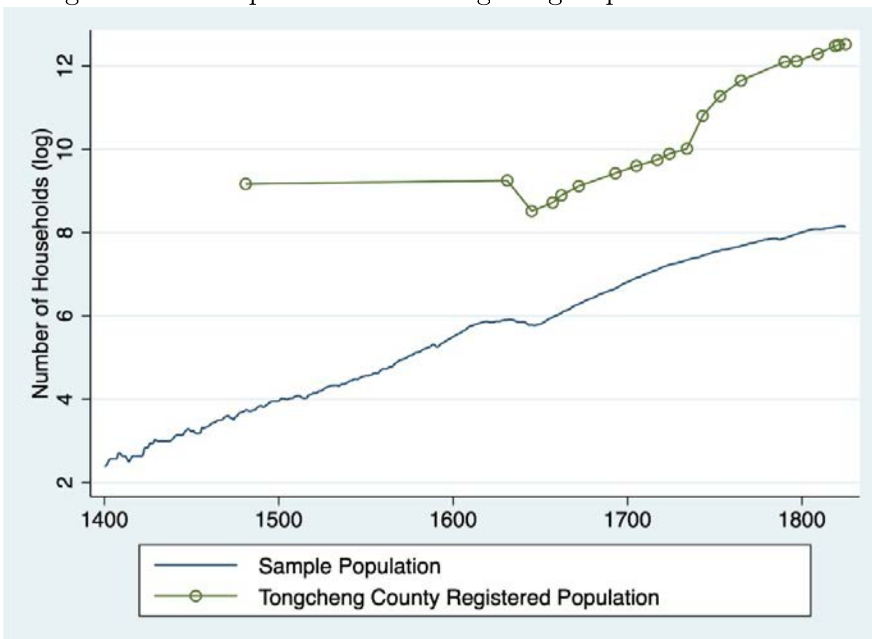
A.2.1 Comparisons to Other Statistics

A first check is to consider mortality rates by age group. Population figures at the regional level are typically based on gazetteers (local histories about a certain place).[25] Telford (1990) compares demographic patterns in the Tongcheng genealogical data and the Eight 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).

As noted in the text, a sample consisting of these seven clans matches national shares of upper status population as defined by Fei (1946) and Chang (1955). One might also ask how the present sample compares in terms of the very top status groups to other evidence. In his classic study based on national lists of *jinshi*, which are extremely reliable, Ho (1967) reports that during the Qing

[25] Three county-level gazetteers about Tongcheng cover the period under analysis; they are *Tongcheng xian zhi* (1490), *Tongcheng xian zhi* (1696), *Tongcheng xuxiu xian zhi* (1827).

Figure A.1: Sample and Total Tongcheng Population over Time



Notes: Lower series is sample household population, measured as log number of male heads-of-households (N = 426); upper series is log number of households, Beattie (1979), Table 3 (N = 21). Also shown the fraction of the sample based on these series; scale on the right.

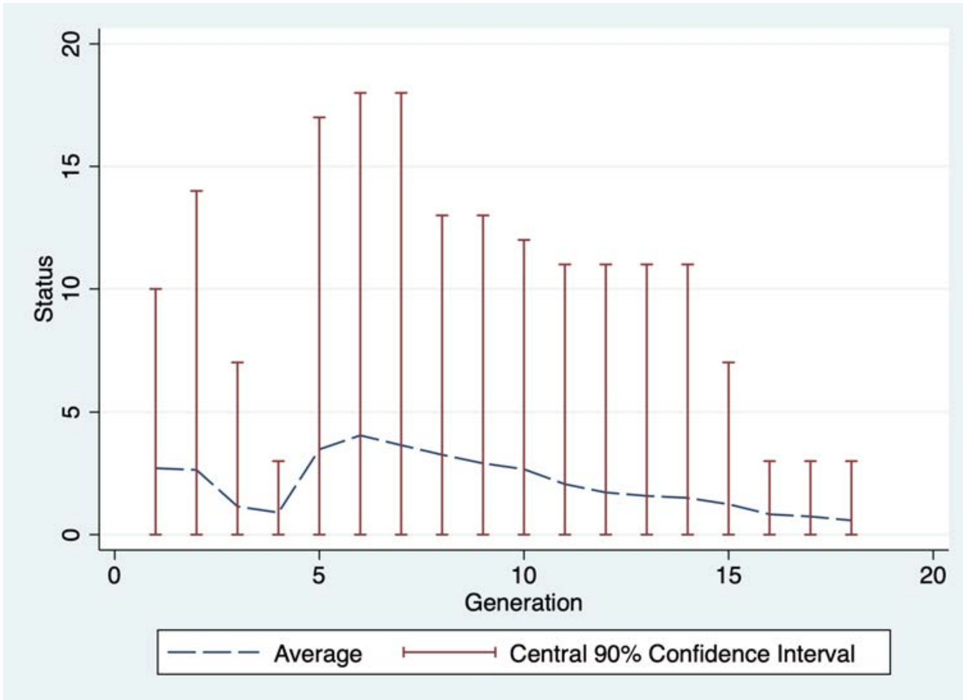
in Anhui there were 41 *jinshi* per one million population, or, 0.0041 percent. There were regional variations, and the province of Anhui was below the provincial average in terms of *jinshi* per capita in Qing China (Ho 1967, p. 228). In the Tongcheng sample, there were a total of 14 *jinshi* during the Qing, which is about 0.045 percent of the population in the data. Thus, there are about ten times more *jin-shi* in the sample than in Qing Anhui overall.

