Early Industrialization and Fertility Patterns in Historical China*

Shuai Zhao† and Arkadiusz Szydłowski‡

February 21, 2024

Abstract

This paper employs the demographic shocks between 1851 and 1880 as the instrumental variable to analyze the causal link between early industrialization and fertility patterns using a unique historical dataset. Our findings present a significantly negative relationship between early industrialization and the number of children but no significant causal link between early industrialization and timing of the first birth in the 19th and early 20th century China. While this indicates that early industrialization could reduce fertility, it also suggests that, in the traditional Chinese society, this reduction does not happen through timing of the first birth being delayed due to late marriage but rather through couples using intra-marital birth control methods. Finally, we find that the opportunity cost of childcare for women and the quality-quantity trade-off are two main channels linking early industrialization to fewer children.

Keywords: Early industrialization, Fertility patterns, Demographic shocks

*We would like to thank Giuseppe De Feo, Chris Vickers and all members in the Theoretical and Applied Microeconomics Cluster at the University of Leicester for offering constructive and critical comments. We are also grateful to Luna Liu, Jiani Zheng, Jiayue Li and Bingyue Tian for their research assistance in transcribing genealogical records.

†Corresponding author, *E-mail*: shuaizhao@email.sdu.edu.cn, *Address*: Center for Economic Research, Shandong University, Jinan, 250100, China

‡*E-mail*: a.szydlowski@kent.ac.uk, *Address*: Sibson Building, School of Economics, University of Kent, Canterbury, CT2 7PE, United Kingdom

1 Introduction

For more than a century, scholars have debated over the relationship between economic growth and demographic transition. The demographic transition theory developed by Landry (1934) and Notestein (1945) suggests the demographic transition is the result of economic growth and the emergence of new institutions in the process of industrialization. According to the demographic transition theory, the industrialization is considered one of the most important prerequisites for the demographic transition. This view has been supported by several empirical studies that find that early industrialization, particularly the expansion of sectors where women had a comparative advantage, may have limited fertility by altering women's fertility patterns, causing them to delay timing of the first birth, marry later, have fewer children, or even choose lifetime celibacy (Spagnoli, 1983; Wanamaker, 2012; Voigtländer and Voth, 2013).

On the other hand, some Western European countries experienced low fertility prior to industrialization which is at odds with the demographic transition theory. Some economic historians argue that delaying timing of the first birth caused by late marriage, fewer children and lifetime celibacy in Europe (coined "European Marriage Pattern") could provide more opportunities for women to accumulate human capital, thus promoting productivity growth (Greif and Tabellini, 2010; Moor and Zanden, 2010; Foreman-Peck and Zhou, 2018). For instance, Foreman-Peck and Zhou (2018) suggest that the existence of the European Marriage Pattern brought about the Industrial Revolution, and it is one of the most important reasons for the 'Great Divergence' between Europe and the rest of the world.[1]

Clearly, the reverse causality between industrialization and fertility patterns is the fundamental problem for empirical tests of demographic transition theory. In order to address this issue, we use a unique instrumental variable, the rate of population change between 1851 and 1880 in China, to investigate the causal link between early industrialization and fertility patterns. Demographic shocks caused by wars, pestilence and natural disasters promote industrialization by changing the factor prices of land and labour and facilitating

[1]The European Marriage Pattern as a contributor to industrialization is questioned by Dennison and Ogilvie (2014), because they suggest that there is not enough evidence for the contribution of the European Marriage Pattern to economic success.

wider production of cash crops used by the industry, thus our instrument is correlated with industrialization levels (also confirmed by our first stage results).[2] Secondly, controlling for the gender ratio, year of birth fixed effects and various regional geographic, socioeconomic and historical characteristics, the population change rate between 1851 and 1880 caused by sudden disasters does not directly affect fertility and timing of the first birth. Our instrumental variables strategy falls into a wide range of studies using exogenous variation due to natural disasters or wars for identification (e.g. Miguel, Satyanath, Sergenti, 2004), in particular using resulting demographic changes as instruments (Hornung, 2014).

In the 19th and early 20th centuries, Chinese society was in stagnation in the Malthusian economy (Chan, 2014; Chen and Kung, 2016). A series of sudden rebellions and natural disasters such as the Taiping Civil War (1851-1864), the Nian Rebellion (1853-1868), the Dungan Revolt (1862-1877), the North China Famine (1876-1879) and so on, led to the most significant population loss Chinese society had ever experienced (Wakins and Menken, 1985; Wakeman, 1997; Cao, 2001; Li and Lin, 2015). Demographic shocks between 1851 and 1880 significantly altered factor prices and increased the land-labour ratio. The development of industry necessitated relatively inexpensive land and adequate labour, which were facilitated by higher labour force participation rates as a result of rising labour wages, as well as falling land prices as a result of more wastelands. Second, the reduced population and food security pressures caused by the demographic shocks allowed farmers to devote more of their arable land to cash crops rather than food crops. This in part provided a relatively more abundant supply of raw materials for early industrial development, especially the light industry. Moreover, more unclaimed land after demographic shocks could attract more immigrants from areas that were minimally affected (Ho, 2013), and the majority of immigrants are young adults which constitute a high quality labour force for industrialization.

As mentioned above, we control for gender ratio and year of birth fixed effects in our instrumental variable regressions, which is important for the validity of our instrument.

[2] Voigtländer and Voth (2013) evidence a similar mechanism in the XIV-XVIII century Europe – wider availability of land due to Black Death facilitated a move from grain production to animal husbandry where women have comparative advantage. As we concentrate on the XIXth century, the livestock farming was not the only alternative to grain production anymore, as production of cash crops for industry constituted an attractive option then. Voigtländer and Voth (2013) also explain why China did not follow the same mechanism as Europe after the plague outbreak in the XIVth century.

Wars and conflicts are able to change the gender ratio in the affected areas, resulting in a relative decline in the male young adult population, thus affecting the marriage market from the supply side. Moreover, wars and natural disasters may have potential trauma effects that impact on fertility decisions. We minimize the importance of these effects by limiting our sample to people born between 1875 and 1925. People usually make fertility decisions in their twenties or later. As a result, even those individuals born in 1875 made their fertility decisions usually after 1895, and this period is far enough from the 1879 (the last year of the North China Famine) to claim that the trauma effects of those extreme events are unlikely to play a role. However, in order to further strengthen the credibility of our identification approach, we include the year of birth fixed effects to control for any potentially remaining trauma effects and any economic policies that could affect fertility during this period.

