

Confucianism and The Persistence of Son Preference*

Chak Sang Chan [†]

June 15, 2023

Abstract

This paper investigates the cultural origins of son preference in China. I hypothesize that Confucianism, with its emphasis on filial piety and male lineage, may have contributed to the persistent preference for sons. To test this hypothesis, I use a unique dataset of the number of Confucian academies at county level from 725 to 1800, normalized by the 1953 population at each county, and examine its relationship with sex ratio at birth at county level in 2020. My results suggest a strong positive correlation between the two, providing evidence that Confucianism may have played a role in shaping son preference in China. My results are further confirmed by exploiting both the distance of a county from Zhu Xi academies and a quota system which was introduced in 1425 by Emperor Yongle to restrict the number of presented scholars from each prefecture.

Keywords: Confucianism, Son Preference, Norms

JEL Codes: key1, key2, key3

*I am grateful to my advisors Tore Ellingsen and Robert Östling for their guidance and support. I also thank participants in the Microwave seminars at IIES, Brown Bag seminar at SSE for their helpful comments. All errors are my own.

[†]Stockholm School of Economics, chaksang.chan@phdstudent.hhs.se

1 Introduction

In recent decades, China has experienced a significant gender imbalance due to a cultural preference for male offspring. This phenomenon, known as son preference, has led to a skewed sex ratio at birth (SRB), with more male births than female births. Despite the importance of gender equality has been emphasized, son preference and the resulting gender imbalance persist in many parts of China. In the 2000 census, 11 provinces fell in the 112–120 range of SRB, and another 12 provinces exceeded 120. In 2010, the SRB of 28 provinces was greater than 112, of which 13 exceeded 120, which is far above the natural level of around 105 boys per 100 girls [Jiang and Zhang \(2021\)](#).

Apart from the “missing girls” problem through gender selective abortions, infantilizing baby girls, refusing to register Hukou for early born female children, and the outright abandonment or selling of female children, the country now faces a severe gender disparity, with a surplus of males. The skewed sex ratio is most prominent among the younger generations, where the number of marriageable men far surpasses that of marriageable women. This imbalance has created what is commonly referred to as the “marriage squeeze” or “marriage market failure” ([Nie \(2020\)](#), [Jiang et al., \(2014\)](#)). With a shortage of women, particularly in rural areas, many men face difficulties in finding suitable partners. By using the data of the 2020 Chinese Census conducted by the National Bureau of Statistics of China, in [Figure 10](#), we can see that there is a positive correlation between ratio of unmarried marriageable men, which is calculated using the number of unmarried men whose age is 15 or above, divided by the total number of men aged 15 or above, and the sex ratio at birth at county level. The effects of this skewed sex ratio on the marriage market are multifaceted. Firstly, it intensifies competition among men, leading to a rise in bride prices and dowries [Anderson \(2007\)](#). Families with sons often have to pay exorbitant amounts to secure a bride, exacerbating financial burdens and contributing to socioeconomic inequalities if some less privileged families cannot afford the bridge price. Secondly, the scarcity of women can result in an increased incidence of trafficking and forced marriages [Xia et al., \(2020\)](#), as vulnerable

women become targets for exploitation. Furthermore, the skewed sex ratio and its impact on the marriage market may have broader social implications. It can potentially fuel social unrest and crime rates [Edlund et al., \(2013\)](#), as frustrated men who are unable to marry may become disenchanted and marginalized within society. Additionally, the demographic imbalance poses challenges for population stability, economic productivity, and long-term social development.

It is true that the one-child policy, which was in effect in China from 1979 to 2015 [Li et al., \(2021\)](#), along with other policies i.e. 1978-84 Land reform [Almond et al., \(2019\)](#), as well as decrease in the cost of abortion due to technological advancement [Chen et al., \(2013\)](#), did attribute to the imbalance sex ratio at birth generally in China, it may not tell the whole story of skewed sex ratio at birth in China that some regional variations in son preferences and sex ratio at birth have persisted for several decades, and the same patterns remain today even the One Child Policy was officially abandoned in 2015. Most of the existing literature try to study the reasons of rising sex ratio at birth in China in general, few of them try to capture the cultural origin of son preference, which is one of the most important driving forces of skewed sex ratio apart from policy change, and also the persistent regional difference in sex ratio at birth. In this paper, I try to fill the gap in this regard by empirically exploring the hypothesis that Confucianism, a traditional Chinese belief system, is the cultural origin of son preference in China ([Almond et al., \(2013\)](#), [Zhou et al., \(2012\)](#)).

Confucianism, which originated in ancient China and became widespread during the Han dynasty, emphasizes the importance of hierarchy, social order, and filial piety. In Confucian thought, the family is the basic unit of society, and maintaining harmony within the family is crucial to the well-being of society as a whole. Confucianism also places great emphasis on the importance of continuing the family lineage through male heirs: “There are three things which are unfilial, and to have no posterity is the greatest of them” .

Although it has long been hypothesized that the Confucianism played an important role in shaping values and norms in Confucian societies, as well as served as an important

determinant of the evolution and persistence of son preference in not just China but also the whole East Asia with strong Confucian traditions (Cleland, Verral, and Vaessen 1983; Cho, Arnold, and Kwon 1982; Freedman and Coombs, 1974), none of the existing research tries to capture the quantitative effect of Confucianism on contemporary son preference, and establish a casual relationship between the two.

In this paper, I document a positive a relationship between the prevalence of Confucianism in a given area from 725 to 1800, and the contemporary sex ratio at birth in that area. To do so, I use data on the number of Confucian academies at the county level, normalized by the population in 1953, as a proxy for the prevalence of Confucianism. I also use the number of Buddhist temples in 1820 as a placebo test, as Buddhism does not place the same emphasis on family and lineage as Confucianism.

My OLS results suggest that there is a positive correlation between the number of Confucian academies at the county level and the sex ratio at birth in 2020. In contrast, I find no relationship between the number of Buddhist temples and the sex ratio at birth. These results provide evidence that the prevalence of Confucianism, in a given area from 725 to 1800, is associated with contemporary son preference in China.

To the best of my knowledge, there are also two papers trying to relate the skewed sex ratio at birth in East Asia with the impact of religions ([Tong et al., \(2020\)](#), [Kim and Song \(2007\)](#)). The former finds that, in China, counties with a greater presence of Daoist temples have more imbalanced (male-biased) sex ratios, whereas a greater presence of Buddhist temples and Islamic mosques is associated with less imbalanced sex ratios by using the locations of existing religious venues. While the latter finds that, in Korea, the prevalence of Protestant and Catholic churches turns out to be negatively related to sex ratio at birth, whereas that of Buddhist temples shows a positive relationship. Although neither Daoism nor Buddhism has a clear stance in gender role, nor stresses the importance of continuing a family line, they argue that the positive correlation is driven by the fact that Daoism and Buddhism absorb the Confucian tradition of son preference in China and Korea respectively,

especially after Confucianism was abandoned in these two places. Moreover, the relationships documented in their studies cannot be seen as casual: by simply looking at the relationship between the prevalence of different religions in different areas and the contemporary sex ratio at birth, the results may be driven by reverse causality i.e. Conservatism of a place affected the existing number of Daoist and Buddhist temples and omitted variables i.e. an unobserved variable like Conservatism, which is driven by other factors affects both the number of temples and the son preference. By looking at its historical influence in different areas, my paper serves as a supporting evidence that Confucianism attributes to contemporary son preference by finding a link between historical influence of Confucianism and contemporary son preferences, which in turn affect the fertility decisions and sex ratios in East Asia.

One concern with my approach is that the number of Confucian academies in the past may have been endogenous to the pre-existing cultural values about son preferences, and these may continue to persist until today. I try to address this causality problem by exploiting both the distance of a county from Zhu Xi academies and a quota system introduced in 1425 by Emperor Yongle as two instrument variables.

The first stage of my first IV estimates suggest that the shortest distance of a county from one of the Zhu Xi academies is negatively correlated with the subsequent number of Confucian academies, and the relationship is robust when I take the log form of the distance instead, which is consistent with the hypothesis that regions close to the Zhu Xi academies were probably influenced sooner by Zhu Xi's teachings. Over time, these regions would have developed a stronger Confucian culture compared with more distant regions.

And the first stage of my second IV estimates suggest that the number of small rivers over a main river in a prefecture is positively correlated with the number of Confucian academies after the introduction of quota system in 1425, which tried to maintain a balance among different regions by restricting the number of presented scholars from a prefecture. Historically, the number of small rivers in a prefecture affected how counties were formed, as it is not efficient to have many rivers within a county due to high administrative costs.

Subsequently, it might also affect the number of Confucian academies built nearby as more quotas were allocated to areas with more counties.

Similar to what previous literature suggest e.g. [Alesina et al., \(2013\)](#); [Alesina and Giuliano \(2010\)](#); and [Fernández \(2007\)](#), my findings provide some evidence that deep-rooted cultural beliefs continue to shape contemporary attitudes and affect family decisions in the context of Confucianism in China, although this paper focuses on within country variation, instead of utilizing the variation across different countries.

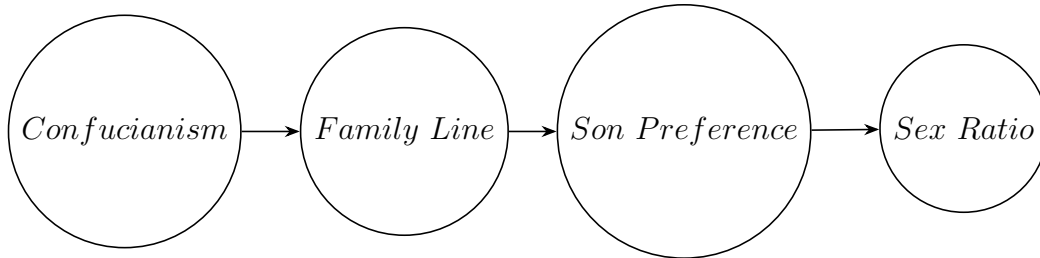
The rest of the paper is organized as follows. In Section 2, I introduce the background of historical Confucianism, the role of historical Confucian academies and also contemporary son preference and sex ratio at birth in China. In Section 3, I describe the data used in this project and the sources of the data. In Section 4, I describe the empirical strategy used and present the OLS results. In Section 5, I discuss the backgrounds of the two instruments I use and present the corresponding results. In Section 6, I include a placebo test by using the number of Buddhist temples at each county in 1820, normalised by the population in 1953. In Section 7, several robustness tests are included. In Section 8, I try to discuss about some potential channels, through which historical Confucianism affects the contemporary son preferences. Section 9 offers concluding thoughts.

2 Background

2.1 Confucianism and Confucian Academies

Confucianism is a philosophy that originated during Zhou dynasty in ancient China and has had a profound impact on not only the country’s history and culture, but also that in other East Asian countries. Confucius was a philosopher and thinker who lived in the 6th century BCE and his teachings became known as Confucianism. Confucianism emphasizes the importance of education, social harmony, respect for authority, and adherence to traditional values. One of the key principles of Confucianism is the importance of family and filial

piety. This emphasis on the family has historically been associated with the development of Confucian clans and genealogy. Since China has always been a patrilineal society, to continue a family line, one should have at least a son, which has gradually developed a preference for sons over daughters. As some suggestive evidence, in the first two columns of Table 5, we can see that historical Confucianism is positively correlated with historical genealogy, which I use the normalised number of genealogy books from 1201 and 1896 as a proxy, consistent with my hypothesis that Confucianism affects the value of people place on the continuation of a family line. To ensure that contemporary sex ratio at birth is not driven entirely by the genealogy in the past, but instead son preference in a region is influenced by historical Confucianism as I propose, in column 3, we can see that the positive correlation between Confucianism and contemporary son preference remains the same in terms of both magnitude and significance, even controlling for historical genealogy.



