ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Hu, Sijie

Working Paper

Celebrating legacy: The intergenerational transmission of reproduction and human capital in Ming-Qing Chinese families

GLO Discussion Paper, No. 1572

Provided in Cooperation with:
Global Labor Organization (GLO)

Suggested Citation: Hu, Sijie (2025) : Celebrating legacy: The intergenerational transmission of reproduction and human capital in Ming-Qing Chinese families, GLO Discussion Paper, No. 1572, Global Labor Organization (GLO), Essen

This Version is available at:
https://hdl.handle.net/10419/312862

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Celebrating legacy: The intergenerational transmission of reproduction and human capital in Ming–Qing Chinese families[*]

Sijie Hu[1,2]

[1] School of Economics, Renmin University of China

[2] Global Labor Organization (GLO)

Abstract

In unified growth models, the evolving nexus between population dynamics and technological change is key to achieving sustained economic growth. This paper uses genealogical records of 23,449 males and their spouses to investigate this interplay—the intergenerational transmission of reproduction and human capital—within six Chinese lineages from 1300 to 1920. Examining the relationship between reproduction and long-run reproductive success, the empirical results reveal an optimal level of reproduction, demonstrating a strong Darwinian trade-off: high reproduction in each generation did not consistently lead to long-term reproductive success. Further analysis of the mechanisms is consistent with a Beckerian trade-off, highlighting the potential costs of excessive reproduction through contrasting outcomes in sons' quality: having more brothers exhibited little apparent impact on marriageability but may have been associated with lower human capital. Together, these findings contribute to a deeper understanding of micro-demographic dynamics in pre-modern China and the persistence of Malthusian constraints.

Keywords Reproduction, Long-run reproductive success, Child quantity-quality trade-off, Ming–Qing China

JEL Classification I25, J13, N35**,** O15

[*] I am grateful to Shuji Cao and Qin Jiang at Shanghai Jiaotong University for kindly sharing the original digitalized genealogies of the Zhou, Que, and Huang lineages with me and Familyserach.org for providing researchers with digital images of the genealogical records of the Gu, Zha, and Zhuang lineages. I acknowledge Rongyu Jian and Xinyu Liu for their research assistance.

I am incredibly grateful to Neil Cummins, Debin Ma, Eric Schneider, James Kai-sing Kung, Noam Yuchtman, Oded Galor, and Carol Shiue. I also wish to thank Gregory Clark, Jan Kok, Zhiwu Chen, Nan Li, Qin Jiang, Zhan Lin, Yu Hao, Ming Lei, Qun Che, and Qing Wang for their extensive feedback and also participants in the LSE Asia Economic History Seminar, LSE Graduate Economic History Seminar, EHS Annual Conference 2021, RES Annual Conference 2021, European Social Science History Conference 2021, Eighth International Symposium on Quantitative History, 2022 World Economic History Congress, Economic History Workshop at Peking University, and the Quantitative History Webinar Series for their comments. I acknowledge the financial support of the National Natural Science Foundation of China (No.72203224). Any errors are my own.

1 Introduction

To explain the long-run economic growth, the "population-idea nexus" is always of the essence (Doepke, 2004; Mokyr & Voth, 2010). Technology, a key element of economic development, does not "just happen" but depends on the population, particularly the accumulation of human capital (Kremer, 1993). In unified growth theories that seek to explain both the Malthusian stagnation and economic take-off, the relationship between reproduction and technology in the long run is carefully traced (Galor & Weil, 2000; Galor & Moav, 2002; Galor, 2011, 2022).

Therefore, two crucial elements come into play: population size and population composition (Galor, 2022). In the era of Malthusian stagnation, higher living standards resulted in larger family sizes, leading to a gradual increase in the proportion of descendants from families with higher living standards within the total population. Over the long run, these two "wheels of change" operated under the influence of natural selection, ultimately increasing the representation of growth-promoting traits in the total population (Galor & Klemp, 2019; Galor, 2022). This dynamic, in turn, promoted investment in human capital and facilitated the demographic transition and transition from stagnation to sustained growth (Galor & Moav, 2002).

Drawing on the framework of Galor (2022) and Galor and Klemp (2019), this paper examines the evolution of the two "wheels of change" in the pre-transitional Chinese context. I use a new genealogical dataset containing 23,449 men and their spouses to exploit multigenerational associations in reproduction across six lineages in Southeast China from 1300 to 1920, a period primarily covers the last two imperial dynasties, the Ming (1368–1644) and Qing (1644–1911). I intend to explore Chinese men's reproductive success within a multigenerational model by examining the patterns and mechanisms of fertility transmission. I focus on two trade-offs: a Darwinian trade-off between reproduction and long-run reproductive success and a Beckerian trade-off between the quantity and quality of children. I aim to show

how reproduction affects the long-run reproductive success of a patriline through the accumulation of human capital.[1]

To begin, I test for a Darwinian trade-off by examining the optimal level of reproduction for long-run reproductive success in the six lineages. I estimate the relationship between the number of sons and the number of patrilineal male descendants in the subsequent three generations to test whether high reproduction in the first generation could translate into high reproduction over the next three. The results of the negative binomial regression indicate a trade-off between reproduction and long-run reproductive success. After controlling for all variables, each additional son, up to six, increased the number of grandsons, great-grandsons, and great-great-grandsons a man could expect. However, having more than six sons reduced the number of patrilineal male descendants over the following three generations.

Next, I analyze the possible mechanisms through which parental reproduction might affect next-generation reproduction. The initial positive but later negative effect of reproduction on long-run reproductive success suggests a potential negative relationship between child quantity and quality, or in other words, a Beckerian trade-off. Family size in the first generation could affect the quality of the second generation, which, in turn, would affect the likelihood of the second generation producing male descendants in subsequent generations.

To test for Beckerian trade-offs between child quantity and quality, I conduct two logistic regressions to assess whether larger family sizes negatively impacted children's quality. Given the patriarchal structure of the genealogical records, this analysis focuses on sons and investigates the relationship between son quantity and quality. For son quantity, I use the number of sons a father had, that is, the number of brothers a man had (including himself), to measure the final family size. For the dependent variables in the logistic models, I employ two

[1] Given that only patrilineal male descents are officially recorded in the genealogical books, this paper focuses on all the male descents within the six lineages. Moreover, in pre-modern China, a typical patriarchal society, only patrilineal male descendants mattered (Freedman, 1966; Harrell, 1985; Song et al., 2015).

dummy variables of son quality. The first is *Married*, which measures whether a man married at least once before his death, as marriage was a prerequisite for leaving male descendants in traditional China. The second indicator, *Degree*, captures a son's human capital related to education, measured by whether he earned an academic degree through the *keju* (national civil examinations)—a widely recognized indicator of formal education and human capital in the Ming–Qing Chinese context (Shiue, 2017). Of the men in the sample, 1,592 men (6.79 percent) obtained *keju* degrees.

The two logistic regressions illustrate that the quantity of sons affected different aspects of son quality. There was a positive correlation between the number of brothers (including himself) a man had and his marriageability, but a negative correlation between this and his likelihood of obtaining a *keju* degree. While sons from larger families were more likely to marry, they did not benefit similarly in academic achievement. After controlling for fathers' and grandfathers' social status, the number of brothers was negatively correlated with educational attainment, though this relationship was statistically insignificant.

Because of potential endogeneity issues from unobservable parental preferences and household characteristics affecting both son quantity and quality, establishing causality remains challenging, and correlations may be biased. Galor and Klemp (2019) address this by using the "protogenesic interval" (PI), the time between marriage and the birth of the first child, as an instrument for fertility. Unfortunately, Chinese genealogical records lack marriage dates and birthdates for the first child, so I adapt this approach by using the mother's age at the birth of her first son as an instrument for the number of brothers a man had.

The instrument is conditionally exogenous. Ming–Qing China maintained a universal and early marriage pattern for females, with the mean age at first marriage consistently between 16 and 19 years (Lee & Wang, 1999; Harrell & Pullum, 1995). Furthermore, Ming–Qing China operated mainly under a "natural fertility" regime with minimal, if any, birth control (Wolf,

1984, 2001). Given this early and nearly fixed marriage age, along with the lack of conscious control over birth timing, the mother's age at the birth of her first son would largely be random and minimally influenced by her husband's socio-economic characteristics. Thus, younger maternal age at first son's birth would extend the mother's reproductive period, with the onset of son-bearing being conditionally exogenous after accounting for cultural and socio-economic factors, independent of the father's preferences regarding son quality. Additionally, this instrument could not directly affect a son's likelihood of marriage or academic achievement.

The instrumental variable results, broadly consistent with the baseline findings, point to the presence of a Beckerian trade-off when measuring son quality as education-related human capital rather than marriageability. The number of brothers had nearly no impact on marriage likelihood but appeared to have reduced the likelihood of obtaining a *keju* degree. Overall, the logistic and IV results are indicative of a Beckerian trade-off, measured by educational attainment, in the six lineages during the Ming–Qing period.

This paper contributes to the existing literature in two ways. On the one hand, it provides some of the earliest quantitative evidence on reproductive success within a multigenerational model for pre-modern China. Biologists studying the trade-off between fertility and various biological traits find that high fertility does not always lead to high survival rates (Lack, 1954; Williams, 1966; Stearns, 1989). Social scientists have then examined different types of Darwinian trade-offs within the human species and suggested different relationships between female fertility and offspring survivorship, parental fertility and next-generation fertility, and the optimal level of fertility and long-run reproductive success in different societies (Kaplan et al., 1995; Hill & Hurtado, 1996; Kaplan, 1996; Borgerhoff Mulder, 2000; Strassmann & Gillespie, 2002).

Specifically, Galor and Klemp (2019) investigate natural selection forces in pre-demographic transition Quebec using genealogical data from half a million residents over four

generations. They identify a hump-shaped relationship between fecundity and long-run reproductive success, with the optimal level of reproduction below the population median. In contrast, the micro-demographic dynamics over multiple generations in historical China are only partially understood. In particular, questions persist regarding Chinese families' strategies to ensure lineage survival over time and the optimal level of reproduction in that context. To secure greater continuity in bloodlines, reproductive success must not only be achieved in a single generation but also transmitted across generations. Song et al. (2015), drawing on data from the Qing dynasty in North China, find that patrilineages with high social origins increased their representation in the overall population not by maximizing male descendants in each generation but by reducing extinction risk.

When comparing optimal fertility levels, the six-son threshold in the six Chinese lineages exceeds the sample's median and mean of two sons, revealing a distinct fertility pattern in Southeast China compared to Quebec, where the optimal fertility was below the median. At the same time, this paper finds a similar pattern across Southeast and North China: while ancestors of high social status did retain productive advantages, the most successful patrilineal lines were not those with the highest fertility in each generation.

On the other hand, this paper extends the empirical literature on the trade-off between the quantity and quality of children in pre-transition societies. Economists have studied this trade-off since Becker (1960; Becker & Lewis, 1973; Becker, Murphy, & Tamura, 1990) first introduced fertility decisions into economic analysis, proposing that parents often sacrifice the number of children they could have for higher quality in the children they had. A considerable amount of empirical literature also tries to support or challenge this Beckerian argument in both historical and modern times (see the examples of Northern Europe: Baudin & De la Croix, 2023; of India: Rosenzweig & Wolpin, 1980; of Thailand: Knodel, Havanon, & Sittitrai, 1990; of Norway: Black, Devereux & Salvanes, 2005; of Brazil: Ponczek & Souza, 2012; of the USA:

Tan, 2019; of Korea: Lee & Park, 2019; of England: Clark & Cummins, 2019; Klemp & Weisdorf, 2019; of contemporary China: Qian, 2005; Rosenzweig & Zhang, 2006; Li, Zhang, & Zhu, 2008; Liu, 2014).[2]

In historical Chinese contexts, the relationship between child quantity and quality remains less conclusive. Shiue (2017) uses genealogical data from Tongcheng County and identifies a negative relationship between sons' quantity and educational attainment during the seventeenth and eighteenth centuries, which disappeared afterward. Meanwhile, Bai et al. (2023) observe a comparable trade-off in North China, but only after 1800. In addition, Song et al. (2015) speculate that high-status founders in the Qing period may have strategically leveraged this trade-off to preserve their lineage.

Alternatively, this paper shifts the geographic focus to the core lower Yangtze region, historically the most developed area in China and central to recent economic history debates (Shiue, 2016). Examining the relationships between reproduction and both marriage and education in a multigenerational model, this paper finds that family size affected child quality to varying extents, an aspect not fully explored in previous literature. Results on educational attainment also align with those of Shiue (2017), indicating a more pronounced Beckerian trade-off before 1800, contrasting with the patterns observed in Northeastern China by Bai et al. (2023).

In addition, reproduction, human capital formation, and long-run reproductive success were interconnected in pre-transitional societies. One primary mechanism through which parental reproduction affected next-generation reproduction was balancing reproduction and investment in child quality. Galor and Klemp (2019) find that, in pre-1800 Quebec, moderate fecundity coupled with higher education levels better ensured lineage continuity and

[2] The evidence from contemporary China is conflicting. Qian (2005) and Li et al. (2008) both exploit the 1990 population census but find a contradictory relationship between family size and children's educational attainment. Liu (2014) finds a strong negative relationship between family size and children's height.

contributed to a secular fertility decline; larger family sizes, conversely, negatively affected survival rates, suggesting a child quantity-quality trade-off. However, limited consensus exists on whether nineteenth-century China's lack of a fertility transition was due to the absence of these trade-offs or other factors. By linking the Darwinian and Beckerian trade-offs, this paper seeks to provide a more comprehensive understanding of micro-demographic dynamics in late imperial China and offers insights into the absence of a fertility transition.

The rest of the paper is structured as follows: Section 2 describes the genealogical data, and Section 3 introduces the empirical strategies. Section 4 reports the results for long-run reproductive success and the child quantity-quality trade-off. Section 5 discusses the results, and Section 6 concludes.

2 Data and main variables

2.1 The lineage sample

In Ming–Qing China, lineages were the most widespread and long-lasting forms of social organization (Feng & Chang, 2001; Zelin, 2009). In a society that attaches great importance to maintaining and expanding the patrilines, keeping genealogical records has become a standard practice for most lineages to remind the offspring of their family history (Zhao, 2001; Feng, 2009). The primary data used in this paper come from the genealogical books of six lineages in Southeast China; the sample includes 23,449 men and their spouses, primarily born between 1300 and 1920 (see Figure 1).[3] Of the six lineages, three—the Huangs, the Ques, and the Zhous—are classified as "common" lineages, while the other three—the Gus, the Zhas, and the Zhuangs—are classified as "elite", distinguished by the absolute number and proportion of

[3] The genealogical records for the six lineages include 36,456 men. To ensure the accuracy of family relationships across multiple generations (fathers, grandfathers, sons, grandsons, great-grandsons, and great-great-grandsons), this paper uses a subset of 23,449 men whose birth years were explicitly recorded. Recording birth years can serve as a proxy for the completeness and reliability of the records, particularly in documenting an individual's reproduction outcomes (Hu, 2023a).

gentry scholars within the male population of each lineage (Hu, 2023a). Online Appendix Figure A1 shows the two provinces and four prefectures where the six lineages are located.

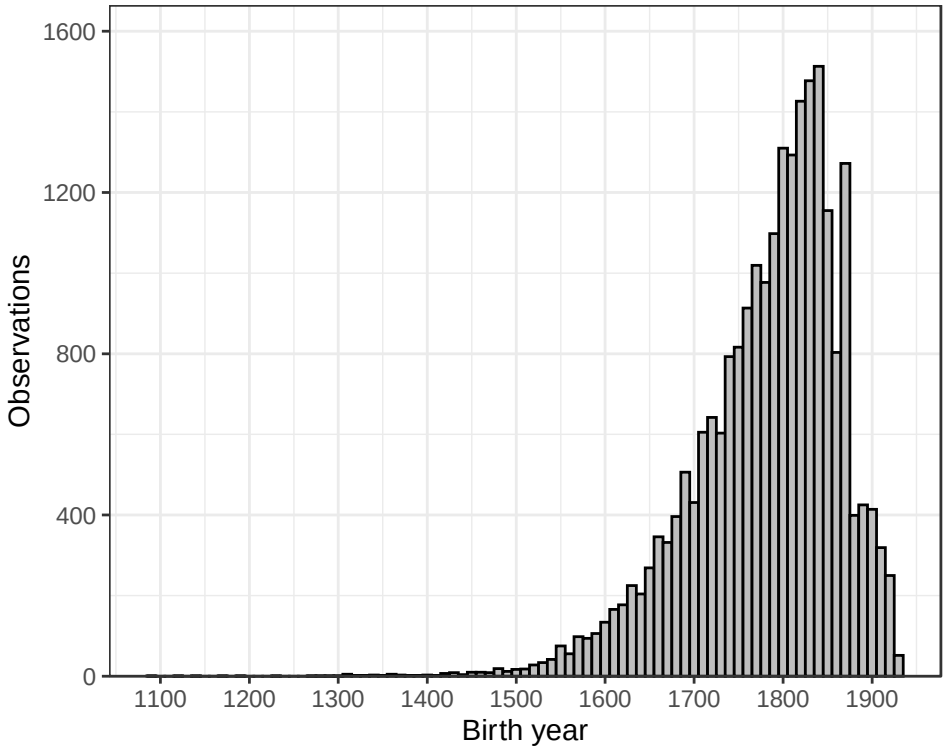

Figure 1 Distribution of men's birth years

Source: The lineage sample.

Genealogies of a lineage always include an introduction to the history of the family, the rules of compilation, the rules and regulations that family members had to follow, a family tree that includes all the male members recorded in the book, and finally a series of detailed entries for each male descendant in the family (see Figure A2 for an example of a male individual's entry in the book). Since lineage size was the primary indicator of a lineage's prosperity, measured mainly by the number of surviving male descendants, all sons who survived infancy were documented in the genealogical books.[4] Each male member's entry includes the number

[4] Campbell and Lee (2002) and Harrell (1987) compare genealogical records with population registers in Liaoning and Taiwan during the mid and late Qing periods, respectively, finding that some male descendants listed in the

of sons who survived infancy as well as their names, allowing for easy linkage of male family members across generations.