At the same time, *jinshi* were rare, with many parts of Anhui province not producing a single *jinshi* over centuries. Also, other regions of China, in particular in Zhejiang and Jiangsu province, produced *jinshi* by an order of magnitude higher than Tongcheng (see Ho 1967). Overall, Tongcheng was a noteworthy place at a local, perhaps provincial level, but it was not an unusual Chinese region. In addition, the list of people who are recorded in the sample as *jinshi* can be compared against other lists of *jinshi* degree holders from the Chinese state (see Fang 2010; Cao 2016; Wang 2017). We have verified that the information on the *jinshi* in the sample is consistent with the information of these official lists.

A.2.2 Inception of the Genealogy

Across the seven clans, the first recorded husband birth year ranges from 1298 (Chen clan) to 1598 (Zhao clan). One might ask why the clan decides to establish its genealogy at a particular time. If the genealogy of a clan is established because one clan member had achieved extraordinarily high status, this might affect intergenerational mobility estimates. To examine this, Figure A.2 shows average clan status by generation.

Figure A.2: Clan Status by Generation



Notes: Shown are clan average status and the central 90% confidence interval of clan status for the first to eighteenth generation.

Figure A.2 shows that average clan status does not monotonically fall across generations following the inception of the genealogy. In particular, average clan status from the 5th to the 10th generation is higher than in the first generation. Thus, the role of the primogenitor of each clan is limited in our sample.

A.2.3 Clan Representation Over Time

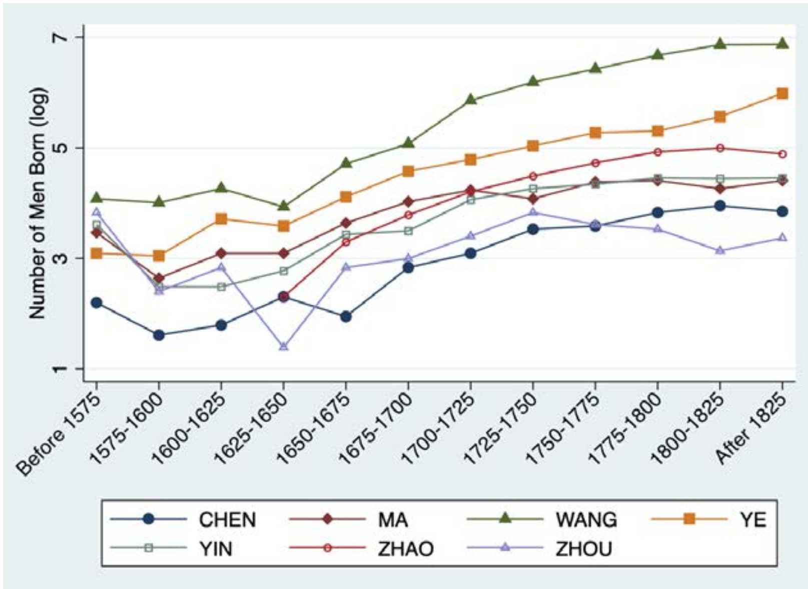
Figure A.3 shows the size of each clan in terms of the number of men born in each birth cohort. All clans are present during the period from 1575 to 1825.

