DISCUSSION PAPER SERIES

DP11614

A CULTURE OF KINSHIP: CHINESE GENEALOGIES AS A SOURCE FOR RESEARCH IN DEMOGRAPHIC ECONOMICS

Carol Hua Shiue

ECONOMIC HISTORY

ISSN 0265-8003

A CULTURE OF KINSHIP: CHINESE GENEALOGIES AS A SOURCE FOR RESEARCH IN DEMOGRAPHIC ECONOMICS

Carol Hua Shiue

Discussion Paper DP11614
Published 07 November 2016
Submitted 07 November 2016

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

A CULTURE OF KINSHIP: CHINESE GENEALOGIES AS A SOURCE FOR RESEARCH IN DEMOGRAPHIC ECONOMICS

Abstract

This paper discusses the use of Chinese genealogies for research on economic demography. I focus both on what is known about the genealogy as a data source, and what are the open questions for future research. Chinese genealogies contain individual level records at the individual level. With the publication of new catalogues and efforts to collect genealogies, the number of genealogies is even larger than previously thought, with most dating to the late Ming (1368-1644) and Qing (1644-1911) Dynasties. These records contain information about the Chinese population history, over a period for which there is no alternative source of information. Yet the source still remains largely unexploited. Although the work of transcribing the data is significant, and selection biases need to be carefully considered, preliminary analysis of the data for a sample of married men for Tongcheng County in Anhui Province suggests these data are a rich source of information for demographic and economics research.

JEL Classification: J10, N00

Keywords: Genealogical data, Chinese historical demography, lineage population, intergenerational linked data

Carol Hua Shiue - carol.shiue@colorado.edu
University of Colorado Boulder and CEPR

Acknowledgements
Thanks to Matthew Butner, Murat Iyigun, and Wolfgang Keller for comments and suggestions. Minghui Li, William Ridley, and Yuwei Jia provided excellent research assistance. Support from NICHD is gratefully acknowledged.

A Culture of Kinship:

Chinese Genealogies as a Source for Research in Demographic Economics

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

Abstract

This paper discusses the use of Chinese genealogies for research on economic demography. I focus both on what is known about the genealogy as a data source, and what are the open questions for future research. Chinese genealogies contain records at the individual level. With the publication of new catalogues and efforts to collect genealogies, the number of genealogies is even larger than previously thought, with most dating to the late Ming (1368-1644) and Qing (1644-1911) Dynasties. These records contain information about the Chinese population history, over a period for which there is no alternative source of information. Yet the source still remains largely unexploited. Although the work of transcribing the data is significant, and selection biases need to be carefully considered, preliminary analysis of the data for a sample of married men for Tongcheng County in Anhui Province suggests these data are a rich source of information for demographic and economics research.

* Thanks to Matthew Butner, Murat Iyigun, and Wolfgang Keller for comments and suggestions. Minghui Li and Yuwei Jia provided excellent research assistance. Support from NICHD is gratefully acknowledged.
Email: shiue@colorado.edu

1. Introduction

Microdata are important for the study of economic processes that take place within families and across social groups. National censuses or household surveys provide information at the individual and household level that have been key to studying diverse topics in fertility, mortality, marriage, and social mobility (Ruggles, McCaa, Sobek, and Cleveland, 2015).[1] Contemporary IPUMS-International (Integrated Public Use Microdata Series) and IPUMS-USA, however, are based on independent surveys and either do not allow or severely limit the tracking of individuals across censuses—names, geographic residence, place of birth, and identifying information are suppressed to protect the privacy of those being surveyed.[2] By contrast, historical microdata are not restrained by confidentiality rules. Recent efforts to link information over census surveys (at the level of names, addresses, and other information that is available) have considerably expanded the amount of microdata for the 18th and 19th centuries—for the United States from 1790 until 1930, and for Britain from 1851 to 1911 (Ruggles, 2014 and 2006).

In this article, I provide an update to sources of Chinese genealogies and discuss new ways that genealogies can contribute to recent research areas in economics. This is important particularly in China, where, like many other countries, national surveys started relatively late—the first Chinese nation-wide household census was carried out only in 1982 (Wang 2000). Chinese genealogies are valuable from at least two perspectives. First, they are the only source that provides individual and household data over a broad area that includes all the major provinces of China, the so-called China Proper area—which has persisted in nearly the same outline from the Ming Dynasty (1368-1643) and the Qing Dynasty (1644-1911) until today. This includes core areas that have been central to some of the most important debates of the last 15 years in the economic history of China. For example, much of the discussion on the long-run divergence in economic growth in China vis-à-vis the West (Pomeranz, 2000) have focused on the more

[1] To the extent that some of the variables overlap, IPUMS-International also contains harmonized cross-country data. In the United States, the first county level population and agricultural statistics appear in 1790—a relatively early date compared to availability elsewhere in the world—and a 1-in-1000 sample of household and person records in begun in 1960.

[2] In addition, low levels of geography are suppressed and geographic areas that are identified are randomized. See https://international.ipums.org/international-action/faq.

advanced areas within China, the Yangzi Delta and surrounding areas (Li, 2000). Over the 18th and 19th centuries, the role of key cities like Shanghai and Guangzhou also became increasingly important, as the economic foundations of the modern Chinese economy are traced to the historical evolution of these nodes for foreign trade and domestic commerce (Keller, Li, Shiue, 2013; Keller, Santiago, Shiue 2016). The value of genealogical data, once culled, is that it can shed some light on these regions with respect to lineage structure, urban to rural migration, family formation, and investments in education.

Second, genealogies provide an unusually long-run temporal perspective that can be potentially exploited together with the regional data contained therein. In some cases, the records reach back to the 14th century, when historical data of this type, globally speaking, is scarce. When information was recorded, as during the reign of the first Ming emperor, Hongwu (1368-1398), who ordered a large-scale survey of land quality and population for tax accounting purposes, it was not often repeated (Ho, 1959). Genealogies, by contrast, were updated, even if the updates were retrospective in nature. We can find in these sources information on 10, 15, or 20 continuous generations. The long-run temporal scope of genealogies opens new areas of empirical research as they may relate to social and intergenerational mobility, or the organization of lineages. These possibilities have yet to explored in full.

The use of Chinese genealogies for research is not new. Yuan (1931) used the genealogy of a southern Chinese family to reconstruct life tables. In 1969, a conference held in Salt Lake City discussed "Scholastic Application of the Chinese Clan Genealogies.[3] Starting in the 1970's, samples from genealogical records were used to estimate population trends, the distribution of mortality, age-specific-fertility, and age-at-marriage in different families—updates of, as well as citations to, this literature can be found in Liu (1978), Telford (1986), Zhao (1997). A few studies also analyzed questions related to lineage formation. For example, Liu (1995) examined 99 men in three- generation families in order to study the likelihood that multiple generations lived under one roof. Self-described as a "small-scale" study, Harrell (1985) draws connections

[3] See Eberhard's (1969).

between the growth and segmentation of wealthy and poor branches of lineages and population growth in three lineages.

What is new, however, is the increased accessibility and availability of genealogies. Particularly with a growing interest in contemporary microdata research, genealogies offer a way to address new research questions with historical data. The rest of the paper is organized as follows. Section 2 presents a brief historical background to the original purposes of the genealogy. Sections 3 and 4 discuss the availability and differences in the types of genealogies that can be found. Section 5 gives an overview of the content, and in Section 6 I present information on data from a set of genealogies, from Tongcheng County in Anhui Province, that have been tabulated. Finally, Section 7 provides some examples of research questions that can be pursued with these data.