Although genealogical data are highly valuable for studying reproduction and survival, they are not free from selection bias. Hu (2023a) has a detailed discussion of selection biases in this six-lineage sample. One of the main biases affecting the empirical estimation in this paper is the lack of daughters' information. Due to a strong preference for sons in imperial China, daughters are highly underreported. The sample contains records for 27,919 sons but only 8,607 daughters, with information on half of the daughters limited to a numerical count under their father's entry, without further detail. Given that the number of sons who survived infancy is fully recorded for each man's entry, I use this number in the present paper to measure reproduction and family size. Moreover, because of the patrilineal structure of Chinese families, the term "descendants" in this paper refers solely to patrilineal male descendants who inherited the lineage surname.[5]

Another potential bias in the genealogical data concerns the lack of information regarding family members who migrated away from the lineage's original location. Genealogy compilers often face difficulties tracing and documenting information about these individuals, particularly across subsequent generations after their relocation. In such cases, genealogies typically annotate an individual's entry with remarks such as "moved out" or, when their information could no longer be traced, "information not traceable". For this analysis, I exclude migrated individuals with incomplete information. For those who moved out of the original

population registers were absent from the genealogical books. However, as Harrell (1987) also notes, these "missing" was intentional, as these male descendants were typically those who died young and were excluded from the genealogies according to the compilation rules for Taiwan's genealogical records. Since Southeastern China lacks comparable population registers, a similar type of verification is not possible. However, it is widely accepted that genealogical books from Southeastern China are of higher quality than those from Taiwan, with a more complete account of male births (Liu, 1992).

[5] As a robustness check, I also explore a subsample of daughters in Section 4.2.2.

town while retaining complete genealogical information, I control for their migration behavior, which will be detailed in Section 3.

2.2 Dependent and independent variables

In the main analysis of the optimal reproduction levels for long-run reproductive success, the dependent variable is the recorded number of patrilineal grandsons, great-grandsons, and great-great-grandsons. The subsample includes 12,734 potential grandfathers (those with at least one son), 7,366 potential great-grandfathers (those with at least one grandson), and 4,398 potential great-great-grandfathers (those with at least one great-grandson). Men with incomplete records of male descendants in the subsequent generations are excluded. The independent variable is the number of sons who survived infancy a male had.

In the analysis examining the child quantity-quality relationship, which is the mechanism through which parental reproduction affects next-generation reproduction, the dependent variables are two dummy variables that measure two quality indicators in males: their marital status (*Married*) and their social status (*Degree*). The independent variable is the number of sons in a family who survived infancy, including the male individual himself and his brothers. In other words, it refers to the total number of recorded male children, both biological and adoptive, listed under the father's entry. This variable represents the final family size in terms of male offspring.

Since long-run reproductive success in this paper is measured by the number of patrilineal male descendants a man could have, the two outcomes of interest reflect a man's quality regarding his ability to produce male heirs. These quality measures are closely linked to the level of parental investment in the individual.

First, marriage (*Married*) is considered a valid measure because it is a prerequisite for having male descendants. The marriage market in Ming–Qing China was severely imbalanced

due to factors such as female infanticide and polygamy, leading to a shortage of women (Lee & Wang, 1999). Marriage thus became a "sensitive measure of privilege" as access to marriage was contingent on having access to resources (Lee & Campbell, 1997; Lee & Wang, 1999). A substantial proportion of men, particularly those from impoverished or low-status families, were unable to marry.[6] Thus, the "quality" of the male individual determined whether he could marry to a great extent.

Second, academic degrees (*Degree*) are a widely recognized measure of human capital in pre-modern China (Shiue, 2017). Academic degrees were crucial because they determined social status, which, in turn, significantly influenced a man's ability to have male offspring (Hu, 2023a). To quantify this, I categorize all men in the sample into 16 sublevels of social status, constructing the variable *Status*. These sublevels are then grouped into seven broader levels to form the variable *Grade*. Individuals within *Status* sublevel 6 onwards, except for sublevels 9 and 11, who held academic degrees are coded as 1 for *Degree* (see Table 1).

Social status in the Ming–Qing period was closely linked to wealth and education, and acquiring academic degrees was one of the key channels through which individuals could enhance their social standing. On the one hand, both formal and informal education enabled individuals to rise to higher social status and considerable wealth through the *keju*, the national civil examination system.[7] Men from various backgrounds, including peasants, artisans, and merchants, were eligible to take the exams. Academic degrees were awarded at three levels — *shengyuan*, *juren*, and *jinshi*—to those who passed the corresponding county-level, provincial-level, and national-level exams (Miyazaki, 1981; Chen, Kung, & Ma, 2020). Candidates were allowed to retake the exams multiple times if they failed, but even after succeeding, not all

[6] In the sample, 8.08 per cent men failed to survive to adulthood, and 23.5 per cent of men who survived to adulthood were unmarried throughout the lifetime.

[7] *Keju* was initiated in 600 AD and abolished in 1905. Its presence changed imperial China into a meritocracy. In the tenth century, Emperor Zhenzong of the Song dynasty (960–1276) once wrote "There is no need to buy farmland, for books will get you a position with a high salary;/There is no need to build a house, for books will bring you a luxurious residence with golden walls."

degree holders secured government positions, particularly those with lower-level degrees (Chang, 1962). On the other hand, wealth could also lead to higher social status by allowing individuals to purchase *keju* degrees and attain official positions (Zhang, 2013). In either case, holding a *keju* degree is a reliable measure of human capital in the Ming–Qing context.

For individuals who were literate but did not obtain a *keju* degree, the *hao* (pen name) is a marker of literacy, distinguishing them from the general population. In the 16-level classification, individuals at level 2 held a *hao* but had no other social status records. In traditional China, a *hao* was typically a pen name adopted by literate individuals to express personal virtues and sentiments, often appearing in their prose or poetry. Although the cost of acquiring basic literacy in Ming–Qing China was relatively modest, literacy remained a privilege (Shiue, 2017). Therefore, the possession of a *hao* itself signifies basic literacy training. *Hao* not only functions as a non-*keju*-related indicator of social status but may also suggest a certain level of household wealth, as it reflects the family's ability to afford literacy training for its children.

Table 1 Degree and social status classification

Status	Grade	Degree	Count	Percent	Description
1	1	0	20,471	87.30%	No official status
2	1	0	1,063	4.53%	No official status, pen name holder
3	2	0	63	0.27%	Lineage chief; donor to the lineage and the county
4	2	0	115	0.49%	Literate but with no academic degree (teacher of the village or editor of genealogical books)
5	2	0	45	0.19%	Awarded honorary titles by the emperor but with no academic degree
6	3	1	497	2.12%	Lower degree holder (normal *shengyuan* and civil *shengyuan*)
7	4	1	550	2.35%	Students at the Imperial Academy (lower degree)
8	5	1	39	0.17%	Intermediate/high degree holder (*juren*, *gongsheng*, *jinshi*), but with no official position
9	5	0	53	0.23%	Prospective officials (*houbu*), with no academic degree
10	5	1	70	0.30%	Prospective officials (*houbu*), with academic degree
11	5	0	47	0.20%	Clerks (*wei'ruliu*) or lowest-ranking official (*zong jiupin*), without academic degree
12	5	1	20	0.09%	Clerks (*wei'ruliu*) or lowest-ranking official (*zong jiupin*), with academic degree
13	6	1	100	0.43%	Low-/medium-ranking local official and low-ranking court official, with normal and civil *shengyuan* degree
14	6	1	153	0.65%	Low-/medium-ranking local official and low-ranking court official, with a degree of studentship at the Imperial Academy
15	6	1	79	0.34%	Low-/medium-ranking local official and low-ranking court official, with an intermediate/ high degree
16	7	1	84	0.36%	High-ranking local official and medium/high-ranking court official

Source: Ho 1962, Chapter 1; Telford 1995, Appendix 3A; Shiue 2017, Table 1.

3 Empirical strategy

3.1 The Darwinian trade-off: The relationship between reproduction and long-run reproductive success

3.1.1 The model

To test the optimal level of reproduction for long-run reproductive success, I apply a multigenerational model, adhering to the approach of Kaplan et al. (1995) and Galor and Klemp (2019). I first test whether the relationships between the number of sons and the log-transformed number of patrilineal male descendants in the next three generations are monotonic. I run a non-parametric analysis by using the LOWESS method. As Figure 2 presents, the relationships are not strictly linear. The man in the sample with the most sons and grandsons is Zhuang Chaosheng from the Zhuang lineage. He married 10 times and had 12 sons and 50 patrilineal grandsons. However, he does not hold the record for the most patrilineal great-grandsons and great-great-grandsons. His father, Zhuang Yinghui, had five sons and had the most patrilineal great-grandsons, 88. Que Qixin from the Que lineage had only four sons, but his male descendants produced 194 patrilineal great-great-grandsons for him, the highest number in the sample.

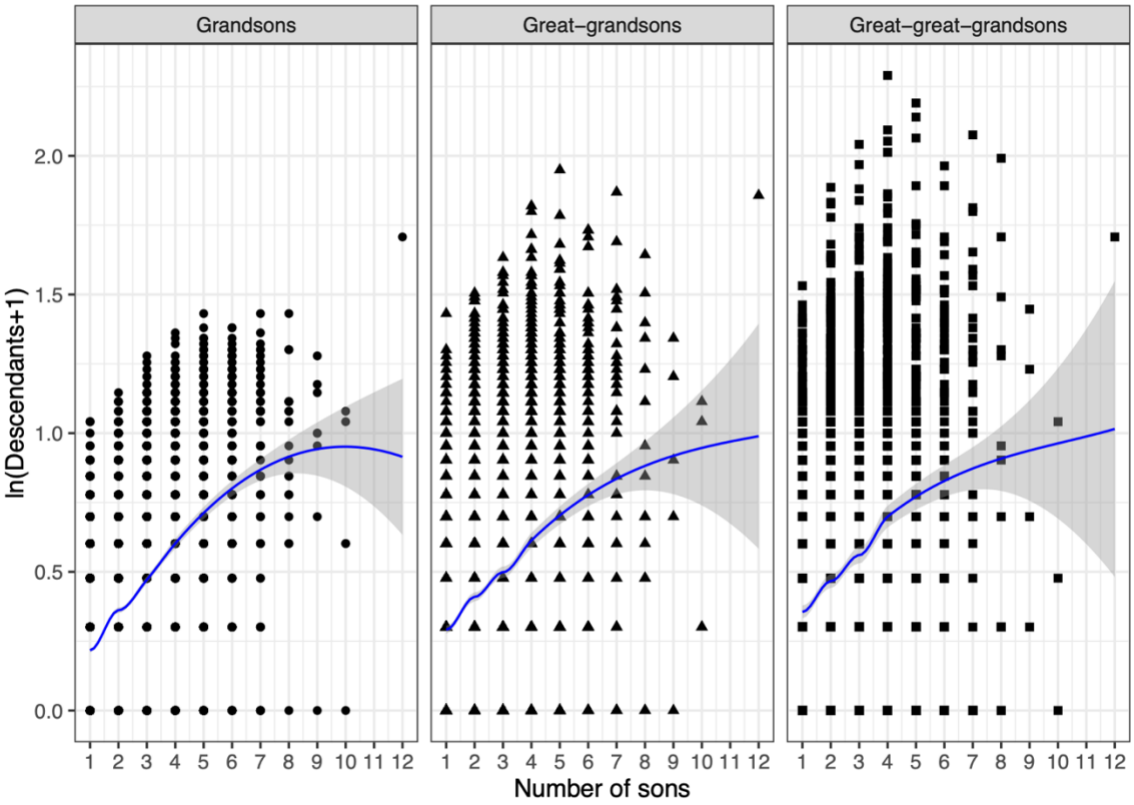

Figure 2 The qualitative pattern between reproduction and long-run reproductive success, the LOWESS method.

Notes: 1. The outcome of interest used in the analysis is the log-transformed number of male descendants in the following three generations. I also add 1 to the origin number to keep all the zero observations. 2. The shaded area represents the 95% confidence interval for the LOWESS smooth curve.

Because the number of grandsons (mean = 2.05, variance = 8.03), the number of great-grandsons (mean = 3.29, variance = 30.54), and the number of great-great-grandsons (mean = 5.18, variance = 110.13) are all count variables and are over-dispersed, I employ negative binomial regression to test the relationship between the number of sons and the number of male descendants in the following three generations based on the following equation:

$$Descendants_i = \alpha + \beta_1 Sons_i + \beta_2 Sons_i^2 + \delta P_i + \varepsilon_i, \quad (1)$$

where *Descendants* refers to the number of patrilineal male offspring that a male individual (Generation 1) had over the three generations following his sons' generation (Generations 3–

5, i.e., grandsons, great-grandsons, and great-great-grandsons); *i* denotes male individuals, specifically potential grandfathers, great-grandfathers, and great-great-grandfathers. *Sons* is the number of sons that the male individual had. Given the non-monotonic relationships shown in Figure 2, I also control for the squared term of *Sons*.

P is a set of control variables that would also affect the number of male descendants in Generations 3–5. ε is the error term. If an optimal level of reproduction did exist, men with a certain number of sons would be expected to have the highest number of grandsons, great-grandsons, and great-great-grandsons. Exceeding this optimal number of sons could result in fewer male descendants in subsequent generations, suggesting a negative β_2 would be expected.

3.1.2 The control variables

I condition on a set of control variables in the model to establish the causal relationship between reproduction and long-run reproductive success.

Grade and *Marriages.* It is widely accepted that social status and wealth significantly affected reproduction in pre-modern settings (Clark & Hamilton, 2006). The number of marriages that a man had was also positively associated with his social status and wealth. Thus, both *Grade* and *Marriages* are controlled for the effects of the male's socio-economic characteristics on his long-run reproductive success. *Grade* refers to the 7-level categorical variable, as outlined in Table 1. *Marriages* denotes the number of times that the man married during his life.

Firstborn. In traditional Chinese culture, the firstborn son enjoyed a "firstborn premium" or "inheritance advantage", as he bore the primary responsibility for continuing the patrilineage (Lee & Campbell, 1997; Li & Zhen, 2015). This role was expected to correlate positively with long-run reproductive success.

Adoptee. In a society that highly values filial piety, having a male heir to continue the bloodline is a priority for every man. However, not every man could fulfill this task, leading to the widespread adoption practice.[8] As an adoptee, usually also the sole heir in the new adoptive family, he would face additional pressure to excel and perform well (Waltner, 1990; Wolf & Huang, 1980). Therefore, I control for whether the man is an adoptee or not.[9]

Mother's marital status. In Ming–Qing China, the marriage system was a hybrid between monogamy and polygamy: men were legally permitted only one wife at a time, yet could concurrently have multiple concubines. After the death of a principal wife, a man could marry secondary wives. However, secondary wives and concubines held a lower social status than principal wives; consequently, sons born to secondary wives or concubines often faced unequal access to resources and privileges within the lineage. Hence, the mother's marital status is included to condition this potential adverse effect. It is coded as one if the man's mother was a secondary wife or a concubine.

Survival to adulthood. Lifespan is closely related to reproduction and human capital formation. About half of the males in the genealogical sample had complete vital records, but males who died before adulthood were also marked in the genealogies. I use the dummy *Survival* to distinguish males who failed to reach adulthood from the other males to control for its negative effects on reproduction.[10]

Migration. Migration may affect an individual's reproductive success. When moving to a new county, a settler might adjust reproductive behavior to adapt to their new environment. I include *Migration* as a control to account for its potential effect on the number of male descendants across generations.

[8] Typically, a man without a male heir would adopt his nephew from his brothers and male cousins (Lee & Wang, 1999; Waltner, 1990; Wolf & Huang, 1980).
[9] Of the 23,449 men in the sample, 1,731 were adoptees.
[10] I condition on lifespan as a proxy for health-related factors as robustness checks in Sections 4.1.2 and 4.2.2.

Birth cohort and *Lineage*. The birth cohort and lineage to which a man belonged would affect his reproduction and long-term reproductive success. I classify all 23,449 men in the sample into twelve birth cohorts, beginning with a "pre-1400" interval (1300–1400) and ending with a "post-1900" interval (1900–1920), with ten half-century cohorts in between. *Lineage* refers to the ancestral lineage to which each man belonged. These two fixed effects are included to account for time-variant and time-invariant factors, thus reducing omitted variable bias.

3.2 The Beckerian trade-off: mechanisms through which reproduction affected long-run reproductive success

3.2.1 Baseline model

I first run probit regressions based on the following equation:

$$P(Quality_i = 1) = \Phi(\alpha + \beta Brothers_i + \delta Z_i + \gamma W_i + \varepsilon_i), \qquad (2)$$

where *Quality* denotes the male individual's quality, and *i* denotes male individuals. Following Galor and Klemp (2019), I employ two indicators to measure quality, *Married* and *Degree*. *Married* equals one if the man was married at least once, and *Degree* equals one if the man was a *keju* degree holder. α is the constant. As the number of daughters is incomplete in the genealogies, I use *Brothers* to measure the quantity of children. It equals the number of brothers (including himself) in a man's generation of his family of origin, in other words, the number of sons who survived infancy that his father had.

Z denotes a set of control variables, most of which are also used in Equation (1), including *Firstborn, Adoptee, Mother's marital status, Survival to adulthood, Migration, Birth cohort,* and *Lineage*. *W* includes a set of additional controls linked to the man's father and grandfather that could affect his marital status and social achievement. ε is the error term. If a trade-off

existed, a negative β would be expected since it represents parents choosing between the quantity and quality of their sons.

The additional control variables in *W* include the man's father's social status and number of marriages, as well as his grandfather's social status, number of marriages, and number of sons.

Father's grade and *Father's marriages*. For each man, I control for his father's social status and number of marriages to reflect the socio-economic characteristics of his family of origin. *Father's grade* uses the same 7-level category in Table 1 to classify the father's social status.