To analyse the relationship between early industrialization and fertility patterns, we construct a unique dataset by transcribing genealogical records mainly from the Chinese Genealogy Database, the China Industrial Survey Report, the Population History of China (Qing dynasty), the Compilation of Historical Population in the Republic of China and so on. We employ the individual-level number of children and the age of wife at birth of the first son as the dependent variable. The key independent variable is the output per capita at the county level, which equals the county-level total output divided by the county-level number of population. We also control for individual characteristics, socioeconomic, geographic and historical characteristics at the regional level, province FEs and year of birth FEs.

There are two potential mechanisms by which industrialization may have altered fertility patterns of Chinese society in the 19th and early 20th centuries. First, the opportunity costs of child-rearing are borne primarily by women, so women often face a trade-off between child-rearing and work (Becker, 1973; Budig and England, 2001; Loughran and Zissimopoulos, 2009; Do, Levcheko and Raddatz, 2016). Demographic shocks lead to higher labour wages and the industrial development, especially the development of sectors where women have a comparative advantage, which generally makes women more inclined to choose work and thus postpone first birth and have fewer children. Second, as technology advances, the wage gap between adult and child labour widens, raising the cost of child-rearing for parents while lowering the cost of schooling for children (Hazan and Berdugo, 2002). Then, adults tend to invest in child quality over quantity by replacing child labour with child education, lowering the number of children in each household or making more women more likely to delay timing of the first birth. This quality-quantity trade-off has been evidenced by Rosenzweig and Zhang (2009) and Angrist, Lavy and Schlosser (2010), among others.

Our findings present a significantly negative association between early industrialization and the number of children, but no significant causal association between early industrialization and timing of the first birth in the 19th and early 20th century China. A 10% increase in output per capita reduces the number of children by 0.0618, while a 10% increase in output per capita in the textile sector, in which women have a comparative advantage, lowers the number of children by 0.0657. This may suggest that, rather than postponing first birth to restrict the number of children, Chinese couples were more likely to choose intra-marital birth control techniques such as contraception, abortion and even infanticides, especially female infanticides (Lee and Feng, 1999). By contrast, in Europe, due to ineffective contraception, religious restrictions on abortion and less than 1% illegitimacy rate, late marriage was the only feasible way to control fertility (Hajnal, 2017).

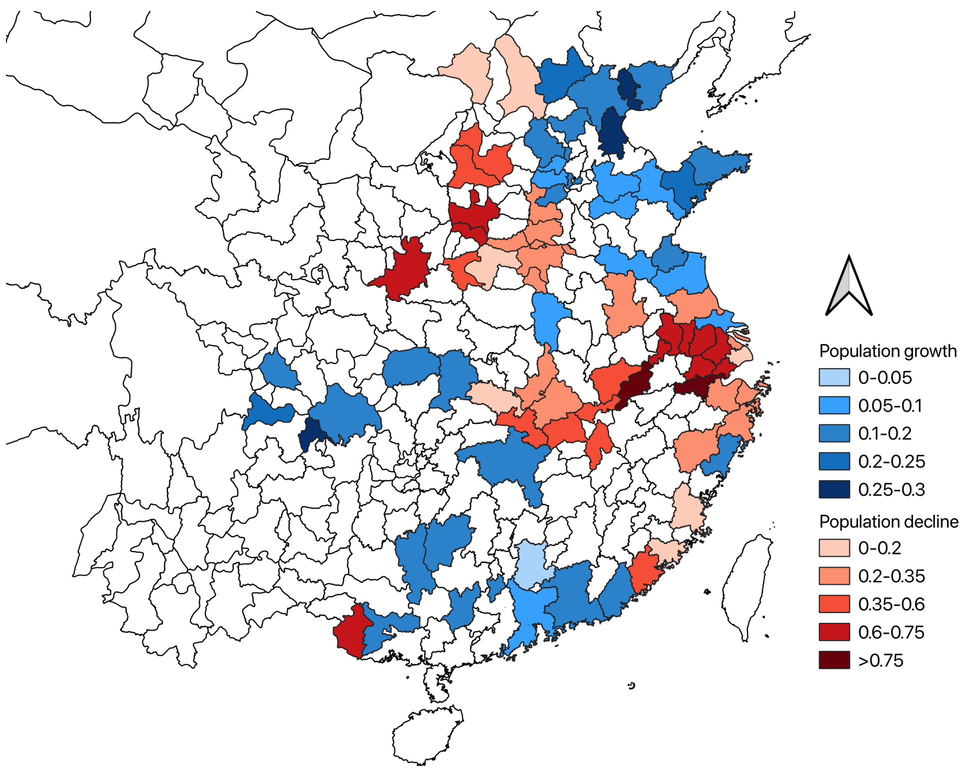
This paper makes at least three contributions to the literature. First, massive transformations in the 19th and early 20th century China give us a very rare opportunity to use demographic shocks between 1851 and 1880 as an instrumental variable. This provides identifying power to analyze the causal link between early industrialization and fertility patterns. Recent articles investigating the relationship between fertility and industrialization include Wanamaker (2012) and Ager, Herz and Brueckner (2020). Both papers establish a link between industrialization and fertility using data from the US. Unlike in these papers, we find that opportunity cost for childcare for women and quantity-quality trade-off are the main channels linking the two, rather than the value of child labour combined with quantity-quality trade-off (Ager, Herz and Brueckner, 2020). Second, in the historical context of China we draw conclusions that depart from many previous studies that focused on other regions (Spagnoli, 1983; Wanamaker, 2012; Voigtländer and Voth, 2013; Hajnal, 2017). In particular, couples can choose to control their fertility in marriage and intra-marital birth control can have similar effects as delaying timing of the first birth by late marriage, a channel predominantly highlighted in the later studies. Third, our study presents potential researchers with a historical dataset that can be used to study reproduction decisions in the early stages of industrialization.

The rest of this paper is organized as follows: Section 2 gives the historical overview and background. Section 3 introduces the identification strategy. Data and empirical results are shown in Section 4 and Section 5, respectively. Finally, Section 6 concludes this paper.

2 Historical overview and background

2.1 Demographic shocks between 1851 and 1880

Figure 1: Prefecture-level rate of population change between 1851 and 1880



Notes: (i) Data source: Prefecture locations are taken from China Historical Geographic Information System (CHGIS) Version 6, and the information on prefecture-level population change rate between 1851-1880 is from Population History of China (Qing Dynasty) compiled by Cao (2001); (ii) The blue color represents prefectures that experienced population growth between 1851 and 1880, while the red color represents prefectures that experienced population decline between 1851 and 1880; (iii) As the hue moves from light to dark, the magnitude of population growth or decline increases; (iv) White regions are those regions where data is missing or cannot be matched to the industrial data.