2. Origin and Purpose of Chinese Genealogies

Chinese genealogies follow the male line of descent, providing the names of the individuals and—to varying degrees—additional information of groups of people related by marriage, by birth, or who otherwise recognize a common ancestor. It is significant as a written record of a family's patrilineal history, meaningful primarily to the people whose forebears are mentioned in the document. By the Ming and Qing dynasties, the genealogical tradition had spread to a broad swath of society (Harrell, 1995). Although the genealogies were compiled privately, lineages followed a traditional template in terms of form and content, based on ritualistic elements important to ancestral worship and clan rules. The physical document itself was kept in ancestral halls, or in the residence of key lineage members, and not shared with outsiders. Anthropological studies have suggested that village settlements, at least in some areas of China, were related to the lineage (Freedman, 1966). The organization and decision-making leadership were also arranged according to the heads of the segments of the lineage (Liu, 1959). In many respects, lineage organization is distinct from Western forms of social organization. As emphasized in Greif (2006) and Greif and Tabellini (2015), initial differences (clan organization in China, versus city organization in Europe) can produce bifurcating patterns in the two societies.

The emphasis on the importance of ancestry and kinship began long before the Ming and Qing. Historians point to the early practice of ancestry veneration and sacrifices performed for ancestors seems to have intensified during the Shang Dynasty, 1766-1045 BC (Li, 1999, Lakos, 2010). Archeological evidence from that period suggests that elites practiced a form of necromancy in which people attempted to communicate with dead ancestors, as well as fairly involved sacrificial rituals to ancestral tombs (Fu, 2003). Markings on burned oracle bones–ox scapulae or turtle shells–were used in these rituals in order to reveal auspicious days for decision-making and to predict future events. Spirits of the dead, it was believed, were able to influence the future–and the rituals may have provided a sense of legitimacy and credibility to the actions of these early states.

Confucius (551-479 BC) moved away from shamanism, but retained the essential importance of kinship as an organizing principle of human relationships that was in place in the ancient dynasties. In the Western Zhou period (1087-222 BC), there were already references to the "law of kindred" or "lineage law" (*zongfa*) (Thatcher, 2004). Thatcher (2004) outlines a ruling house-lineage power continuum that envisions lineages sharing power with rulers, identifying sometime around 550 BC as the benchmark date when non-royal families began to form new surnames and founding lineages after having been given wealth and prestige through their service to the rulers—a development that suggests competing interests between the lineage and the state may have started early on.

Some of the earliest written genealogies probably date to before the Han to Sui Dynasties (206 BC - 220 AD), and were compiled by the royal families and clans. Although the original copies were destroyed, these genealogies have been referred to in later works. The first great historian of China, Sima Qian (145-87 BC)—who has been compared to the Greek Herodotus in terms of historical analytical method—claims in the *Historical Records* (*Shi Ji*) of having read genealogies dating to the Spring Autumn period (770 - 477 BC) (Lo, 1969; Martin, 2010).

Centuries later, many of the titles listed in the Tang (618-907) Dynastic History are clearly still of royal houses, prime ministers and officials; and it appears that the genealogical tradition

continued over the Song Dynasty (960-1279). While it is possible that genealogies of non-elites simply were not preserved, it seems more likely that it was not until the Ming that there was a turning point after which a larger number of genealogies from elites as well as ordinary citizens were compiled.

A couple of factors likely contributed to these trends. First, an attempt to mimic aristocratic behavior might have been behind the change. Still, the timing of the emergence is relevant and was likely related to the institutional changes of late imperial China. The Civil Service Examination system—the gateway into the upper strata of society—was always dominated by the political elite. But in comparison to prior dynasties, over the transition from the Ming to the Qing Dynasty the system became less discretionary and more open to men from families with no aristocratic or royal connections (Ho, 1962). In the context of these institutional shifts, upwardly mobile literati would have found it attractive to adopt a convention popular to royal houses and important families as a way of establishing their continued social, economic, and political presence over many generations.[4] A second factor that strengthened the genealogical tradition was a new Ming Dynasty policy requiring land and population registrations, which were in turn the basis of taxation quotas. Faure (1989, 2007) has suggested that genealogy played a role in the lineage members' claims of rights to settle on the land. In a tax collection system where local heads were delegated with the responsibility of collecting the required tax quota, lineage organization could play a fiscally functional role—not only in fulfilling tax obligations, but also in evading taxes.

In sum, the popular convention of putting one's family members into a written compilation is emblematic of an interest that many people had in their family origins, lineage continuity, and intra-lineage member connections. Beyond cultural interest, lineage organization may also have had practical functions, including village settlement rights, sheltering land-holdings from taxation, defense, and public goods provision within the lineage. In addition, the late imperial

[4] New ruling families, or new settlement families, may have had ulterior objectives, e.g. to create a connection to royalty or to lay claim to a right to resources. It appears that the existence of the genealogical document tended to carry credibility, and some tried to take advantage of that fact (Kurz, 2014).

state had by this time settled into a mode of governance such that there was the need for lineages to organize local defense in order to settle inter-lineage conflicts. Membership in a lineage group could make allegiances and responsibilities clearer and more explicit. Given that genealogies were costly to print and produce, the argument that they served cultural and practical purposes is plausible.

An important implication is that while the biological tie was key to the father-son relationship, the practical considerations of the lineage as a group opens the door to voluntarily formed relationships. Anecdotally, there was the possibility of males not originally related by blood entering the lineage. This could occur if adult males married a daughter of the lineage and accepted her surname, or, as is sometimes recorded in the genealogy, through adoption. Conversely, biological relationships in the lineage tree could be cut off, perhaps as a result of out-migration from the region, or, voluntary household division. This raises the broader issue of how "family" should be defined. The presence of marriage, male adoption, and household division could all be viewed as different strategies of household formation, and they show how membership in lineage was adaptable and somewhat flexible. Seen in this light, we need to recognize the voluntary aspect to membership in a lineage group.

Further research will allow a greater understanding of the full dimensions and the fluidity of genealogical membership. For example, since there are many genealogical compilations that are updates to existing genealogical titles, one could compare earlier and later editions to see if there were any revisions to the family tree over time, and whether there are clues as to what the nature of those revisions might have been. Examining these changes would help us understand the population included in genealogies.

At the turn of the 20th century, with the collapse of the last imperial dynasty and the abolition of the civil service examinations and other structures of the imperial period, it is not clear how much genealogical activity was maintained. The rise of the Chinese Communist Party and Mao's movement came with active efforts to break down the paternal authority embedded in lineage organization (Yang, 1959a and 1959b). Further research on the temporal and geographical scope

of currently available genealogies is also necessary to establish the extent to which this occurred, and the extent to which there is a revival of the genealogical tradition today.

3. Availability

Traditionally, genealogies were compiled and updated by members of the lineage and a limited number of copies were produced and kept in the hometown. The only way to know the full scope of publically available titles is to consider the union of holdings in different libraries and institutions around the world. Public holdings of genealogies exist around the world, including the Shanghai Library, the National Library of China, provincial libraries in China, the Genealogical Society of Utah, collections in Taiwan, Singapore, the Library of Congress, the Tōyō Bunko, as well as collections at Columbia University, Harvard University, Peking University, and other institutions. As of 2003, a new count revealed around 51,000 genealogies around the world (Wang, 2008), including some 7000 genealogies in private hands.[5] As far as the distribution of the largest of the public collections, the count showed over 11,000 genealogies available in Shanghai Library; over 6400 in the Genealogical Society of Latter Day Saints in Utah, around 1200 in the National Library of China, and an equal number in Hunan Library. Within these titles, a total of some 608 surnames, covering the most common surnames as well as many of the rare surnames, can be identified.[6] Since there are undoubtedly still more genealogies held in private hands, it is likely the number of titles will continue to rise as these works enter the public domain.