Grandfather's grade, *Grandfather's marriages*, *Grandfather's sons*. Moreover, I account for the grandfather's effects in addressing endogeneity concerns to a certain degree. Establishing a causal relationship between family size and child quality is challenging because of omitted variable bias, as unobserved household characteristics and parental preferences might obscure any trade-off between child quantity and quality. In a male-dominant society that values filial piety, a couple's preferences would be influenced by the decisions of the husband's parents. Furthermore, research on natural populations with minimal social gradients in fertility shows an intergenerational correlation in reproductive fitness (Pluzhnikov et al., 2007). Therefore, to partly address endogeneity, I control for each man's grandfather's social status, number of marriages, and number of sons. Table A1 in the Online Appendix reports the summary statistics of the sample.

3.2.2 Instrumental variable model

As mentioned before, the unobservable parental preference and household features that affected both the quality and quantity of children would bias the effects of family size on child quality. To address the concern of omitted variable bias, besides including the grandfathers'

effects, I also construct an instrumental variable, the mother's age at the birth of her first surviving son. The onset of son-bearing and, consequently, the duration of a mother's reproductive period are directly correlated with family size and are also conditionally exogenous.

To investigate the trade-off between fertility and offspring quality, Galor and Klemp (2019) employ the concept of the "protogenesic interval" (PI), which represents the duration between the date of marriage and the birth of the first child, as a measure of fertility. Since the genealogical records do not include specific marriage dates or the birth dates of the first child but only the birth dates of the first surviving son, I modify this PI measure to another variable, namely, the mother's age at the birth of her first surviving son.

A consensus in terms of historical Chinese marriage pattern is that although Chinese men's age at first marriage varies, and a moderate yet significant proportion of men remained single throughout their lifetime, marriage was early and universal for females, irrelevant of the social standing of their families of origin (Wolf & Huang, 1980; Liu, 1985, 1992; Harrell & Pullum, 1995; Lee & Wang, 1999).[11] In terms of the marital fertility pattern for women, Ming–Qing China primarily adhered to a "natural fertility regime", where deliberate fertility controls—late starting, longer spacing, and early stopping—were largely absent (Wolf, 1984, 2001; Hu, 2023b). Hence, if a woman had a surviving son at an earlier age, this suggests that she could have a longer reproductive span, resulting in more surviving sons throughout her lifetime. As depicted in Figure 3, most mothers had their first surviving sons by their early 20s.

[11] According to Barclay et al. (1976), Liu (1985) and Harrell and Pullum (1995), the mean age at first marriage for women was around 16 to 19 years old. In the records of the *Gazetteer of Wujin*—the county where the Zhuang lineage was located—23 chaste women have documented ages at first marriage, with an average of 16.5 years. Similarly, in the *Gazetteer of Chuzhou*, which covers the prefecture where the Huang, Que, and Zhou lineages were located, 60 chaste women have recorded ages at first marriage, with an average of 17.5 years.

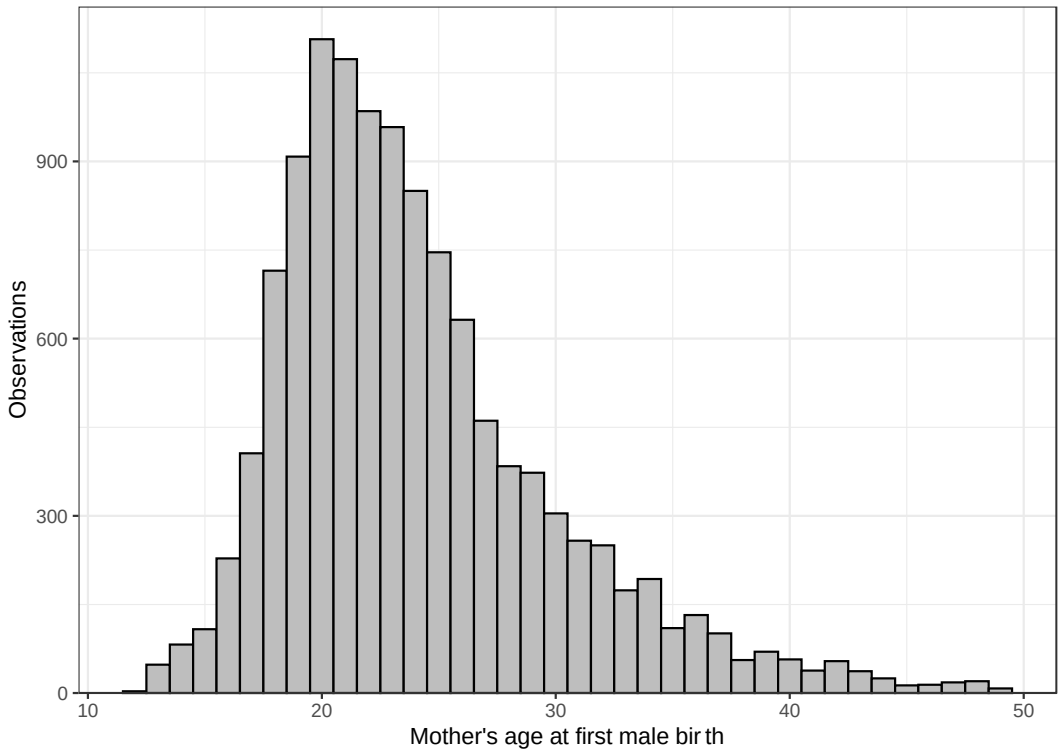

Figure 3 The distribution of mothers' age at first male birth.

Source: The lineage sample.

To be a valid instrument, the timing of a mother's first male birth needs to be conditionally exogenous. As previously mentioned, in Ming–Qing China, women's marriage age was uniformly early across socio-economic backgrounds, suggesting that the timing of first sons' births was largely contingent on random factors rather than marriage timing. However, this timing might not have been entirely random; cultural and socio-economic influences could still have shaped fertility patterns to some extent (Galor & Klemp, 2019). For instance, women marrying into wealthy or higher-status families might experience different fertility patterns. To mitigate potential socio-economic confounding factors affecting child quality, I control for the complete set of family characteristics discussed in the previous two sections, particularly the father's and grandfather's social status and number of marriages, to better isolate the effect of random variations in the mother's age at her first male birth.

Hence, I then instrument the number of brothers that a man had by the conditionally exogenous variation induced by his mother's age at first son's birth. I use an OLS regression and a probit regression to estimate Equations (3) and (4), respectively,

$$Brothers_i = \alpha_2 + \beta_2 MotherAge_i + \delta_2 Z_i + \gamma_2 W_i + \varepsilon_i, \qquad (3)$$

$$P(Quality_i = 1) = \Phi(\alpha_3 + \beta_3 \widehat{Brothers}_i + \delta_3 Z_i + \gamma_3 W_i + \varepsilon_i), \qquad (4)$$

in which the variable *MotherAge* equals the mother's age at the birth of her first surviving son. *Brothers* still denotes the final family size. The other notations are the same as in Equation (2).

4 Results

4.1 The Darwinian trade-off: The relationship between reproduction and long-run reproductive success

4.1.1 Baseline results

Table 2 reports the results of negative binomial regressions based on Equation (1).[12] As expected, the number of descendants in generation 2, *Sons*, had a strong positive effect on the number of patrilineal male descendants in generations 3–5. The coefficient on the square term of *Sons* maintains less than one, demonstrating its negative and statistically significant effect on reproduction in the subsequent three generations in all columns in all three specifications. The fitness for long-run reproductive success could diminish beyond a certain level of reproduction.

[12] All coefficients are presented as incidence rate ratios. A coefficient greater than one indicates a positive effect, while a coefficient less than one indicates a negative effect.

Table 2 The effects of *Sons* on the number of male descendants for males born between 1300 and 1920, negative binomial regression

	Dependent Variable: Number of male descendants in								
	Gen.3	Gen.4	Gen.5	Gen.3	Gen.4	Gen.5	Gen.3	Gen.4	Gen.5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sons	2.088***	1.840***	1.866***	1.885***	1.580***	1.576***	1.903***	1.610***	1.592***
	(0.076)	(0.081)	(0.104)	(0.062)	(0.063)	(0.077)	(0.046)	(0.054)	(0.074)
$Sons^2$	0.956***	0.966***	0.962***	0.962***	0.975***	0.971***	0.957***	0.968***	0.964***
	(0.006)	(0.006)	(0.007)	(0.005)	(0.006)	(0.007)	(0.003)	(0.005)	(0.006)
Grade 2							1.590***	1.386***	1.366***
							(0.095)	(0.121)	(0.173)
Grade 3							1.401***	1.586***	1.451***
							(0.058)	(0.103)	(0.139)
Grade 4							1.541***	1.538***	1.376***
							(0.070)	(0.124)	(0.137)
Grade 5							1.215**	1.193*	1.222
							(0.085)	(0.117)	(0.172)
Grade 6							1.403***	1.388***	1.659***
							(0.070)	(0.108)	(0.192)
Grade 7							1.997***	2.505***	2.444***
							(0.169)	(0.282)	(0.289)
Marriages							1.074***	1.082***	1.099***
							(0.018)	(0.026)	(0.036)
Controls									
Firstborn	N	N	N	Y	Y	Y	Y	Y	Y
Adoptee	N	N	N	Y	Y	Y	Y	Y	Y
Mother's marital status	N	N	N	Y	Y	Y	Y	Y	Y
Migration	N	N	N	Y	Y	Y	Y	Y	Y
Survival	N	N	N	Y	Y	Y	Y	Y	Y
Birth cohort FE	N	N	N	Y	Y	Y	Y	Y	Y
Lineage FE	N	N	N	Y	Y	Y	Y	Y	Y
N	12,734	7,366	4,398	12,734	7,366	4,398	12,734	7,366	4,398
Pseudo R^2	0.070	0.034	0.023	0.106	0.084	0.068	0.120	0.089	0.073

Notes: 1. Coefficients are incidence rate ratios (IRR) for the negative binomial regression, and robust standard errors are in parentheses, clustered on fathers. 2. *p<0.1; **p<0.05; ***p<0.01.

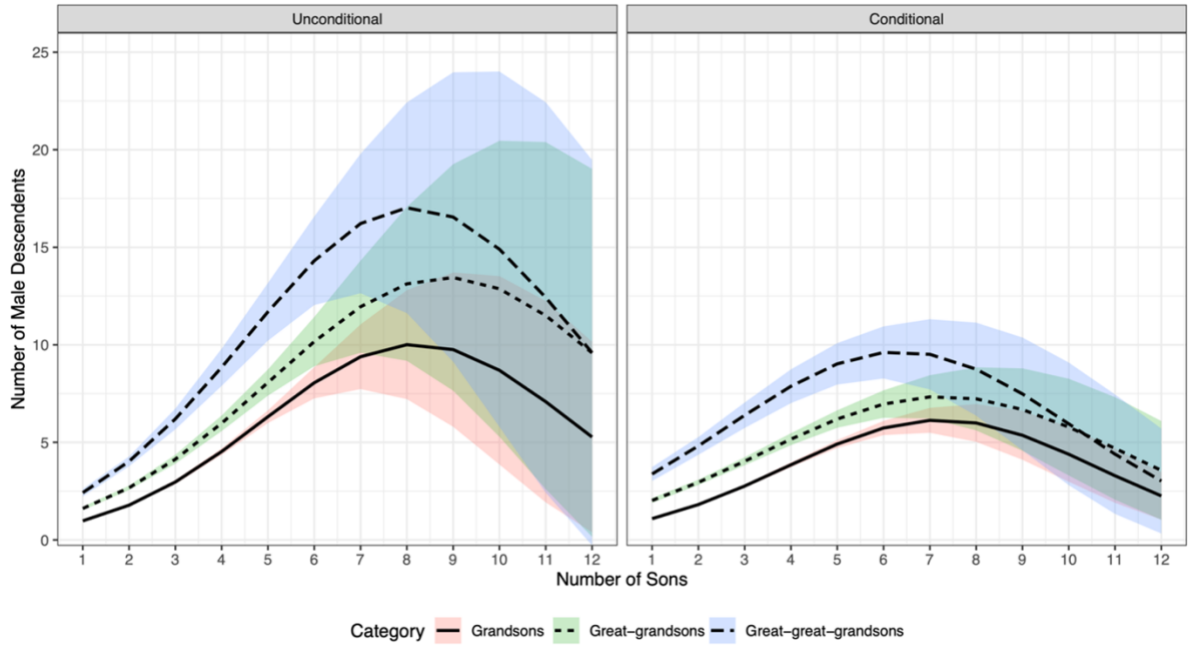

Figure 4 Predicted number of patrilineal male descendants in the three generations at each level of reproduction

Notes: 1. The predicted values for the "Unconditional" plot are calculated from the results in columns 1–3 of Table 2, and the predicted values for the "Conditional" plot are calculated from the results in columns 7–9. 2. The shaded area represents the 95 percent confidence interval for the predicted curve.

The two plots in Figure 4 display the predicted numbers of patrilineal male descendants for each number of sons, derived from the regression results in Table 2. In the left plot of Figure 4, which reflects results from columns 1 to 3, the predicted numbers of grandsons, great-grandsons, and great-great-grandsons all peak at around eight to nine sons. After conditioning on all factors, including the man's characteristics and lineage and birth cohort fixed effects, a hump-shaped pattern persists (see the right plot of Figure 4, based on the regression results in columns 7 to 9 of Table 2). Without control variables, the optimal number of sons for maximizing great-great-grandsons is 8, yielding a predicted 17 great-great-grandsons. When all control variables are applied, the optimal number falls to around 6, with a predicted 10 great-great-grandsons. In terms of absolute fertility, the optimal level was relatively high, given that the median and mean number of sons among these potential grandfathers, great-grandfathers, and great-great-grandfathers was only about two. Nevertheless, both sets of

results indicate an optimal level of reproduction that maximizes long-run reproductive success, underscoring a robust Darwinian trade-off.[*]

The coefficients for *Grade* and *Marriages*, which are both statistically significant and greater than one in columns 7 to 9, underscore the important role that social status and wealth played in long-term reproductive success. High-status men were able to leave more male descendants than their low-status counterparts in subsequent generations. A man's social status and wealth influenced his reproduction and the reproductive success of at least the next three generations. Figure 5 illustrates this social gradient in long-run reproductive success: men in Grades 2–7 had more patrilineal grandsons, great-grandsons, and great-great-grandsons than men in Grade 1. The persistent effect of social status observed in the six lineages aligns with findings by Song et al. (2015) in North China from 1725 to 1875 and Lee and Park (2019) in pre-modern Korea.

[*] To rigorously test and rule out non-hump-shaped relationships, I conduct the test proposed by Lind & Mehlum (2010), which assesses the statistical significance of the hump shape. The results in Table A2 in the Online Appendix provide strong evidence for hump-shaped patterns. Specifically, for each category of male descendants analyzed (grandsons, great-grandsons, and great-great-grandsons), the relationship exhibits a significant positive slope at the lower extreme, followed by a negative slope at the higher extreme, with optimal fertility levels for each category being 7, 7, and 6 sons, respectively. These optimal levels align with the predictions in columns 7–9 of Table 2 and the right panel of Figure 4. The statistical significance of these slopes at both extremes, confirmed by their associated t-values (all $p < 0.001$), strongly suggests that the number of male descendants follows a robust hump-shaped pattern.

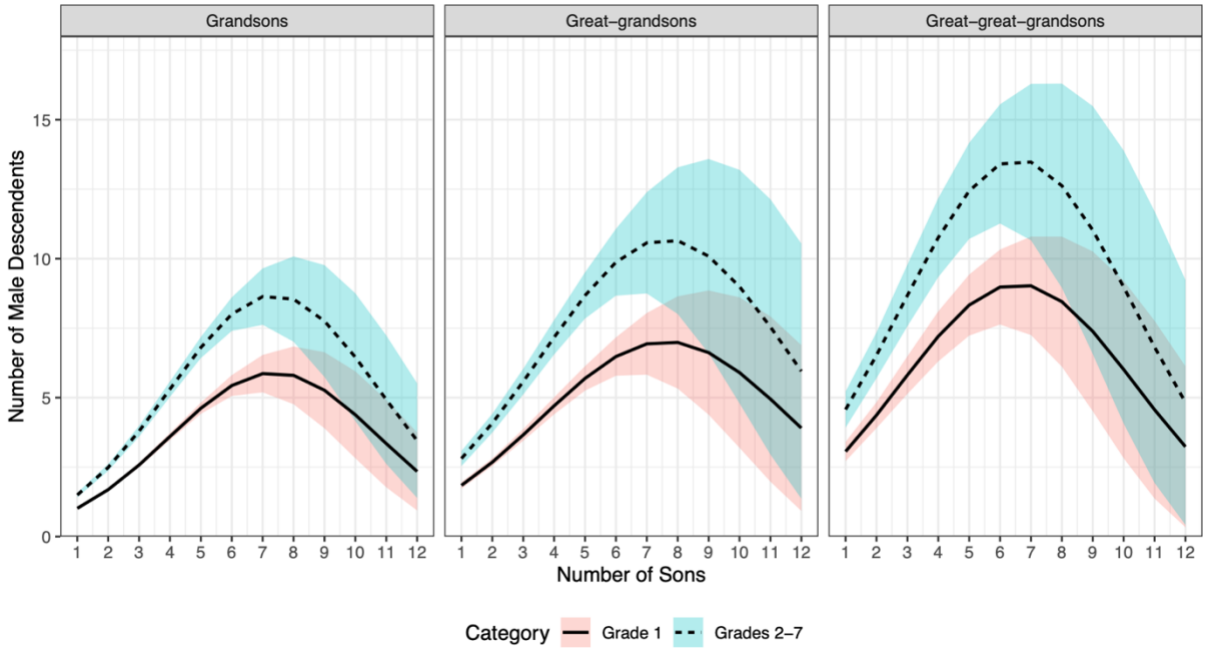

Figure 5 Predicted number of male descendants in the three generations at each level of reproduction by social status

Notes: 1. The predicted values are based on the adjusted results in columns 7–9 of Table 2, with men from Grades 2–7 grouped as a high-status group in the model. 2. The shaded area represents the 95% confidence interval for the predicted curve.