Chinese society experienced massive population losses due to wars or natural disasters during the period from 1851 to 1880, such as the Taiping Civil War (1851-1864), the Nian Rebellion (1853-1868), the Dungan Revolt (1862-1877), the North China Famine (1876-1879) and so on. In this subsection, we mainly focus on the the Taiping Civil War and the Northern China Famine. This is primarily due to the fact that no other events that caused demographic shocks in the late Qing Dynasty caused a larger scale of population loss and affected a wider area than the Taiping Civil War and the Northern China Famine. Secondly, the industrial data in our study does not include Gansu, Ningxia, and the vast majority of Shaanxi province, which are mainly affected by the Dungan Revolt.

The Taiping Civil War, commanded mostly by Hong Xiuquan, began in early 1851 as an armed uprising against the Qing Dynasty government. The Taiping Army, created by Xiuquan Hong on the basis of God-worshippers[3], was the major armed force. The Taiping Army formally established the Taiping Heavenly Kingdom after capturing Yong'an Prefecture in Guangxi in the autumn of 1851. Following that, the Taiping army fought across most of China, eventually capturing Nanjing city in 1853, renaming it Tianjing, and establishing its capital there. After more than ten years of confrontation with the local forces of the Qing Dynasty government, the Taiping regime came to an end in 1864 when Nanjing, the capital of the Taiping Heavenly Kingdom, was attacked and occupied by the Qing Dynasty army.

The Taiping Civil War wreaked massive economic, demographic, and cultural havoc, seriously undermining the Qing government's foundations. In this subsection, we focus on the impact of the Taiping Civil War on the population in the late Qing Dynasty. In recent years, some scholars have attempted to estimate the population loss caused by the Taiping Civil War. Perkins (2013), for example, claims that the Taiping Civil War resulted in the deaths of over 40 million people in five provinces: Hubei, Zhejiang, Jiangsu, Anhui, and Jiangxi. Cao (2001) evaluated the population loss in the seven provinces most affected by the Taiping Civil War (Hubei, Zhejiang, Jiangsu, Anhui, Jiangxi, Fujian, and Hunan) and concluded that 73.3 million deaths happened as a result of the civil war. Li and Lin (2015) argue that Cao (2001) overestimated the population loss in the seven provinces by extrapolating the causal effect of the war on population loss. According to their findings,

[3] God-worshippers are believers of God-worshipping Society, which is a unique secret association formed by Hong Xiuquan by the absorption of Christian teachings.

the Taiping Civil War resulted in 53 million deaths in the above seven provinces, while an estimated 71 million people across all major war zones were killed. Although scholars disagree on the estimated number of people who died in the Taiping Civil War, it is widely acknowledged that the war resulted in a large number of casualties.

The North China Famine, also known as the Ding-wu Disaster, in addition to the Taiping Civil War, was another event that caused a demographic shock. The North China Famine was caused by a major drought that struck a large area of China's Yellow River basin between 1876 and 1879. In just four years, the drought swept through Shanxi, Henan, Shaanxi, Zhili, and Shandong provinces, and even had some impact on northern Jiangsu, northern Anhui, eastern Gansu and northern Sichuan. The famine claimed the lives of at least 9.5 to 20 million people out of a population of about 108 million in five major provinces in northern China and also led to 20 million people fleeing for food (Watkins and Menken, 1985; Ho, 2013). Shanxi and Henan were the most affected of the five northern provinces. For example, less than 50,000 people remained in Taiyuan Prefecture, Shanxi Province, out of a population of one million (Li et al., 1994).

Figure 1 presents the prefecture-level rate of population change between 1851 and 1880 for those prefectures that can be matched to the counties in industrial data. Hubei, Hunan, Jiangxi, Anhui, Jiangsu, Zhejiang and Fujian Provinces suffered the most severe population losses in southern China. The geographical distribution of population losses in southern China also overlaps highly with the main battlefield areas of the Taiping Civil War. In northern China, the population losses in Shanxi, Henan and Shaanxi provinces, which were affected by the North China Famine, were the highest. However, we do not find huge population losses in Zhili and Shandong provinces. Even in the period 1851-1880, there was a slight increase in population in many prefectures of Zhili and Shandong provinces. This is because the drought began later in Zhili and Shandong provinces, and while it had an impact, it was considerably weaker than in Shanxi, Henan, and Shaanxi provinces, resulting in a population increase between 1851 and 1880 that was greater than the population loss due to draught in these provinces.

2.2 Early industrialization in the 19th and early 20th centuries

2.2.1 The evolution of industrial enterprises in the 19th and early 20th centuries

Early industrialization in China started fairly late and at a lower level than in the West. The process of development took many twists and turns. Early industrialization in China started in the middle of the nineteenth century. By 1949, when the People's Republic of China was founded, industrial output accounted for only 17% of total industrial and agricultural output (Bo, 1990).

Table 1: Top 20 counties in terms of output per capita

All		Textile	
County	Output per capita	County	Output per capita
Shanghai	211.425	Wuxi	63.992
Guangzhou	133.573	Shanghai	53.018
Wuxi	85.916	Wuchang	47.848
Jinan	81.969	Yuci	33.711
Qingdao	58.418	Zhengxian	31.007
Tianjin	53.695	Jixian	24.431
Wuchang	51.523	Xinjiang	28.104
Hangxian	48.744	Hangxian	24.431
Ninghe	36.884	Jinan	19.975
Yuci	35.183	Huolu	19.330
Zhengxian	33.250	Changshan	17.247
Shantou	32.623	Chengdu	17.096
Hankou	32.242	Wujin	15.365
Cangwu	31.585	Taicang	13.127
Xinjiang	30.028	Chongming	13.030
Yangqu	28.957	Nantong	13.002
Jixian	28.535	Qidong	12.946
Nanjing	28.187	Qingdao	11.801
Wujin	26.777	Nanhai	11.781
Huolu	22.544	Wuxian	11.459

Notes: (i) Data source: Both total output value and output value for textile industry are taken from the China Industrial Survey Report, and the county-level population in 1930 is from the Compilation of Historical Population in the Republic of China; (ii) This table reports the top 20 counties in terms of total output per capita and the top 20 counties in terms of output per capita for textile industry in 1930; (iii) County-level output per capita is calculated by dividing the county-level output values of a specific industry by the population of that county; (iv) The output value of cotton, linen, silk, clothing, towels, and other associated products makes up the total output value of the textile sector; (v) The above counties are mainly based on the local administrative divisions of the Republic of China.

In the 19th and early 20th centuries, China's industrial enterprises were divided into three types based on ownership: foreign-owned enterprises, government-supervised merchant undertakings and private enterprises. After losing the first Opium War with Britain in 1842, the Qing government was compelled to sign the Treaty of Nanjing, which opened the ports of Guangzhou, Xiamen, Fuzhou, Ningbo, and Shanghai to international trade, ending a long-standing policy of sealing the country to foreign trade.