Concurrently, images of genealogies have been placed online. This has increased the number of records that can be searched by Province or County location, year, title, or family name. The website of FamilySearch.org has, for example, as of June 2016, over 3.2 million digital images of

[5] The difference can be attributed to the fact that in numerous cases lineages updated their genealogy over time, so for example, the genealogy with the title "Wang Lineage Genealogy" may have a couple of updated editions, spearheaded by different authors.
[6] A majority of the Chinese population is covered by just 100 of the more popular surnames.

genealogies from a total of 5,018 records, which span 775 years over 1239 to 2014.[7] Thus, increasingly, genealogies are being made more widely available in the form of scanned digital images, in addition to microfilm or microfiche, making it is possible to access the information in these sources more easily. However, the work of entering the information into usable data formats still remains to be done.

Although genealogies are available for major regions of China, there is regional heterogeneity in the number of titles that are available and differences in the holdings of each library. For example, the current collection of genealogies in Shanghai Library, which holds approximately one-third of the total genealogies publically available, contains 2,512 genealogies with the popular surname Wang, of which 822 are just from Zhejiang Province alone. The library also has 225 Wang genealogies from the nearby Jiangsu Province, 168 from Anhui Province, and 269 from Hunan; but only 38 Wang genealogies from Sichuan, 50 from Hubei, and 33 from Guangdong, and none from regions such as Inner Mongolia, Tibet, and Xinjiang. Some of the regional distribution of surnames may reflect geographic settlement patterns of particular clans and their migration patterns—for example, the surname Boerjijite is more likely to be found in Inner Mongolia. However, another force behind the observed geographical distribution of overall numbers of titles could be the influence of the extent of local interest in curating genealogical titles. A more comprehensive analysis of the distribution of available genealogies would help to clarify these issues, but since the number of genealogies is related to the strength of efforts in curating, an alternative strategy for assessing systematic patterns in the records is to consider differences in the distribution of contents within the currently available pool of compilations.

4. Types of genealogies

Among the titles that are available today, the terms *jiapu*, *zupu*, and *zongpu*, are all associated with genealogies. In addition, because of regional and idiosyncratic differences, it is difficult to make broad generalizations about content and coverage. However, as a first cut, the size, or the

[7] See https://familysearch.org/ for more details.

number of people mentioned in the genealogy can be compared. Telford (1986) suggests three classifications for genealogies, based on observed differences in size. Smaller “family” or “branch” genealogies tend to be relatively short in length, and cover a relatively small number of people, or a limited group of nuclear families, and there may be limited generational depth. Numerous branch genealogies might be connected to a larger lineage or clan, but on its own, branch genealogies tend to cover a number of people who lived in proximity to each other geographically and over time. Since one has to be literate to write the names of the members, the smaller the group, the more likely it would seem to be that these are selected towards more educated branches of the lineage. Only future research, however, can confirm the extent to which this may be true, and whether or not the sampling is significantly skewed towards more educated families, on average.

By contrast, the "lineage genealogy" could cover the descendants of several branches of the descent line, all of whom share a common surname and typically reside in neighboring villages or towns. Typically, the progenitor is the man who the members recognize as the one who first migrated or settled in a particular location. There tends to be more generational depth that can be found in these types of genealogies—upwards of a dozen generations appears to be not unusual, and thousands of individuals are included (Telford, 1986).

Finally, the largest type that is the “clan genealogy”, which can include upwards of ten thousand names, listed in tables that delineate the multiple lineages and branches sharing a common ancestor. Many people are included who could reside across various distant areas and adjacent provinces, and in the case of these clan genealogies there appears to be a tradeoff in the amount of detail included, so that less individual-level information is available.

5. Content of Genealogies

Although each genealogy was privately compiled, Ming and Qing genealogists were familiar with and adhered to popular rules of compilation: which are either stated at the beginning of the genealogy or implicitly adopted (Harrell, 1987). A convention in genealogies, including the large

clan genealogies that encompass many lines of descent, is to begin with the progenitor, after which the names of male members (who survived childhood) are included. Their relationships are mapped in an extensive diagram depicting the precise generational and relational links in a pedigree chart. In a separate section, the lineage genealogy will typically contain an entry for each male household, organized according to birth order of the male members. These entries may provide vital statistics on each individual in the household such as birth and death dates, and potentially data on the in-marrying wives, their surnames, vital statistics, and hometown origins. Not all genealogies contain these data in equal detail, however. There also exists small genealogies consisting of handwritten sheaves that may contain not much more than the pedigree chart.

Lineage genealogies that cover several hundred to a few thousand male descendants, and as Harrell (1995) and Telford (1995) have pointed out, are the most promising of all the genealogical types for demographic-economic study. These mid-sized genealogies provide details on the interfamilial relationships of each generation and across generations through a tree diagram, and contain a biographical section on men that includes not just vital statistics, but also additional demographic and socioeconomic information. In order to provide future generations with the ability to continue the rituals, compilers might also include a record of the location of graves, diagrams on ancestral halls, texts relating to grave worship, family rules of conduct, biographies of prominent members, a record of lineage lands, and an overall history of the family.

From the information given on family relationships and related variables of interest, one can construct a range of additional variables, for example, on whether the relationship to the father is adoptive, the order of birth, and the native place (county and village of origin) of in-marrying kin—this includes, in some genealogies, the spouses of daughters, and the wives of sons. The birth order of the sons is almost always present when vital statistics are available, and the birth order of the daughters is sometimes included as well. Since the descendants of a particular segment of the lineage oftentimes will tend to settle together residentially, residential segments can be identified. Some genealogies also provide the locations of internment. These location identifiers can help us track the spatial pattern of different members of the lineage and

movements over time, and future analysis of migration patterns using these data seems to be quite possible.

Despite the presence of similarities, there is heterogeneity across genealogical content. Some of the heterogeneity in content could be related to systematic differences in the functions of the lineage itself; even within China, there might have been significant differences in the extent to which lineages were a dominant form of social organization (Watson, 1982; Ebrey, 1983; Ebrey and Watson 1986; Cohen, 1990). A better understanding of the settlement patterns of different lineages and land use rights in different regions of China is another area that will benefit from more research in the future.

6. Sample Population—Tongcheng Genealogies

The key difference to note between the basic construction of genealogies and census data is that census data typically seek to record the observed population at a certain date, either at the time of registration or in retrospect. For applications where permanent earnings are preferred, such as in the estimate of intergenerational elasticity, this can be a problem as imposing the assumption that current earnings are equal to permanent earnings will typically result in life-cycle biases (Jäntti et al. 2006). By contrast, genealogical data presents one entry per person in biographical format, so each observation implicitly summarizes a lifetime observation. When the birth and death dates are given, the achievement listed in that individual's record can be considered the highest personal achievement over the lifetime of the individual, capturing an aspect that would have required many census observations as an alternative.