Confucianism was spread throughout ancient China by a network of Confucian academies (*Shuyuan* 書院). The first recorded establishment of Confucian academies dates back to 725 during the Tang dynasty. They were places where scholars could teach and study the classics, and where books collected from around the country could be preserved. By the late Tang dynasty, Confucian academies had appeared all over China though most of them were of private uses at that time. During the Song Dynasty, the imperial government established a network of state-sponsored academies, known as the Guozijian, to train officials for the civil service. As a result of this expansion, many private academies were eventually transformed into public institutions, supported by the state and tasked with training civil servants for the imperial bureaucracy. Since then, Confucian academies became important institutions in China, and they played a crucial role in the spread of Confucianism and its associated

values in ancient China.

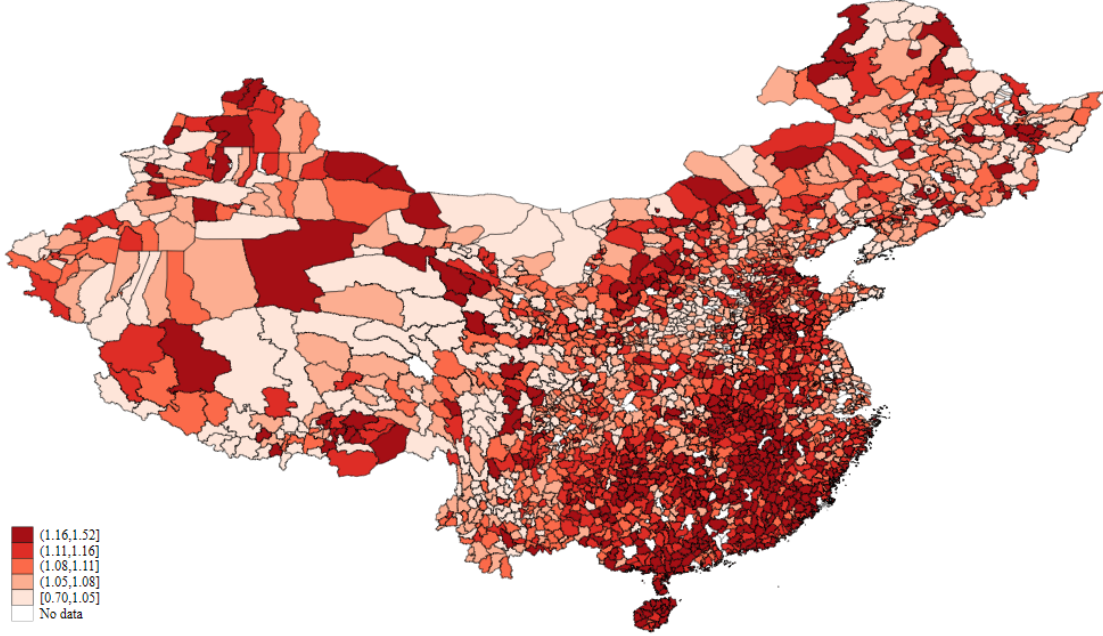
3 Data Sources

3.1 Confucian Academies from 725 to 1800

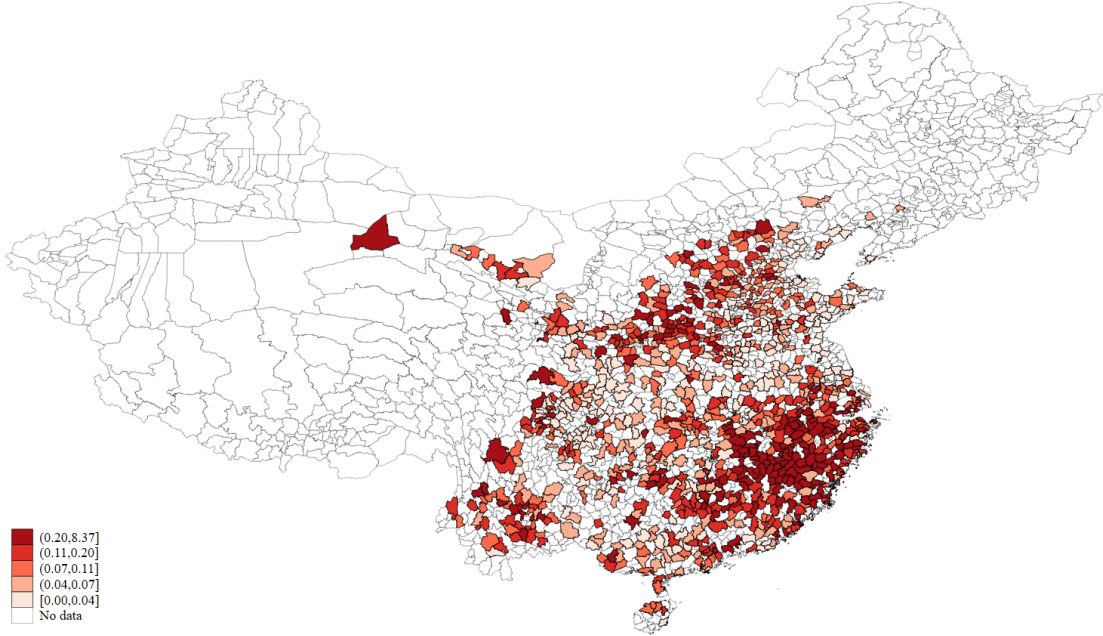
To measure the influence of historical Confucianism at county level, I use the number of Confucian academies at each county as a proxy. I first collect information about Confucian academies from 725 to 1800 from [Bai \(2012\)](#), which was gathered from extensive gazetteers and historical documents. The book records all the Confucian academies and their corresponding locations at county or prefecture level in the past as well as their building years [Figure 9](#). I first digitise the book and record every academy based on their location and the period built. I then count the number of Confucian academies at each county in the past, and link the past location to the location nowadays using my database built up with multiple sources e.g. Google, Wikipedia and Baidu. To ensure that the number of Confucian academies was not driven by population density, I also normalise its value with the population in 1953 at county level from the 1953 Chinese Census. The reason why I use the population in 1953 is that it is the first Census data conducted at county level since the establishment of the People's Republic of China (PRC) in 1949. In [Figure 13](#), we can see the correlation between population in 1953 and population in 2020 at county level and the correlation between the two is 0.42, so the relative growth rate of population in different counties should arguably not be a great concern. As a robustness check, in [Table 15](#) and [Table 13](#), I run the same regressions on Confucian academies without normalisation and Confucian academies by population in 2020, and the results are still significant.

3.2 Sex Ratio at Birth in 2020

The county level data of the sex ratio at birth in 2020, along with other contemporary socioeconomic controls are collected from the 2020 Chinese Census, which is conducted by



(a) Sex Ratio at Birth in 2020 at the county level



(b) The number of Confucian Academies built from 725 to 1800 at the county level normalized by the 1953 population

Figure 1: Contemporary Sex Ratio at Birth and Historical Distribution of Confucian Academies

the National Bureau of Statistics of China and covers the entire population, and provides the most comprehensive demographic information.

3.3 Quota System

The data of the quota system, along with other historical controls are collected from [Bai and Jia \(2016\)](#). However, the data set is at the prefecture level, and does not cover the entire territory of China. Thus, I code the missing values as zeros and add another dummy indicating missing values of those variables.

3.4 China Family Panel Studies

The China Family Panel Studies (CFPS) is a longitudinal survey conducted by the Institute of Social Science Survey (ISSS) at Peking University and the Institute of Sociology (IOS) at the Chinese Academy of Social Sciences (CASS). The CFPS aims to comprehensively study various aspects of family dynamics, intergenerational relationships, and social change in China.

I use the survey wave in 2010, as this is the only wave containing the genealogy information. This nationally representative survey covers 33,600 adults and 14,960 households in 25 provinces,³¹ representing 95% of the population in China.

To see if historical Confucianism does affect how people value the importance of the continuation of the family line nowadays, I use the question “How important is having children to carry on the family name”, from 1 to 5, where 5 represents that it is very important and 1 represents that it is not important at all. Moreover, since they also include information about the birth order and genders of the respondents’ children, I can explore the relationship between the value respondents placed on the norm and sex ratio among their children. Consistent with my hypothesis, in [Figure 11](#), we can see that the higher value one places on the norm, the more likely he or she has a more skewed sex ratio among children.

3.5 Buddhist Temples in 1820

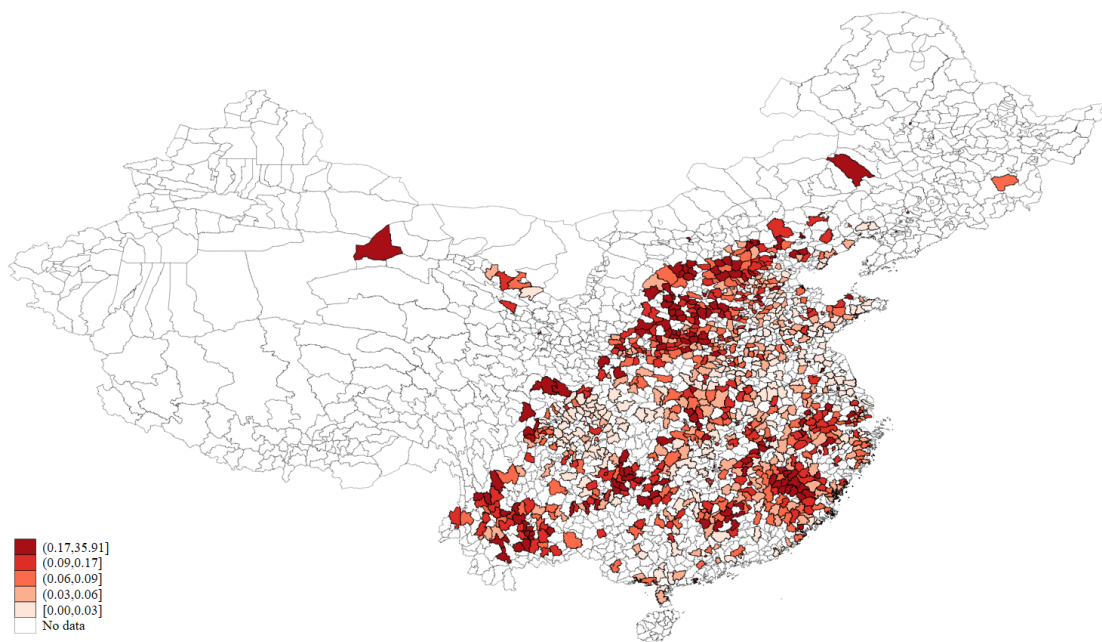
The data on Buddhist sites originates from the National Gazetteer of the Qing Dynasty (1820). I collect it from the website of WorldMap. It contains the precise locations of 2407 Buddhist sites in China in 1820. I then linked their coordinates to the county they belong to nowadays, and count the total number of Buddhist temples in each county.

3.6 Other Variables

The data of population at county level in 1953 is collected from the 1953 Chinese Census, which is the first census data at county level ever conducted in China. The data of historical genealogy between 1201 and 1896 proxied by the number of genealogy books at prefecture level is from [Chen et al., \(2022\)](#).