4.1.2 Robustness

This section tests the robustness of the previous results of the Darwinian trade-off to (1) an alternative estimation model, (2) potential heterogeneity across lineages with different social strata, (3) controlling for lifespan, (4) converting the number of sons into a categorical variable, and (5) differences between male and female reproductive success.

First, using a negative binomial regression model, the baseline analysis establishes a hump-shaped relationship between the number of sons and long-run reproductive success. As shown in Table B1 in the Online Appendix, the coefficients on the squared term of *Sons* are negative across all three specifications, indicating that this hump-shaped relationship is robust to using an OLS regression model.[†]

[†] To better fit the OLS regression model, I log-transform *Descendants*, and, to keep all the zero observations, make the outcome variable used in the OLS model is $ln\ (Descendants_i+1)$.

Second, I estimate the baseline regression separately for the two subsamples—the three elite lineages and the three common ones—to examine whether the trade-off is present in only one lineage type. Table B2 and Figure B2 in the Online Appendix report the results. The hump-shaped pattern emerges in both groups, suggesting that the baseline result is not driven by a specific group. However, as expected, Figure B2 shows that the optimal fertility level differs, with elite lineages reaching a higher optimal level of fertility compared to common lineages. Additionally, the negative effect of having a larger family size is stronger in common lineages than in elite ones.[‡]

Third, men's lifespan might significantly affect their reproduction and, thus, long-run reproductive success. To address this concern, I further condition the analysis on men with complete birth and death records, controlling for lifespan. The results are reported in Online Appendix Table B3. Lifespan positively affected the number of male descendants a male could have in generations 3–5. When lifespan is included as a control, the coefficients of *Sons* and $Sons^2$ are highly comparable to the results in columns 7–9 of Table 2, affirming that the baseline results are robust.

Fourth, although the coefficients on $Sons^2$ in Tables 2, B1, B2, and B3 and the results in Table A2 consistently suggest a hump-shaped relationship between reproduction and long-run reproductive success, this could still be an artifact of using the squared term. To validate the non-monotonic nature of this relationship, I conduct an additional regression with the same estimation equation, converting the number of sons from a continuous variable to a categorical variable and excluding the square term of the number of sons.[§]

The results in Table B4 and Figure B1 in the Online Appendix affirm the persistence of the hump-shaped pattern, particularly in the relationship between the number of sons and the

[‡] I conduct the same Lind & Mehlum (2010) test and confirm the validity of the hump-shaped pattern.
[§] As only one man in the sample, Zhuang Chaosheng, had more than 10 sons, I combine him with the men who had 10 sons, forming a group of men with 10 or more sons.

number of great-great-grandsons. The optimal level of reproduction remains at about seven sons, underscoring the robustness of the baseline results.

Fifth, while the baseline analysis focuses on male reproductive success, female reproductive success may not exhibit a similar pattern. In fact, serial monogamy and polygamy (such as concubinage) were uncommon in the six lineages: only 2,517 men (10.7 percent) in the sample had more than one marriage, with 2,022 of these involving only remarriages rather than concubines. However, focusing solely on male reproductive patterns may still overstate the role of income and wealth—and thus the role of men—in reproductive success, while overlooking the significant contributions of women, who often took primary responsibility for child rearing and education.

Unfortunately, information on women's education and literacy is unavailable in genealogies. To partially address this gap concerning female reproductive success, I conduct two additional analyses: 1) controlling for the social class of fathers-in-law (categorized as degree-/officie-holders or non-degree holders) as a proxy for women's human capital, and 2) restricting the sample to men who were married only once in their lifetime, so that their reproductive success directly corresponds to that of their wives.

Table B5 presents the results from these two analyses. Additionally, controlling for the social class of fathers-in-law and restricting the sample to men with only one marriage did not significantly alter the coefficients on *Sons* or the squared term of *Sons*, implying that female reproductive success was comparable to that of males to a certain extent, given the low prevalence of serial monogamy and polygamy in traditional China.

4.2 The Beckerian trade-off: Mechanisms through which reproduction affected long-run reproductive success

The decline in the number of grandsons, great-grandsons, and great-great-grandsons after reaching the optimal fertility level suggests a Beckerian trade-off. This indicates that an excessively large family size can diminish each son's quality, ultimately impacting reproductive outcomes in subsequent generations. Galor and Klemp (2019) propose that moderate fecundity could enhance child quality by increasing the likelihood of marriage and educational attainment, thus enabling children to leave more descendants. Accordingly, in this section, I examine two types of Beckerian child quantity-quality trade-offs in the six lineages: the effects of son quantity on the probability of sons marrying and achieving a *keju* degree.

4.2.1 Logistic estimation results

Tables 3 and 4 report the logistic regression results based on Equation (2).[**] In Table 3, where quality is measured by marital status, the key independent variable, *Brothers*, consistently retains coefficients greater than one across all five specifications.[††] Having more brothers increased a man's likelihood of marriage. Additionally, being the firstborn or an adopted son conferred an "inheritance advantage" (columns 3–5), as these roles were ritually significant, often resulting in greater resource allocation to continue the patriline, further enhancing marriage likelihood. In contrast, being the son of a secondary wife or concubine posed a disadvantage, as indicated by the coefficient on *Mother's marital status* falling below one in columns 4 and 5. Columns 4 and 5 affirm that the positive correlation between child quantity and quality persists even when controlling for paternal and grandpaternal effects. Compared to the number of brothers, the social status of fathers and grandfathers was more influential in affecting an individual's marriageability.

[**] Online Appendix Tables A3 and A4 also report the OLS regression results based on Equation (2).
[††] All coefficients in Tables 3 and 4 are presented as odds ratios. A coefficient greater than one indicates a positive effect, while a coefficient less than one indicates a negative effect.

Table 3 Relationship between family size and marital status, logistic regression

	Dependent Variable: Married				
	(1)	(2)	(3)	(4)	(5)
Brothers	1.051***	1.058***	1.138***	1.092***	1.090***
	(0.015)	(0.017)	(0.022)	(0.022)	(0.022)
Firstborn			1.535***	1.509***	1.513***
			(0.057)	(0.057)	(0.057)
Adoptee			1.193**	1.304***	1.298***
			(0.084)	(0.096)	(0.096)
Mother's marital status			1.293***	0.839**	0.831**
			(0.090)	(0.068)	(0.068)
Father's grade 2				2.060***	1.940***
				(0.425)	(0.402)
Father's grade 3				2.493***	2.574***
				(0.418)	(0.449)
Father's grade 4				2.961***	3.028***
				(0.399)	(0.431)
Father's grade 5				1.715***	1.789***
				(0.326)	(0.362)
Father's grade 6				2.031***	2.196***
				(0.330)	(0.402)
Father's grade 7				6.957***	7.112***
				(3.530)	(3.595)
Father's marriages				1.389***	1.386***
				(0.078)	(0.078)
Grandfather's grade 2					1.007
					(0.169)
Grandfather's grade 3					1.394***
					(0.175)
Grandfather's grade 4					1.058
					(0.116)
Grandfather's grade 5					0.707*
					(0.131)
Grandfather's grade 6					0.610***
					(0.087)
Grandfather's grade 7					1.812*
					(0.564)
Grandfather's marriages					1.003
					(0.039)
Grandfather's sons					0.990
					(0.015)
Migration	N	Y	Y	Y	Y
Survival	N	Y	Y	Y	Y
Birth cohort FE	N	Y	Y	Y	Y
Lineage FE	N	Y	Y	Y	Y
N	23,281	23,246	23,246	23,246	23,106
Pseudo-R^2	0.001	0.227	0.233	0.245	0.248

Notes: 1. If the man is an adoptee, then his father- and grandfather-related characteristics denote the stepfather's and step-grandfather's characteristics, respectively. 2. Coefficients are the odds ratios for the logistic regression, and robust standard errors are in parentheses, clustered on fathers. 3. *p<0.1; **p<0.05; ***p<0.01.

Table 4 Relationship between family size and degree, logistic regression

	Dependent Variable: Degree				
	(1)	(2)	(3)	(4)	(5)
Brothers	1.229***	1.219***	1.270***	1.011	0.988
	(0.030)	(0.028)	(0.031)	(0.028)	(0.030)
Firstborn			1.509***	1.517***	1.536***
			(0.087)	(0.105)	(0.113)
Adoptee			1.296**	1.306*	1.244
			(0.154)	(0.189)	(0.183)
Mother's marital status			1.776***	0.808*	0.785*
			(0.163)	(0.100)	(0.102)
Father's grade 2				7.079***	5.195***
				(1.400)	(1.069)
Father's grade 3				13.167***	6.739***
				(1.727)	(0.988)
Father's grade 4				15.296***	8.120***
				(1.858)	(1.162)
Father's grade 5				19.810***	8.642***
				(3.346)	(1.654)
Father's grade 6				32.383***	14.433***
				(4.753)	(2.450)
Father's grade 7				72.243***	29.759***
				(15.754)	(7.081)
Father's marriages				1.191***	1.170**
				(0.076)	(0.078)
Grandfather's grade 2					2.507***
					(0.507)
Grandfather's grade 3					3.649***
					(0.527)
Grandfather's grade 4					3.058***
					(0.454)
Grandfather's grade 5					4.408***
					(0.999)
Grandfather's grade 6					4.343***
					(0.693)
Grandfather's grade 7					6.829***
					(1.410)
Grandfather's marriages					1.018
					(0.044)
Grandfather's sons					0.964
					(0.025)
Migration	N	Y	Y	Y	Y
Survival	N	Y	Y	Y	Y
Birth cohort FE	N	Y	Y	Y	Y
Lineage FE	N	Y	Y	Y	Y
N	23,281	22,436	22,436	22,436	22,304
Pseudo-R^2	0.016	0.161	0.170	0.365	0.392

Notes: 1. If the man is an adoptee, then his father- and grandfather-related characteristics denote the stepfather's and step-grandfather's characteristics, respectively. 2. Coefficients are the odds ratios for the logistic regression, and robust standard errors are in parentheses, clustered on fathers. 3. *p<0.1; **p<0.05; ***p<0.01.

Table 4 suggests a somewhat different role that family size played. Before including the paternal and grandpaternal effects, columns 1–3 suggest that for a one-unit increase in *Brothers*, the odds of a male achieving an academic degree versus not achieving one would increase by a factor of about 1.2 to 1.3. However, after including the effects of fathers and grandfathers in columns 4–5, the coefficients on *Brothers* lose statistical significance and become smaller than one in column 5, implying a potential negative correlation. Having a larger number of brothers might be negatively associated with the likelihood of a male attaining a *keju* degree.

As expected, fathers' social status had the most substantial effect on sons' quality. Figure 6 demonstrates the predicted probability of being married and holding a *keju* degree for families of different sizes, conditioned on fathers' social status. The two plots in Figure 6 show that the considerable difference in sons' quality arises from the difference in fathers' social status. For a man raised in a family of two sons, the predicted probability of marriage was about 80.4 percent if he had a father who belongs to Grades 2–7, while that for a man with a Grade 1 father was about 68.4 percent. The convergence of these two percentages becomes more pronounced as family size increases. The differences are more expansive for the probabilities of achieving academic degrees. The predicted probability of holding degrees for men with Grades 2–7 fathers in a two-son family was 16.3 percent, and that for men with Grade 1 fathers was only 2.7 percent.

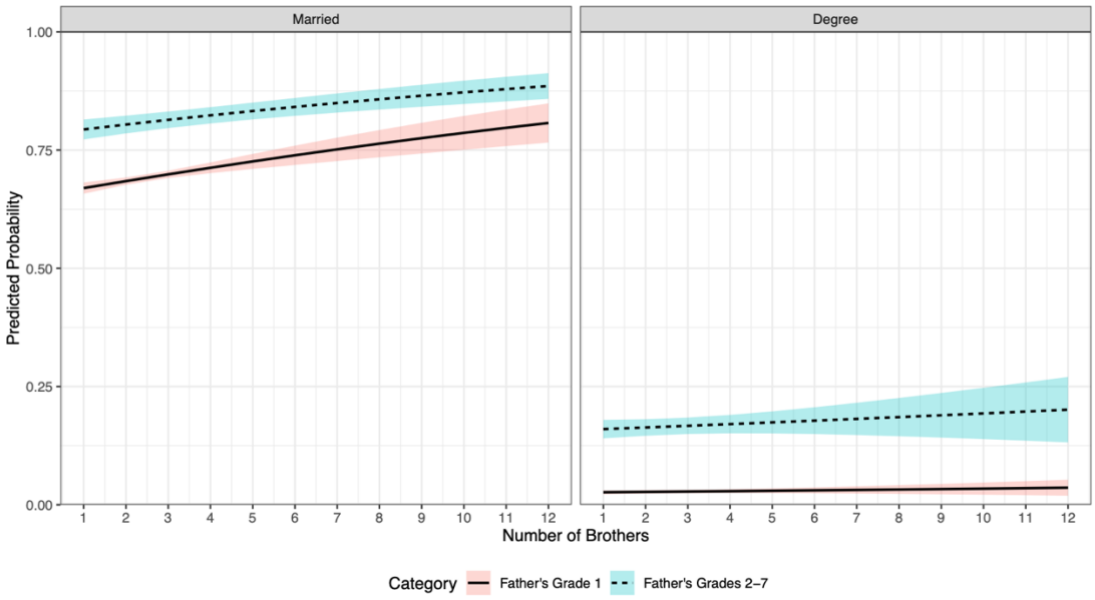

Figure 6 Predicted probability of being married and earning a degree for families of different sizes, with a low-status and a high-status father.

Notes: 1. The predicted values are calculated from the adjusted results in columns 5 of Table 3 and 5 of Table 4, respectively, with men with fathers higher than Grade 1 grouped in Grades 2–7. 2. The shaded area represents the 95% confidence interval for the predicted curve.

The baseline results reveal the varying effects of the number of brothers on sons' quality. When measured by marriageability, men raised in relatively large families appeared to have an advantage. However, when the attainment of *keju* degrees measures quality, this positive effect disappears and even has the potential to turn negative. This distinction corresponds with historical and social patterns in Ming–Qing China: although marriage for men was not universal, getting married was significantly easier than obtaining a *keju* degree. The large family size rarely created a barrier to marriage in families with many sons. However, it could become an obstacle to producing a *keju*-degree holder, as obtaining such a degree—whether by examination or purchase—required substantial financial resources and considerable time. Consequently, the results suggest that the Beckerian trade-off, as reflected in marital status, was absent in the six lineages, while it may have been present in terms of human capital formation. In addition, paternal and grandpaternal social status exerted a much stronger influence than the number of brothers.

4.2.2 Robustness

This section presents the robustness of the previous results of the Beckerian trade-off to (1) different subperiods, (2) potential heterogeneity across lineages with different social strata, (3) controlling for lifespan, (4) alternative human capital indicators, (5) the 16-level social status classification, and (6) a subsample of daughters.‡‡

I first divide the sample into three sub-samples that refer to three distinct cohorts and report the results in Table B6. For males born between 1300 and 1650, the number of brothers they had and their quality measures were not significantly correlated. However, after 1650, *Brothers* displayed a positive correlation between with *Married*. When measuring quality based on *Degree*, the coefficients on *Brothers* are not statistically significant across all three periods. Overall, similar to the baseline results in Tables 3 and 4, family size positively correlated with the likelihood of marriage. In contrast, a weak and even negative correlation emerged with the likelihood of attaining a *keju* degree.

Second, I compare the correlations between the quantity and quality of sons in elite and common lineages. Table B7 and Figure B3 show the heterogeneous effects. As expected, sons from elite lineages have a higher probability of marriage and academic achievement, regardless of the number of brothers, compared to those from common lineages. In contrast, in relative terms, family size has a stronger positive association with marriageability in common lineages than in elite ones. Meanwhile, although the negative association between *Brothers* and *Degree* is quantitatively larger in elite lineages compared to common lineages, it still remains statistically insignificant.

I then incorporate the lifespan of each man and his father into the model. Similar to the previous reasoning, lifespan is a proxy for health, potentially influencing an individual's

‡‡ Summary statistics for the additional variables related to daughters are also provided in Table A1.

probability of marrying and holding a *keju* degree. Therefore, I include the male individual's and his father's age at death, with results in Table B8. After accounting for health-related factors, the coefficients on *Brothers* fall below one in columns 1 and 2 when quality is measured by marriageability. In columns 3 and 4, when quality is measured by obtaining a *keju* degree, the coefficients on *Brothers* remain below one and reach statistical significance. These findings suggest that, among men with recorded lifespans for themselves and their fathers, the number of brothers negatively impacted their human capital formation once health characteristics were considered. The results thus continue to support the existence of a Beckerian trade-off in these six lineages, affirming the baseline findings.

I also use whether a man had a *Zi* (courtesy name) or *Hao* (pen name) as two alternative measures for human capital. I use *Zi* and *Hao* to measure rudimentary literacy because *Degree* measures the most "upper-tail" human capital in traditional China. *Zi*, a courtesy name, is a name given to a male by his father or his teacher when he reaches adulthood, and *hao*, as mentioned earlier, is a pen name that a literate male would give himself. Having a *zi* or *hao* means that the man could read at least a rudimentary literacy. These two measures are similar to signature-based literacy, the conventional indicator for human capital (Schofield, 1968; Ogilvie, Edwards, & Küpker, 2022). Therefore, I change the outcome of interest from *Degree* to *Zi* and *Hao*, and Table B9 reports the results.

As shown in columns 1 and 3, before conditioning on lifespan, coefficients on the number of brothers a man had were greater than one, albeit statistically insignificant. After conditioning on the lifespan of the male individual and his father, the coefficients in columns 2 and 4 remain statistically insignificant but fall to be smaller than one, suggesting a potential negative impact of *Brothers* on literacy. The results are comparable to those in Table 4, which suggests that the baseline results are robust. The results of the three indicators of human capital, upper-tail

human capital (*Degree*), and rudimentary literacy (*Zi* and *Hao*), all demonstrate the presence of a Beckerian trade-off of children in the six lineages in Ming–Qing China.