In terms of who was included in the genealogy, it is useful to distinguish between the groups that were eligible to be included in the genealogy, but not included because of inadvertent selection, and the groups that were not eligible due to overt selection. On the latter, according to Harrell (1995), it was the ritualistic aspect, not wealth and status, that dictated who was included in the genealogy. Genealogies tend to produce the most complete data around adult men in the lineage, since all men were ritualistically significant. Wives and mothers of the men are also

important, and there is often vital data on these women. Daughters, thus, will have more complete data in the genealogy of their husbands, rather than in the genealogy of their fathers. Other groups, for example, those who died young, were for the most part not considered ritualistically significant, explaining why there is less data on them, yet even here there are exceptions to this rule that vary across genealogies.[8]

Even if high status was not a requirement for membership in a lineage, status could still have been a contributing factor that made it more likely to be in the genealogy due to inadvertent selection. The concern is that the voluntary nature of genealogical data assembly could induce selection. Also, the retroactive updating of the genealogy might lead to recall bias, and there may also be survivor bias. Successful fathers may have had more numerous successful sons, who could better provide information on the lineage genealogy.

To investigate these issues, I use the tabulated data from one county in China, Tongcheng County of Anhui Province. These data are drawn from seven lineage genealogies. Summaries of the sample population are in Table A. Combining the seven lineages, the sample consists of over 9,787 married men, and a total of 11,378 marriages. Due to the fact that some men married more than once, there are more married women in the sample than there are married men. These couples produced a recorded 20,177 sons and 11,150 daughters (Table B). Although the birth order of the daughters can be seen in the genealogy, there are fewer daughters compared to sons. Potentially, as more genealogies become digitized, it may become possible to build up the data on daughters by following the genealogy of their husbands.

The socio-economic data consist of status-related positions, noteworthy achievements, degrees, and titles earned by the male lineage members, or the male relatives of in-marrying wives. These markers could include official titles, degrees, and other areas of socio-economic prominence.[9] The nature of the available socioeconomic information in genealogies differs from modern

[8] At the same time, sometimes the very young are observed, so there are exceptions. See Liu (1995) for studies that provide more detail on mortality rates.

[9] For further details on these 22 categories, see Shiue (2016), and Telford (1995).

notions of assets or income, but nevertheless, the status positions are correlated with financial well-being. In the case of the details on specific official titles or degrees received, the relative positions can be exactly ranked. The correlation may be imperfect, however, as merchants who purchased a vanity degree or who contributed to the charitable interests of the lineage would have likely been noted, but similarly wealthy merchants who did not make a status enhancing action might not have been similarly noted. In these instances, additional clues, such as whether or not the individual had multiple wives, or was the son or close kin of a person with status, could also be used to make inferences about the status of a person.

The multiple ranks within the national civil service examinations and the official titles and appointments in place during the Qing Dynasty were maintained uniformly with relatively minor variation. Thus, degrees and official ranks provide a clear ranking of status based on educational achievements and official positions. Although it is possible to obtain estimates of the earnings for these different categories of status, it may be preferable to assume a ranking across the various types of achievements and contributions that are most frequently mentioned. To the extent that it is status, rather than income that is recorded, genealogies share with the commonly used occupational index the notion that status not only can be ranked but are also effective proxies for measuring individual standing within the population distribution (Hauser and Warren, 1997; Hauser, Warren, Huang and Carter, 2000).

In the majority of cases, there is information on the name, the generation, vital statistics, birth order, marriage and children, but no mention of any status. A reasonable assumption is that if there are no notations of status, the lack of any information on an individual, other than vital statistics, implies that the individual had no status that was especially notable. The absence of notations of status thus also carries information about the individual. Table C gives demographic details for a linked sample of fathers and sons on a few additional variables by clan, which emerged from the information given in genealogies in Tongcheng County. The variable, "Education", is an index that equals 1 if the genealogy indicates that the individual participated at any level of the civil service examination system, and 0 otherwise. Given the examinations required very high levels of preparation, it is not surprising that the averages are very low.

This is an important observation because it confirms previous findings in, for example, Liu (1978) that there is scant information in genealogies on status. Given the high standards for what is considered significant status, this is what we would expect in a larger lineage with a broad socio-economic distribution. It also shows that that the majority of the people included in the genealogies were far from being elites or literate individuals with degrees and official ranks. Who were these men who had nothing noted next to their vital statistics, who had one wife over their lifetime, some of who died early, and, sometimes ended up with just one surviving son or none at all? These were most likely the craftsmen, the small peddlers, the farmers, and other commoners who made up the bulk of the population.

Thus, despite the fact that the lineage would have had to rely on the literacy and wealth of the literate individuals to compile the genealogy, a genealogy that covers thousands of people will include large numbers of non-elites and commoners. It is the latter, as a group, who in fact appear as the large majority population in a mid-sized genealogy. Smaller genealogies that focus only on specific branches of the clan may be more susceptible to sampling issues however.

Selection biases of genealogies

Because the Chinese genealogy is based on the male line, it may seem impossible to use it for reconstructing the overall population profile. Nevertheless, the concerns about the selection biases of genealogies may be addressed in a few different ways. Telford (1990) suggested that one approach is to compare genealogical data with other data in order to check for the external validity of sample averages, while noting the differences. One set of statistics that offers a more complete record of the population are the Qing population registers. These data should be more representative of the population being recorded because they were a product of the Qing Eight Banner System registration system.[10] These data are only available for areas in China's

[10] The lands were were organized under the Imperial Household Agency and the Jilin Military Yamen, an office in the General Office of the Eight Banner Command. See https://www.icpsr.umich.edu/icpsrweb/ICPSR/series/265. For the imperial household lineage, there are observations going back to the seventeenth century (Lee et al. 1993).

northeast, in today's Liaoning and Heilongjiang Provinces, but nevertheless, comparing the Tongcheng data with that of the Banner populations for 1774 to 1873, when the latter starts to become available, Telford (1990) finds a very similar variation in the probability of dying for different age categories across both of the two populations (see Telford, 1990, Figure 2). The level of mortality is generally recorded as being higher in the Tongcheng sample, especially for the 50 to 70 year olds. A plausible reason noted by Telford is the Taiping Rebellion (1850-64) affected Tongcheng but not Liaoning. More recently, Campbell and Lee (2002) compared data from genealogies of Liaoning to the household registers. They also find evidence of elevated mortality in genealogies, the opposite of what would be expected if simple demographic selection on more privileged and educated men was an important determinant of the genealogical sample.

For many research questions, it is not necessary to reconstruct the entire population distribution, but rather, to make comparisons within subsamples. To the extent that some individuals, such as unmarried men who died early, are less likely to be included in both population subsamples, the comparison of population samples still provides valuable insights. One can also go further and assess the results in terms of the direction of bias and bounds. For example, since education is one of the variables that can be found in the biographical section, this is another area where one can check to see if the reported measures are reasonable. The most systematic evidence on education in China during Ming-Qing is related to the state examinations. In particular, the number of licentiates (*sheng-yuan*), individuals that passed the initial state examination, was about 500,000 in the year 1700 (Elman, 2000), or roughly 0.3% of the population. In the Tongcheng sample, about 0.76% of the men around the year 1700 were licentiates. Accounting for women, children, and elderly indicates that the fraction of licentiates in Tongcheng was similar, or perhaps somewhat lower than that in China as a whole.

If we were to consider the highest degree holders (*jinshi*), the numbers for which were controlled by quotas, Ho (1962) reports that during the Qing in Anhui there were 41 *jinshi* per one million population, or, 0.0041 percent. The province of Anhui, it should be noted, was below the provincial average in terms of *jinshi* per capita in Qing China (Ho, 1962, p. 228). In comparison,

Tongcheng county in Anhui had 14 *jinshi* during the Qing as per the Tongcheng sample, which comes to about 0.045 percent of the population.[11] Thus, there are about ten times more *jinshi* in the Tongcheng sample than in Qing Anhui overall, but Tongcheng was also an important urban center that had a well-known reputation for producing famous literati of the Qing (Beattie, 1979).