(a) The distribution of Buddhist Temples in 1820



(b) The number of Buddhist Temples in 1820 at the county level normalized by the 1953 population

Figure 2: Historical Distribution of Buddhist Temples

4 OLS Estimation

In this section, I investigate the relationship between historical Confucianism, which is measured by the number of Confucian academies, normalised by population in 1953, and the sex ratio at birth in 2020 at county level. Specifically, the county level estimating equations take the following form:

$$Y_c = \alpha_c + \beta \text{Confucianism}_c + X_p^G \phi + X_p^H \tau + X_c^C \eta + \epsilon_c \quad (1)$$

where Y_c is the sex ratio at birth in 2020 in county c ; Confucianism_c is the historical number of Confucian academies, normalised by population in 1953; the coefficients of main interest β captures the relationship between historical Confucianism and contemporary sex ratio at birth; X_p^G is a vector of geographical control variables, which includes a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9, as these variables might affect historical transportation and thus the spread of Confucianism among people within a region; X_p^H is a vector of historical control variables, which includes a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level, as population density and economic prosperity in the past might affect the demand for education and thus the number of Confucian academies built; X_c^C is a vector of contemporary socioeconomic controls at county level, which includes average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020. Although it should be noted that controlling for contemporary variables may be a bad control given that Confucianism may shape other kinds of behaviours and values, apart from son preference, although in [Table 17](#) and [Table 18](#), we can see that historical Confucianism is not correlated with various contemporary socioeconomic variables e.g. average education year, urban rural ratio, proportion of immigrants from other provinces and fertility rate.

Table 1: Confucianism and Contemporary Sex Ratio (OLS)

	(1) SRB	(2) SRB	(3) SRB	(4) SRB
Confucianism	0.00918*** (0.00253)	0.00970*** (0.00258)	0.0107*** (0.00263)	0.0115*** (0.00277)
Geographic controls		✓	✓	✓
Historic controls			✓	✓
Contemporary controls				✓
Mean dependent vars	1.11	1.11	1.11	1.11
R^2	0.018	0.086	0.114	0.136
Observations	1136	1136	1136	1132

Notes: County-level OLS regressions. Dependent variable is the sex ratio at birth in 2020 at county level. Confucianism is the historical number of Confucian academies from 725 to 1800, normalised by population in 1953. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardized. Standard errors in parentheses are clustered at the prefecture level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The ordinary least squares (OLS) estimation shows that the Confucianism has a significantly positive effect on contemporary sex ratio at birth, and this effect becomes greater after controlling for other factors in other columns. Focusing on column (4), the coefficient of 0.0114 implies that a one standard deviation increase in Confucianism is associated with an increase in contemporary sex ratio at birth by 1.14 percentage points.

5 IV Estimation

Although the previous results suggest that there is a strong positive correlation between the historical Confucianism and the contemporary son preference revealed by the sex ratio at birth, and Buddhism is used as a placebo test to show that the relationship between historical Confucianism and contemporary son preference is unique i.e. son preference is not simply driven by the intensity of religions, regardless their stances on the importance of continuation of family line, the relationship may be driven by some omitted variables which affected both the decisions of building Confucian academies in the past and the son preferences nowadays i.e. conservatism in the past affected both the number of Confucian academies built and the conservatism nowadays. To further explore the potential casual effect of Confucianism on contemporary son preference. I exploit both the distance of a county from Zhu Xi academies and the introduction of the quota system in 1425 and use the number of quota as instrument variables to address the concern of endogeneity.

5.1 Distance from Zhu Xi Academies

5.1.1 Background

Zhu Xi (1130-1200) was one of the most influential Confucian philosopher, scholar, and writer who lived during the Song Dynasty. He is best known for his contributions to Confucianism, particularly his interpretation of the works of Confucius and Mencius. Zhu Xi's philosophy, known as Neo-Confucianism, combined elements of Confucianism, Buddhism, and Taoism,

and had a significant impact on Chinese intellectual and cultural history.

Zhu Xi’s most significant contribution to Confucianism was his systematization of the teachings of Confucius and Mencius. He believed that Confucianism was not just a moral code, but a comprehensive system of thought that could be used to understand and interpret the world. Zhu Xi emphasized the importance of studying the classics and the need for scholars to engage in self-cultivation to achieve moral and intellectual excellence.

His philosophy developed and spread during his teaching at Yuelu Academy in Changsha (in Hunan province), Hanquan Academy in Jianyang (Fujian) and Bailudong Academy in Jiujiang (Jiangxi). These academies served as centers of learning for Confucian scholars and were instrumental in the dissemination of Neo-Confucian thought throughout China.

Distance to the Zhu Xi academies should arguably be orthogonal to economic development in the early twentieth century. Our two main justifications are as follows. First, Zhu Xi did not choose to teach in the three academies but was there by coincidence. Specifically, Zhu Xi built Hanquan Academy in Jianning, Fujian province, because his mother was buried there. Under Confucian mourning rituals, a son must stay near his mother’s tomb for at least one year. During this period, Zhu Xi established Hanquan Academy and resided there for 10 years before he was appointed by the Prime Minister as the Prefect of Nankang (in Jiujiang, Jiangxi) in 1178. One year later, Zhu Xi rebuilt the prestigious Bailudong Academy there, developing new curricula based on his neo-Confucian principles. In 1193, Zhu Xi was appointed a governor of relief in Changsha, Hunan province, where he seized the opportunity to introduce his philosophy and curricula to Yuelu Academy in Changsha (Shu, 2003). Given that his official appointments and mourning rituals were not chosen by Zhu Xi, the locations of three academies were arguably ‘selected’ by exogenous forces instead of chosen endogenously. More specifically, in [Table 8](#), we can see that the distance of a county from Zhu Xi Academies is not correlated with the number of Confucian Academies built before the birth of Zhu Xi. Second, the Zhu Xi academies were located in neither economic centres nor backward regions. For instance, we can see in [Table 6](#) that the distance of a

county from Zhu Xi academies is correlated with neither its economic prosperity, nor its importance in transportation in eighteenth century. Also in Table 7, we cannot see any significant relationship between the distance of a county from Zhu Xi academies and factors that potentially affect the son preference of that county in 2020.

5.1.2 Results

The IV estimation results are presented in Table 2, consistent with my hypothesis, the number of Confucian academies built from 1127 to 1800 normalised by the population in 1953 is negatively correlated with the distance of a county from Zhu Xi academies in all three specifications at one percent significance level in the first stage regressions. The second stage results document that historical Confucianism instrumented by the log distance of a county from Zhu Xi Academies is positively correlated with the contemporary SRB in a county. The results are robust across all specifications in terms of significance and magnitude of coefficients. The magnitude of effects of historical Confucianism is larger than the OLS estimations, suggesting that the IV estimations might measure the local average treatment effect of counties for which the historical Confucianism was affected by the distance from Zhu Xi Academies.

5.2 Quota System

5.2.1 Background

Before the introduction of the quota system, the education level of scholars in the southern part of the Ming Dynasty was significantly higher than that of the northern part. During the Hongwu period of Ming Taizu’s reign, a total of 867 scholars passed the imperial examination, of which 620 were from the southern region, accounting for 71% of the total. In the 30th year of the Hongwu reign of the Ming dynasty (1397), the ‘Nanbei Bang Case,’ also known as the ‘Spring and Summer List Case,’ served as a trigger point for a reform in the imperial examination. During this exam, all the northern candidates failed. To narrow the gap

Table 2: Confucianism and Contemporary Sex Ratio (1st IV)

	(1)	(2)	(3)
	SRB	SRB	SRB
Confucianism	0.0743** (0.0320)	0.0711** (0.0276)	0.0680*** (0.0255)
Mean dependent vars	1.11	1.11	1.11
	(1)	(2)	(3)
	Confucianism	Confucianism	Confucianism
Ln Distance(km)	-0.0886*** (0.0329)	-0.0915*** (0.0333)	-0.0989*** (0.0351)
Geographic controls	✓	✓	✓
Historic controls		✓	✓
Contemporary controls			✓
Mean dependent vars	0.46	0.46	0.46
Observations	1121	1121	1117

Notes: County-level IV regressions. Dependent variable is the sex ratio at birth in 2020 at county level. Confucian is the historical number of Confucianism academies from 1127 to 1800, normalised by population in 1953. I employ the log distance of a county from Zhu Xi Academies as the IV for the number of Confucian academies. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardized. *t* statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

between the southern part and other regions, in September of the first year of the Hongxi reign (1425), the imperial examination was divided into North and South sections. In the second year of the Xuande reign (1427), Emperor Xuande of the Ming Dynasty issued an edict to formally implement the division of the imperial examination into North and South sections, and added a Middle section. It was stipulated that if 100 people were admitted in the imperial examination, 55 would be selected from the South section. The southern section covers the areas of Zhejiang, Jiangxi, Fujian, Hubei and Guangdong; the northern section covers the areas of Shuntian, Shandong, Shanxi, Henan and Shaanxi; and the central section covers Sichuan, Guangxi, Yunnan, Guizhou, Fengyang, Luzhou, Lushan, and Hezhou. In 1436, the central government began to appoint government officials to each prefecture to select candidates for the province-level exam (Li (1989)). To confirm whether the quota system did serve the purpose of balance the opportunity to pass the imperial exam in different regions, in [Figure 4](#), we can see the percentage change in the number of presented scholars from the three regions, from the period of 1368-1435 to the period of 1436-1505, which are the periods before and after the implementation of the quota system. We can see that there were significant growths in the central and the northern regions in terms of the number of presented scholars. Conversely, there was a decline in the number of presented scholars in the southern region as soon as the quota system was imposed. To see the overall change in the distribution of the presented scholars after the quota system, I include a pie chart of the distribution of presented scholars, based on regions that they represented in [Figure 5](#). To see if this may potentially be correlated with the distribution of the Confucian academies built after the quota system, I divide my data set of Confucian academies into two parts: before and after the quota system. In [Figure 6](#), we can see that there was a significant increase in the proportion of Confucian academies being built in the central and northern areas, suggesting that the quota system may potentially change the distribution of Confucian academies being built afterwards.

Conceptually, for two prefectures with the same population size, the prefecture with more

counties tended to have higher quotas per capita, because the quota assignment followed the stepwise role and each county would get some quota. Therefore, it would be optimal to find an instrument that affected the formation of counties in a prefecture but did not affect other factors that are potentially related to son preferences or contemporary sex ratio at birth, through any other channels such as economic development. Therefore, by borrowing the insight from [Bai and Jia \(2016\)](#), in which they use the number of small rivers divided by the river length as an instrument for the number of quotas, exploiting the fact that it was not efficient to have many rivers within a county due to high administrative costs, therefore the number of small rivers affected the number of counties, and thus the availability of quotas. In [Figure 7](#), we can see the relationship between the number of small rivers divided by river length and the number of quotas divided by river length, after controlling the river length and whether there is a main river near the prefecture. Since the availability of quotas affected the number of Confucian academies built after the quota system i.e. it would be not efficient to build either too many academies in places with less quotas or too little academies in places with too many quotas. In [Figure 8](#), we can see that the normalised number of Confucian academies built after the quota system and the number of quotas divided by river length are positively correlated, after controlling the length of the main river and the population in 1880.