A more refined social status classification is employed to more accurately control for the socio-economic characteristics of the men's family of origin. In the regression results shown in Table B10, the father's and grandfather's grades are categorized using the more detailed 16-level social classification presented in Table 1. This finer control does not significantly alter the effect of family size. Before accounting for lifespan, the number of brothers remained positively correlated with marriageability but negatively, albeit weakly, correlated with acquiring a *keju* degree. After conditioning on the individual's and his father's lifespan, the coefficients for *Brothers* became less than one and statistically significant, reinforcing the presence of a Beckerian trade-off.

Finally, I test the child quantity-quality trade-off using the daughters' sample. Although daughters were underrecorded in Chinese genealogical books, 2,414 fathers in the sample had records of their 4,122 daughters' marriages. Using the recorded son-in-law information, I run an ordinal logistic regression based on Equation (5):

$$P(Husband\ class_j \leq k) = \Phi(\alpha_4 + \beta_4 Sisters_j + \beta_5 Brothers_j + \delta_4 D_j + \delta_5 F_j + \varepsilon_j), \quad (5)$$

Husband class is an ordinal variable representing the social class of the daughter's husband and is used here to measure the daughters' "quality". k identifies the four categories of *Husband class*, ranging from 0 to 3. If the daughter was unmarried, this variable equals 0; if she married a non-degree holder, a degree-holder, or an officeholder, it equals 1, 2, or 3, respectively. [§§] j refers to individual daughters, and α is the constant. *Daughter* measures the

[§§] Among the 4,122 daughters, 351 (8.5 per cent) remained celibate, either due to early death before marriage or lifelong singlehood; 3,164 (76.8 per cent) married non-degree holders; 314 (7.6 per cent) married degree holders; and 293 (7.1 per cent) married officeholders.

number of sisters (including herself), which corresponds to the total number of daughters her father had. *Brothers* represents the number of brothers she had. *D* encompasses the daughter's characteristics, including whether she was the firstborn among her sisters and the lineage to which she belonged. *F* includes a set of controls that capture features of the daughter's natal family, such as father effects (social grades and lifespans of fathers, as well as social grades of paternal grandfathers), mother's marital status (whether she was a concubine or not), and her father's 50-year birth cohort. ε is the error term.

Table B11 exhibits the results. As expected, among all factors, a father's and grandfather's social status were the decisive factors affecting the daughter's marriage, while the number of sisters and brothers she had had only a minimal effect on her marital outcomes. Nevertheless, as suggested in column 4 of Table B11, controlling for the father's lifespan changes the coefficients on *Sisters* and *Brothers* to less than one, still implying the presence of a Beckerian trade-off. An additional increase in family size could potentially have an adverse effect on the probability of a daughter marrying a degree holder or an office holder.

4.2.3 Causal identification: Using the mother's age at first son's birth as an instrument

This section turns to the instrumented evidence on the trade-off between child quantity and quality. The logistic regression results in Section 4.2.1 could be biased because of the unobserved parental preference and household features, although controlling for the grandpaternal effects could partially capture the unobserved features.

The IV estimation results based on Equations (3) and (4) are shown in Tables 5 and 6.[***] The *F*-statistic on the instrument remains higher than the Stock and Yogo (2005) critical value

[***] Because information on mothers' birth years is unavailable for every man in the sample, the instrumented analysis is limited to a subsample. I also use this subsample to conduct the same baseline analysis, with results reported in Table A5 of Online Appendix A. The findings for this subsample are consistent with the baseline results.

in Table 5 and in the specifications of periods after 1650 in Table 6, ruling out the weak instrument concern.

Table 5 details the full sample results, comprising all males born between 1300 and 1920 who had records of the mother's age at the birth of the first son across the six lineages. Before I report the 2SLS results, I report those of the reduced-form estimates in columns 1 and 2 of Table 5. The OLS results for the first stage are reported in column 3, and the average marginal effects of the probit regressions for the second stage are reported in columns 4 and 5, where the dependent variables are *Married* and *Degree*, respectively. Table 6 reports the results for three sub-cohorts, 1300 to 1650, 1650 to 1800, and 1800 to 1900. Columns 1, 4, and 7 report the first-stage results, and the remaining columns report the second-stage results. In all specifications, a complete set of control variables discussed in Sections 3.1.2 and 3.2.1 is included, along with birth cohort fixed effects and lineage fixed effects.

The instrumented results in Table 5 display a broadly consistent pattern with those in Tables 3 and 4, with coefficients pointing in the same directions. As shown in columns 1 and 4, when considering marriageability, the coefficients for the mother's age at first son's birth and the number of brothers— estimated using the instrument—imply that the quantity of sons may have had a potentially positive effect on their quality. However, this effect appears minimal in quantitative terms and lacks statistical significance. In contrast, when considering *keju* degree attainment, evidence of a quantity-quality trade-off emerges (columns 2 and 5). As shown in column 5, an increase of one son in family size is associated with a 1.4-percentage-point decrease in the likelihood of obtaining a *keju* degree for men in the six lineages. Additionally, men with high-status fathers exhibit a significant advantage: for example, a man

with a Grade 7 father would experience a 23.8-percentage-point and 53.8-percentage-point increase in the probability of marriage and degree attainment, respectively.[†††]

Table 5 Impact of the quantity of children on the quality of children: Instrumented results

	Reduced-form		2SLS		
	Married	Degree	Brothers	Married	Degree
	(1)	(2)	(3)	(4)	(5)
Mother's age at first son's birth	-0.001	0.001**	-0.057***		
	(0.001)	(0.0003)	(0.004)		
Brothers				0.015	-0.013**
				(0.014)	(0.006)
Father's grade 2	0.081**	0.071***	0.102	0.080**	0.073***
	(0.037)	(0.020)	(0.122)	(0.037)	(0.020)
Father's grade 3	0.169***	0.140***	0.514***	0.163***	0.163***
	(0.026)	(0.024)	(0.170)	(0.027)	(0.029)
Father's grade 4	0.189***	0.130***	0.425***	0.184***	0.146***
	(0.020)	(0.021)	(0.142)	(0.020)	(0.024)
Father's grade 5	0.099**	0.107***	-0.107	0.099**	0.101***
	(0.046)	(0.032)	(0.165)	(0.046)	(0.032)
Father's grade 6	0.128***	0.226***	0.614**	0.118***	0.260***
	(0.040)	(0.046)	(0.277)	(0.042)	(0.050)
Father's grade 7	0.244***	0.504***	0.479	0.238***	0.538***
	(0.080)	(0.098)	(0.309)	(0.085)	(0.101)
Father's marriages	0.055***	0.006	0.211***	0.052***	0.009**
	(0.012)	(0.004)	(0.057)	(0.013)	(0.004)
Observations	11,606	10,891	11,632	11,606	10,891
Number of clusters	5,902	5,557	5,916	5,902	5,557
Pseudo R^2	0.232	0.401			
F-statistic on instrument			249.65		

Notes: 1. Robust standard errors clustered by fathers in parentheses. 2. All specifications shown include the mother's marital status, out-migration, survival to adulthood, grandfather's grade, grandfather's marriages, grandfather's sons, birth cohort FE, and lineage FE. 3. Columns 1–2 are subjected to probit regression. Coefficients are average marginal effects. 4. Columns 3–5 present the results of the Ivprobit estimation: the first stage being OLS (columns 3), and the second stage involves probit regression on the predicted values from the first stage (columns 4 and 5). The two models have the same first stage. The coefficients in columns 4 and 5 are average marginal effects. The *F*-statistic on the instrument is derived from a 2SLS estimate. 5. *p<0.1; **p<0.05; ***p<0.01.

[†††] To further evaluate the validity of my chosen instrument, I implement the local-to-zero approach proposed by Conley et al. (2012), which allows for a small but nonzero correlation between the instrument and the error term. Table A6 in the Online Appendix reports the results under four different priors on this correlation parameter. Among them, the most conservative case assumes a normal prior with a mean of 0 and a standard deviation of 0.01, and it produces the tightest bounds (-1.78, 1.62), suggesting slight potential bias. A more relaxed assumption using a uniform prior with support [-0.1, 0.1] allows for greater potential endogeneity. It results in wider bounds (-8.34, 8.29). Although the broader confidence intervals imply that stronger violations of the exclusion restriction could affect the estimates, the overall results from Table A6 show that the estimates do not shift significantly under plausible deviations from being strictly exogenous.

Table 6 Two-stage regression with mother's age at first son's birth by period, measured by *Married* and *Degree*

	Born in 1300–1650			Born in 1650–1800			Born in 1800–1920		
	1st Stage	2nd Stage		1st Stage	2nd Stage		1st Stage	2nd Stage	
	Brothers	Married	Degree	Brothers	Married	Degree	Brothers	Married	Degree
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Father's age at first son's birth	-0.032**			-0.066***			-0.054***		
	(0.014)			(0.007)			(0.004)		
Brothers		-0.059	0.052		0.017	-0.021**		0.026	-0.003
		(0.144)	(0.060)		(0.021)	(0.011)		(0.018)	(0.007)
Observations	232	210	200	3,900	3,865	3,900	7,500	7,500	7,263
Number of clusters	133	121	114	1,963	1,947	1,963	4,042	4,042	3,935
F-statistic on instrument	4.89			90.89			191.55		

Notes: 1. Robust standard errors clustered by fathers in parentheses. 2. All specifications shown include firstborn, adoptee, mother's marital status, migration, survival to adulthood, father's grade and marriages, and grandfather's grade, marriages, and sons, and lineage FE. 3. Ivprobit estimation: the first stage being OLS, the second stage probit, regressed on the predicted values from the first stage. The two models have the same first stage. 4. The coefficients in columns 2, 3, 5, 6, 8, and 9 are average marginal effects. The *F*-statistic on the instrument is derived from a 2SLS estimate, which has the same first stage. 5. *p<0.1; **p<0.05; ***p<0.01.

Table 6 shows that the relationships differ by cohort. Across all three cohorts, the number of brothers did not significantly impact marriageability. Nonetheless, for men born between 1650 and 1800, during the early and mid-Qing dynasty, each additional brother was associated with an approximate 2-percentage-point decrease in the probability of attaining a degree.

The inconsistent statistical significance between the instrumented and logistic estimates underscores the challenges of establishing causal relationships between child quantity and quality in historical contexts. It also implies that both models may be biased because of inherent limitations. The logistic regression suffers from endogeneity bias due to omitted variables, such as unobserved family characteristics and parental preferences. While the IV estimation mitigates endogeneity to some extent, it introduces potential instrument-induced bias, as the instrument is unlikely to be perfectly exogenous and efficient.

Despite these limitations, the two models yield consistent results in terms of coefficient direction and magnitude, providing informative findings. Overall, the results suggest that family size did not have a clear negative impact on a son's marriageability but may have reduced his likelihood of attaining a *keju* degree, pointing to the presence of a Beckerian trade-off in the six southeastern Chinese lineages, particularly for the 1650–1800 cohort. This pattern also corroborates Shiue's (2017) findings for Tongcheng County of Anhui Province: the trade-off of children, as reflected in *keju* attainment, was evident in the Qing dynasty before 1800 but disappeared thereafter, primarily due to the shift in returns to human capital.

5 Discussion

Section 4 demonstrates the presence of the two trade-offs within the six lineages. The findings from the two trade-offs collectively illustrate a nuanced relationship between reproduction and human capital in a multigenerational context. On the one hand, high-status men of Grades 2–7

in the six lineages were able to produce more patrilineal male descendants over at least four consecutive generations. This persistence stemmed from the inheritance of privileges by their sons, grandsons, and great-grandsons, which, in turn, amplified their probability of marriage and *keju* degree attainment. On the other hand, high-status men—who generally had relatively large families—experienced diminishing returns when family size became excessively large. An additional son could reduce the number of patrilineal grandsons, great-grandsons, and great-great-grandsons, as well as lower their sons' chances of acquiring a *keju* degree.

Nonetheless, for most high-status families with moderate family sizes (averaging 2.21 sons), the dynamics still favored a "winner-take-all" outcome. As shown in Table 2 and Figure 4, for families within the range of up to six sons, larger family sizes tended to extend the family line by increasing the likelihood of raising accomplished male descendants, with limited drawbacks. Despite the two-trade-offs, fathers of Grades 2–7 could likely offset any adverse effects of larger family sizes on their son's quality. Degree-holders and officeholders possessed sufficient resources to support relatively large numbers of sons and secure their education. By leveraging their social capital and providing "cultural capital" to their sons, their sons had greater ease in achieving *keju* success (Ho, 1962; Jiang & Kung, 2021). [*] For these fathers, having more sons within a reasonable range, as suggested by the six-son limit in Figure 4, was akin to buying additional "lottery tickets" for success in the *keju* competition. Although budget constraints were present, they were generally high enough that trade-offs between reproduction and long-run reproductive success or between child quantity and quality were less burdensome; instead, these families were positioned to afford both.

In contrast, Grade 1 fathers encounter the two trade-offs in a fundamentally different manner. Their family sizes were limited—not as a deliberate choice to invest in their sons'

[*] In Ming–Qing China, large lineages each maintained their own lineage schools (Zelin, 2009). The existence of these schools led to a relatively low marginal cost for rearing additional sons.

human capital but due to strict resource constraints. While rudimentary literacy was accessible, full literacy and *keju* degrees were far more challenging.[†] Although the social mobility incentive drove low-status men to pursue *keju* degrees, high opportunity costs and intense competition prevented most of them from making consistent attempts (Ebrey, 1993; Shiue, 2017). Elman (2000, p. 240) points out that attaining classical literacy required "substantial investments of time, effort, and training", always demanding "financial and labor sacrifices". Given these high costs and low probability of success, maintaining a large family size, along with consistently supporting their sons in the *keju* pursuit, was largely impractical, making the trade-offs less relevant to these families.

Considering how Chinese families navigated the two trade-offs, the absence of a fertility transition in nineteenth-century China becomes more understandable. Coale (1974, pp. 352–353) argues that a crucial condition for a society to experience a secular decline in fertility is that "perceived social and economic circumstances must make reduced fertility seem an advantage to individual couples".[‡] While many Western European societies fulfilled this condition in the nineteenth century, with affluent families leading in limiting family size compared to poorer ones (see, for example, Clark & Cummins, 2015), nineteenth-century China did not. The Malthusian mechanism continued to operate throughout this period. The empirical investigation in this paper directly responds to this condition, implying that reduced fertility was not viewed as advantageous by parents, particularly those of high status, as having more sons correlated with having more high-status sons. Without a secular fertility decline, the absence of a modern economic take-off is unsurprising (Galor & Weil, 2000).

6 Conclusion

[†] In terms of rudimentary literacy, Rawski (1979, p. 23) estimates that in nineteenth century China, the literacy rates of the male population ranged from 30 to 45 per cent.
[‡] The other two conditions that Coale (1974, pp. 352–353) mentioned are "fertility must be within the calculus of conscious choice" and "effective techniques of fertility reduction must be available."

Transitioning from Malthusian stagnation to sustained economic growth necessitates a transitioned demographic pattern. This paper uses a new genealogical dataset to examine the demographic dynamics of patrilineal lines within six Chinese lineages from 1300 to 1920. The empirical investigation provides evidence that points to the presence of Darwinian and Beckerian trade-offs. Greater reproduction could impede long-term reproductive success. A close analysis of mechanisms suggests that increased paternal reproduction had minimal impact on sons' marital outcomes but appeared to have negatively affected their human capital. Instrumenting family size with the variation induced by the mother's age at the birth of her first son further confirms the presence of a trade-off between the quantity of sons and their human capital in the six lineages. Notably, paternal social status, rather than family size, had a more sustained and significant effect on lineage continuity and sons' social achievements.

These findings thus illuminate the hierarchical structure of Chinese society, where social continuity was reinforced through reproduction and inheritance. Fei (1992, p. 120) observes that traditional China "maintains structural stability by using the biological process underlying reproduction as the medium to establish social continuity". This consanguineous inheritance pattern aligns closely with Chinese families' strategies to ensure lineage survival. The enduring demographic dynamics in these six Chinese lineages also shed light on the persistence of China's historical demography pattern through the late imperial period, illustrating the absence of a following transition to modern economic growth.

References

Bai, Y., Li, Y., & Lam, P.H. (2023). Quantity-quality trade-off in Northeast China during the Qing dynasty. *Journal of Population Economics*, 36, 1657–1694. https://doi.org/10.1007/s00148-022-00933-x

Barclay, G.W., Coale, A.J., Stoto, M. A., & Trussell, T. J. (1976). A Reassessment of the Demography of Traditional Rural China. *Population Index*, 42(4), 606–635.

Baudin, T., & De la Croix, D. (2023). The Emergence of the Child Quantity-Quality Tradeoff – Insights from Early Modern Academics. *LIDAM Discussion Papers No. 2023015.* Université catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES).

Becker, G. S. (1960). An Economic Analysis of Fertility. In Universities-National Bureau Committee for Economic Research (Eds.), *Demographic and Economic Change in Developed Countries* (pp. 209–240). New York, NY: Columbia University Press.

Becker, G. S., & Lewis, H. G. (1973). On the Interaction Between the Quantity and Quality of Children. *Journal of Political Economy*, 81(2, Part 2), S279-S288. https://doi.org/10.1086/260166

Becker, G.S., Murphy, K.M., & Tamura, R. (1990). Human Capital, Fertility, and Economic Growth. *Journal of Political Economy*, 98(5), 12–37. https://doi.org/10.1086/261723

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2005). The More the Merrier? The Effect of Family Size and Birth Order on Children's Education. *The Quarterly Journal of Economics*, 120 (2), 669–700. https://doi.org/10.1093/qje/120.2.669

Borgerhoff Mulder, M. (2000). Optimizing Offspring: The Quantity–Quality Tradeoff in Agropastoral Kipsigis. *Evolution and Human Behavior*, 21 (6), 391–410. https://doi.org/10.1016/S1090-5138(00)00054-4

Campbell, C., & Lee, J. (2002). State Views and Local Views of Population: Linking and Comparing Genealogies and Household Registers in Liaoning, 1749–1909. *History and Computing*, 14(1–2), 9–29. https://doi.org/10.3366/hac.2002.14.1-2.9

Chang, C. L. (1962). *The Income of the Chinese Gentry. A Sequel to the Chinese Gentry: Studies on Their Role in Nineteenth-Century Chinese Society.* Seattle, WA: University of Washington Press.