Foreign-invested industrial enterprises began to establish and develop in China after the opening of the above five ports for international trade. Foreign-owned enterprises concentrated in shipping, terminal warehousing, ship repair and construction, natural resource extraction, raw material processing and light industry. The distribution of industries served the needs of foreign companies to develop and expand their predatory import and export trade in China (Liang, 2011). In addition to foreign-owned firms, government-supervised merchant undertakings and private enterprises were the two most essential components of early Chinese industry in the 19th and early 20th centuries.

In terms of government-supervised merchant undertakings, it was essential for the Qing government to use private capital to set up modern civil industry. It was financed by merchants and the government-appointed officials managed it (McElderry, 1986). Following the defeat of Britain and France in the Second Opium War (1856-1860), the rulers of the Qing dynasty decided to develop the country and strengthen the army by learning advanced technology in the West to compete with western countries. Hence, the westernisation group, led by Hongzhang Li, Guofan Zeng, and others, began the Self-Strengthening Movement, which lasted more than 30 years. The movement initially concentrated on the development of the military industry through the introduction of cutting-edge Western machinery and equipment, and it gradually extended to the production of civilian goods. Some representative enterprises include the Anqing Ordnance Facility, the Hanyang Ironworks, the Jiangnan Manufacturing Center, the China Merchants Steamship Navigation Company and so on. In 1894, China was defeated in the Sino-Japanese War, the Beiyang Navy was wiped out, and the Self-strengthening Movement came to an end. Nevertheless, those government-supervised merchant undertakings founded during the Self-strengthening Movement were preserved (Yoon, 2015).

Private enterprises typically had intermediaries employed by foreign-owned companies and overseas Chinese as their main investors. The first private companies were established after the beginning of the Self-strengthening Movement and the end of the Taiping Civil War. For example, the Fachang machinery plant was established in 1866, and the Jichanglong silk reeling plant was founded in 1873. The capital size of private industrial enterprises generally did not exceed 100,000 yuan, and in some cases, only a few thousand yuan, and was mainly concentrated in the light industry (Shen, 1999).

2.2.2 Early industrialization after demographic shocks

Early industrialization in China had its first high point in the 1880s. There are four reasons for the first high point of early industrialization, and the first three reasons are directly or indirectly related to demographic shocks between 1851 and 1880. The first reason was a change in factor prices as a result of demographic shocks. Between 1851 and 1880, there was a sharp decline in population in some areas, and the labour shortage raised labour prices, increasing labour participation rates, particularly among women. A large amount of wasteland following the sudden population decline brought down the price of land and provided cheap land for industrial development (Voigtländer and Voth, 2013).

The second reason was the reduction in population pressure. Before the population shocks of the late Qing Dynasty, the registered population of the Qing Dynasty experienced a dramatic increase from the early Qing Dynasty due to land tax reforms and the introduction and spread of high-yielding crops from America. In the late Qing, Chinese society was in stagnation in the Malthusian economy (Chan, 2014; Chen and Kung, 2016). The sharp decline in population relieved population pressure, allowing more arable land to be turned over to cash crops such as cotton, flax, and silkworms, which were all important raw materials for the light industry.

The third reason was the increased number of immigrants. A large amount of land left untouched by natural and man-made disasters attracted many immigrants from places that were minimally affected by the population shocks, and the majority of immigrants were predominantly young adults, who were a good labour force for industrial development. For example, according to a report in Shenbao (one of the oldest newspapers in China), many migrants immigrate to Hangzhou, Jiaxing and Huzhou in 1882 from those provinces minimally affected by the Taiping Civil War (1851-1864). More than half of the resettlers were soldiers from the Xiang Army or the Huai Army[4], therefore the majority of them were

[4]The Xiang Army or Huai Army was a local army established by Zeng Guofan or Li Hongzhang in the late

young and strong labourers (Li, 2015).

The fourth reason, which is largely unrelated to demographic shocks, was the collapse of the subsistence farming economy[5], facilitated by dumping of industrial goods from Europe. Following the end of the First Opium War in 1842, when the Qing government was forced to abandon its closed-door policy, many cheap industrial goods from Europe were imported into China, contributing to the collapse of the subsistence farming economy. This, in turn, freed up much of the labour force bound to the land (Bradby, 1975).

Through these channels, demographic shocks enabled industries, especially those industries in which women have a comparative advantage, to grow. Table 1 and Table 2 present top 20 counties in terms of output per capita and capital per capita for all sectors and just textile sector. The majority of counties are located in those prefectures experiencing large demographic shocks. For example, in Table 1, Guangzhou, Jinan, Qingdao, Tianjin, Ninghe, Shantou, Jixian and Huolu out of the top 20 counties in terms of output per capita measured by product in all sectors did not experience demographic shocks, while only Jixian, Huolu, Qingdao and Nanhai out of the top 20 counties in terms of output per capita measured by textile sector product did not experience demographic shocks between 1851 and 1880.

For our analysis, we select four representative sectors from registered factories in the China Industrial Survey Report, which are the textile, printing, grain processing and metal smelting. In the early stage of industrialization, the development of heavy industry, including steel, non-ferrous metallurgy, machine building and so on, was limited by a higher barrier to entry due to the inadequate accumulation of technology, capital and skilled workers. By contrast, light industry, such as textile, grain processing, printing and similar industries, had a lower entry barrier and were therefore more widely distributed at the county level in the Republic of China than heavy industry. For instance, 92 out of 135 counties had the textile sector (i.e. at least one textile factory), and 80 out of 135 counties had the grain processing sector in Table 3. While the percentage of counties with a printing sector

Qing Dynasty to quell the Taiping Civil War (1851-1864). Each army derives its name from the place of origin of its top generals and soldiers (Bai, Jiang and Yang, 2023).

[5] In imperial China, the subsistence farming economy has three main characteristics. First, the subsistence farming economy is primarily family-based, with a small production scale. Second, it features a very straightforward labour division, with men farming and women weaving. Finally, with a combination of agriculture and cottage industry and little commercial exchange, it is self-sufficient.