In general, size grows for all clans, although growth rates vary by period; for example, the growth rate is negative for three clans in the birth cohort 1625-1650, which includes the dynastic transition from the Ming to the Qing (1644). Some clans grow faster than others, however, sustained overtaking in terms of clan size is rare.

A.2.4 Recall Bias

Other concerns result from the retrospective nature of how genealogies are compiled, in particular recall bias. Genealogies are typically updated after every two or three generations. One might be concerned that the updating is correlated with the particularly high status of a clan member, and that it leads to the reporting of many new members. In that way, the clan would use its resources to confirm the significance of its achievements ex-post. Another channel for clan size growth might be

Figure A.3: Sample Size over Time



Notes: Shown is the size of each clan as measured by the number of men born in each of the twelve birth cohorts shown.

that people who were not previously part of the clan might try to establish an ancestry relationship to the high status individual. We employ panel regression analysis to provide evidence on this and related hypotheses (seven clans by twelve birth cohorts, as shown in Figure A.3). Table A.1 shows the results.

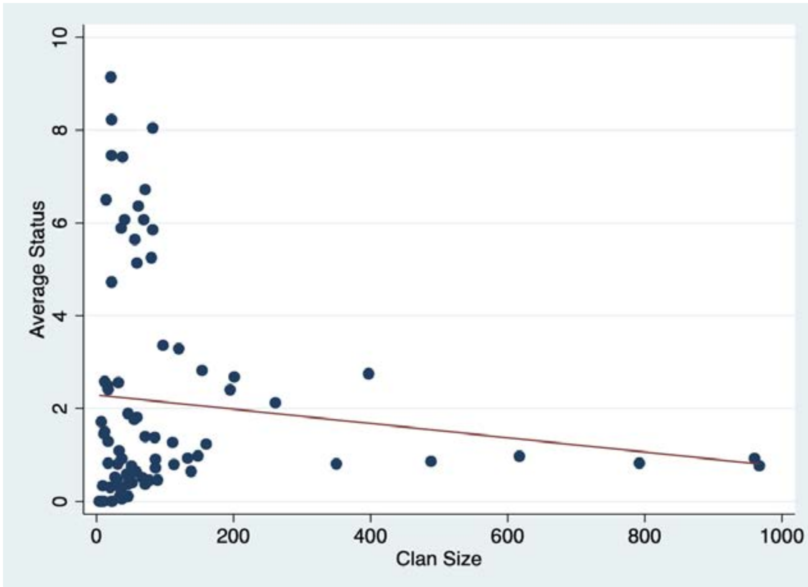
There is a negative point estimate for lagged average clan status (not significant, see Table A.1, column (1)), which does not support the hypothesis that notable achievements in the recent past trigger the inclusion of new members into the clan's genealogy. This does not change with the inclusion of birth cohort and clan fixed effects, see Table A.1, columns (2) and (3). Overall, we do not find evidence that recall bias plays a major role for the composition and characteristics of the sample.

Table A.1: Clan Success and Recall Bias

	Clan Size		
	(1)	(2)	(3)
Clan Status (-1)	−0.098 (0.096)	−0.026 (0.098)	−0.061 (0.157)
Birth Cohort Fixed Effects		Y	Y
Clan Fixed Effects			Y
N	74	74	74

Notes: Dependent variable is clan size, measured by the number of men born in a cohort (in units of 100s). Clan Status (-1) is average clan status lagged by one birth cohort. Estimation by OLS. Standard errors in parentheses.

Figure A.4: Clan Size and Status: Birth Cohort Analysis



Notes: Shown is the size of each clan as measured by the number of men born in each of twelve birth cohorts, versus average clan status in that birth cohort; N = 81.

A.2.5 Using Cross-Clan Variation to Assess Potential Bias

Because the creation and maintenance of a genealogy requires resources and literacy, the key selection concern for the sample is wealth bias. We can shed direct evidence on this using cross-clan variation, given that the seven clans differ substantially in their average wealth as proxied by status, see Table 3. It was shown in the text that clan size does not simply reflect a clan's resources, measured either by status, by wives per man, or by number of children (Figure 5). There is no evidence that richer clans systematically report more new members than poorer clans.