While this suggests that lineage members covered in Tongcheng genealogies had a higher level of education than Anhui's population on average, clearly, *jinshi* were rare, with many parts of Anhui not producing a single *jinshi* over centuries. The strong influence of aggregation in these comparisons becomes clear when noting that a single prefecture could have as many as 1,004 *jinshi* during the Qing (Ho, 1962, p. 247). With seven counties to a prefecture, this means that the average county of that prefecture had 1,004/7 = 143 *jinshi* during the Qing, or an order of magnitude higher than the number in Tongcheng county. Therefore, while the number of men with the highest levels of human capital in Tongcheng was higher than in the local surrounding area, Tongcheng was not among the top human capital areas in China; rather, it was noteworthy at a local, perhaps provincial level.

Moreover, it is important to note that the variation in *jinshi* across lineages in the Tongcheng sample dwarfs the difference between the sample variation in *jinshi* in the genealogies and what we know about the population (see Table C). At the top of the list, the *Ma* lineage had 9 *jinshi* relative to 627 men, a ratio of 1.4%, whereas other lineages in the Tongcheng sample do not have a single *jinshi*. As a consequence, sample variation across lineages is far more important than sample composition of who is in a genealogy and who is not. Thus, while lineages with a *jinshi* may be more likely to have a genealogy, a count of the actual numbers of *jinshi* in the Tongcheng genealogies show that the percentages of these degree holders within the clan population are within the bounds of the minimum and maximum that were produced in an

[11] There are 8,291 married men during the Qing in the sample. To convert this into a population figure, I assume that 20% of all men did not marry, and that the Qing population was composed of below-age-of-marry/men/women to one-third each. This gives a scaling factor of 3.75: 14 *jinshi*/(8,291 x 3.75) = 0.045 percent. If there are 20% of men not marrying, and there is universal marriage of women, then there must be 20% fewer daughters than sons, and 20% fewer women than men.

average county. Put simply, the potential sample selection in the genealogy, as a genre, seems minor in comparison to the rather large and pronounced differences that we see in the achievements of different families.

We can also look at the representation of different status levels in the sample. There may be a tendency to exaggerate the status of people in the genealogy, or to drop the poorest segments of the lineage--in both cases there would be many more officials or educated people in the genealogical sample than in the society at large. While there exists no generally agreed-upon status classification for Qing China, in the Liaoning Eight Banner sample 98% of males had "No Status" while 2% were "Officials".[12] This compares to about 71% of men having "No Status" in the Tongcheng sample, while about 1.4% have an official position. The relatively high fraction of "Officials" in the Eight Banner population might be related to the fact that it was a less densely populated area. In addition, according to estimates from Ho (1962), during the Qing Dynasty, the number of Bannermen who attained *jinshi* status as a percentage of the population ranked first among all provincial regions of China. The status distinction in the Liaoning data set also appears to have less gradation than what can be obtained from the genealogies. This could help to explain why the fraction of "No Status" men in the Tongcheng sample is lower.

In sum, when we are able to examine many lineages and sub-branches of lineages from these mid-sized genealogies, we can confirm that a majority of the observations are of commoners; it appears that these genealogical records are not exclusively dedicated to elite status groups. Taking only seven lineages, the profile of key variables in the Tongcheng sample gives reasonable ranges in comparison to other sources, and to the extent there are differences, they can be explained by the regional differences. This suggests that combining the data across many lineages is useful as a strategy for revealing regional demographic and economic patterns. The large number of genealogies increases the feasibility of constructing a larger data set over more regions.

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

Regression to the mean, recall bias, and survivor bias

Variation across lineages is useful for examining the factors that might affect whether a particular person, or observation, is included in the genealogy. As noted earlier, there were a number of different reasons why the genealogical tradition emerged. However, one concern is that genealogies might often begin with a particularly noteworthy man, who then becomes the progenitor of the lineage. Part of his noteworthiness might come from a high level of education, which is one of the most important signs of status and one of the most consistently reported characteristics of noteworthy persons. Alternatively, perhaps later generations were more likely to select a noteworthy progenitor. In either case, the implication would be a trend of declining status over time. If this is applied as a general rule for all genealogies, then we would consistently find regression to the mean. The tabulated sample gives us an opportunity to examine whether that happens.

In the Tongcheng genealogical sample, there are three lineages whose records begin with an educated progenitor: the *Chen* (progenitor born in 1298), the *Wang* (1358), and the *Ma* (1408). However, the highest lifetime attainment of these three progenitors was not more than what might be considered an intermediate level of human capital and not the highest possible (e.g *jinshi* degree holders). For the other four Tongcheng lineages, the highest levels of education are typically found nine generations after the inception of the lineage. Controlling for lineage differences in average education, there is no evidence that education levels in a lineage fall in later generations. Thus, a simple regression-to-the-mean effect is not present, suggesting that initial presumption about noteworthy founders is unlikely to be the driving characteristic of the Qing genealogical tradition. Again, this points to the diversity of lineages that possessed genealogies.

Positive selection would also arise if genealogical records contain many more entries of success compared to failure. I consider first, the correlation between average education and the number of observations over the seven Tongcheng lineages, which, at -0.10 is not significant. This suggests there is no evidence that on average, more successful lineages have included more entries in their genealogies. A related concern is that periods during which a lineage is successful

are also those when relatively many lineage members are recorded. However, the correlation between average education and number of observations is relatively small and negative (-0.07). We know that the updating of the genealogy often occurred retrospectively by the relatively resourceful lineage members. One might therefore believe that periods *right after* a lineage had been relatively successful are those when substantial resources would be devoted to updating the genealogy and the number of records would be relatively high. However, the correlation of -0.04 between average education in the previous generation and the number of lineage members recorded in the Tongcheng genealogy does not confirm this notion either.

Finally, one can examine whether strong groups of individuals tend to be overrepresented towards the end of the sample (survivor bias). If survivor bias is strong, one would expect that lineages with high achievements account for a relatively large share of the post-1800 observations in the data. Because Chinese genealogies deteriorate in their record-keeping in the 20th century (e.g. Harrell, 1987), I focus instead on the fraction of observations for the 19th century, versus before. Using these data, a regression of the lineage's share of post-1800 observations on (the log of) the lineage's average education yields a coefficient of 0.018 (s.e. = 0.062, R^2 = 0.02). This suggests, by this measure, that major survivor bias effects are not present.

The advantage of genealogies is that there is no issue with match rates over the group who self-profess to be members of the lineage—it is 100%—rather, the issue is only about the representativeness of the genealogies. Further digitization of the genealogical data would allow us to compare samples over a greater number and variety of regions. The next section considers how these data could be used for economics research.

7. Genealogical Data and New Research Directions

Human capital accumulation in previous centuries

By exploring household level information, genealogies can provide insights on the interplay between fertility, lineage success, and human capital formation. In Shiue (2013), I focus on the

characteristics of married men and dynamics over time in Anhui Province, examining the relationship between child quantity and quality in the period before modern economic growth. Over the Ming and Qing Dynasties, the number of high-level political positions stayed roughly the same while the population increased significantly. This lowered the probability for any one individual of passing the examinations and helped to intensify competition over the available slots. While at the beginning of the Qing Dynasty there is evidence for a quantity-quality tradeoff, the tradeoff weakens over the Qing, consistent with a lowering of incentives to educate. Decreasing education incentives may in turn have reduced income per-capita through increasing family size.