Chen, T., Kung, J.K., & Ma, C. (2020). Long Live keju! The Persistent Effects of China's Imperial Examination System. *The Economic Journal*, 130 (631), 2030–2064. https://doi.org/10.1093/ej/ueaa043

Clark, G., & Cummins, N. (2015). Malthus to Modernity: Wealth, Status, and Fertility in England, 1500–1879. *Journal of Population Economics*, 28, 3–29. https://doi.org/10.1007/s00148-014-0509-9.

Clark, G., & Cummins, N. (2019). The Child Quality-Quantity Tradeoff, England, 1780–1879: A Fundamental Component of the Economic Theory of Growth is Missing. *CEPR Discussion Paper 11232*.

Coale, A.J. (1974). The Demographic Transition Reconsidered. In *The Population Debate: Dimensions and Perspectives. Papers of the World Population Conference, Bucharest 1* (pp. 347–355). New York, NY: United Nations.

Conley, T. G., Hansen, C. B., & Rossi, P. E. (2012). Plausibly Exogenous. *Review of Economics and Statistics*, 94 (1), 260–272. https://doi.org/10.1162/REST_a_00139

Doepke, M. (2004). Accounting for Fertility Decline during the Transition to Growth. *Journal of Economic Growth*, 9 (3), 347–383. https://doi.org/10.1023/B:JOEG.0000038935.84627.e4

Ebrey, P. B. (1993). *Chinese Civilization: A Sourcebook* (2nd ed). Mumbai: Free Press.

Elman, B. A. (2000). *A Cultural History of Civil Examinations in Late Imperial China*. Berkeley, CA: University of California Press.

Fei, X. (1992). *From the Soil: The Foundations of Chinese Society*. Berkeley, CA: University of California Press.

Feng, E. (2009). Zongzu buduan bianxiu zupu de tedian jiqi yuanyin [Features and Reasons for Lineages to compile and update genealogies]. *Journal of Huaiyin Normal College*, 31(5), 638–647.

Feng, E., & Chang, J. (2001). Qingren shehui shenghuo [*Social Life in the Qing Dynasty*]. Shenyang: Shengyang Press.

Freedman, M. (1966). *Chinese Lineage and Society: Fukien and Kwangtung*. London: Athlone Press.

Galor, O. (2011). *Unified Growth Theory*. Princeton, NJ: Princeton University Press.

Galor, O. (2022). *The Journey of Humanity: The Origins of Wealth and Inequality*. Berkely, CA: Penguin Publishing Group.

Galor, O., & Klemp, M. (2019). Human Genealogy Reveals a Selective Advantage to Moderate Fecundity. *Nature Ecology & Evolution*, 3 (5), 853–857. https://doi.org/10.1038/s41559-019-0846-x

Galor, O., & Moav, O. (2002). Natural Selection and the Origin of Economic Growth. *The Quarterly Journal of Economics*, 117(4), 1133–1191. https://doi.org/10.1162/003355302320935007

Galor, O., & Weil, D.N. (2000). Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond. *American Economic Review*, 90 (4), 806–828. https://doi.org/10.1257/aer.90.4.806

Harrell, S. (1985). The Rich Get Children: Segmentation, Stratification, and Population in Three Chekiang Lineages, 1550-1850. In S. B. Hanley, & A. P. Wolf (Eds.), *Family and Population in East Asian History* (pp. 81–109). Stanford, CA: Stanford University Press.

Harrell, S. (1987). On the Holes in Chinese Genealogies. *Late Imperial China*, 8(2), 53–79. https://doi.org/10.1353/late.1987.0014

Harrell, S., & Pullum, T.W. (1995). Marriage Mortality and the Developmental Cycle in Three Xiaoshan Lineages. In S. Harrell (Ed.), *Chinese Historical Microdemography* (pp. 141–162). Berkeley, CA: University of California Press.

Hill, K., & Hurtado, A. M. (1996). *Ache Life History: The Ecology and Demography of a Foraging People*. New York, NY: Taylor and Francis.

Ho, P. (1962). *The Ladder of Success in Imperial China: Aspect of Social Mobility, 1368–1911*. New York, NY: Columbia University Press.

Hu, S. (2023a). Survival of the Literati: Social Status and Reproduction in Ming–Qing China. *Journal of Population Economics*, 36, 2025–2070. https://doi.org/10.1007/s00148-023-00960-2

Hu, S. (2023b). Descendants over 300 Years: Marital Fertility in Five Lineages in Qing China. *Asia-Pacific Economic History Review*, 63 (2), 200–224. https://doi.org/10.1111/aehr.12269

Jiang, Q., & Kung, J. K. S. (2021). Social Mobility in Late Imperial China: Reconsidering the "Ladder of Success" Hypothesis. *Modern China*, 47(5), 628–661. https://doi.org/10.1177/0097700420914529

Kaplan, H. (1996). A Theory of Fertility and Parental Investment in Traditional and Modern Human Societies. *American Journal of Physical Anthropology: The Official Publication*

of the American Association of Physical Anthropologists, 101 (S23), 91–135. https://doi.org/10.1002/(SICI)1096-8644(1996)23+<91::AID-AJPA4>3.0.CO;2-C

Kaplan, H. S., Lancaster, J. B., Johnson, S. E., & Bock, J. A. (1995). Does Observed Fertility Maximize Fitness among New Mexican Men? *Human Nature*, 6(4), 325–360. https://doi.org/10.1007/BF02734205

Klemp, M., & Weisdorf, J. (2019). Fecundity, Fertility and the Formation of Human Capital. *The Economic Journal*, 129 (618), 925–960. https://doi.org/10.1111/ecoj.12589

Knodel, J., Havanon, N., & Sittitrai, W. (1990). Family Size and the Education of Children in the Context of Rapid Fertility Decline. *Population and Development Review*, 16 (1), 31–62. https://doi.org/10.2307/1972528

Kremer, M. (1993). Population Growth and Technological Change: One Million B.C. to 1990. *Quarterly Journal of Economics*, 108 (3), 681–716. https://doi.org/10.2307/2118405

Lack, D. (1954). *The Natural Regulation of Animal Numbers*. Oxford: Clarendon Press.

Lee, J.Z., & Campbell, C. (1997). *Fate and Fortune in Rural China: Social Organization and Population Behavior in Liaoning 1774–1873*. New York, NY: Cambridge University Press.

Lee, J.Z., & Wang, F. (1999). *One Quarter of Humanity: Malthusian Mythology and Chinese Realities*. Cambridge, MA: Harvard University Press.

Lee, S., & Park, J. H. (2019). Quality over Quantity: A Lineage-Survival Strategy of Elite Families in Premodern Korea. *Social Science History*, 43 (1), 31–61. https://doi.org/10.1017/ssh.2018.38

Li, H., Zhang, J. & Zhu, Y. (2008). The Quantity-Quality Trade-off of Children in a Developing Country: Identification Using Chinese Twins. *Demography*, 45, 223–243. https://doi.org/10.1353/dem.2008.0006

Li, N., & Zhen, M. (2015). Fenjia xichan, caifu chongji yu shengyu xingwei [Partible Inheritance, Birth Order, and Fertility: Historical Investigation from the South-eastern Village in Qing and Republican China]. *Economic Research Journal*, 2, 145–159.

Lind, J. T., & Mehlum, H. (2010). With or Without U? The Appropriate Test for a U-shaped Relationship. *Oxford Bulletin of Economics and Statistics*, 72(1), 109–118. https://doi.org/10.1111/j.1468-0084.2009.00569.x

Liu, H. (2014). The Quality-Quantity Trade-off: Evidence from the Relaxation of China's One-Child Policy. *Journal of Population Economics*, 27(2), 565–602. https://doi.org/10.1007/s00148-013-0478-4

Liu, T. (1985). The Demographic Characteristics of Two Clans in Hsiao-shan, 1650–1850. In S. B. Hanley & A. P. Wolf (Eds.), *Family and Population in East Asian History* (pp. 13–61). Stanford, CA: Stanford University Press.

Liu, T. (1992). Mingqing Shiqi Jiazurenkou Yu Shehuijingji [*Lineage Populations and Socio-economic Changes in the Ming-Ch'ing Periods*]. Taipei: Institute of Economics, Academia Sinica.

Miyazaki, I. (1981). *China's Examination Hell: The Civil Service Examinations of Imperial China*. New Haven and London: Yale University Press.

Mokyr, J., & Voth, H. (2010). Understanding Growth in Europe, 1700–1870: Theory and Evidence. In S. Broadberry, & K. H. O'Rourke (Eds.), *The Cambridge Economic History of Modern Europe: Volume 1, 1700–1870* (pp. 8–42). Cambridge, UK: Cambridge University Press.

Ogilvie, S., Edwards, J., & Küpker, M. (2022). Economically Relevant Human Capital or Multi-purpose Consumption Good? Book Ownership in Pre-modern Württemberg. *Explorations in Economic History*, *83*, 101418. https://doi.org/10.1016/j.eeh.2021.101418

Pluzhnikov, A., Nolan, D. K., Tan, Z., McPeek, M. S., & Ober, C. (2007). Correlation of Intergenerational Family Sizes Suggests a Genetic Component of Reproductive Fitness. *The American Journal of Human Genetics*, 81(1), 165–169. https://doi.org/10.1086/518446

Ponczek, V., & Souza, A. P. (2012). New Evidence of the Causal Effect of Family Size on Child Quality in a Developing Country. *Journal of Human Resources*, 47 (1), 64–106. https://doi.org/10.1353/jhr.2012.0006

Qian, N. (2005). Quantity-Quality: The Positive Effect of Family Size on School Enrolment in China. *Working paper, Department of Economics, Brown University.*

Rawski, E. (1979). Education and Popular Literacy in Ch'ing China. Ann Arbor, MI: The University of Michigan Press.

Rosenzweig, M.R., & Wolpin, K.I. (1980). Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment. *Econometrica*, 48 (1), 227–240. https://doi.org/10.2307/1912026

Rosenzweig, M.R., & Zhang, J. (2006). Do Population Control Policies Induce More Human Capital Investment? Twins, Birthweight, and China's "One Child" Policy. *IZA Discussion Paper No. 2082*. Institute for the Study of Labor, Bonn, Germany.

Schofield, R. S. (1968). The Measurement of Literacy in Pre-industrial England. In J. Goody (Ed.), *Literacy in Traditional Societies*. Cambridge, UK: Cambridge University Press.

Shiue, C.H. (2016). A culture of kinship: Chinese genealogies as a source for research in demographic economics. *Journal of Demographic Economics*, 82 (4), 459–482. https://doi.org/10.1017/dem.2016.24

Shiue, C.H. (2017). Human capital and fertility in Chinese clans before modern growth. *Journal of Economic Growth*, 22 (4), 351–396. https://doi.org/10.1007/s10887-017-9148-9

Song, X., Campbell, C.D., & Lee, J.Z. (2015). Ancestry matters: Patrilineage growth and extinction. *American Sociological Review*, 80 (3), 574–602. https://doi.org/10.1177/0003122415576516

Stearns, S. C. (1989). Trade-offs in Life-history Evolution. *Functional Ecology*, 3 (3), 259–268. https://doi.org/10.2307/2389364

Stock, J., & Yogo, M. (2005). Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments. In D. W. K. Andrews, & J. H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (pp. 109–120).

Strassmann, B. I., & Gillespie, B. (2002). Life–history Theory, Fertility and Reproductive Success in Humans. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269 (1491), 553–562. https://doi.org/10.1098/rspb.2001.1912

Tan, H.R. (2019). More is Less? The Impact of Family Size on Education Outcomes in the United States, 1850–1940. *Journal of Human Resources*, 54 (4), 1154–1181. https://doi.org/10.3368/jhr.54.4.0517.8768R1

Telford, T.A. (1995). Fertility and Population Growth in the Lineages of Tongcheng County, 1520–1661. In S. Harrell (Ed.), *Chinese Historical Microdemography* (pp. 48–93). Berkeley, CA: University of California Press.

Waltner, A. (1990). *Getting an Heir: Adoption and the Construction of Kinship in Late Imperial China*. Honolulu, HI: University of Hawaii Press.

Williams, G. C. (1966). *Adaptation and Natural Selection.* Princeton, NJ: Princeton University Press.

Wolf, A. P. (1984). Fertility in Prerevolutionary Rural China. *Population and Development Review*, 10(3), 443–470. https://doi.org/10.2307/1973515

Wolf, A. P. (2001). Is There Evidence of Birth Control in Late Imperial China? *Population and Development Review*, 27(1), 133–154. https://doi.org/10.1111/j.1728-4457.2001.00133.x

Wolf, A. P., & Huang, C. (1980). *Marriage and Adoption in China, 1845–1945*. Stanford, CA: Stanford University Press.

Zelin, M. (2009). The Firm in Early Modern China. *Journal of Economic Behavior and Organization*, 71(3), 623–637. https://doi.org/10.1016/j.jebo.2009.03.002

Zhang, L. (2013). Legacy of success: office purchase and state-elite relations in Qing China. *Harvard Journal of Asiatic Studies*, 73(2), 259–297. https://doi.org/10.1353/jas.2013.0020.

Zhao, Z. (2001). Chinese Genealogies as a Source for Demographic Research: A Further Assessment of Their Reliability and Biases. *Population Studies*, 55 (2), 181–193. https://doi.org/10.1080/00324720127690

Online Supplementary Appendices

Celebrating legacy: The intergenerational transmission of reproduction and human capital in Ming–Qing Chinese families

Sijie Hu

School of Economics, Renmin University of China

159 Zhongguancun Street, Beijing, China

Email: sijiehu@ruc.edu.cn

ORCID ID: 0000-0002-3680-8592

Online Appendix A: Additional tables and figures

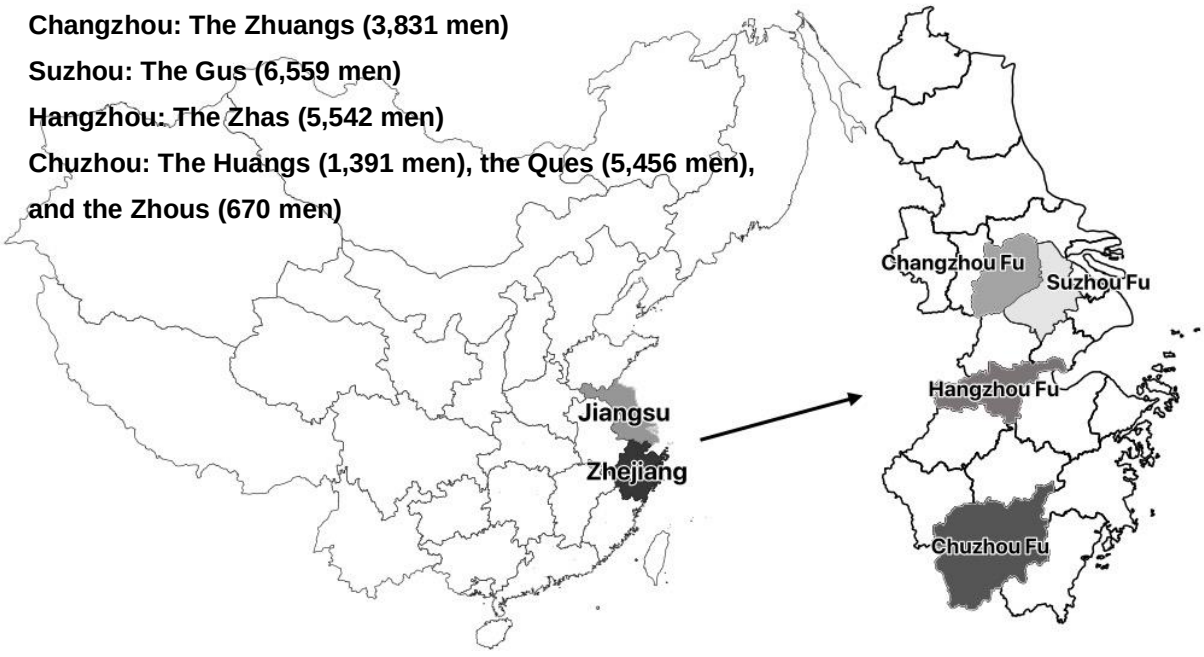

Figure A1 Map of Jiangsu and Zhejiang, China in 1820.

Source: CHGIS 2007.