Wealth Bias: Additional Results One might be concerned that even though there is no cross-sectional correlation between a clan's size and its resources, there could be such a correlation in the temporal dimension. To provide evidence on this, the following examines such temporal correlation between status and clan size across twelve birth cohorts.[26] We begin with a scatterplot, see Figure A.4.

Figure A.4 shows that there is no strong relationship between status and clan size across birth cohorts, and to the extent that there is a relationship it is negative. The following extends this in two ways. First, in addition to average status, we consider the relationship between clan size and other indicators of resources, namely fertility and female partner per man. Second, we include various sets of fixed effects, which are employed in some parts of the analysis in the text to reduce omitted variables bias. Table A.2 shows the results.

[26] The birth cohorts are those of Figure A.3.

Table A.2: Clan Size versus Status, Fertility, and Female Partners per Man

	Status			Fertility			Female Partners per Man		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Clan Size	−0.154	−0.074	−0.078	0.012	0.016	−0.019	0.005	0.001	−0.009
	(0.142)	(0.165)	(0.109)	(0.026)	(0.028)	(0.037)	(0.007)	(0.008)	(0.009)
Birth Cohort FE		Y	Y		Y	Y		Y	Y
Clan FE			Y			Y			Y
N	81	81	81	81	81	81	81	81	81

Notes: Results from OLS regressions of the variable on top of the column on the number of men born (size), both averaged by clan and by birth cohort. Clan size in 100s of men. Fertility is measured as the number of male children. FE stands for fixed effects. Standard errors in parentheses.

A.2 confirms that there is no positive relationship between status and clan size. In particular, accounting for changes that are common to all clans by including birth cohort fixed effects does not yield a significant relationship (A.2, column (2)). Furthermore, there is no significant within-clan relationship between status and size either, as A.2, column (3) indicates.

There is a positive coefficient for fertility and clan size, however, it is not significant at standard levels (A.2, column (4)). Moreover, the positive point estimate turns negative (still insignificant) when time-invariant differences across clan size are accounted for by including clan fixed effects (column (6)). Qualitatively similar results are obtained for the relationship between clan size and the number of female partners per man, see columns (7) to (9). Overall, there is little evidence that a clan's representation in the sample is positively related to cross-clan variation in resources.

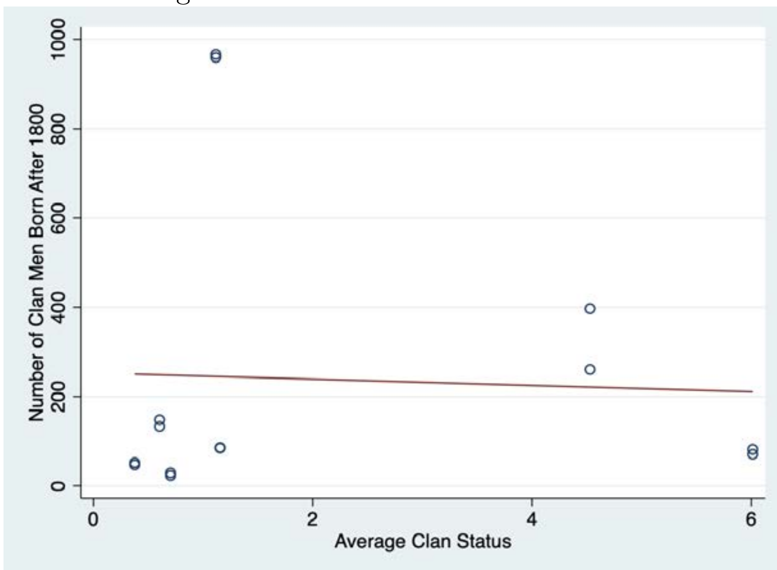
Survivorship Bias A related concern is survivorship bias: over time, this type of bias could result in a disproportionately fraction of high-achieving (high skill) individuals compared to low-achieving (low skill) individuals. One implication is that towards the end of the sample period the distribution becomes skewed towards relatively high-achieving clans. As Figure A.5 shows, however, there is little evidence that the distribution of the sample towards the end of the sample period is becomes skewed towards high-status clans.