To see if the number of small rivers over a main river in a prefecture is a valid instrument, i.e. it is not correlated with other factors that potentially affected the number of buildings in general, as a placebo test, I also run the same regressions of the number of Confucian academies built before the quota system, and the number of Buddhist temples in 1820 on the number of small rivers divided by river length, with different types of controls. In [Table 9](#) and [Table 10](#), we can see that the number of small rivers divided by river length affected neither the number of Buddhist temples nor the number of Confucian academies built before the quota system, which is consistent with the hypothesis that it was only relevant once the quota system was imposed as it was more efficient to build more academies in places where

more quotas were available.

5.2.2 Results

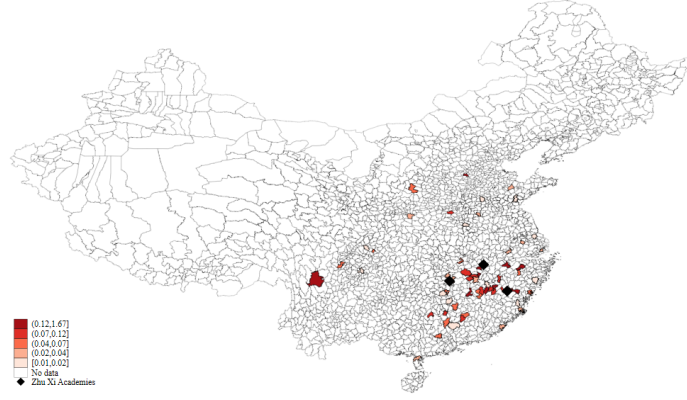
The IV estimation results are presented in [Table 3](#), consistent with my hypothesis, the number of Confucian academies built from 1425 to 1800 normalised by the population in 1953 is positively correlated with the log number of small rivers in all three specifications at one percent significance level in the first stage regressions. The second stage results document that historical Confucianism instrumented by the log number of small rivers is positively correlated with the contemporary SRB in a county. The results are robust across all specifications in terms of significance and magnitude of coefficients. The magnitude of effects of historical Confucianism is again larger than the OLS estimations, suggesting that the IV estimations might measure the local average treatment effect of counties for which the historical Confucianism was affected by the number of quotas assigned. As a robustness test, in [Table 19](#), I use directly the log number of quotas divided by the river length at each prefecture as an instrument instead of the log value number of small rivers divided by the river length, the results are still consistent and significant.

Table 3: Confucianism and Contemporary Sex Ratio (2nd IV)

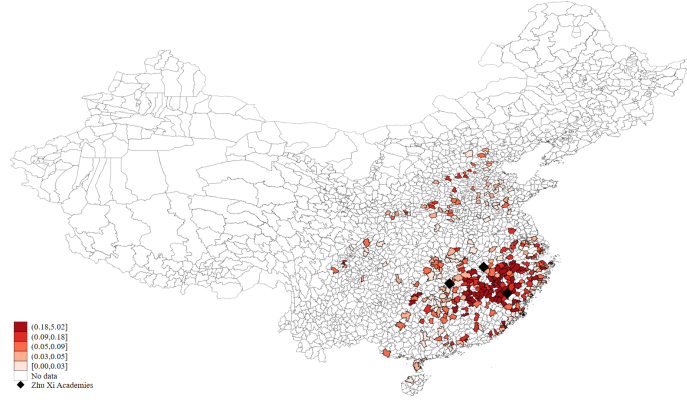
	(1) SRB	(2) SRB	(3) SRB
Confucian Post	0.0477** (0.0223)	0.0568** (0.0248)	0.0630** (0.0270)
Mean dependent vars	1.11	1.11	1.11
	(1) Confucian Post	(2) Confucian Post	(3) Confucian Post
Ln #small rivers/River length	0.569*** (0.163)	0.538*** (0.157)	0.538*** (0.166)
Geographic controls	✓	✓	✓
Historic controls		✓	✓
Contemporary controls			✓
Mean dependent vars	0.697	0.697	0.697
Observations	982	982	979

Notes: County-level IV regressions. Dependent variable is the sex ratio at birth in 2020 at county level. Confucianism is the historical number of Confucian academies from 1425 to 1800, normalised by population in 1953. I employ the log value of the number of small rivers with length less than 90km divided by the river length as the IV for the number of Confucian academies. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardised. *t* statistics in parentheses.

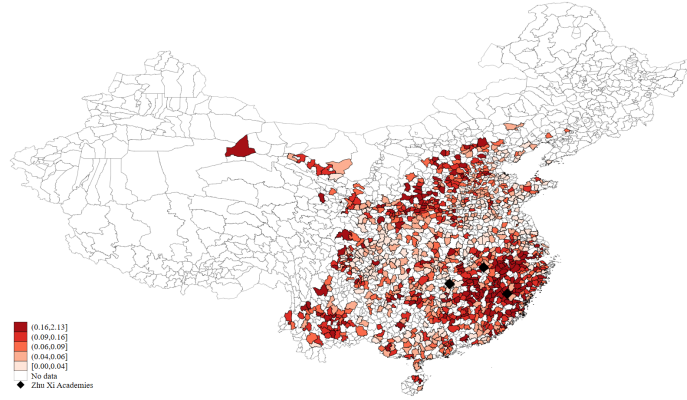
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$



(a) The number of Confucian Academies built before the birth of Zhu Xi at the county level normalized by the 1953 population



(b) The number of Confucian Academies built from 1127 to 1425 at the county level normalized by the 1953 population



(c) The number of Confucian Academies built from 1425 to 1800 at the county level normalized by the 1953 population

Figure 3: The number of Confucian Academies built before Zhu Xi, built from 1127 to 1425 and built after 1425 at the county level normalized by the 1953 population and the locations of Zhu Xi Academies

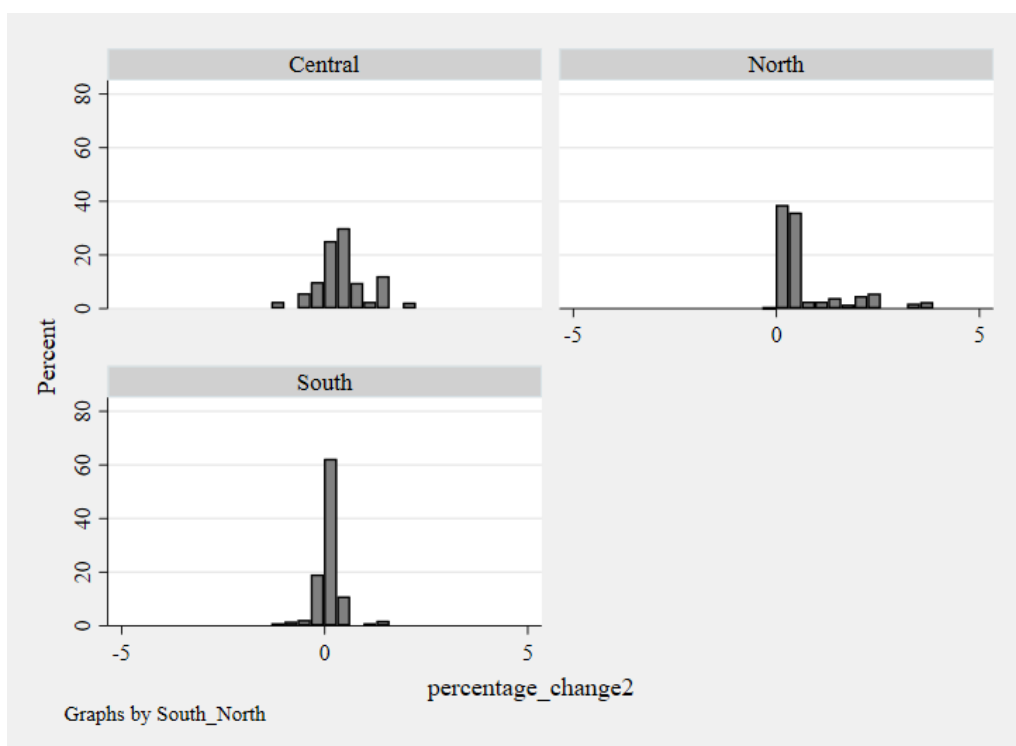
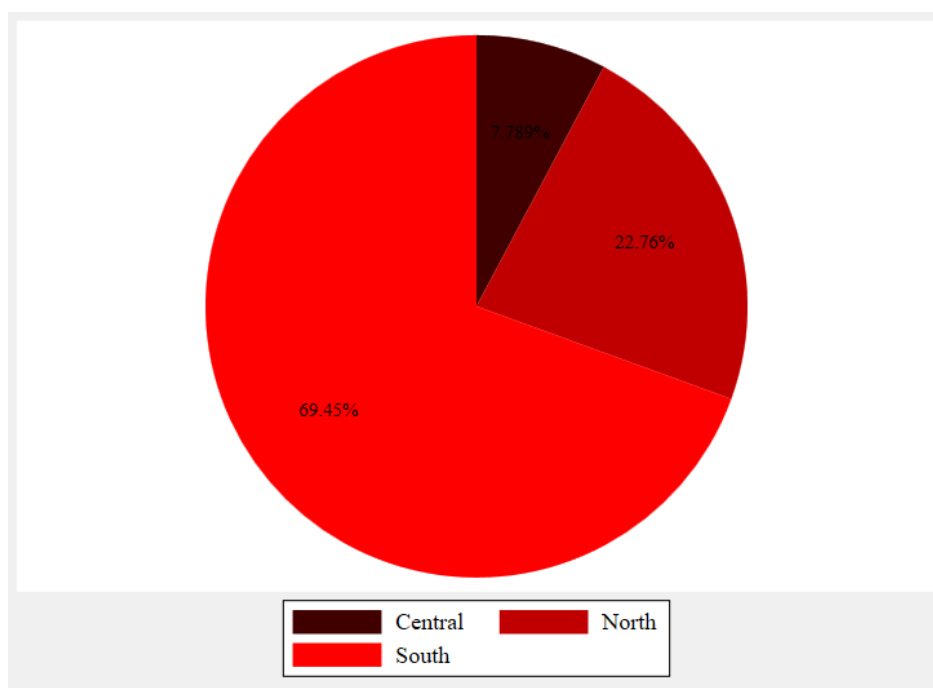
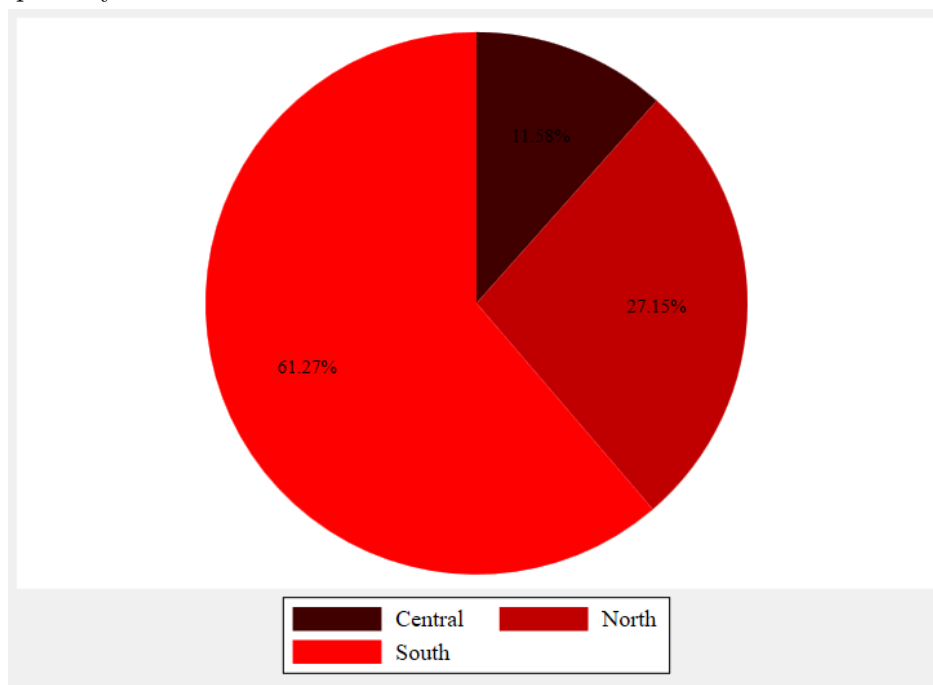