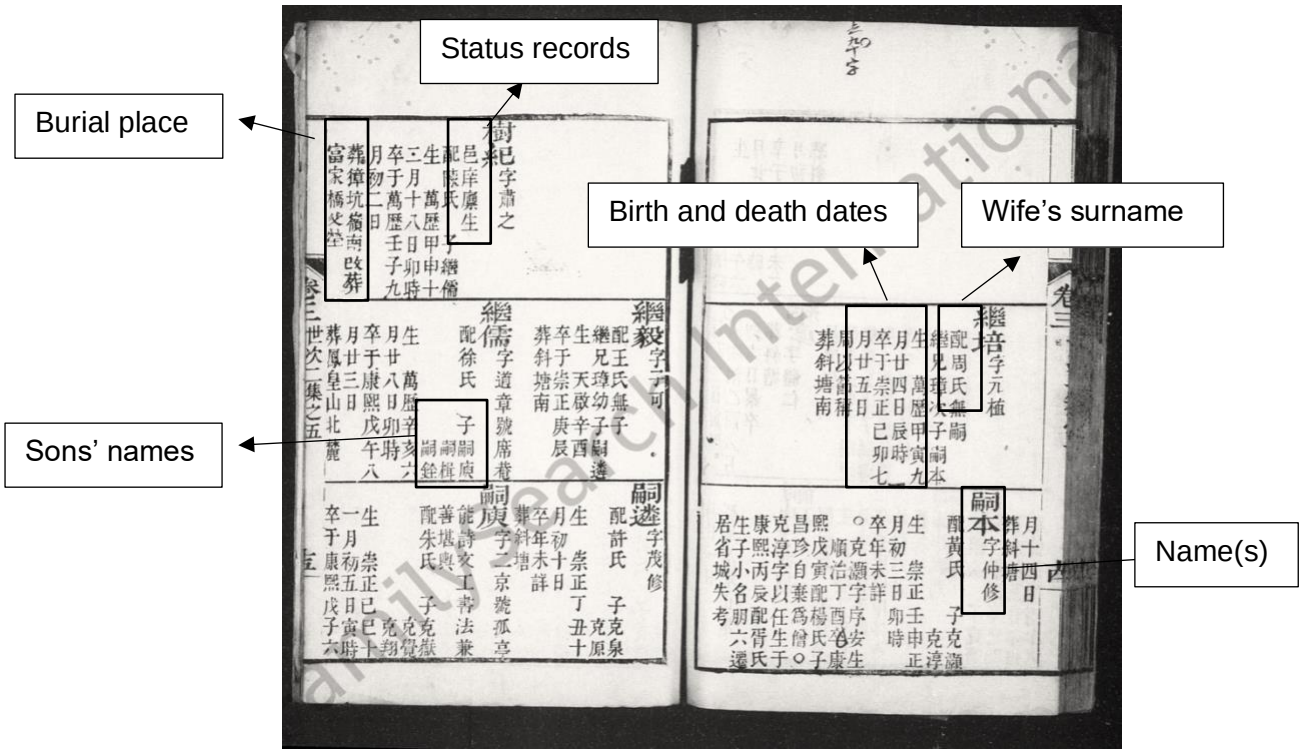

Figure A2 Example of a page from a Chinese genealogical book.
Source: Familysearch.org.

Table A1 Summary Statistics

Statistics	N	Mean	Std.	Min	Max
Main sample					
Number of grandsons	12,741	2.05	2.83	0	50
Number of great-grandsons	7,366	3.29	5.53	0	88
Number of great-great-grandsons	4,398	5.18	10.49	0	194
Number of sons	23,449	1.19	1.37	0	12
Rank	23,449	1.25	0.94	1	7
Number of marriages	23,449	0.84	0.67	0	10
Married	23,449	0.71	0.45	0	1
Degree	23,449	0.07	0.25	0	1
Number of brothers (Father's sons)	23,281	2.69	1.56	1	12
Firstborn	23,449	0.54	0.5	0	1
Adoptee	23,449	0.07	0.26	0	1
Mother's marital status	23,449	0.12	0.32	0	1
Father's grade	23,281	1.48	1.27	1	7
Father's marriages	23,284	1.21	0.6	0	10
Grandfather's grade	23,142	1.66	1.47	1	7
Grandfather's marriages	23,143	1.27	0.75	0	10
Grandfather's sons	23,142	2.82	1.63	1	12
Migration	23,449	0.02	0.13	0	1
Survival to adulthood	23,449	0.92	0.27	0	1
Birth cohort	23,449	1757.7	83.74	1350	1900
Lineage	23,449	3.89	1.54	1	6
Lifespan	11,791	50.1	17.5	1	105
Father's lifespan	15,498	57.17	14.14	1	105
Zi	23,449	0.48	0.5	0	1
Hao	23,449	0.09	0.29	0	1
Father's social status	23,281	2.16	2.93	1	16
Grandfather's social status	23,142	2.58	3.40	1	16
Mother's age at first son's birth	11,713	24.08	5.86	12	49
Daughter sample					
Husband class	4,122	1.13	0.65	0	3
Number of sisters	4,122	2.31	1.32	1	9
Number of brothers	4,122	2.20	1.62	0	12
Firstborn	4,122	0.58	0.49	0	1
Father's birth cohort	4,122	1732.28	74.52	1350	1900
Father's marriages	4,122	1.33	0.77	0	10
Father's grade	4,122	2.02	1.76	1	7
Grandfather's grade	4,090	2.34	2.00	1	7
Mother's marital status	4,122	0.05	0.21	0	1
Lineage	4,122	4.67	0.47	1	6
Father's lifespan	3,224	58.75	14.51	18	105

Notes: 1. The "number of brothers" includes the male individual himself. 2. Father's grade and marriages, and grandfather's grade, marriages, and sons all denote the stepfather's and step-grandfather's if the individual is an adoptee. 3. "Father's age at death" denotes all individuals' biological father's age at death.

Table A2 Results of the hump shape test for the number of male descendants (Lind & Mehlum, 2010)

Dependent Variable	Extreme Point	Lower Bound Slope (t-value)	Upper Bound Slope (t-value)	Overall Test t-value
Number of grandsons	7.25	0.64*** (26.36)	-0.42*** (-6.66)	6.66
Number of great-grandsons	7.29	0.48*** (14.19)	-0.31*** (-3.71)	3.71
Number of great-great-grandsons	6.36	0.47*** (9.96)	-0.41*** (-3.82)	3.82

Notes: 1. The test follows the methodology proposed by Lind & Mehlum (2010). 2. The t-values are reported in parentheses. 3. *p<0.1; **p<0.05; ***p<0.01.

Table A3 Relationship between family size and marital status, OLS regression

	Dependent Variable: Married				
	(1)	(2)	(3)	(4)	(5)
Brothers	0.010***	0.009***	0.019***	0.013***	0.012***
	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
Firstborn			0.064***	0.060***	0.060***
			(0.006)	(0.006)	(0.006)
Adoptee			0.031***	0.036***	0.036***
			(0.010)	(0.01)	(0.01)
Mother's marital status			0.035***	-0.015	-0.017
			(0.009)	(0.011)	(0.011)
Father's grade 2				0.089***	0.083***
				(0.022)	(0.023)
Father's grade 3				0.109***	0.107***
				(0.015)	(0.016)
Father's grade 4				0.148***	0.151***
				(0.015)	(0.016)
Father's grade 5				0.08***	0.084***
				(0.024)	(0.025)
Father's grade 6				0.091***	0.102***
				(0.017)	(0.019)
Father's grade 7				0.107***	0.115***
				(0.021)	(0.024)
Father's marriages				0.032***	0.032***
				(0.007)	(0.007)
Grandfather's grade 2					-0.004
					(0.023)
Grandfather's grade 3					0.042***
					(0.015)
Grandfather's grade 4					0.014
					(0.017)
Grandfather's grade 5					-0.047*
					(0.027)
Grandfather's grade 6					-0.052***
					(0.019)
Grandfather's grade 7					0.026
					(0.022)
Grandfather's marriages					-0.001
					(0.005)
Grandfather's sons					-0.002
					(0.002)
Migration	N	Y	Y	Y	Y
Survival	N	Y	Y	Y	Y
Birth cohort FE	N	Y	Y	Y	Y
Lineage FE	N	Y	Y	Y	Y
N	23,281	23,281	23,281	23,281	23,281
Pseudo-R^2	0.001	0.261	0.265	0.245	0.248

Notes: 1. If the individual is an adoptee, then "Father's grade" and "Grandfather's grade" denote his stepfather's and step-grandfather's social status. Grade 1 is the reference group. 2. Robust standard errors are in parentheses, clustered on fathers. 3. *p<0.1; **p<0.05; ***p<0.01.

Table A4 Relationship between family size and degree, OLS regression

	Dependent Variable: Degree				
	(1)	(2)	(3)	(4)	(5)
Brothers	0.015***	0.015***	0.019***	0.001	-0.0002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Firstborn			0.028***	0.018***	0.017***
			(0.004)	(0.003)	(0.003)
Adoptee			0.016**	0.015**	0.010*
			(0.007)	(0.006)	(0.006)
Mother's marital status			0.048***	-0.022***	-0.022***
			(0.008)	(0.008)	(0.008)
Father's grade 2				0.113***	0.095***
				(0.020)	(0.019)
Father's grade 3				0.252***	0.198***
				(0.020)	(0.020)
Father's grade 4				0.239***	0.186***
				(0.017)	(0.017)
Father's grade 5				0.298***	0.210***
				(0.032)	(0.032)
Father's grade 6				0.439***	0.341***
				(0.026)	(0.027)
Father's grade 7				0.652***	0.529***
				(0.040)	(0.042)
Father's marriages				0.016***	0.011***
				(0.005)	(0.005)
Grandfather's grade 2					0.033**
					(0.015)
Grandfather's grade 3					0.089***
					(0.014)
Grandfather's grade 4					0.048***
					(0.012)
Grandfather's grade 5					0.122***
					(0.030)
Grandfather's grade 6					0.135***
					(0.022)
Grandfather's grade 7					0.262***
					(0.033)
Grandfather's marriages					0.001
					(0.005)
Grandfather's sons					-0.001
					(0.001)
Migration	N	Y	Y	Y	Y
Survival	N	Y	Y	Y	Y
Birth cohort FE	N	Y	Y	Y	Y
Lineage FE	N	Y	Y	Y	Y
N	23,281	23,281	23,281	23,281	23,281
Pseudo-R^2	0.009	0.088	0.094	0.284	0.308

Notes: 1. If the individual is an adoptee, then "Father's grade" and "Grandfather's grade" denote his stepfather's and step-grandfather's social status. Grade 1 is the reference group. 2. Robust standard errors are in parentheses, clustered on fathers. 3. *p<0.1; **p<0.05; ***p<0.01.

Table A5 Relationship between quantity and quality of sons, the subsample of men who had records of having mother's age at first son's birth

	Dependent Variable:			
	Married		Degree	
	(1)	(2)	(3)	(4)
Brothers	1.099***	1.058**	1.166***	0.928*
	(0.0298)	(0.0285)	(0.0504)	(0.0416)
Firstborn	1.606***	1.602***	1.398***	1.425***
	(0.0769)	(0.0785)	(0.131)	(0.174)
Adoptee	1.175	1.140	1.531	1.128
	(0.179)	(0.180)	(0.400)	(0.373)
Mother's marital status	1.286***	0.826*	2.167***	0.920
	(0.121)	(0.0906)	(0.318)	(0.214)
Father's grade 2		1.597*		6.581***
		(0.412)		(1.962)
Father's grade 3		3.175***		13.24***
		(0.829)		(3.342)
Father's grade 4		4.222***		12.27***
		(0.865)		(3.007)
Father's grade 5		1.849*		9.722***
		(0.624)		(3.422)
Father's grade 6		2.277**		25.33***
		(0.764)		(8.258)
Father's grade 7		9.656		112.4***
		(14.99)		(66.02)
Father's marriages		1.391***		1.181
		(0.106)		(0.173)
Grandfather's grade 2		1.045		1.936**
		(0.242)		(0.581)
Grandfather's grade 3		1.650***		4.092***
		(0.276)		(0.992)
Grandfather's grade 4		1.189		2.033***
		(0.193)		(0.529)
Grandfather's grade 5		0.789		7.460***
		(0.230)		(3.025)
Grandfather's grade 6		0.717		4.215***
		(0.213)		(1.305)
Grandfather's grade 7		1.492		5.314***
		(1.004)		(2.374)
Grandfather's marriages		1.037		1.147*
		(0.0612)		(0.0889)
Grandfather's sons		0.965*		0.922*
		(0.0205)		(0.0421)
N	11,683	11,606	10,961	10,891
Pseudo R^2	0.216	0.235	0.142	0.396

Notes: 1. If the individual is an adoptee, then "Father's grade" and "Grandfather's grade" denote his stepfather's and step-grandfather's social status. Grade 1 is the reference group. 2. All specifications are controlled for migration, survival to adulthood, birth cohort, and lineage fixed effects. 3. Robust standard errors are in parentheses, clustered on fathers. 3. *p<0.1; **p<0.05; ***p<0.01.

Table A6 Sensitivity analysis of instrument validity using the local-to-zero (LTZ) approach (Conley et al., 2012)

Prior Distribution Assumption	Estimates for the Plausibly Instrumented Variable	
	Lower Bound	Upper Bound
Normal (0, 0.01)	-1.78	1.62
Uniform (-0.05, 0.05)	-4.15	4.14
Uniform (-0.1, 0.1)	-8.34	8.29
Normal (0, 0.1)	-17.07	16.93

Note: The test follows the "Plausibly Exogenous" methodology proposed by Conley et al. (2012).

Online Appendix B: Tables for Robustness Checks

Table B1 Effects of sons on patrilineal male descendants, OLS regression

	Dependent Variable: ln (Number of male descendants+1) in								
	Gen.3	Gen.4	Gen.5	Gen.3	Gen.4	Gen.5	Gen.3	Gen.4	Gen.5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sons	0.368***	0.302***	0.318***	0.310***	0.238***	0.234***	0.305***	0.235***	0.231***
	(0.018)	(0.029)	(0.039)	(0.017)	(0.026)	(0.033)	(0.015)	(0.022)	(0.030)
$Sons^2$	-0.015***	-0.012***	-0.016***	-0.012***	-0.010**	-0.012**	-0.013***	-0.012***	-0.014***
	(0.003)	(0.005)	(0.006)	(0.003)	(0.004)	(0.005)	(0.003)	(0.004)	(0.005)
Grade 2							0.323***	0.208***	0.187**
							(0.046)	(0.066)	(0.095)
Grade 3							0.215***	0.290***	0.197***
							(0.033)	(0.052)	(0.068)
Grade 4							0.306***	0.222***	0.228***
							(0.032)	(0.053)	(0.070)
Grade 5							0.132***	0.126*	0.126
							(0.041)	(0.071)	(0.106)
Grade 6							0.236***	0.202***	0.276***
							(0.039)	(0.057)	(0.092)
Grade 7							0.540***	0.732***	0.756***
							(0.066)	(0.093)	(0.122)
Marriages							0.039***	0.085***	0.121***
							(0.011)	(0.024)	(0.027)
Firstborn	N	N	N	Y	Y	Y	Y	Y	Y
Adoptee	N	N	N	Y	Y	Y	Y	Y	Y
Mother's marital status	N	N	N	Y	Y	Y	Y	Y	Y
Migration	N	N	N	Y	Y	Y	Y	Y	Y
Survival to adulthood	N	N	N	Y	Y	Y	Y	Y	Y
Birth cohort FE	N	N	N	Y	Y	Y	Y	Y	Y
Lineage FE	N	N	N	Y	Y	Y	Y	Y	Y
N	12,734	7,366	4,398	12,734	7,366	4,398	12,734	7,366	4,398
R^2	0.219	0.111	0.081	0.365	0.324	0.318	0.381	0.340	0.331

Notes: 1. Robust standard errors are in parentheses, clustered on fathers. 2. Grade 1 is the reference group. 3. *p<0.1; **p<0.05; ***p<0.01.

Table B2 Effects of sons on patrilineal male descendants by lineage type

	Dependent Variable: Number of male descendants in					
	Gen.3	Gen.4	Gen.5	Gen.3	Gen.4	Gen.5
	Common Lineages			Elite Lineages		
	(1)	(2)	(3)	(4)	(5)	(6)
Sons	2.026***	1.757***	1.917***	1.781***	1.526***	1.426***
	(0.081)	(0.128)	(0.198)	(0.048)	(0.050)	(0.063)
$Sons^2$	0.950***	0.954***	0.939***	0.965***	0.976***	0.979***
	(0.005)	(0.009)	(0.011)	(0.004)	(0.005)	(0.006)
N	3,470	1,876	1,092	9,264	5,490	3,306
Pseudo R^2	0.084	0.050	0.039	0.153	0.124	0.103

Notes: 1. The coefficients are the incidence rate ratios (IRR) for the negative binomial model. 2. All specifications shown include grade, marriages, firstborn, adoptee, mother's marital status, migration, survival to adulthood, birth cohort FE, and lineage FE. 3. Robust standard errors are in parentheses, clustered on fathers. 4. *p<0.1; **p<0.05; ***p<0.01.

Table B3 Effects of sons on patrilineal male descendants, conditioning on lifespan

	Dependent Variable: Number of male descendants in		
	Gen.3	Gen.4	Gen.5
	(1)	(2)	(3)
Sons	1.750***	1.574***	1.606***
	(0.045)	(0.055)	(0.078)
$Sons^2$	0.963***	0.969***	0.963***
	(0.004)	(0.005)	(0.006)
Lifespan	1.008***	1.005***	1.002
	(0.001)	(0.001)	(0.002)
N	8,685	6,015	3,746
Pseudo R^2	0.127	0.096	0.083

Notes: 1. The coefficients are the incidence rate ratios (IRR) for the negative binomial model. 2. All specifications shown include grade, marriages, firstborn, adoptee, mother's marital status, migration, survival to adulthood, birth cohort FE, and lineage FE. 3. Robust standard errors are in parentheses, clustered on fathers. 4. *p<0.1; **p<0.05; ***p<0.01.

Table B4 Effects of sons on patrilineal male descendants, with alternative independent variable

	Dependent Variable: Number of male descendants in		
	Gen.3	Gen.4	Gen.5
	(1)	(2)	(3)
2 sons	1.788***	1.498***	1.465***
	(0.043)	(0.059)	(0.085)
3 sons	2.560***	1.958***	1.831***
	(0.072)	(0.088)	(0.115)
4 sons	3.594***	2.645***	2.507***
	(0.118)	(0.147)	(0.216)
5 sons	4.652***	3.136***	2.640***
	(0.230)	(0.245)	(0.324)
6 sons	5.154***	3.251***	2.498***
	(0.340)	(0.370)	(0.358)
7 sons	5.830***	4.310***	3.360***
	(0.560)	(0.768)	(0.636)
8 sons	4.479***	2.174**	2.275*
	(0.599)	(0.661)	(0.973)
9 sons	5.954***	2.410***	1.808
	(1.207)	(1.045)	(1.139)
≥ 10 sons	4.593***	3.151***	1.232
	(1.373)	(1.214)	(0.537)
N	12,734	7,366	4,398
Pseudo R^2	0.120	0.089	0.073

Notes: 1. The coefficients are the incidence rate ratios (IRR) for the negative binomial model. 2. All specifications shown include grade, marriages, firstborn, adoptee, mother's marital status, migration, survival to adulthood, birth cohort FE, and lineage FE. 3. Robust standard errors are in parentheses, clustered on fathers. 4. *p<0.1; **p<0.05; ***p<0.01.