Table 2: Top 20 counties in terms of capital per capita

All		Textile	
County	Capital per capita	County	Capital per capita
Shanghai	55.453	Yuci	28.898
Qingdao	38.049	Xinjiang	27.870
Cangwu	29.974	Shanghai	18.308
Yuci	29.397	Wuchang	14.140
Daye	27.089	Wuxi	13.585
Yangqu	25.767	Zhengxian	13.253
Hangxian	21.307	Tianjin	11.514
Jinan	21.268	Jixian	11.118
Tangshan	20.301	Huolu	10.603
Jiujiang	18.783	Hangxian	8.241
Shantou	17.558	Jiujiang	6.929
Tianjin	17.442	Jinan	6.906
Guangzhou	17.128	Qingdao	6.437
Ninghe	16.986	Qidong	5.902
Wuxi	15.646	Haimen	5.262
Wuchang	15.149	Nantong	4.189
Zhengxian	14.731	Chongming	3.889
Jiangpu	13.959	Chengdu	3.396
Fushan	12.981	Wuhu	3.129
Xiamen	12.511	Taicang	3.104

Notes: (i) Data source: Both total capital value and capital value for textile industry are taken from the China Industrial Survey Report, and the county-level population in 1930 is from the Compilation of Historical Population in the Republic of China; (ii) This table reports the top 20 counties in terms of total capital per capita and the top 20 counties in terms of capital per capita for textile industry in 1930; (iii) County-level capital per capita is calculated by dividing the county-level capital values of a specific industry by the population of that county; (iv) The capital value of cotton, linen, silk, clothing, towels, and other associated products makes up the total capital value of the textile sector; (v) The above counties are mainly based on the local administrative divisions of the Republic of China.

is slightly lower, it is still more widely distributed than the percentage of counties with a metal smelting sector.

The share of female workers for all sectors was 26.7%, while the share of child labour is 15.5%. These numbers suggest that the employment of female workers, and even child labour, was common in the early stage of industrialization in the 19th and early 20th centuries after demographic shocks in China. Despite this, there are significant differences in female and child labour participation rates across sectors. For example, the textile sector has 40.2% female workers, while the printing industry has only 0.72%. In terms of child labour, the participation rate of child labour in the printing industry is 37.4%, compared to 2.8% for grain processing.

Table 3: Composition of workforce across four main sectors

	Share of female workers (%)		Share of child labour (%)		
Sector name	Mean	St. dev.	Mean	St. dev.	Counties
Textile	40.191	32.743	19.978	22.864	92
Printing	0.720	4.153	37.350	20.736	52
Grain processing	0.808	4.224	2.871	11.361	79
Metal smelting	1.010	4.116	41.466	14.869	40
All	26.916	27.337	15.649	17.311	134

Notes: (i) Data source: Both share of female workers and share of child labour in total workers are taken from the China Industrial Survey Report; (ii) The data above for each sector is aggregated from the data for the product to which it corresponds.

3 Data

To study the relationship between early industrialization and fertility patterns by using the demographic shocks in the late Qing dynasty as the instrumental variable, we matched genealogical records primarily from the Chinese Genealogy Database[6] with industrial data from the China Industrial Survey Report and population data from both Compilation of Historical Population in the Republic of China and Population History of China (Qing Dynasty). We also transcribed education data from Education Statistics of the Republic of China for analyzing mechanisms behind the relationship between industrialization and fertility patterns. Finally, our main sample covers 5,054 genealogical records found in 134 counties or 85 prefectures. To preserve representativeness of our sample, regression analyses are population weighted based on county-level population in 1930 from the Compilation of Historical Population in the Republic of China. Table 4 summarizes the main sources of data used.

The main source of our genealogical records is the Chinese Genealogy Database, which was created by the Central China Normal University, and it was officially opened to the public in 2018. The Chinese Genealogy Database has collected 8.44 TB of digital genealogy data (digital images), a total of 120,893 volumes, up to now. Collected genealogies in this database span over 400 years from the mid-Ming Dynasty to the current period, and this database covers plenty of lineages in the majority of counties in China. The construction of

[6]See http://gd.ccnu.edu.cn. Chinese genealogies have been used in previous studies by Telford (1992), Zhao (1997) and Shiue (2016).

this database has provided a great convenience for genealogical research. As some of the counties could not be matched with industrial data using the Chinese Genealogy Database, thus we supplemented a small portion of genealogical records from other online resources. Namely, we supplemented our data with some of genealogies preserved in libraries or local archives by using the information on the Shanghai Library Chinese Genealogy Knowledge Service Platform[7].

A sample page of a typical Chinese genealogy is shown in Figure 2. As can be seen there, a typical Chinese genealogy is usually presented in the form of male mini-biographies. Each male mini-biography contains individual's name, birth order of the individual among all his brothers, birth and death dates, father's name, wives' and concubines' surnames, wives' and concubines' birth and death dates, the number of sons and the number of daughters. A typical Chinese genealogy does not record details of daughters other than the number of daughters. The location and completion date of the genealogy are also disclosed. The information in the mini-biographies enables us to connect family members from different generations in the same lineage to one another based on the names of the father, making it feasible to calculate the age of couples at birth of the first son. The location of genealogy allows us to match the genealogical records in different counties to the county-level industrial data, county-level educational data, county-level and prefecture-level population data and other controls used in our study. We randomly selected lineages in each county for the majority of lineages in our sample, but a small portion of lineages in some counties were selected mainly on the basis of availability.

Due to the limitations of our sample to the birth dates between 1875 and 1925, for a small number of counties we could not match the industrial data to any of the records in the existing genealogies in the county. In these instances we imputed records from neighbouring counties which lay in the same prefecture as counties within the same prefecture are generally homogeneous in terms of their characteristics.

[7]See https://jiapu.library.sh.cn. The Shanghai Library Chinese Genealogy Knowledge Service Platform mainly displays genealogical information, such as the names, dates and storage locations, rather than digitized images in the Chinese Genealogy Database, so the majority of the genealogies we use to supplement our dataset are not directly accessible from this platform. However, the information available on this platform provides solid assurance of the high quality and reliability of our genealogical records from other online sources.

As mentioned above, all individuals of our genealogical records were born between 1875 and 1925. The choice of this period is motivated by two factors. First of all, this is due to the fact that early industrialization was a process that started in China following the First Opium War (1842). The early industrialization process had an impact on the fertility patterns of those who were born after this war, as well as those of their spouses. Second, this choice ensures that the fertility or childbearing decisions made by the individuals and their spouses in our sample occurred after the demographic shocks of 1851 to 1880.

In addition, in case where an individual only had one wife, we did not include those individuals whose wife deceased before the age of 35 or who had not yet reached the age of 35 when the genealogy was finished in order to obtain a more complete fertility history. We also made sure that if a man had more than one spouse (through remarriage or concubinage), at the time the genealogy was finished, at least one of the spouses had reached the age of 35. Although women usually lose their fertility in their late forties or early fifties, we lowered the age restriction to 35 due to the low average life expectancy at that time, driven by the fact that modern medical care was not sufficiently widespread in the late 19th and early 20th century China and that the majority of typical families could not afford it.