Figure 4: The percentage change of the number of presented scholars from the period of 1368-1435 to the period of 1436-1505

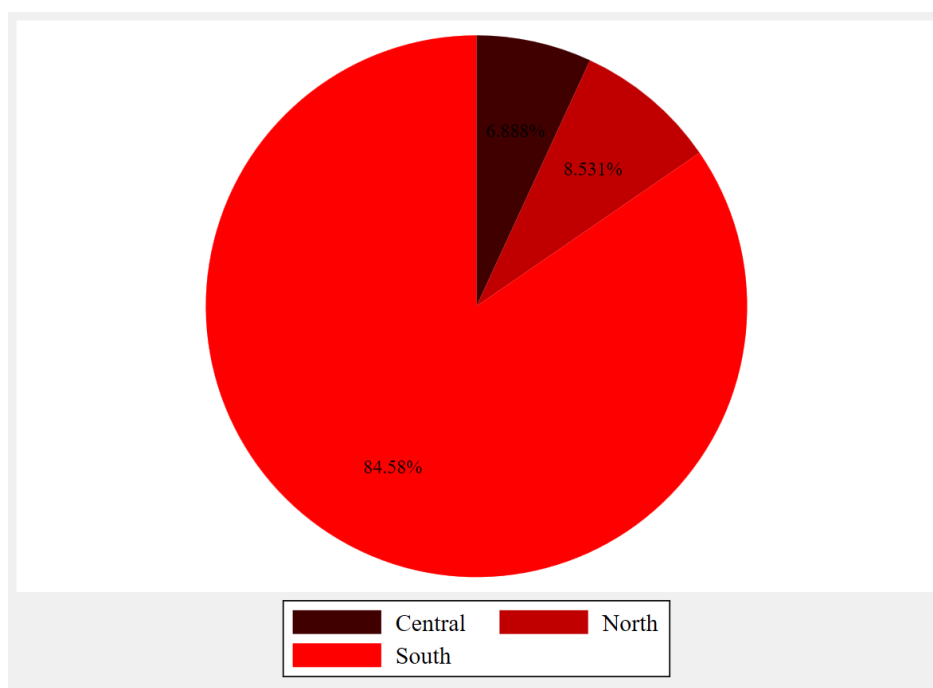


(a) The distribution of presented scholars before the introduction of the quota system

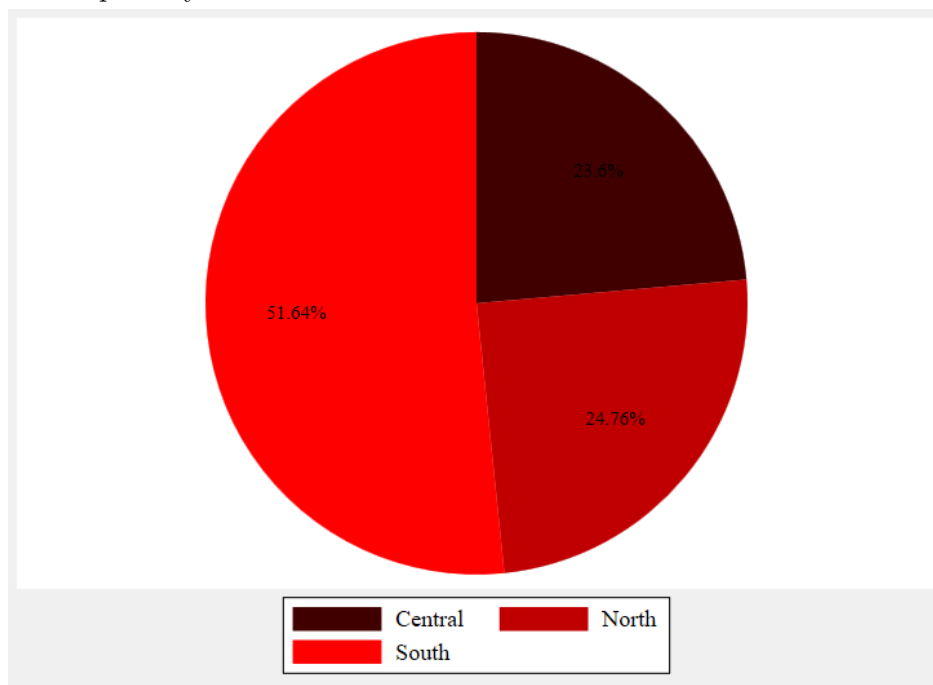


(b) The distribution of presented scholars after the introduction of the quota system

Figure 5: The distribution of presented scholars before and after the introduction of the quota system during Ming dynasty



(a) The distribution of Confucian academies built before the introduction of the quota system



(b) The distribution of Confucian academies built after the introduction of the quota system

Figure 6: The distribution of Confucian academies built before and after the introduction of the quota system from 1271 to 1800, normalized by the 1953 population

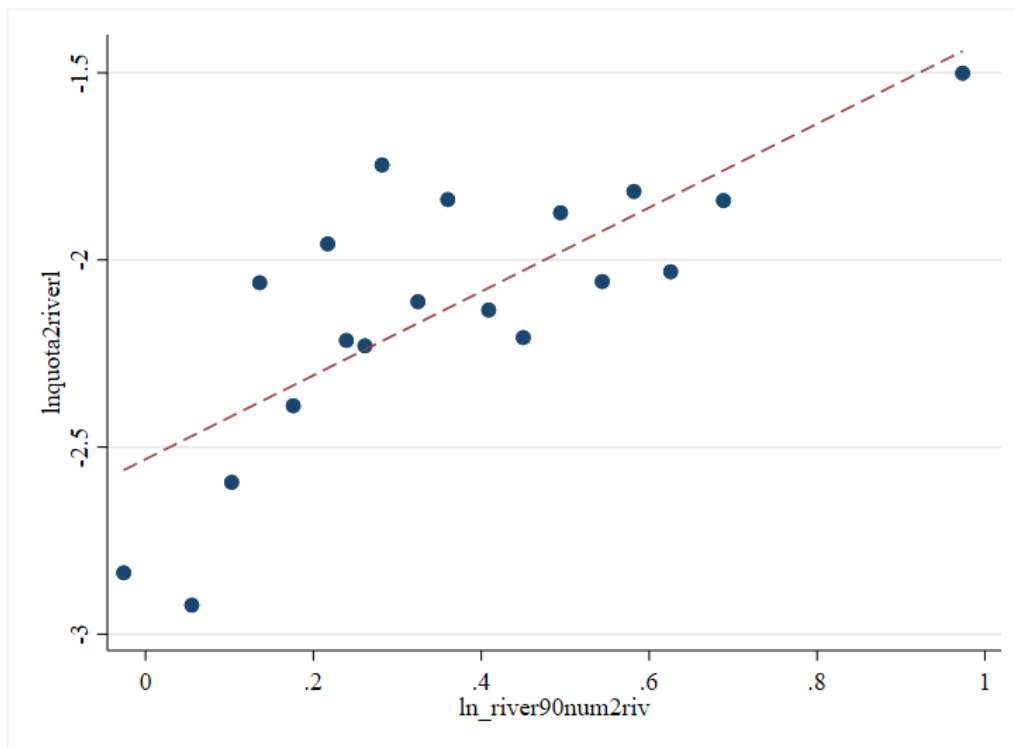


Figure 7: The relationship between the number of quotas divided by river length and the number of small rivers divided by river length, after controlling the length of the main river and the population in 1880

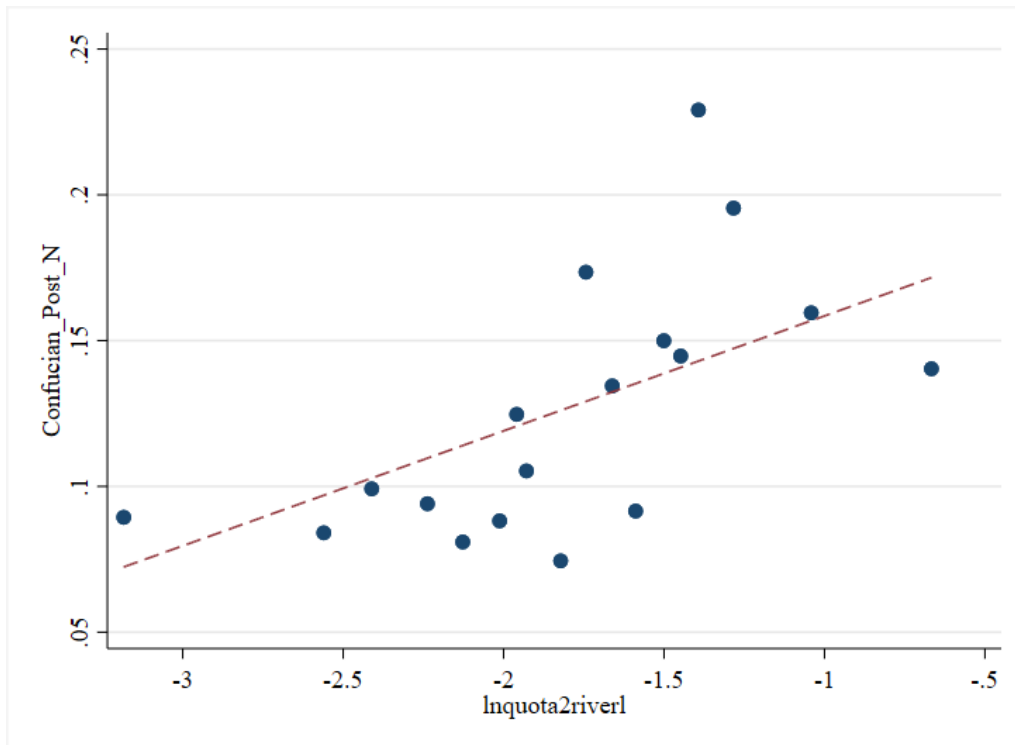


Figure 8: The relationship between the normalised number of Confucian academies built after the quota system and the number of quotas divided by river length, after controlling the length of the main river and the population in 1880

6 Placebo Test

The use of a placebo test is an important methodological consideration in this study. In this case, the number of Buddhist temples in 1820 at the county level is used as a placebo test. This is because Confucianism and Buddhism were two major religions in ancient China, and they have different beliefs and teachings regarding gender roles and son preference. Confucianism places greater emphasis on the importance of family lineage and the role of sons in continuing the family line, while Buddhism has a more neutral stance on gender roles and preference for sons.

By using the number of Buddhist temples as a placebo, the study aims to rule out alternative explanations for the observed relationship between Confucianism and son preference. If there were a significant correlation between the number of Buddhist temples and the sex ratio at birth, this would suggest that the observed relationship between Confucianism and son preference could be attributed to factors other than Confucianism, such as religious affiliation or general cultural norms. Conversely, if no significant relationship between the number of Buddhist temples and the sex ratio at birth is found, it may provide some evidence that the observed relationship between Confucianism and son preference is specific to Confucianism and not due to other factors.