A.2.6 Subsample with Status on Female's Father: Sample Re-weighing

As noted in section 3.2, there are differences for the subsample of the seven genealogies for which information on the status of the wife's father is available. Table A.3 provides additional information. In particular, for the 54 percent of the sample for which status information on the female's father is available, the average percentile rank status of the wife is 0.505, in contrast to 0.494 for those observations for which information on the female's father's status is not available (right column).

Despite its statistical significance, the difference in average status, about one percentage point, does not appear to be large in a quantitative sense. However, if the availability of status information

Figure A.5: Evidence on Survivor Bias



Notes: Figure shows number of clan men born in periods 1800-1825 and post-1825, both versus average clan status.

Table A.3: Status of Female's Father Data Availability

	N	Wife Status
Information Available	6,178	0.505 (0.002)
Information Not Available	5,200	0.494 (0.002)
Difference		0.011
p-value		< 0.01
All	11,378	

Notes: Table gives tests for equality of means of wife's status percentile rank by subsample; standard error given in parentheses.

for the female's father varies systematically with other status measures employed in the analysis, the sample size restriction might affect the results. In particular, status of the female's father information is available for 59% of the first wives, but only for 54% of the fathers of the females that are not first wives.

To address this issue, the following analysis re-weights the estimation sample by the inverse of the probability that each type of female role in the household is present in the restricted sample (see Table A.5). This ensures that the sample weights add to N = 11,378, as in the unrestricted sample (see Table A.3). Results for the reweighted sample are shown on the left side of Table A.4, with the corresponding unweighted results shown on the right.

Table A.4: Sample Restriction – Reweighted Sample Results

	(1)	(2)	(3)	(4)	(5)	(6)
	Sample Re-weighing			No Re-weighing		
Father Status	0.121** (0.015)	0.087** (0.017)		0.127** (0.016)	0.090** (0.018)	
Father in Law Status		0.056** (0.013)			0.060** (0.014)	
Father Status <1700			0.008 (0.044)			0.006 (0.047)
Father Status 18th c			0.109** (0.023)			0.112** (0.024)
Father Status 19th c			0.085** (0.029)			0.088** (0.030)
Father-in-Law Status <1700			0.142** (0.037)			0.151** (0.039)
Father-in-Law Status 18th c			0.054** (0.017)			0.059** (0.018)
Father-in-Law Status 19th c			0.029 (0.023)			0.032 (0.024)
Sum of Sample Weights	11,378	11,378	11,378	6,146	6,145	6,145
N	6,146	6,145	6,145	6,146	6,145	6,145

Notes: Dependent variable is female status; columns (1) to (3) apply weights to account for difference in observing status of female's father by female role in household; all specifications include birth year fixed effects. Robust standard errors in parentheses.

Comparing results with reweighted samples on the left with the unweighted results on the right of Table A.4 indicates that the restriction to limit the sample to observations for which the status of the female's father is observed does not qualitatively affect the results. The is the case for the father-daughter mobility finding (compare columns (1) and (4), respectively), for the role of marital sorting for female intergenerational mobility (columns (2) and (5), respectively), and also the results by cohort are similar (columns (3) and (6), respectively). We conclude that the influence of the sample restriction to observations for which the status of the wife's father is observed is quite limited.

A.2.7 Estimation of Vital Statistics

While the genealogies underlying the sample for this study report vital statistics for roughly 80 percent of the individuals, the data is estimated or cleaned in other ways in some cases. Estimation would typically use auxiliary information, such as a particular individual "would die the year that the Taipings came to Tongcheng", which would have to be year 1634 or 1635, or life tables, a standard tool of demographic analysis. To examine the potential influence of this for the results, we allow for separate coefficients in the estimation of the intergenerational mobility coefficient β_1^b for estimated or adjusted data:

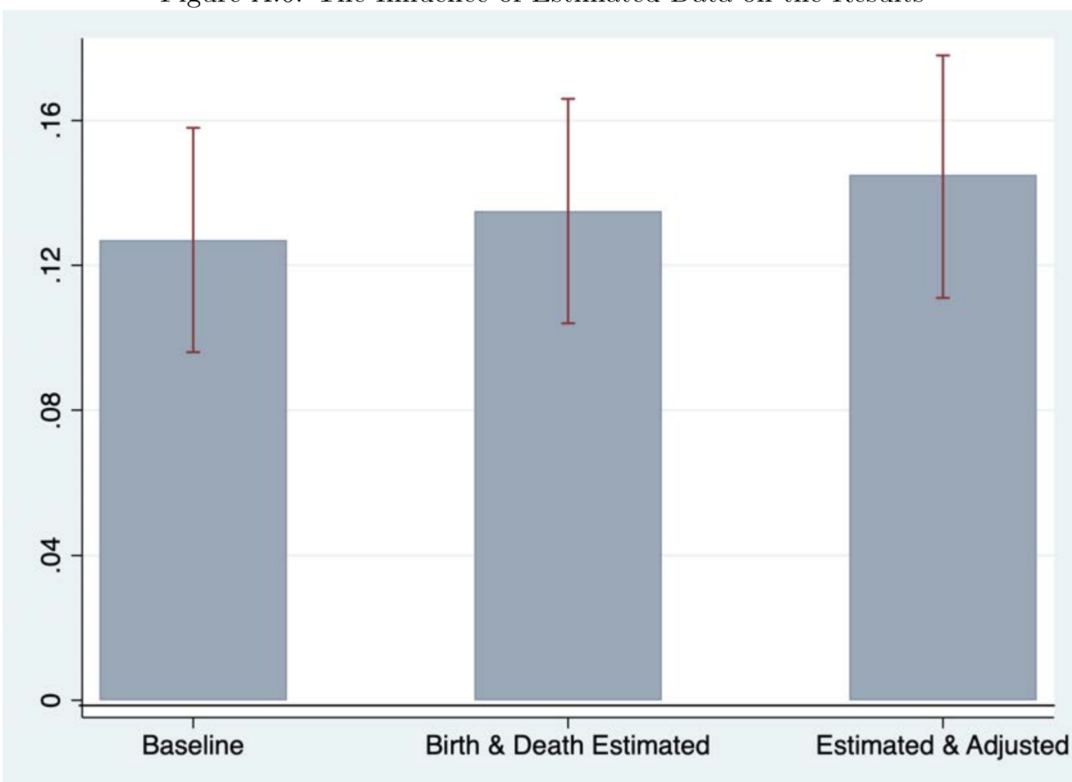
$$daug_{i,t}^{fm} = \beta_1^b father_{i,t}^{fm} + \beta_2^b [I^e = 1] \times father_{i,t}^{fm} + \beta_3^b [I^e = 1] + \mu_t + \varepsilon_{i,t}^{fm}, \tag{6}$$

where I^e is an indicator whether vital statistics for this observation are estimated or adjusted; equation (6) is based on equation (1). Figure A.6 shows results from this analysis.

Results on the left in whichβ_2^b and β_3^b are constrained to zero indicate that the slope coefficient β_1^b is equal to 0.127, confirming the earlier finding in column (2), Table 4. When differential slope coefficients are allowed for for estimated and non-estimated data, the estimate of β_1^b becomes somewhat larger (middle of the figure), and it rises further once we allow for the estimation or adjustment of vital statistics (on the right). This suggests that the estimation of vital statistics leads to some classical measurement error that pushes the slope coefficient estimate β_1^b towards zero. This would translate to an overestimate of mobility. At the same, there is no significant difference between the slope estimate β_1^b with or without distinguishing estimated & adjusted data or not, and quantitatively the difference between estimates does not appear to be large.

Overall, the analysis of the representativeness of the sample has revealed that it cannot be conclusively established that the sample is representative of the population living during the sample period in Tongcheng, primarily because reliable representative statistics that are unrelated to clan records are not yet available. Comparisons of the sample with other well-known facts about China during the sample period suggest that the sample consisting of seven clans is broadly representative, and in moreover, cross-clan evidence from within the sample rejects the presence of wealth as well as recall and survivorship bias. The estimation sample for which there is information on the status of the female's father has somewhat higher average status, however, the influence of this on our estimation results is limited. We have also seen that the influence of the estimation of vital statistics for key results is limited.