Out of 5,054 transcribed genealogical records, the number of genealogical records with the number of children is 4,015, which is due to some volumes of genealogies lacking information on the number of daughters, which is essential for us to calculate the total number of children in the family. Due to the fact that some males were lifetime single, or the couple had no son, or we failed to trace the information regarding the birth dates of wife or concubine or the first son, 4,139 records with the age of wife at birth of the first son were obtained.

The second source of our data is the China Industrial Survey Report, which was published by the Military Commission of the Government of the Republic of China. This report covers 146 counties or cities and almost all industrial products that registered factories could produce in the Republic of China in 1930s. From the China Industrial Survey Report, we could transcribe information on the output value, capital value, number of workers, number of male workers, number of female workers and number of child labour at the county level for each product. Additionally, we chose the sectors of textiles, printing, grain processing, and metal smelting that made up a sizable portion of Republic of China production in the 1930s, and we added up the data for all of the products in each sector. For example, the capital value corresponding to production of cotton, linen, silk, clothing, towels, and other associated products makes up the total capital value of the textile sector. Furthermore, we evaluate whether the chosen sector has a comparative advantage for women by calculating the ratio of female workers to total workers. By merging the industrial data with county-level population data from the Compilation of Historical Population in the Republic of China, we derived output per capita and capital per capita for each representative sector. To attenuate the effect of outliers, we took the natural logarithm of output per capita and capital per capita (plus 1)[8] for each representative sector as key measures of the level of early industrialization.

The two main sources of population data are the Compilation of Historical Population in the Republic of China and Population History of China (Qing Dynasty). The Compilation of Historical Population in the Republic of China is compiled by Yin and Tian (2009). It mainly covers population surveys conducted by the Ministry of the Interior of the Republic of China in various provinces, population and household surveys conducted by provincial governments, relevant periodicals run by the Ministry of the Interior and so on. We collect the population size, gender ratio, population density and other relevant information at the county level. Because the completion dates of the population data collection in different provinces vary, the data collection spans over 10 years, but all population data is for the period of 1930s, the survey year of the China Industrial Survey. For example, population density and population size are mainly derived from the National Land and Population Survey of Cities and Counties, which was published in 1935, although the majority of survey was completed in 1931 or 1932. Furthermore, gender ratio was transcribed from the Household and Population Statistics, which the Ministry of Interior issued in 1938 and supplemented and re-estimated population statistics from earlier surveys. Based on the aforementioned data, male and female population sizes were calculated.

Another source of population data is the Population History of China (Qing Dynasty) compiled by Cao (2001). This volume restores prefecture-level population data for various

[8]We have also tried $\log(\cdot + 0.1)$ and our main conclusions remain unchanged with that normalisation.

provinces in 1776, 1820, 1851, 1880, and 1910 using official Qing dynasty history books, Qing literary collections, notes, local chronicles, and archival materials. In addition, estimated population losses caused by the Taiping Rebellion (1851-1864), the Nian Rebellion (1853-1868), the Dungan Revolt (1862-1877), the North China Famine (1875-1879) and other historical events in different prefectures are discussed in detail in the Population History of China (Qing Dynasty), which allows us to make better use of demographic shocks over the period 1851-1880 to construct an instrumental variable to address the endogeneity of early industrialization. We transcribed the prefecture-level population data in 1851, 1880 and 1910 and employed the population change rate between 1851 and 1880 as the main instrumental variable. For the robustness checks, we calculate the population change rate between 1851 and 1910 as an alternative instrumental variable.

The source of education data is the Education Statistics of the Republic of China compiled by Wang (2010). It covers many aspects of higher, secondary and primary education[9] and social education in the Republic of China. As higher education and even secondary education were not sufficiently widespread in China at that time, many counties lacked universities, and even secondary school students were scarce. Therefore, we mainly transcribed the number of students enrolled in junior primary and senior primary schools at the county level in 1918. We use the educational data mainly to verify the mechanism of quality-quantity trade-off theory. Finally, 116 counties could be matched from education data to our master dataset.

The above are the main data sources and their corresponding descriptions. Descriptions of the other control variables used and their corresponding sources are given in subsection 4.3 and Appendix B.

[9]Primary education in the Republic of China was undertaken by junior primary schools and senior primary schools. Junior primary schools require students to complete four years of study, after which they decide whether to progress to a higher primary school for three years. Separate diplomas are awarded for both.

Table 4: Main data sources

Type	Date	Level	Main sources
Genealogy	1875-1925	Individual	Chinese Genealogy Database
Industry	1930s	County	China Industrial Survey Report
Population	1851, 1880, 1910	Prefecture	Population History of China (Qing Dynasty) compiled by Cao (2001)
Population	1930s	County	Compilation of Historical Population in the Republic of China compiled by Yin and Tian (2009)
Education	1918	County	Education Statistics of the Republic of China compiled by Wang (2010)

4 Identification strategy

4.1 Empirical model

In this section, we employ both ordinary least squares (OLS) regression and two-stage least squares (2SLS) regression to investigate the relationship between early industrialization and fertility patterns. Our ordinary least squares (OLS) regression is shown as follows:

$$M_{i,j} = \beta_0 + \beta_1 Industrialization_{i,j} + \beta_2 X_i + \beta_3 W_j + \mu_p + \gamma_t + \epsilon_{i,j} \tag{1}$$

where $M_{i,j}$ is the number of children for individual i in county j or the age of his wife at birth of the first son, and $Industrialization_{i,j}$ denotes output per capita for county j. X_i contains individual i' s number of marriages, an indicator for whether individual i is the first son and an indicator for whether individual i remarried with other women after he divorced with his wife or his wife deceased. W_j includes a range of county and prefecture characteristics. μ_p are province fixed effects, and γ_t are year of birth fixed effects. Standard errors are clustered at the county level. To preserve representativeness of our sample, regression analyses are population weighted.

In addition, the first stage of two-stage least squares (2SLS) regression is shown as follows:

$$Industrialization_{i,j} = \theta_0 + \theta_1 populationchange_{i,j} + \theta_2 X_i + \theta_3 W_j + \mu_p + \gamma_t + v_{i,j} \quad (2)$$

where $populationchange_{i,j}$ denotes the population change rate during the period between 1851 and 1880, which is employed as the instrumental variable in our model. Other variables in Equation (2) are defined previously. $Industrialization_{i,j}$ is constant at the county level, and $populationchange_{i,j}$ is constant at the prefecture level.

4.2 Potential reverse causality and exclusion restrictions

The critical aspect of the empirical strategy is how to deal with the reverse causality problem. Early industrialization will increase the number of occupations available to women and expose them to higher opportunity costs for childbearing than in the previous era, especially as sectors develop in which they have a comparative advantage. Thus, women are more likely to make the decision to delay timing of the first birth by late marriage or have fewer children (Spagnoli, 1983; Wanamaker, 2012; Voigtländer and Voth, 2013). Some researchers, on the other hand, argue that women delaying childbearing by late marriage, having fewer children, and living a life of celibacy could promote productivity and industrialization because women constitute a crucial workforce for industrial development (Greif and Tabellini, 2010; Moor and Zanden, 2010; Foreman-peck and Zhou, 2018).