Table B5 Effects of sons on patrilineal male descendants, conditioning on female effects

	Dependent Variable: Number of male descendants in					
	Gen.3	Gen.4	Gen.5	Gen.3	Gen.4	Gen.5
	(1)	(2)	(3)	(4)	(5)	(6)
Sons	1.904***	1.613***	1.598***	2.001***	1.674***	1.673***
	(0.045)	(0.054)	(0.074)	(0.052)	(0.076)	(0.106)
$Sons^2$	0.956***	0.968***	0.963***	0.949***	0.962***	0.957***
	(0.003)	(0.004)	(0.006)	(0.004)	(0.007)	(0.009)
Father-in-law's	1.282***	1.325***	1.288**	1.259***	1.364***	1.277**
social class	(0.067)	(0.092)	(0.126)	(0.086)	(0.110)	(0.155)
N	12,734	7,366	4,398	10,404	5.916	3.473
Pseudo R^2	0.120	0.090	0.073	0.113	0.082	0.070

Notes: 1. The coefficients are the incidence rate ratios (IRR) for the negative binomial model. 2. The specifications shown in columns 1 to 3 include controls for grade, marriages, firstborn, adoptee, mother's marital status, migration, survival to adulthood, birth cohort FE, and lineage FE; the specifications shown in columns 4 to 6 include all the aforementioned controls, except marriages. 3. Robust standard errors are in parentheses, clustered on fathers. 4. *p<0.1; **p<0.05; ***p<0.01.

Table B6 Reproduction-marriage and reproduction-degree relationships by period, logistic regression

	Dependent Variable:					
	Married	Degree	Married	Degree	Married	Degree
	Born in 1300–1650		Born in 1650–1800		Born in 1800–1920	
	(1)	(2)	(3)	(4)	(5)	(6)
Brothers	1.239*	0.959	1.072**	0.995	1.090***	0.931
	(0.153)	(0.069)	(0.032)	(0.037)	(0.028)	(0.054)
N	1,485	1,532	10,078	10,078	11,433	10,260
Pseudo R^2	0.358	0.334	0.129	0.383	0.276	0.323

Notes: 1. The coefficients are the odds ratio for the logistic model. 2. All specifications shown include father's and grandfather's grades, father's and grandfather's marriages, grandfather's sons, firstborn, adoptee, mother's marital status, migration, survival to adulthood, birth cohort FE, and lineage FE. 3. Robust standard errors are in parentheses, clustered on fathers. 4. *p<0.1; **p<0.05; ***p<0.01.

Table B7 Reproduction-marriage and reproduction-degree relationships by lineage type

	Dependent Variable:			
	Married	Degree	Married	Degree
	Common Lineages		Elite Lineages	
	(1)	(2)	(3)	(4)
Brothers	1.074**	0.975	1.073***	0.973
	(0.030)	(0.055)	(0.027)	(0.034)
N	7,321	6,453	15,643	15,781
Pseudo R^2	0.107	0.281	0.330	0.415

Notes: 1. The coefficients are the odds ratio for the logistic model. 2. All specifications shown include father's and grandfather's grades, father's and grandfather's marriages, grandfather's sons, firstborn, adoptee, mother's marital status, migration, survival to adulthood, birth cohort FE, and lineage FE. 3. Robust standard errors are in parentheses, clustered on fathers. 4. *p<0.1; **p<0.05; ***p<0.01.

Table B8 Reproduction-marriage and reproduction-degree relationships, conditioning on lifespan

	Dependent Variable:			
	Married		Degree	
	(1)	(2)	(3)	(4)
Brothers	0.971	0.944*	0.942*	0.918**
	(0.026)	(0.030)	(0.033)	(0.033)
Lifespan	1.048***	1.048***	1.025***	1.025***
	(0.003)	(0.003)	(0.003)	(0.003)
Father's lifespan		1.014***		1.008**
		(0.003)		(0.003)
N	11,431	9,314	11,449	9,377
Pseudo-R^2	0.275	0.265	0.388	0.373

Notes: 1. The coefficients are the odds ratios for the logistic regression model. 2. All specifications shown include father's and grandfather's grades, father's and grandfather's marriages, grandfather's sons, firstborn, adoptee, mother's marital status, migration, survival to adulthood, birth cohort FE, and lineage FE. 3. Robust standard errors are in parentheses, clustered on fathers. 4. *p<0.1; **p<0.05; ***p<0.01.

Table B9 Relationship between family size and rudimentary literacy, logistic regression

	Dependent Variable:			
	Zi		*Hao*	
	(1)	(2)	(3)	(4)
Brothers	1.026	0.965	1.011	0.989
	(0.022)	(0.033)	(0.029)	(0.036)
Lifespan		1.017***		1.027***
		(0.002)		(0.002)
Father's lifespan		1.004		1.004
		(0.003)		(0.003)
N	23,138	9,485	23,138	9,184
Pseudo-R^2	0.315	0.407	0.277	0.268

Notes: 1. The coefficients are the odds ratio for the logistic model. 2. All specifications shown include father's and grandfather's grades, father's and grandfather's marriages, grandfather's sons, firstborn, adoptee, mother's marital status, migration, survival to adulthood, birth cohort FE, and lineage FE. 3. Robust standard errors are in parentheses, clustered on fathers. 4. *p<0.1; **p<0.05; ***p<0.01.

Table B10 Relationship between sons' quantity and quality, conditional on father's and grandfather's social status

	Dependent Variable:			
	Married		Degree	
	(1)	(2)	(3)	(4)
Brothers	1.090***	0.941*	0.966	0.901***
	(0.0222)	(0.030)	(0.0269)	(0.031)
Father's status				
Status 2	1.386***	1.405**	4.294***	3.554***
	(0.138)	(0.233)	(0.691)	(0.666)
Status 3	3.492***	1.432	5.667***	3.028**
	(1.500)	(0.812)	(2.445)	(1.690)
Status 4	1.509	1.743	8.173***	5.998***
	(0.403)	(0.600)	(2.060)	(1.800)
Status 5	2.264*	1.103	5.536***	3.816**
	(1.023)	(0.618)	(2.941)	(2.144)
Status 6	2.739***	2.689***	10.107***	7.784***
	(0.490)	(0.663)	(1.618)	(1.366)
Status 7	3.186***	3.087***	12.050***	9.676***
	(0.459)	(0.613)	(1.891)	(1.653)
Status 8	3.127*	3.259**	22.123***	17.509***
	(2.095)	(1.764)	(7.320)	(6.555)
Status 9	1.910	2.537	1.907	4.290*
	(0.927)	(2.007)	(1.514)	(3.584)
Status 10	1.754	2.092*	11.868***	9.875***
	(0.621)	(0.903)	(3.580)	(3.279)
Status 11	1.636	1.977	8.454***	11.902***
	(0.575)	(1.405)	(3.402)	(6.681)
Status 12	2.482***	2.020	14.197***	11.793***
	(0.810)	(0.910)	(4.376)	(3.961)
Status 13	1.988***	2.205**	24.923***	21.743***
	(0.500)	(0.749)	(5.138)	(4.942)
Status 14	3.025***	6.109***	24.383***	20.838***
	(1.133)	(3.389)	(6.961)	(6.948)
Status 15	7.454***	5.908***	48.634***	41.538***
	(3.802)	(2.645)	(11.593)	(11.08)
Status 16	1.967	0.988	22.042***	10.629***
	(1.046)	(0.833)	(10.589)	(5.296)
Lifespan		1.048***		1.025***
		(0.003)		(0.003)
Father's lifespan		1.014***		1.008**
		(0.003)		(0.003)
N	23,106	9,314	22,304	9,372
Pseudo R^2	0.249	0.267	0.411	0.389

Notes: 1. The coefficients are the odds ratio for the logistic model. 2. All specifications shown include father's and grandfather's marriages, grandfather's status and sons, firstborn, adoptee, mother's marital status, migration, survival to adulthood, birth cohort FE, and lineage FE. 3. Father's status 1 and grandfather's status 1 are the reference groups. 4. Robust standard errors are in parentheses, clustered on fathers. 5. *p<0.1; **p<0.05; ***p<0.01.

Table B11 Impacts of family features on daughters' marriages, ordinal logistic regression

	Dependent Variable: Husband's social class			
	(1)	(2)	(3)	(4)
Number of sisters	1.101***	1.050	1.032	0.977
	(0.033)	(0.037)	(0.038)	(0.041)
Number of brothers	1.172***	1.045*	1.040	0.985
	(0.029)	(0.027)	(0.027)	(0.029)
Firstborn daughter		1.394***	1.363***	1.274**
		(0.120)	(0.120)	(0.127)
Father's grade				
Grade 2		2.314***	1.938***	1.843**
		(0.506)	(0.426)	(0.487)
Grade 3		4.424***	2.902***	3.065***
		(0.875)	(0.622)	(0.694)
Grade 4		5.682***	3.783***	4.322***
		(0.869)	(0.667)	(0.766)
Grade 5		4.044***	2.456**	5.072***
		(1.685)	(1.026)	(1.592)
Grade 6		8.812***	4.784***	5.561***
		(1.678)	(1.125)	(1.238)
Grade 7		15.57***	8.966***	7.974***
		(3.991)	(2.769)	(2.549)
Father's marriages		0.949	0.979	1.096
		(0.053)	(0.059)	(0.0682)
Mother's marital status			0.637*	0.610*
			(0.158)	(0.154)
Father's lifespan				1.004
				(0.004)
Grandfather's grade	N	N	Y	Y
Lineage FE	Y	Y	Y	Y
Father birth cohort FE	Y	Y	Y	Y
N	4122	4122	4090	3207
Pseudo R^2	0.061	0.121	0.129	0.131

Notes: 1. The coefficients are odds ratios. 2. Father's grade 1 and grandfather's grade 1 are the reference groups. 3. Robust standard errors are in parentheses. 4. *p<0.1; **p<0.05; ***p<0.01.

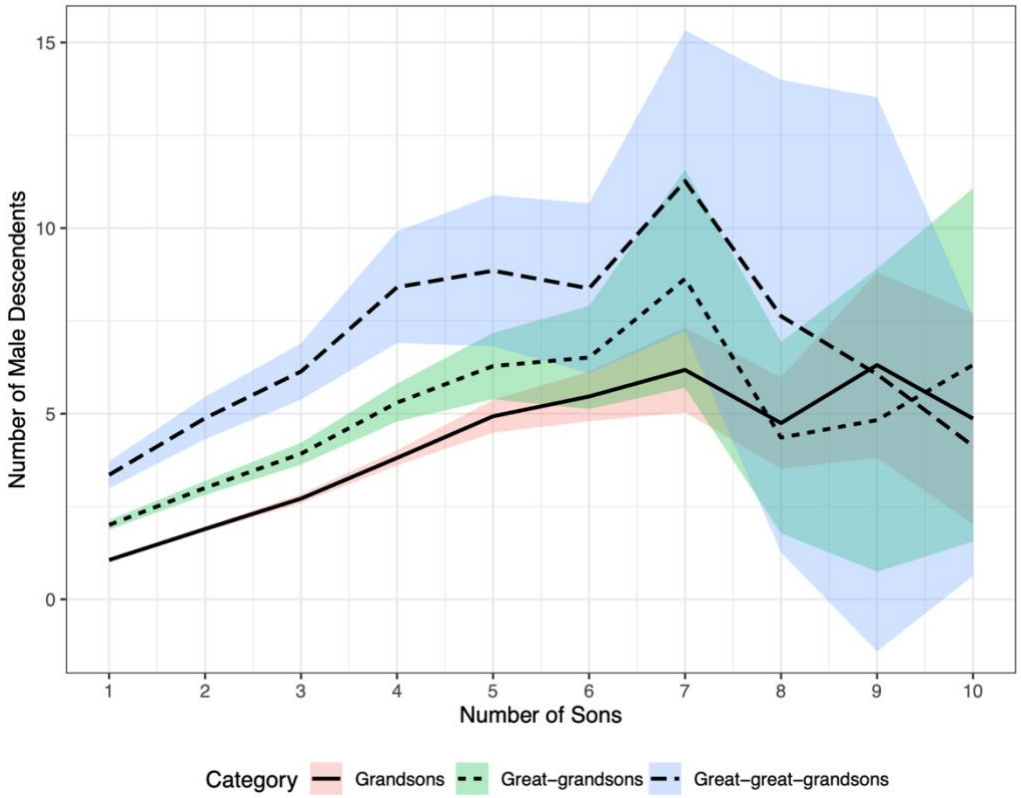

Figure B1 Predicted number of male descendants in the three generations at each level of reproduction after conditioning on all factors.

Notes: 1. The predicted values are calculated from the results in columns 1–3 of Table B3. 2. The shaded area represents the 95% confidence interval for the predicted curve.

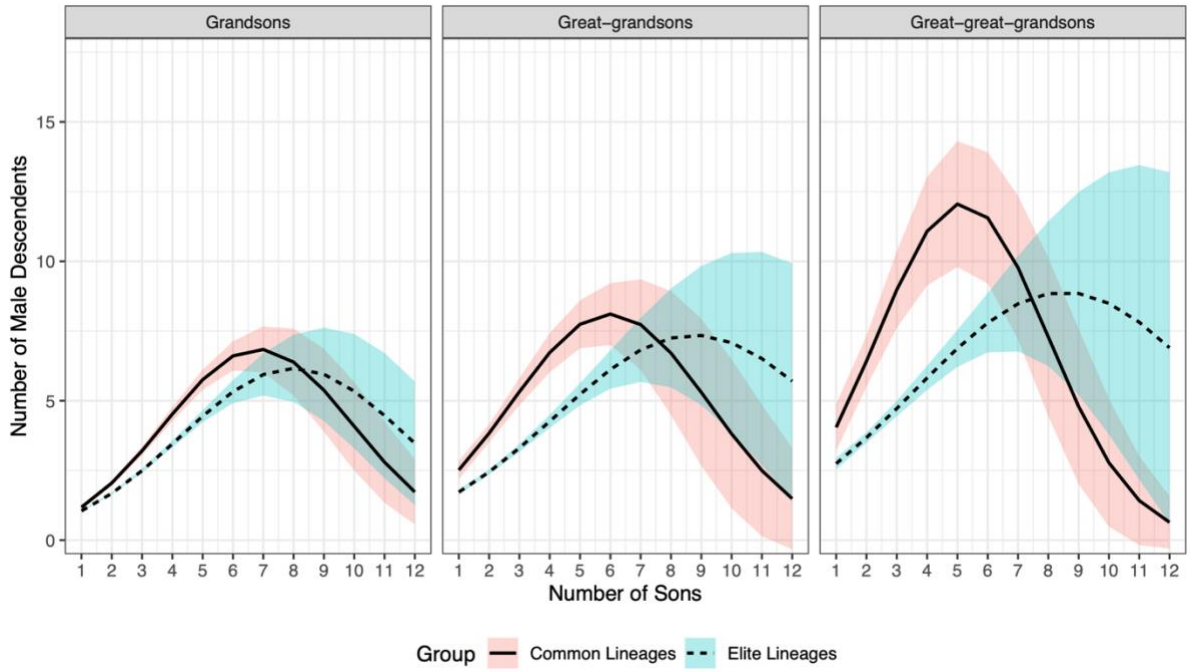

Figure B2 Predicted number of male descendants in the three generations at each level of reproduction by lineage type

Notes: 1. The predicted values are based on the results in Table B2. 2. The shaded area represents the 95% confidence interval for the predicted curve.

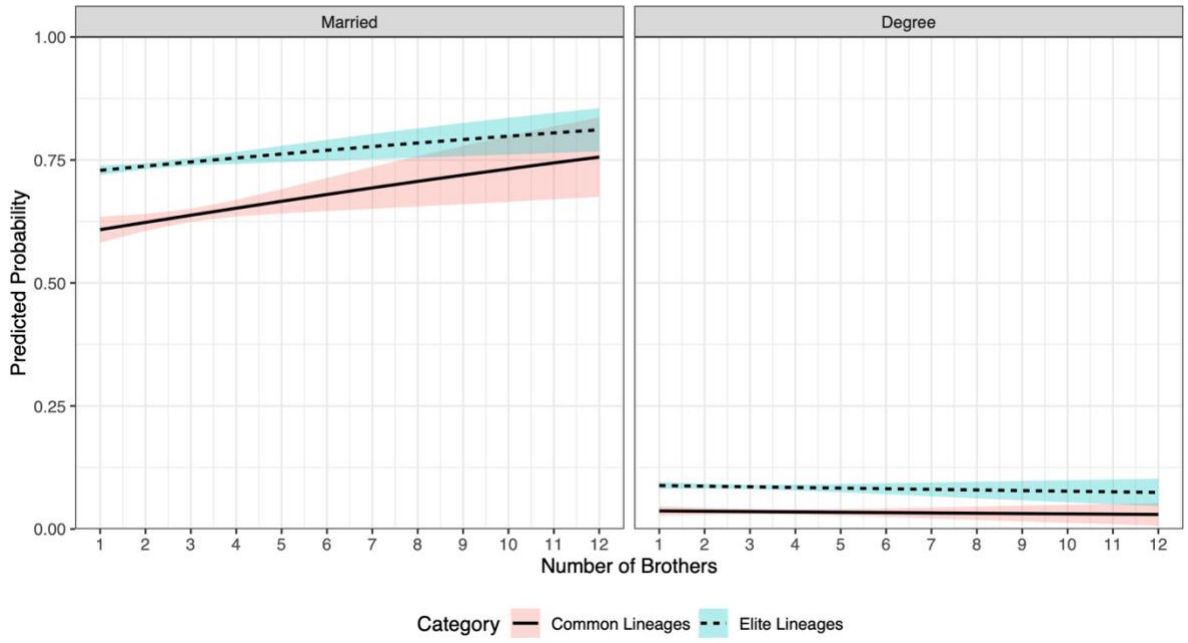

Figure B3 Predicted probability of being married and earning a degree for elite and common lineages.

Notes: 1. The predicted values are calculated from the results in Table B7. 2. The shaded area represents the 95% confidence interval for the predicted curve.