Figure A.6: The Influence of Estimated Data on the Results



Notes: On the left, the slope coefficient β_1^b and 95% confidence interval is shown from estimating equation (6) with $\beta_2^b = 0$ and $\beta_3^b = 0$; in the middle, results for β_1^b are shown for non-zero β_2^b and β_3^b in the case of that both birth year and death year are estimated (N = 262), while on the right side results for β_1^b are shown for non-zero β_2^b and β_3^b in the case of that both birth year and death year are estimated or adjusted (N = 1,605).

Table A.5: Female Roles in Household

	Status	N	Percent
First Wife	1	813	13.2
Single Wife	0	4,659	75.4
Second and Subsequent Wife	0	678	11.0
Concubine	0	2	< 0.1
Betrothed	0	26	0.4
Total		6,178	100.0

Notes: Shown is female role in household for sample for which status of female's father is observed (N = 6,178).

A.3 Status Measures

This section describes the female and male status measured employed in this paper in more detail.

A.3.1 Female Status

We assume that the status of a female depends on her role in the household. For about three quarters of the observations in the estimation sample with N = 6,178, a man has only one female partner during his lifetime. In Table A.5, this case is labeled as "Single Wife".[27] If a man over his lifetime has more than one female partner, the genealogies provide information on which of the females was the wife the husband married. This is the case for 13 percent of the sample, see Table A.5. As described in the text, we code these women as having elevated status (column (1) gives this indicator variable).

Eleven percent of the women are identified as second or subsequent wives. The reason that this share is lower than the share of first wives is that status information on the female's father is available at a higher rate for first wives than for other females. Table A.4 shows that the influence of this on key results is small. Not all females are married to their husband; a relatively small number are concubines (female partners without being married) and others are betrothed only, which means that marriage was not consumated.

A.3.2 Male Status

This section describes the status classification for males in this paper. It applies to the husband, his father, his father-in-law, as well as to the husband's paternal and maternal grandfather. The

[27] For the full sample of N = 11,378 observations from seven clans, the fraction of single wives is similar, at 74%.

genealogies report up to about 30 descriptors that summarize each man's highest lifetime achievements that are relevant for his status; they are presented in Table A.6. One important aspect is whether the man ever participated in China's civil service examination, which gives elevated status. Men who would pass the civil service examination at different levels (local, provincial, and national) would be eligible for public office. Preparing and passing the civil service examination had a high return (Chang 1962, Ho 1967). The sample includes also wealthy mean who are not also *jinshi* and top-level officials, but their number is relatively small. This is because status and income through official position was so high that even members of the wealthiest sought to acquire the human capital to pass the civil service exam. Passing the civil service examinations has been called "the ultimate source of power" (Ho 1967).

If the genealogy does not mention anything beyond a man's vital statistics, wife, and children, this person had no evidence of wealth, degrees, or titles and he would be allocated to the lowest status group, level 1 in Table A.6. Among all men in the genealogies of the seven male Tongcheng clans, about 70% have no particular descriptors; they would have lived close to subsistence, and one can think of them as commoners. In the present analysis, we combine this group of commoners with several groups that had relatively low status, including men that held honorary titles, official students (who would be supported to prepare for the civil service exam), and men who had close relatives with some level of status. For robustness, we aggregate men into four distinct status groups (see Shiue 2019 for analysis with more than 20 different status levels).

Status level 2 includes men whose genealogy entry indicates substantial levels of wealth, including from certain occupations such as being a merchant. This status level also includes higher-level official students as well as men who held military offices at local, provincial, or the national level. Men with the highest level were those who passed the civil service examination at the provincial and national level (*juren* and *jinshi*, respectively, in levels 3 and 4). A juren working in the civil administration of China would typically be in the top 1.5 percent of the status distribution (status level 3), while a civil jinshi would belong to the top 0.2% of the status distribution.[28]

Status in society translated into income and wealth. There is no systematic information on the income or wealth of individual men, but having passed a certain level of degree made a person eligible for a certain level of official position. For example, there were nine levels of civil positions during the late 19th century (Chang 1962, Table 1). A district magistrate would be seventh-ranked civil official, while a provincial governor would be a second-level civil official. The mapping between degree and official position was not deterministic, however, the level of office was increasing in the degree that a man had obtained. Becoming a top-level official in the imperial bureaucracy with only

[28]The levels of China's civil service exam are building up on each other, that is, in order to have the *jinshi* degree one must have the *juren* and the *shengyuan* (graduate of the local exam), and in order to be *juren* one must have passed the *sheng-yuan* examination. 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 *jinshi* degree would be in the low twenties, and it was not unheard of for a man in his fifties to still be a *shengyuan*. Not all *shengyuan* 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).