To study the causal link between early industrialization and fertility patterns, we employ a unique instrumental variable, the rate of population change between 1851 and 1880. This is because demographic shocks such as wars, plagues, and natural disasters can accelerate industrialization and economic expansion by altering land and labour factor prices (Voigtländer and Voth, 2013), and the population change rate caused by sudden disasters should not directly affect fertility or childbearing patterns when controlling for gender ratio, year of birth fixed effects and regional characteristics.

Demographic shocks between 1851 and 1880 significantly altered factor wages and increased the land-labour ratio. The industrial development necessitated relatively inexpensive land and adequate labour, which were facilitated by higher labour force participation rates as a result of rising labour prices, as well as falling land prices as a result of more wastelands. Second, the reduced population and food security pressures caused by the demographic shocks allowed farmers to devote more of their arable land to cash crops rather than food crops. This, in part, provided a relatively more abundant supply of raw materials for early industrial development, especially the light industry. Without the demographic shocks, the huge population of agricultural China would consume the vast majority of its resources, leaving no surplus to develop industry. In addition, more unclaimed land after demographic shocks could attract more immigrants from areas that were minimally affected (Ho, 2013), and the majority of immigrants are young adults that are good labour force for industrialization.

We control for gender ratio and year of birth FEs to guarantee that the population change rate between 1851 and 1880 is uncorrelated with the omitted variables. Wars and conflicts could alter the gender ratio in affected areas, leading to a drop in the number of young male adults and, as a result, influencing the marriage market on the supply side. Furthermore, the year of birth FEs from 1875 to 1925 could control for the possible trauma effects of wars and natural catastrophes on the fertility decisions of individuals experiencing demographic shocks between 1851 and 1880, and any potential economic policies that could affect fertility during this period. Additionally, the interval of birth years of individuals in our genealogical sample enables the instrumental variables to be valid. The individuals in our sample were born between 1875 and 1925. Fertility decisions are typically made in one's twenties, and in some circumstances much later. As a result, even for people born in 1875, fertility decisions were typically made after 1895, and this period is long from 1879 (the final year of the North China Famine); thus, trauma effects from such extreme events are unlikely to play a role.

4.3 Description of variables

4.3.1 Dependent variables and independent variables

Number of children and age of wife at birth of the first son. We mainly investigate the effect of early industrialization on fertility patterns. Our dependent variables are, therefore, mainly the number of children within marriage and age of wife at the first birth, where the timing of the first birth is replaced by the age of wife at birth of the first son, which is highly related. Although the age of wife at birth of the first son will systematically overestimate the true age of wife at the first birth, the resulting measurement error should be small and not systematically related to individual characteristics. Thus, overall the age of wife at birth of the first son should serve as a good proxy for the age of wife at the first birth. Another reason for using the age of wife at birth of the first son is that most genealogies only briefly record the number of daughters but do not record details such as the date of birth. In contrast, information on sons as male members in each household is complete. It should be noted that when the first son is not born to the first wife, but to another wife or concubine, the age of the wife at birth of the first son is actually the age of the biological mother at birth of the corresponding first son.

Output per capita. We use output per capita to measure the level of industrialization, which is calculated by the county-level total output values divided by the number of population in each county in 1930.

4.3.2 Baseline controls

It is necessary to control for a number of covariates that are likely to impact upon both fertility and marriage age. We classify our baseline controls into individual controls and socioeconomic controls.

Gender ratio. We control for gender ratio defined as the ratio of men to women in the county population. Demographic shocks, particularly those resulting from the Taiping Civil War, can lead to changes in the county-level population gender ratio, causing the male population to decline as a proportion of the total population in those counties affected by the Taiping Civil War or the Northern China Famine. The scarcity of male resources in the marriage market can help women delay their timing of the first birth due to late marriage while also giving men an advantage in the marriage market (Abramitzky, Delavande and Vasconcelos, 2011). Hence, controlling for gender ratios at the county level is important for the validity of the instrumental variables in our model.

Number of marriages and remarriage. In Ancient China, the marriage system is a system between monogamy and polygamy. An adult man can have a wife and multiple concubines at the same time, according to the law. Although a concubine's social status is lower than that of a wife, all sons born to concubines and wives are recorded in the genealogy.

First-born son. The existence of an effect of birth order on fertility has been confirmed by several empirical articles (Murphy and Knudsen, 2002; Hatton and Martin, 2010; Kulu et al., 2017). In China, children in the highest birth order, particularly the first son, are expected by their parents to assume greater responsibility for the family's succession, which may affect their fertility. We include an indicator if the husband is the first-born son as the control variable. Unfortunately the dataset does not allow us to identify if the husband is the first-born child, i.e. eldest among all siblings. Firstly, this is because part of the genealogy only records the male members of the family and their wives and concubines, while the sisters of the male members are not recorded. Secondly, although the vast majority of genealogies record the names of the male member's sisters, other information about the female member, especially the date of birth, is not recorded. This makes it possible to identify whether male members are first born or not only by their date of birth in relation to their brothers, without knowing their actual birth order among all their siblings.

Share of child labour in population. Share of child labour is defined as the number of children employed in the county divided by county's population size. The effect of child labour on fertility varies at different stages of industrialization (Hazan and Bedugo, 2002). The wage income child labour generates for households may act as a motivator for families to have more children when child labour legislation is non-existent (Basu and Van, 1998). However, after technology advances, the wage gap between adults and children could promote parents to invest in children quality over quantity by replacing child labour with child education, then lowering the number of children parents choose to have. This quality-quantity trade-off has been proved by many scholars (Rosenzweig and Zhang, 2009; Angrist, Lavy and Schlosser, 2010).

Population density. We control for county-level population density in 1930, which is measured by the number of people per square li (li is a traditional Chinese unit of distance, which is standardized as a half-kilometre). Population density is an important measure of the marriage market density as it affects the meeting rate between partners (see e.g. Bellou 2015, Battistin, Becker and Nunziata 2022).

Urbanization rate. We additionally control for the 1920 urbanization rate because urbanization is typically associated with rising pay rates and declining fertility rates, and women are more likely to postpone childcare or marriage during this period of urbanization and development due to more employment possibilities and improved women's economic status (Sato and Yamamoto, 2005).

4.3.3 Geographic and historical controls

In addition to the baseline controls described above, we include some geographic and historical control variables, which have an effect on fertility decisions and, at the same time, may also be correlated with our instrument as they affect to which extent wars and natural disasters translate into population losses. This helps to support validity of our instrument.