Further research is needed to address differences in the economic growth and divergence in growth trajectories between China and countries leading modern growth. For example, Galor and Klemp (2014) use genealogical data from families who resided in Quebec to investigate whether natural selection favored individuals with a larger predisposition to human capital formation and greater child quality in the 17th and 18th centuries, finding that moderate fecundity was conducive to long-run reproductive success. By considering similar factors that relate to child investment, it would be useful to examine a similar prediction within and across lineages in China, and contrast the findings with what is known about the history of the Western growth experience.

Intergenerational mobility

Genealogies are particularly useful in studying intergenerational mobility, where research interest has intensified recently. In this literature, measuring and comparing the degree to which there is persistence between parent and child outcomes, as well as assessing the causes of differences in this persistence across different regions or over time have been the dominating themes. In cases where the parent-child or father to son identity is not explicitly captured, surname linking has been used as a way to empirically construct data. These methods have

allowed researchers to investigate some of the potential reasons driving observed differences in social mobility across countries (Solon 1999, 2002; Black and Devereaux 2010).[13]

Moreover, since Becker-Tomes's (1979) theoretical analysis on the role of previous generations beyond the father—the grandfather, great-grandfather, and earlier generations—researchers have considered the empirical evidence on intergeneration transmission (Behrman and Tarban 1985, Long and Ferrie, 2012, Lindahl et al. 2012; Clark and Cummins, 2014, Solon 2015). Although much of it has focused on the father-son relationship, Olivetti and Paserman (2015) have also extended this recent literature to estimate the intergenerational income elasticity for both sexes. This not only gives a more accurate depiction of the average intergenerational transmission of status from the father to his children, but if mothers are important to child education then assortative mating could be another path by which status is preserved from one generation to the next. It is worth noting that in general, when working with census surveys, the linking from one generation to another is determined by the dates of the census and the ability of the researcher to find matching names across survey dates. In the case of the U.S. for example, where censuses were carried out at ten year intervals, the transcription of the 1880 census would allow its linking to 1860, 1870, 1900, to form 4 snapshots. Sample size typically declines as we condition on more generations being linked—due to the fact that surname linking loses 60-80% of the population with each iteration because of "linking failures" (Ruggles, 2006) attributable to migration and other reasons.

For linking historical censuses, Ferrie (1996) constructed a sample of males linked on the U.S. Public Use Micro Sample in the 1850 federal census of population to the 1860 federal census of population. Historical censuses are hard to read and errors occur through transcription and matching failures, with efficiency and accuracy in competition. Feigenbaum (2015) links historical micro data on earnings of fathers and sons, from a 1918-1919 Bureau of Labor Statistics sample with 1920 IPUMS sample, reaching a higher matching rate of 56% (with 90% accuracy) between

[13] Data from Nordic, European, and North American samples have yielded numerous estimates of intergenerational mobility for the 20th century; see surveys by Bjorklund and Jantti (2009) and Jantti et al. (2006).

father and son than has been generally obtained. In comparison, the percent of individuals who can't be linked in the Chinese Eight Banner Registers is 14.54% or 24.76% for one generation, depending on whether one considers Liaoning or Heilongjiang Province (Dong et al. 2015). As Ruggles (2006), points out, instead of maximizing the percentage linked across censuses, an alternative approach is to try to minimize selection biases in the linked data.

Genealogies are complementary to current efforts that link information in census surveys for studying long-run issues in demographic economics [(Björkland and Jäntti (2009); Campbell and Lee, (2011)]. However, genealogies provide a direct approach to the linking of individual relationships: the pedigree charts lay out the relationship between different generations of each lineage, and each person in the pedigree chart also has data on vital statistics and other socioeconomic measures. Since genealogies include not only the upper and most educated strata, but also many men who were commoners, the sample can be used to compare differences in mobility patterns across the distribution.[14] It should also be possible to aggregate the population sample across genealogies and in this way trace a daughter from her father's genealogy to her husband's genealogy. These data can be used to study assortative mating as a source of intergenerational status persistence. Further, the relationships of the extended kin group recorded in genealogies reveals networks of contemporaries—people who are related as siblings and cousins—and the status of these individuals could also play a role in intergenerational persistence.

Location versus lineage effects in China

In educational achievements, the emphasis is often on the role of nearby kin, but another factor may be due to locational or geographical spillovers from being near to more educated non-kin. Assuming the village of burial is also the place where the individual lived during his lifetime, this location information allows us to examine not only related individuals living across dozens of scattered locations in and around the major urban center, but also unrelated individuals in different lineages. This information on location of one person relative to other people in his clan,

[14] See Shiue (2016).

versus non-kin who live in the same village, can be used to examine competing roles of human capital spillovers across unrelated people in the same locality versus within lineage. These differences can shed light on whether Chinese society is more influenced by lineage organization according to the particular spread of each lineage across different villages, or whether location and distance is what matters most, regardless of lineages' residential patterns. Moreover, over the time period, there were famines, wars, and dynastic transition. To take the dynastic transition from the Ming to the Qing as an example of an exogenous shock that affected some villages much more than others, we can examine the short-run and long-run impact that this had on lineage structure, human capital spillovers, as well as residential patterns and migration (see Butner, Keller, and Shiue, 2016).

8. Concluding summary

Looking into the past not only adds significantly to our understanding of economic development, but it is also in historical microdata that researchers can find information that is restricted in modern microdata. Chinese genealogies offer rich individual level information for a broad set of regions. These are individuals who profess to belong to distinct lineages, which, based on the information on burial and settlement, have rights of resource use in those regions. More than that, genealogies provide information on individuals for an exceptionally early and long time period. The full extent of the source has yet to be delineated since only certain samples have been digitized, but genealogical data are exceptional in the sense that they include information on intergenerational linkages and extended family relationships of the patrilineal line. Secondly, while multiple census observations over the lifetime of an individual would be necessary to potentially determine the maximum wealth or status of an individual, the information in genealogies are lifetime measures.

An important question has to do with who is included in the genealogy. While further research is needed, a reasonable point of departure is that membership in the genealogy signifies access to resources, land, and responsibility for taxes and lineage contributions. In addition, because genealogies are privately compiled, it is relevant to understand the objectives and origins of the

genealogical tradition and potential areas of selection bias, recall bias, regression to the mean, and survivor bias—and how the severity of the biases can be assessed and potentially addressed.

Both overt omission of non-ritualistically significant individuals and inadvertent selection of those who are ritualistically significant may be present, and because of the heterogeneity across genealogies, each new sample should consider these aspects anew. As the number of publically available genealogies has increased dramatically in the last decade, it will become more and more feasible to do so. The example of Tongcheng County of Anhui Province shows that combining just seven lineage genealogies that are available within one locality yields over 1.5% of the population who were alive in the 18th century, above the fraction of the population that is often registered in census sampling. Furthermore, within a sample of 9,786 married men, we can find a broad representation of status groups. Continued digitization of the data will provide valuable links between contemporary and historical patterns that could allow us to understand long-run economic and social change.

References

Beattie, Hilary J. (1979), *Land and Lineage in China A Study of T'ung-Ch'eng County, Anhwel, in the Ming and Ch'ing Dynasties.* Cambridge: Cambridge University Press.

Becker Gary S., Tomes Nigel (1979), "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility." *Journal of Political Economy,* 87(6):1153–89.