Table A.6: Social Status Descriptors

Status	List of Descriptors
1	No title, degree, and evidence of wealth; honorary or posthumous title; village head; other honors; multiple wives in consecutive marriage; moderate wealth of 1st degree family, incl. minor and expectant official, lower level degree *(shengyuan*, *jiansheng*), and official student; Wealthy family member 1st and 2nd degree, incl. official, *juren*, *gongsheng*, and *jinshi*
2	Educated, scholar, no degrees or office; editor of genealogy; refused office, or prepared but did not pass exam; two or more wives or concubines at the same time; substantial evidence of wealth and property; set up lineage estates, large donations, philantrophy; wealthy farmer, landowner, or merchant; official provincial student, student of Imperial Academy; military *shengyuan*, military office, military juren, jinshi; purchased *jiansheng* and/or purchased office civil *shengyuan*; minor civil office; expectant official/civil official with no degrees, minor degree, or purchased degree
3	*juren*, *gongsheng*, with no/expectant office
4	*jinshi* (no office, expectant office, provincial or Imperial office)

Notes: Descriptors of four-level status ranking; based on Telford (1986, 1992), Chang (1955, 1962), Ho (1967), and Eberhard (1962); the sample distribution across status levels is provided in Table 1.

a *shengyuan* degree was almost impossible, and conversely, most *jinshi* had better-paid positions than being a district magistrate. The level of degree is useful because there is a relatively small number of them, and they are consistently mentioned in the data.

B Empirical Results

B.1 Intergenerational Mobility and Marital Sorting: Results with Clan Fixed Effects

Table A.7 shows results on the role of marital sorting for intergenerational mobility when clan fixed effects are included. Comparing Tables 5 and A.7, the patterns of the results are similar whether or not clan fixed effects are included.

The share of in-law father overall is 40% based on Panel A, column (3), which is the same as without clan fixed effects (see Table 5). Accounting for father-in-law status lowers the intergenerational mobility by 26% (=(13-10.3)/10.3) relative to only considering the biological father (18% without clan fixed effects). For specific cohorts, the extent of this can be considerably larger; in particular, for the first cohort, the sum of β_1^b and β_2^l estimates is more than 75% higher than the β_1^b estimate from the corresponding specification with biological father status alone ((0.138+(-0.009)-0.073)/0.073 = 0.76).

Table A.7: Female Intergenerational Mobility and Marital Sorting: Accounting for Clan Differences

	(1)	(2)	(3)	(4)	(5)	(6)
	Daughter Status			Son Status		
Panel A: Intergenerational Correlation						
Father Status	0.103**		0.079**	0.361**		0.287**
	(0.017)		(0.019)	(0.015)		(0.011)
Father-in-Law Status		0.078**	0.051**		0.355**	0.216**
		(0.013)	(0.014)		(0.019)	(0.020)
Panel B: Intergenerational Correlation Over Time						
Father Status <1700	0.073+		−0.009	0.393**		0.336**
	(0.039)		(0.049)	(0.034)		(0.043)
Father Status 18th c	0.126**		0.109**	0.378**		0.305**
	(0.023)		(0.025)	(0.020)		(0.022)
Father Status 19th c	0.083**		0.075*	0.326**		0.244**
	(0.029)		(0.031)	(0.025)		(0.028)
Father-in-Law Status <1700		0.124**	0.138**		0.343**	0.160**
		(0.033)	(0.041)		(0.041)	(0.050)
Father-in-Law Status 18th c		0.086**	0.049*		0.363**	0.214**
		(0.018)	(0.019)		(0.025)	(0.027)
Father-in-Law Status 19th c		0.051*	0.026		0.350**	0.240**
		(0.023)	(0.024)		(0.031)	(0.035)
N	6,101	6,101	6,101	6,101	6,101	6,101

Notes: Dependent variable on top of colum. Estimation of equation (2) and equation (4). Birth year fixed effects included. Status variables employ percentile ranks; robust standard errors reported in parentheses; **/*/+ indicates significance at 1%/5%/10% level.