The results are consistent with my hypothesis that historical Buddhism exerted no influence on contemporary sex ratio at birth. In [Table 4](#), we can see that there is no significant relationship between the number of Buddhist temples in 1820 and the sex ratio at birth at county level, with or without different types of controls.

Also, as a robustness check, I also include an OLS result of regressing contemporary sex ratio at birth on both Confucianism and Buddhism, in [Table 12](#), we can see that, even controlling the influence of Buddhism, the effect of Confucianism on contemporary sex ratio at birth is still significant.

Table 4: Buddhism and Contemporary Sex Ratio (OLS)

	(1) SRB	(2) SRB	(3) SRB	(4) SRB
Buddhism	-0.000349 (-0.31)	-0.000211 (-0.17)	-0.0000653 (-0.05)	0.0156 (1.14)
Geographic controls		✓	✓	✓
Historic controls			✓	✓
Contemporary controls				✓
Mean dependent vars	1.11	1.11	1.11	1.11
R^2	0.00	0.04	0.06	0.09
Observations	951	951	951	925

Notes: County-level OLS regressions. Dependent variable is the sex ratio at birth in 2020 at county level. Buddhism is the historical number of Buddhist temples in 1820, normalised by population in 1953. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Buddhism is standardised. t statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

7 Robustness Test

8 Mechanism

9 Conclusion

It has long been hypothesized that the Confucianism played an important role in shaping values and norms in Confucian societies, as well as served as an important determinant of the evolution and persistence of son preference in East Asia. I formally test this hypothesis by combining historical data of Confucian academies at county level with contemporary data, measuring both individuals' attitudes towards family continuation and sex ratio at birth, which implicitly reflects son preferences in China. Specifically, this paper shows that in counties where Confucianism had a strong influence, people living there nowadays tend to have stronger son preferences, by placing higher value on the continuation of the family line, and having a more skewed sex ratio at birth.

By employing instrumental variables and considering historical factors, this paper tries to address potential endogeneity issues and provide stronger support for the causal link between Confucianism and son preference. The findings contribute to the understanding of how deeply ingrained cultural beliefs can influence contemporary attitudes and family decisions.

References

References

Alesina Alberto, Giuliano Paola and Nunn Nathan. 2013. "On the Origin of Gender Roles: Women and the Plough." *The Quarterly Journal of Economics*, 128. 10.2139/ssrn.1856152.

- Alesina Alberto and Giuliano Paola. 2010. "The power of the family." *Journal of Political Economy*, University of Chicago Press, vol. 127(2), pages 560-585.
- Almond, D., Edlund, L., and Milligan, K. 2013. "Son Preference and the Persistence of Culture: Evidence from South and East Asian Immigrants to Canada." *Population and Development Review*, 39(1), 75–95.
- Almond, D., Hongbin Li and Shuang Zhang. 2019. "Land Reform and Sex Selection in China." *Journal of Political Economy*, University of Chicago Press, vol. 127(2), pages 560-585.
- Anderson Siwan. 2007. "The Economics of Dowry and Brideprice." *The Journal of Economic Perspectives*, Vol. 21, No. 4 (Fall, 2007), pp. 151-174 (24 pages).
- Bai, Xinliang. 2012. "The study of academies in the Ming and Qing dynasties (Mingqing Shuyuan Yanjiu)." *The Forbidden City Press (Beijing)*.
- Bai, Ying and Ruixue Jia. 2016. "Elite recruitment and political stability: the impact of the abolition of China's civil service exam." *Econometrica*, 84(2):677–733.
- Chen, Yuyu, et al. 2013. "Prenatal Sex Selection and Missing Girls in China: Evidence from the Diffusion of Diagnostic Ultrasound." *The Journal of Human Resources*, vol. 48, no. 1, 2013, pp. 36–70.
- Chen, Zhiwu, Chicheng Ma and Andrew J Sinclair. 2022. "Banking on the Confucian Clan: Why China Developed Financial Markets so Late." *The Economic Journal*, vol. 132(644), pages 1378-1413.
- Edlund L., Li H., Yi J., Zhang J. 2013. "Sex ratios and crime: Evidence from China." *Review of Economics and Statistics*, 95(5), 1520–1534.
- Fernández, Raquel. 2007. "Women, Work, and Culture." *Journal of the European Economic Association*, Volume 5, Issue 2-3, 1 May 2007, Pages 305–332.

- Jiang, Q., Feldman, M.W. and Li, S. 2021. “Marriage Squeeze, Never-Married Proportion, and Mean Age at First Marriage in China.” *Popul Res Policy Rev*, 33, 189–204.
- Jiang, Q. and Zhang, C. 2021. “Recent sex ratio at birth in China.” *BMJ Global Health*, 2021;6:e005438.
- Kim, Doo-Sub and Yoo-Jean Song. 2007. “Does religion matter? A study of regional variations in sex ratio at birth in Korea.” *The Growing Demographic Female Deficit in Asia*, pp.183–207.
- Li, Hongbin, et al. 2011. “Estimating the Effect of the One-Child Policy on the Sex Ratio Imbalance in China: Identification Based on the Difference-in-Differences.” *Demography*, vol. 48, no. 4, 2011, pp. 1535–57.
- Myong, S., Park, J. and Yi, J. 2018. “Social Norms and Fertility.” *Journal of the European Economic Association*, 19 (5), 2429-2466.
- Nie, Guangyu. 2020. “Marriage squeeze, marriage age and the household savings rate in China.” *Journal of Development Economics*, Volume 147, 2020, 102558, ISSN 0304-3878.
- Tong, Yunping, Christie Sennott, Fenggang Yang. 2020. “Religious Geography and County-Level Sex Ratios in China.” *Journal for the scientific study of religion*, Volume60, Issue1, Pages 113-130.
- Xia Y., Zhou Y., Du L., Cai T. 2020. “Mapping trafficking of women in China: Evidence from court sentences.” *Journal of Contemporary China*, 29, 238–252.
- Zhou, C., Wang, X.L., Zhou, X.D. et al. “Son preference and sex-selective abortion in China: informing policy options.” *Int J Public Health*, 57, 459–465.

Tables

Table 5: Confucianism and Genealogy

	(1) Genealogy	(2) SRB	(3) SRB
Confucianism	142.7** (60.52)	0.0114*** (0.00376)	0.0115*** (0.00380)
Genealogy			-9.34e-08 (0.00000428)
R^2	0.066	0.029	0.026
Observations	359	350	350

Notes: Prefecture-level OLS regressions. Dependent variables are Number of genealogy books in 1201-1896, normalised by population at prefecture level in 1851 and the sex ratio at birth in 2020 at prefecture level. Confucianism is the historical number of Confucian academies from 725 to 1800 at prefecture level, normalised by population in 1851. Confucianism is standardized. t statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Balance Check of the 1st Instrumental Variable.

	(1) Tax Per Capita	(2) Transport	(3) Business	(4) Transportation C
Ln Distance	0.000840 (0.67)	0.000176 (0.02)	0.00712 (1.10)	-0.00370 (-0.78)
Geographic controls	✓	✓	✓	✓
Mean dependent vars	0.0475	0.586	0.536	0.263
Observations	1143	1143	1143	1143

Notes: County-level OLS regressions. Dependent variables are Land tax per capita in 1820, Important in transportation, Important in business, average transportation among counties in 18th century respectively. Ln Distance is the log distance of a county from Zhu Xi Academies in kilometers. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. t statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Balance Check of the 1st Instrumental Variable.

	(1) Fertility	(2) Education	(3) Urban Rural Ratio	(4) Immigration
Ln Distance	0.00947* (1.89)	-0.0225 (-1.24)	-0.121 (-0.81)	-0.193 (-1.17)
Geographic controls	✓	✓	✓	✓
Mean dependent vars	1.657	8.437	2.044	0.695
Observation	1136	1136	1132	1136

Notes: County-level OLS regressions. Dependent variables are fertility rate, average education year, urban rural ratio and immigrants from other provinces ratio in 2020. Ln Distance is the log distance of a county from Zhu Xi Academies in kilometers. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. *t* statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Number of Confucian academies built before the birth of Zhu Xi and the log distance of a county from Zhu Xi academies

	(1) Confucianism	(2) Confucianism	(3) Confucianism	(4) Confucianism
Ln Distance	-0.0433 (-1.24)	-0.0404 (-1.09)	-0.0422 (-1.13)	-0.0476 (-1.15)
Geographic controls		✓	✓	✓
Historic controls			✓	✓
Contemporary controls				✓
Mean dependent vars	0.339	0.255	0.293	0.584
Observations	63	63	63	63

Notes: County-level OLS regressions. Dependent variable is the historical number of Confucian academies before the birth of Zhu Xi (from 725 to 1127), normalised by population in 1953. Ln Distance is the log distance of a county from Zhu Xi Academies in kilometers. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardized. *t* statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Confucianism (Pre) and the number of small rivers

	(1)	(2)	(3)
	Confucian Pre	Confucian Pre	Confucian Pre
Ln #small rivers/River length	0.238 (1.43)	0.237 (1.49)	0.245 (1.39)
Geographic controls	✓	✓	✓
Historic controls		✓	✓
Contemporary controls			✓
Mean dependent vars	0.351	0.351	0.351
Observations	383	382	376

Notes: County-level OLS regressions. Dependent variable is the historical number of Confucian academies from 725 to 1425, normalised by population in 1953. Ln #small rivers/River length is the log value of the number of small rivers with length less than 90km divided by the river length. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardized. *t* statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Buddhism and the number of small rivers

	(1)	(2)	(3)
	Buddhism	Buddhism	Buddhism
Ln #small rivers/River length	0.0191 (1.43)	0.0134 (1.49)	0.0234 (1.39)
Geographic controls	✓	✓	✓
Historic controls		✓	✓
Contemporary controls			✓
Mean dependent vars	0.138	0.138	0.138
Observations	956	956	925

Notes: County-level OLS regressions. Dependent variable is the historical number of Buddhist temples in 1820, normalised by population in 1953. Ln #small rivers/River length is the log value of the number of small rivers with length less than 90km divided by the river length. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Buddhism is standardized. *t* statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Balance Check of the 2nd Instrumental Variable

	(1) Pop	(2) Tax Per Capita	(3) Transport	(4) Business	(5) Transport C
Ln #small rivers/River length	-61263.9 (-1.02)	0.000753 (0.16)	-0.00250 (-0.12)	-0.0304 (-1.43)	-0.0161 (-1.16)
Geographic controls	✓	✓	✓	✓	✓
Mean dependent vars					
Observation	984	984	826	984	826

Notes: County-level OLS regressions. Dependent variables are population in 1880, Land tax per capita in 1820, Important in transportation, Important in business, average transportation among counties in 18th century respectively. Ln #small rivers/River length is the log value of the number of small rivers with length less than 90km divided by the river length. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. *t* statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Confucianism, Buddhism and Contemporary Sex Ratio (OLS)

	(1) SRB	(2) SRB
Confucianism	0.00889*** (4.84)	0.0139*** (5.10)
Buddhism	-0.00168 (-1.43)	-0.0352*** (-3.50)
Contemporary controls		✓
Mean dependent vars	1.11	1.11
R^2	0.018	0.05
Observations	1448	1421