Behrman, Jere and Tarbman, Paul, (1985), "Intergenerational Earnings Mobility in the United States: Some Estimates and a Test of Becker's Intergenerational Endowments Model", *The Review of Economics and Statistics*, 67(1): 144-51.

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

Björklund, Anders and Markus Jäntti (2009), "Intergenerational income mobility and the role of family background", in Wiemer Salverda, Brian Nolan, and Time Smeeding, eds., *Oxford Handbook of Economic Inequality.* Oxford University Press.

Butner, Matthew, Wolfgang Keller, and Carol H. Shiue (2016), "Location and Lineage in Human Capital Acquisition", Economic History Association conference, 2016.

Campbell, Cameron and James Z. Lee (2011), "Kinship and the Long-Term Persistence of Inequality in Liaoning, China, 1749-2005", *Chinese Sociological Review*, 44(1):71-103.

Campbell, C. and J. Z. Lee (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

Clark, Greg and Neil Cummins (2014), "Intergenerational Mobility in England, 1858-2012: Surnames and Social Mobility", *Economic Journal*, 125(582).

Cohen, Myron L. 1990. "Lineage Organization in North China." *Journal of Asian Studies* 49(3): 509- 534.

Dong, H., Campbell, C., Kurosu, S. et al. (2015), "New Sources for Comparative Social Science, Historical Population Panel Data from East Asia", *Demography* 52: 1061.

Eberhard, Wolfram (1969), "Chinese Genealogies as a Source for the Study of the Chinese Society", World Conference on Records and Genealogical Seminar, Salt Lake City, Utah.

Ebrey, Patricia (1983), "Types of Lineages in Ch'ing China: A Re-examination of the Changs of T'ung-ch'eng," *Ch'ing shih wen-t'I*, 4: 1-20.

Ebrey, Patricia and Watson, James L., eds. (1986), *Kinship Organization in Late Imperial China, 1000-1940*. Berkeley: University of California Press.

Elman, Benjamin (2000), *A Cultural History of Civil Service Examinations in Late Imperial China.* University of California Press.

Faure, David (2007), *Emperor and Ancestor.* Stanford: Stanford University Press.

Faure, David (1989), "The Lineage as a Cultural Invention: The Case of the Pearl River Delta", *Modern China*", 15(1): 4-36.

Ferrie, Joseph, (1996), "A New Sample of Males Linked from the Public Use Micro sample of the 1850 U.S. Federal Census of Population to the 1860 U.S. Federal Census of Manuscript Schedules" *Historical Methods,* 29: 141-156.

Feigenbaum, James J. (2015), "Intergenerational Mobility during the Great Depression", Working paper.

Freedman, Maurice (1966), *Chinese Lineage and Society: Fukien and Kwangtung.* New York: Humanities Press.

Fu, Pei-jung (2003), "Religions", in *Encyclopedia of Chinese philosophy,* Antonio S. Cua, ed. New York: Routledge.

Genealogical Society of the Church of Jesus Christ of Latter-Day Saints (1974), "Chinese Local Histories as a Source for the Genealogist." Series J, No. 3.

Galor, Oded and Marc Klemp (2014), "The Biocultural Origins of Human Capital Formation", NBER Working Paper 20474.

Greif, Avner (2006), "Family Structure, Institutions, and Growth: The Origins and Implications of Western Corporations", *American Economic Review P&P,* 96(2): 308-312.

Greif, Avner and Guido Tabellini (2015), "The Clan and the City: Sustaining Cooperation in China and Europe", Working paper.

Hao Dong, Cameron Campbell, Satomi Kurosu, Wenshan Yang, and James Z. Lee (2015), "New Sources for Comparative Social Science: Historical Population Panel Data From East Asia", *Demography*, 52(3): 1061–1088.

Harrell, Stevan, ed. (1995), *Chinese Historical Microdemography*. Berkeley: University of California Press.

Harrell, Stevan (1987), “On the Holes in Chinese Genealogies”, *Late Imperial China,* 8(2): 53-79.

Harrell, Stevan (1985), “The rich get children: segmentation, stratification, and population in three Chekiang lineages.” In *Family and Population in East Asian History,* eds. Susan Hanley and Arthur Wolf. Stanford: Stanford University Press, 81-109.

Hauser, Robert M. John Robert Warren (1997), “Socioeconomic Indexes for Occupations: A Review, Update, and Critique”, *Sociological Methodology* 27(1): 177-298.

Hauser, Robert M. John Robert Warren, Min-Hsiung Huang and Wendy Y. Carter (2000), “Occupational Status, Education, and Social Mobility in the Meritocracy”, in *Meritocracy and Economic Inequality*, Kenneth Arrow and Samuel Bowles, eds. Princeton: Princeton University Press.

Ho, P’ing-ti (1959), *Studies on the Population of China, 1368-1952.* Cambridge: Harvard University Press.

Ho, P’ing-ti (1962), *The ladder of success in Imperial China : aspects of social mobility, 1368-1911.* New York: Columbia University Press.

Jäntti, Markus, Bernt Bratsberg, Knut Røed, Oddbjørn Raaum, Robin Naylor, Eva Österbacka, Anders Björklund, Tor Eriksson (2006), “American Exceptionalism in a New Light: A Comparison of Intergenerational Earnings Mobility in the Nordic Countries, the United Kingdom and the United States”, IZA Discussion Paper no. 1938.

Keller, Wolfgang and Javier Andres Santiago, and Carol H. Shiue (2016), “Trade in China During the Treaty Port Era”, mimeo.

Keller, Wolfgang, Ben Li, and Carol H. Shiue (2013), “Shanghai’s Trade, China’s Growth: Continuity, Recovery, and Change since the Opium War”, *International Monetary Fund Economic Review,* 61: 336-378.

Kurz, Johannes (2014), “On the Southern Tang Imperial Genealogy”, *Journal of the American Oriental Society,* 133(4):601-621.

Lakos, William (2010), *Chinese Ancestor Worship: a practice and ritual oriented approach to understanding Chinese culture.* Newcastle upon Tyne: Cambridge Scholars Publishing.

Lee, James, Cameron Campbell and Wang Feng. 1993. "An introduction to the demography of the Qing imperial lineage, 1644-1911." Schofield, Roger and David Reher eds. Old and New Methods in Historical Demography. Oxford: Oxford University Press, 361-382.

Li, Bozhong (2000), Jiangnan de zaoqi gongyehua, 1550-1850 [Early industrialization in the Yangzi Delta, 1550-1850]. Beijing: Zhongguo shehui kexue wenxian chubanshe.

Li, Liu (1999), “Who were the ancestors? The origins of Chinese ancestral cult and racial myths”, *Antiquity* 73(281): 602-06.

Lindahl, Mikael, Mårten Palme, Sofia Sandgren Massih, and Anna Sjögren (2012), “The Intergenerational Persistence of Human Capital: An Empirical Analysis of Four Generations”, IZA Discussion Paper No. 6463.

Liu, Hui-chen Wang, (1959), *The Traditional Chinese Clan Rules.* Locust Valley, N.Y.: J.J. Augustin

Liu Ts’ui-jung (1978), “Chinese genealogies as a source for the study of historical demography.” In *Studies and essays in commemoration of the golden jubilee of Academia Sinica* [中央研究院成立五十週年紀念譜文集]. Taipei: Academia Sinica.

Liu Ts’ui-jung (1995), “Demographic Constraint and Family Structure in Traditional Chinese Lineages, ca. 1200-1900”. In *Chinese Historical Microdemography,* ed. Stevan Harrell. Berkeley: University of California Press.

Lo, Hsiang-lin (1969), *Scholastic application of the Chinese clan genealogies : history and arrangement of Chinese genealogies.* Salt Lake City, Utah: Genealogical Society of Utah.

Long, Jason, and Joseph Ferrie (2012), “Grandfathers Matter(ed): Occupational Mobility across Three Generations in the U.S. and Britain, 1850-1910, mimeo, Northwestern University.

Martin, Thomas R., (2010), *Herodotus and Sima Qian : the first great historians of Greece and China, a brief history with documents.* Boston: Bedford.

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

Pomeranz, Kenneth (2000), The Great Divergence: China, Europe, and the Making of the Modern World Economy. Princeton: Princeton University Press.

Ruggles, Steven, and Robert McCaa, Matthew Sobek, and Lara Cleveland (2015), "*The IPUMS Collaboration: Integrating and Disseminating the World's Population Microdata,*"*Journal of Demographic Economics,* 81: 203-216

Ruggles, Steven (2014), " Big Microdata for Population Research." *Demography,* 51: 287-297

Ruggles, Steven (2006), "Linking Historical Censuses: A New Approach" *History and Computing* 14: 213-224.

Shiue, Carol H. (2013), "Human Capital and Fertility in Chinese Clans Before Modern Growth", NBER Working Paper 19661.

Shiue, Carol H. (2016), "Social Mobility in the Long-Run: An Analysis with Five Linked Generations", mimeo.

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

Solon, G. (2002) Cross-Country Differences in Intergenerational Earnings Mobility. *The Journal of Economic Perspectives* 16(3), 59-66.

Solon, Gary (2015), "What do we know so far about multigenerational mobility?", NBER # 21053.

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

Telford, Ted A. (1990), "Patching the holes in Chinese genealogies." *Late Imperial China.* 11(2): 116- 35.

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

Thatcher, Melvin (2004), "Kinship and government in Chu during the Spring and Autumn era, 722-453 B.C." Ph.D. Dissertation, University of Washington.

Wang, Feng (2000),"Censuses of 1982 and 1990", in *Handbook of International Historical Microdata for Population Research*, eds. Patricia Kelly Hall, Robert McCaa, Gunnar Thorvaldsen. Minneapolis: Minnesota Population Center.

Watson, James L. (1982), "Chinese Kinship Reconsidered: Anthropological Perspectives on Historical Research", *China Quarterly* 92: 589-622

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

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

Wang Heming, ed., (2008), Zhongguo jiapu zongmu [Comprehensive catalogue of Chinese genealogies], Shanghai: Shanghai gu ji chu ban she.

Yuan, I-chin, (1931), “Life Tables for a Southern Chinese Family from 1365 to 1849,” *Human Biology*, 3.2: 157-179.

Zhongwei Zhao” (1997) Long-term Mortality Patterns in Chinese History: Evidence from a Recorded Clan Population”, *Population Studies*, 51, 2:

Zeng, Zhen and Yu Xie (2014), “The Effects of Grandparents on Children’s Schooling: Evidence from Rural China”, *Demography,* 51(2): 599-617.

Table A. Characteristics of men and their wives

Men		Variable	Obs	Mean	Std.	Min	Max
	Demographics						
		Birth year	9,787	1760.55	74.69	1298	1885
		Birth month	9,787	6.90	3.51	1	12
		Death year	8,142	1796.64	72.51	1348	1929
		Death month	8,142	6.53	3.41	1	12
	Education						
		Education	9,787	0.07	0.25	0	1
		Father's education	9,787	0.11	0.31	0	1
	Wealth						
		Status	9,786	1.65	3.67	0	22
		No. of wives	9,783	1.17	0.44	1	5
Women		**Variable**	**Obs**	**Mean**	**Std.**	**Min**	**Max**
	Demographics						
		Birth year	11,378	1766.24	75.31	1300	1887
		Birth month	11,378	6.75	3.47	1	12
		Death year	8,708	1797.10	73.66	1355	1930
		Death month	8,708	6.54	3.39	1	12
	Other						
		Father's education	6,179	0.08	0.28	0	1
		Father's status	6,179	1.86	4.74	0	22

Table B. Characteristics of sons and daughters

	Variable	Obs	Mean	Std. dev.	Min	Max
Sons						
Demographics						
	Birth order	20,177	2.17	1.33	1	11
	Birth year	20,176	1792.21	74.18	1330	1909
	Birth month	20,175	6.79	3.49	1	12
	Death year	14,849	1806.75	67.67	1380	1929
	Death month	14,858	6.62	3.43	1	12
Their fathers						
	Birth year	20,177	1759.80	73.56	1298	1866
	Birth month	20,177	6.95	3.51	1	12
	Age at death	16,448	55.40	13.12	17	91
	Education	20,177	0.08	0.26	0	1
Their mothers						
	Birth year	20,177	1763.88	74.02	1300	1887
	Birth month	20,177	6.86	3.49	1	12
	Age at death	15,165	55.70	15.15	16	96

	Variable	Obs	Mean	Std. dev.	Min	Max
Daughters						
Demographics						
	Birth order	11,150	1.81	1.09	1	10
	Birth year	2,443	1857.55	24.22	1615	1903
	Birth month	2,443	6.46	3.42	1	12
	Death year	1,418	1850.85	23.63	1654	1908
	Death month	1,420	7.06	3.31	1	12
Their fathers						
	Birth year	11,151	1770.63	66.09	1365	1866
	Birth month	11,151	6.99	3.53	1	12
	Age at death	8,541	55.37	13.42	15	90
	Education	11,151	0.09	0.28	0	1
Their mothers						
	Birth year	11,151	1774.78	66.55	1369	1870
	Birth month	11,151	6.86	3.50	1	12
	Age at death	7,784	55.87	15.44	11	96

Table C: Summary statistics by lineage

Lineage name		Education	Brothers	Total Siblings	Share female siblings	Father's education	Mother's age at death
CHEN	Observations	291	291	291	291	291	250
	Mean	0.003	3.716	5.495	0.266	0.017	61.340
	Std.	0.059	1.831	2.732	0.207	0.130	13.631
MA	Observations	627	627	627	627	627	621
	Mean	0.327	2.691	4.396	0.302	0.415	60.359
	Std.	0.469	1.389	2.177	0.221	0.493	17.188
WANG	Observations	4681	4681	4681	4681	4681	4333
	Mean	0.034	3.435	4.996	0.267	0.075	57.707
	Std.	0.181	1.640	2.183	0.200	0.263	14.113
YE	Observations	1607	1607	1607	1606	1607	1480
	Mean	0.101	3.103	4.711	0.274	0.156	58.911
	Std.	0.302	1.617	2.121	0.216	0.363	15.908
YIN	Observations	604	604	604	604	604	567
	Mean	0.026	3.194	4.796	0.271	0.040	59.087
	Std.	0.161	1.523	2.234	0.229	0.195	13.824
ZHAO	Observations	769	769	769	769	769	693
	Mean	0.017	3.331	4.831	0.265	0.051	58.245
	Std.	0.129	1.495	2.151	0.193	0.220	14.184
ZHOU	Observations	314	314	313	313	314	304
	Mean	0.022	2.834	3.738	0.174	0.038	58.270
	Std.	0.148	1.568	2.100	0.220	0.192	12.883
Total	Observations	8893	8893	8892	8891	8893	8248
	Mean	0.063	3.285	4.846	0.267	0.106	58.394
	Std.	0.244	1.621	2.208	0.208	0.308	14.656