Notes: County-level OLS regressions. Dependent variable is the sex ratio at birth in 2020 at county level. Confucian is the historical number of Confucian academies from 1271 to 1800, normalised by population in 1953. Buddhism is the historical number of Buddhist temples in 1820, normalised by population in 1953. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Both Confucianism and Buddhism are standardised. *t* statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Confucianism and Contemporary Sex Ratio (OLS): Normalised by Population in 2020

	(1) SRB	(2) SRB	(3) SRB	(4) SRB
Confucianism	0.00755** (0.00366)	0.00837** (0.00365)	0.00949*** (0.00361)	0.00861** (0.00362)
Geographic controls		✓	✓	✓
Historic controls			✓	✓
Contemporary controls				✓
Mean dependent vars	1.11	1.11	1.11	1.11
R^2	0.012	0.077	0.102	0.116
Observations	1136	1136	1136	1132

Notes: County-level OLS regressions. Dependent variable is the sex ratio at birth in 2020 at county level. Confucianism is the historical number of Confucian academies from 725 to 1800, normalised by population in 2020. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardized. Standard errors in parentheses are clustered at the prefecture level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Confucianism and Contemporary Sex Ratio (OLS): Standard Error Clustered at Province level

	(1)	(2)	(3)	(4)
	SRB	SRB	SRB	SRB
Confucianism	0.00918** (0.00383)	0.00970** (0.00399)	0.0107*** (0.00342)	0.0115*** (0.00282)
Geographic controls		✓	✓	✓
Historic controls			✓	✓
Contemporary controls				✓
Mean dependent vars	1.11	1.11	1.11	1.11
R^2	0.018	0.086	0.114	0.136
Observations	1136	1136	1136	1132

Notes: County-level OLS regressions. Dependent variable is the sex ratio at birth in 2020 at county level. Confucianism is the historical number of Confucian academies from 725 to 1800, normalised by population in 1953. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardized. Standard errors in parentheses are clustered at the province level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Unnormalised Confucianism and Contemporary Sex Ratio (OLS)

	(1) SRB	(2) SRB	(3) SRB	(4) SRB
Confucianism	0.0134*** (0.00279)	0.0125*** (0.00301)	0.0134*** (0.00285)	0.0140*** (0.00288)
Geographic controls		✓	✓	✓
Historic controls			✓	✓
Contemporary controls				✓
Mean dependent vars	1.11	1.11	1.11	1.11
R^2	0.038	0.096	0.121	0.139
Observations	1136	1136	1136	1132

Notes: County-level OLS regressions. Dependent variable is the sex ratio at birth in 2020 at county level. Confucianism is the historical number of Confucian academies from 725 to 1800. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardized. Standard errors in parentheses are clustered at the prefecture level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Confucianism and Contemporary Sex Ratio between 1 to 4 (OLS)

	(1)	(2)	(3)	(4)
	Sex ratio 1-4	Sex ratio 1-4	Sex ratio 1-4	Sex ratio 1-4
Confucianism	0.00663** (0.00223)	0.00720** (0.00234)	0.00820*** (0.00239)	0.00898*** (0.00251)
Geographic controls		✓	✓	✓
Historic controls			✓	✓
Contemporary controls				✓
Mean dependent vars	1.1	1.1	1.1	1.1
R^2	0.015	0.091	0.124	0.143
Observations	1136	1136	1136	1132

Notes: County-level OLS regressions. Dependent variable is the sex ratio from 1 to 4 in 2020 at county level. Confucianism is the historical number of Confucian academies from 725 to 1800, normalised by population in 1953. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardized. Standard errors in parentheses are clustered at the prefecture level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 17: Confucianism and Contemporary Social Economic Factors (OLS)

	(1)	(2)	(3)	(4)
	Education	Education	Urban Rural Ratio	Urban Rural Ratio
Confucianism	0.0199 (0.0321)	0.0235 (0.0325)	0.0576 (0.0628)	0.134 (0.0989)
Geographic controls		✓		✓
Historic controls		✓		✓
Contemporary controls				
Mean dependent vars	9.13	9.13	10.39	10.39
R^2	0.000	0.151	0.001	0.003
Observations	1136	1136	1132	1132

Notes: County-level OLS regressions. Dependent variables are the average education year and the urban rural ratio in 2020 at county level. Confucianism is the historical number of Confucian academies from 725 to 1800, normalised by population in 1953. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardized. Standard errors in parentheses are clustered at the prefecture level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 18: Confucianism and Contemporary Social Economic Factors (OLS)

	(1)	(2)	(3)	(4)
	Migration	Migration	Fertility	Fertility
Confucianism	0.161 (0.212)	0.238 (0.208)	0.00573 (0.00746)	0.00339 (0.00790)
Geographic controls		✓		✓
Historic controls		✓		✓
Contemporary controls				
Mean dependent vars	5.9	5.9	1.39	1.39
R^2	0.000	0.177	0.000	0.153
Observations	1136	1136	1136	1136

Notes: County-level OLS regressions. Dependent variables are the internal migration rate in percentage from other provinces and the fertility rate in 2020 at county level. Confucianism is the historical number of Confucian academies from 725 to 1800, normalised by population in 1953. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardized. Standard errors in parentheses are clustered at the prefecture level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 19: Confucianism and Contemporary Sex Ratio (2nd IV):
Using the number of quotas divided by the river length directly

	(1) SRB	(2) SRB
Confucian Post	0.0525*** (0.0200)	0.0475** (0.0190)
Mean dependent vars	1.11	1.11
	(1) Confucian Post	(2) Confucian Post
Ln #quotas/River length	0.321*** (0.0960)	0.332*** (0.0982)
Geographic controls	✓	✓
Historic controls	✓	✓
Contemporary controls		✓
Mean dependent vars	-2.141445	-2.141445
Observations	984	979

Notes: County-level IV regressions. Dependent variable is the sex ratio at birth in 2020 at county level. Confucianism is the historical number of Confucian academies from 1425 to 1800, normalised by population in 1953. I employ the log value of the number of quotas with length less than 90km divided by the river length as the IV for the number of Confucian academies. Geographic controls include a dummy variable indicating whether a prefecture is located on the coast, a dummy variable indicating whether it is located on a major river, a dummy variable indicating whether latitude of it is greater than 26.9. Historic controls include a dummy variable indicating whether a prefecture is a large city, population density and tax per capita in 1820 at prefecture level. Contemporary controls include average education year, minority rate, proportion of immigrants from other provinces and urban-rural ratio in 2020 at county level. Confucianism is standardised. *t* statistics in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figures

朝代	省别	书院名称	续表
	贵州 (11所)	平越石壁书院 (七年)、平越中峰书院 (十三年)、贵阳明书院 (十四年)、兴隆卫南山书院 (十五年)、贵阳正学书院 (二十一年)、都匀麟楼书院、镇远紫阳书院、贵阳渔矶书院、△兴隆卫月潭书院、普定卫xx书院、安住卫xx书院	
	四川 (18所)	汉州南轩书院 (元年)、乐山九峰 (高标) 书院 (二年)、富顺学易 (西湖、景阳、江阳) 书院 (四年)、遂宁书台 (武信、斗城、鱼山) 书院、眉州鹤山书院 (九年)、合州合宗 (濠溪) 书院 (十年)、井研崇正书院 (二十一年)、屏山麓山书院 (二十八年)、巴县凝道书院、阆中锦屏书院、长宁文明书院、万县集贤书院、开县三贤书院、中江斗山书院、东乡石鼓书院、△南溪云台书院、南溪南溪 (凤翔) 书院、洪雅修文 (雅江) 书院	

其二、正德、嘉庆两朝修复、重建前代书院 74 所。(见表七、表八)

Period Province

表七 正德朝各地修复、重建前代书院情况表

朝代	省别	书院名称
正德朝 (28所)	直隶 (2所)	毛公书院 (十年)、景县董子书院
	河南 (1所)	宝丰明道书院 (七年)
	山西 (1所)	黎城沧溪书院
	陕西 (1所)	凤翔岐阳 (崇正) 书院
	山东 (1所)	曲阜洙泗书院
	江苏 (2所)	华亭九峰书院、金坛龙山书院 (十一年)
	安徽 (1所)	施州齐山书院 (九年)
	浙江 (2所)	绍兴旧稽山书院、江山逸平 (南塘、正学) 书院

表八 嘉靖朝各地修复、重建前代书院情况表

朝代	省别	书院名称
嘉靖朝 (46所)	河南 (5所)	光山洙水书院 (三年)、禹县儒林书院 (七年)、永城涂蒙书院 (二十一年)、商丘应天书院、登封嵩阳 (太乙、太室) 书院
	山西 (1所)	榆次源池书院 (三年)
	江苏 (6所)	吴县学道书院、长洲和靖书院 (二年)、丹徒准海书院 (三年)、泰州胡公 (安定) 书院 (十七年)、上元南轩书院、上元明道书院
	安徽 (2所)	婺源紫阳 (晦庵) 书院 (九年)、阜阳西湖书院 (十四年)
	浙江 (10所)	镇海湖山书院 (五年)、浦江月泉书院 (九年)、瑞安心极书院 (三十三年)、淳安蜀阜 (旭峰) 书院 (四十三年)、归安长春书院、新昌石鼓 (鼓山) 书院、黄岩南峰书院、永嘉东山书院、奉化广平书院、黄岩九溪书院
	江西 (8所)	新淦高峰书院 (元年)、庐陵白鹭洲书院 (五年)、丰城龙光书院、德兴柳湖书院 (十年)、临川临汝 (南湖) 书院 (三十七年)、上饶御书院 (三十八年)、德兴蒙斋书院、德兴归轩书院

Figure 9: An example of a list of Confucian academies recorded in the book.

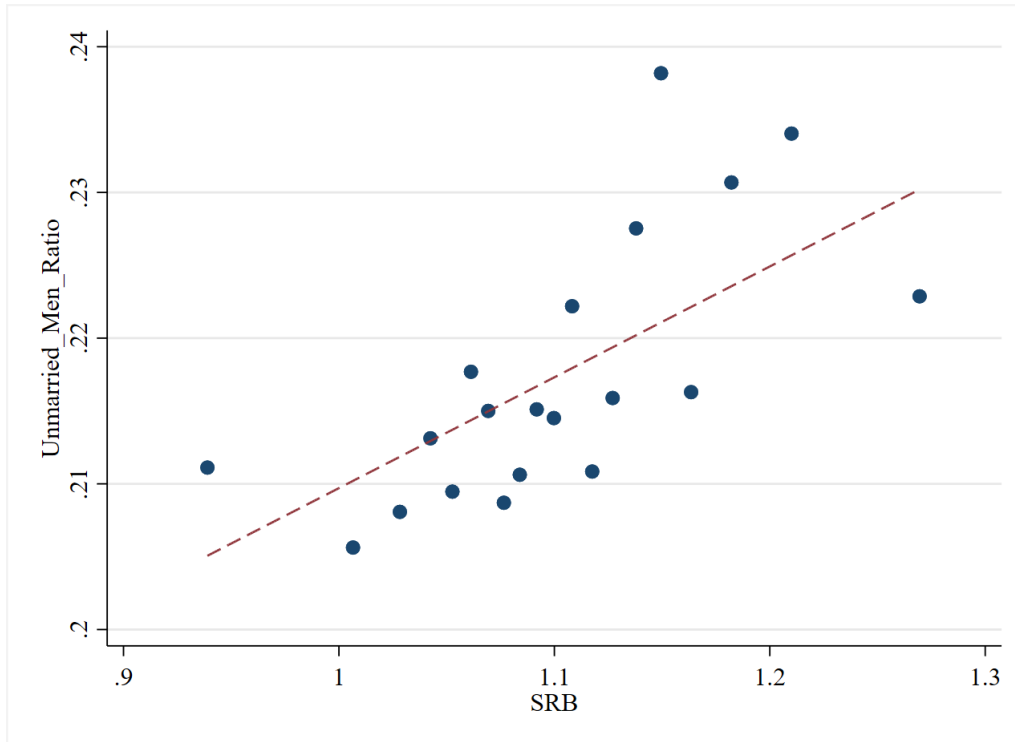


Figure 10: The relationship between the ratio of unmarried marriageable men and the sex ratio at birth at county level.

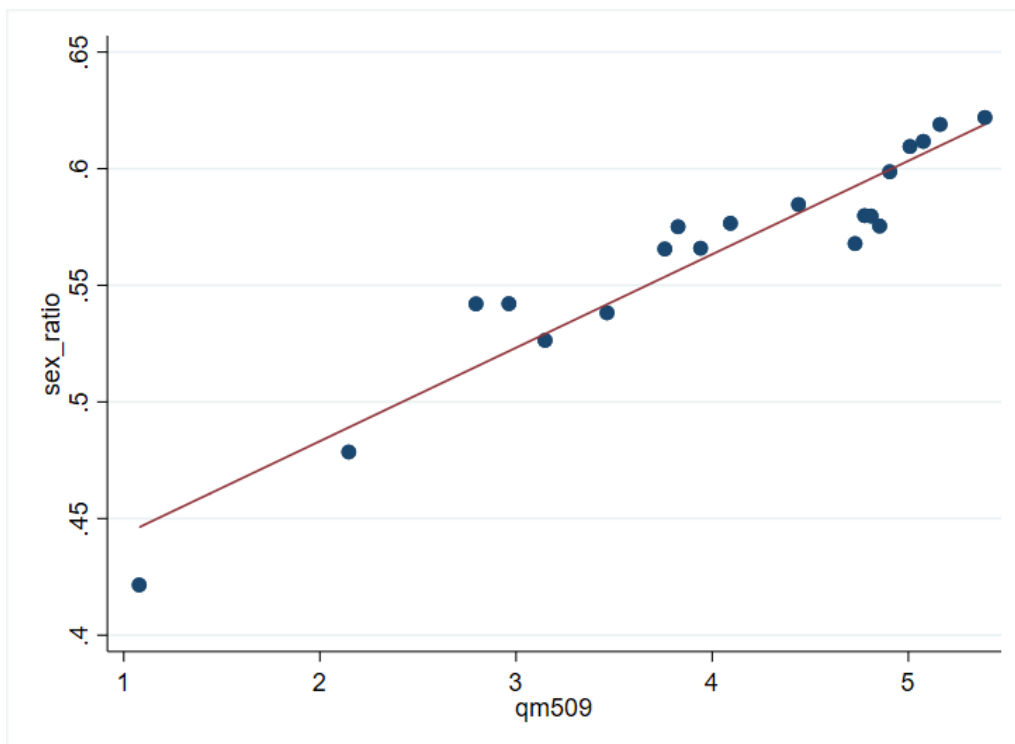
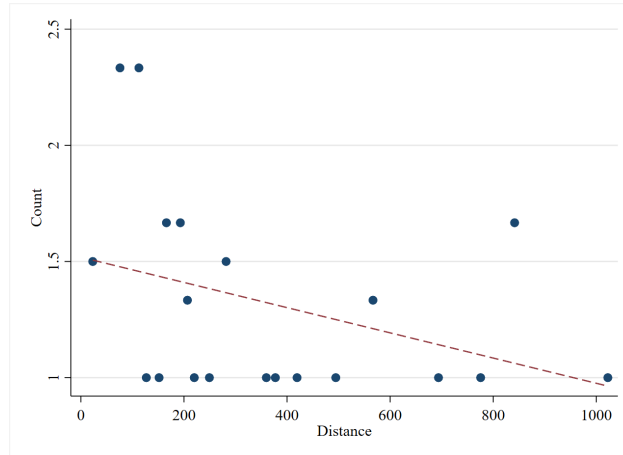
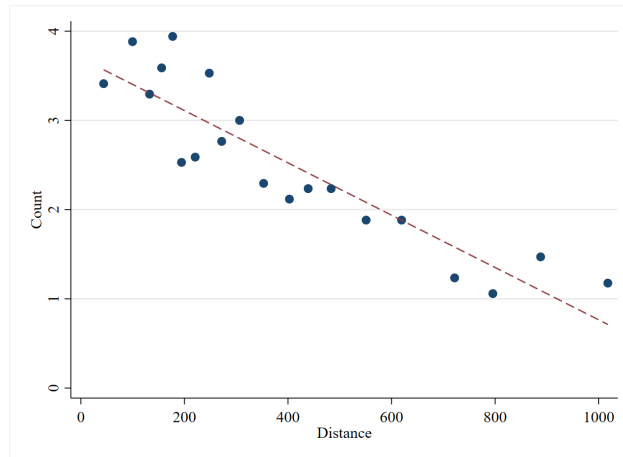


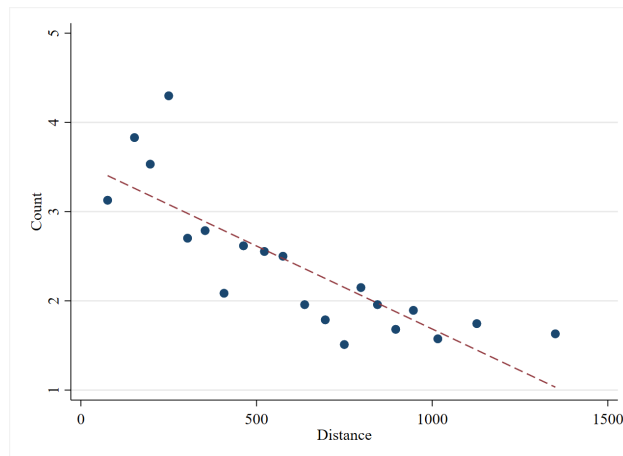
Figure 11: The relationship between the importance of having children to carry on the family name and the sex ratio among children at individual level.



(a) Distance effect on the number of Confucian academies built before Zhu Xi



(b) Distance effect on the number of Confucian academies built from 1127 to 1425



(c) Distance effect on the number of Confucian academies built from 1425 to 1800

Figure 12: Distance effect on the number of Confucian academies

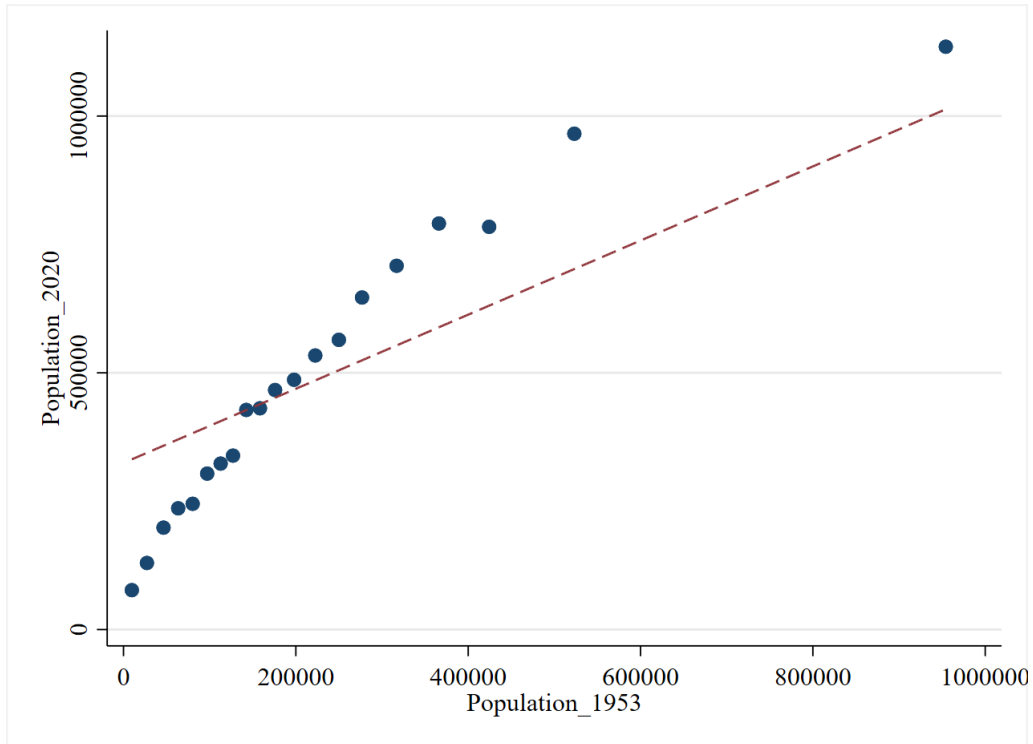


Figure 13: The relationship between the population in 2020 and the population in 1953 at county level.

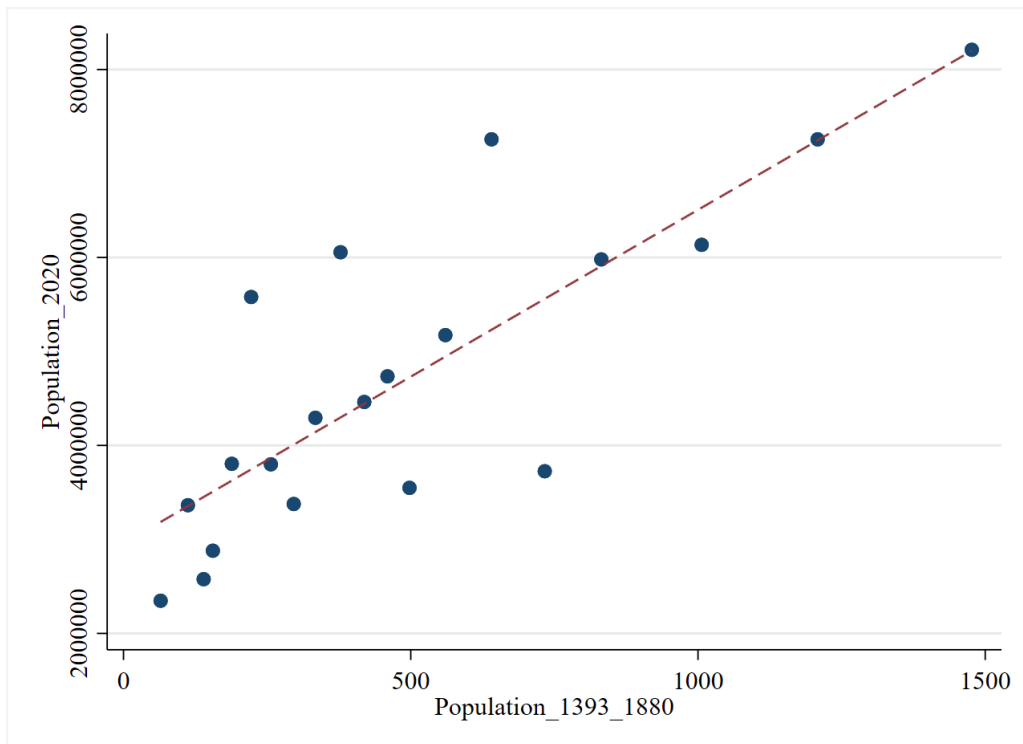


Figure 14: The relationship between the population in 2020 and the average population between 1393 and 1880 at prefecture level.

Appendix A. Placeholder