Distance to coast and distance to major navigable rivers. The late Qing dynasty, after losing a series of wars with the West, gradually opened its doors at the behest of the victorious West. The order is from the opening of coastal port cities to the opening of riverine cities on important navigable rivers. For this reason, the closer a city is to the coastline and major navigable waterways, the more likely it is to have been influenced by Western thought and culture earlier. This will challenge the dominance of traditional Confucianism as well as clan culture, thereby affecting the way families behave in terms of child-rearing and women's decision-making in marriage (Stasson, 2021). Distance to the coast is measured as the distance between a prefecture's centroid to the closest point on the coast, whereas distance to major navigable rivers is the distance between a prefecture's centroid to the closest point on major navigable rivers.

Rice suitability, wheat suitability, maize suitability, sweet potato suitability and economic crops suitability. The construction of these crop suitability indices takes into account the plant's eco-physiological characteristics, climate and environmental requirements. A higher value for a specific crop usually suggests better conditions for growth. Rice, wheat, maize and sweet potatoes are the most important food crops for the Chinese, and the suitability of these food crops can reflect the level of the agricultural economy of a prefecture. Meanwhile, farming is a physically demanding work, and field ploughing is best performed by males. In light of this, adult females were more inclined to choose to marry adult males at a younger age in areas more suitable for production of these crops, in addition to parents favoring sons and having more children (Fan and Wu, 2023). In addition, we controlled for the cash crops suitability, which matters for industrialization and the fertility decision of females.

Tea center in Ming-Qing period. Tea was one of the most valuable internationally traded commodities in ancient China. Whether or not a place was historically a centre for tea manufacturing can give us an insight into the historical economic prosperity of the area, which had a significant impact on the number of children a family could afford in an agricultural society. Furthermore, women have a comparative advantage in tea cultivation and harvesting, which can greatly influence women's economic status and thus their marital and reproductive decisions (Qian, 2008). To determine if a prefecture was a tea center during the Ming and Qing eras, we employ a binary variable.

Clan strength in Ming-Qing period. During a period of comparatively low productivity, particularly in the countryside, small families' or individuals' ability to cope with risks was limited. In the face of natural disasters or wars, individuals or small families usually suffered devastating blows. The clan, a collection of blood or karmic ties of many small families with the same or different surnames, played a role in the ancient Chinese countryside, where the influence of imperial power could not reach below the county level, in the provision of public goods and in the sharing of risks among small families. Due to the clan's function, it also had the authority to regulate family size and marriage patterns. Clans as a social structure persisted in rural China after the fall of the empire, and they still have an impact on decisions made by members of the clan regarding pregnancy and marriage in early 20th-century republican China and now (Zhang, 2019). We control for the (logged) number of genealogies standardized by population in each prefecture to measure the prefecture-level clan strength. Similar proxy variables can be found in Chen et al. (2022) and Fan et al. (2023).

Confucian academies in Ming-Qing period. Culture is an informal institution that has a long-lasting and profound effect on people within a society. This influence is inherited by subsequent generations and has a significant impact on how individuals behave and make decisions. Confucianism, one of the core elements of traditional Chinese culture, has shaped the traditional Chinese institution of marriage and fertility patterns, with the belief that one of the primary goals of marriage is to perpetuate the family line and thus preserve the family heritage (Ko, Haboush and Piggott, 2003). The logarithm of prefecture-level number of Confucian academies is used as the control variable in our study.

Distance to national capital and distance to provincial capitals in Ming-Qing period. Distance to the capital or a provincial capital is a proxy for historical political influence. Industrialization and female decision-making over marriage and procreation are both impacted by this influence. The impacts of this influence were long-lasting and many of them persisted after the Qing dynasty collapsed. For example, in the late Qing Dynasty, the Self-strengthening Movement initiated by the upper echelons of the ruling class established many government-supervised merchant undertakings, which were an important part of China's national industry and occupied a monopoly position in many sectors (Yoon, 2015). Furthermore, most of these government-supervised merchant undertakings were not compelled to close due to the collapse of the Qing Dynasty; rather, they functioned as the foundation and backbone of the nation's industrial development, both in the early twentieth century and to nowadays.

5 Empirical results

5.1 Descriptive analysis

In this subsection, descriptive statistics for main variables used are present in Table 5, while the information about other control variables is given in Table 14 in Appendix B. For those individuals born between 1875 and 1925 in Table 5 , 68% of children is the first son of their parents. In terms of marriage, after the dissolution of their marriage to their wife or concubine or after the passing of their wife or concubine, 10% of men remarry. Moreover, the average number of marriages was 1.09, which includes those who remarried. Despite the fact that a man could have a wife and multiple concubines in traditional Chinese society at that time, men with multiple spouses were a small minority in that society. In addition, individuals born between 1875 and 1925 have an average of 3.1 children, with 1.8 sons and 1.32 daughters. The difference between the number of sons and the number of daughters

Table 5: Summary statistics

Variable	Mean	St. dev.	Min	Max	Observations
		Panel A. Genealogical data			
Individual level:					
First son (1=yes, 0=no)	0.68	0.47	0	1	5,054
Remarriage (1=yes, 0=no)	0.10	0.30	0	1	5,054
Number of marriages	1.09	0.39	0	6	5,054
Number of sons	1.80	1.33	0	9	5,054
Number of daughters	1.32	1.35	0	10	4,015
Number of children	3.10	2.07	0	12	4,015
Number of children per woman	2.99	1.98	0	12	3,902
Age of wife at birth of the first son	24.39	5.67	12	54	4,139
Age at birth of the first son	27.28	6.79	11	73	4,219
		Panel B. Population data			
County level:					
Gender ratio	1.22	0.16	0.92	1.89	134
Population (in 10,000)	50.95	40.16	4.65	344.20	134
Population density (per sq. li)	189.12	678.06	10	6924	133
Prefecture level:					
Population in 1851 (in 10,000)	305.07	172.22	59.8	798.1	85
Population in 1880 (in 10,000)	250.02	175.16	38.4	684.7	85
Population change rate between 1851 and 1880 (%)	−15.72	32.64	−84.32	29.02	85
		Panel C. Education data			
County level:					
Number of current senior primary students	486.24	511.04	21	3195	116
Number of current junior primary students	5043.38	4186.26	256	23720	116
Percent of current senior primary students (%)	0.103	0.071	0.007	0.345	116
Percent of current junior primary students (%)	1.257	0.966	0.088	5.168	116

Notes: (i) The population density in 1930 is measured in number of people per square li (li is a traditional Chinese unit of distance, which is standardized as a half-kilometre).

reflects the son preference in traditional Chinese society. The age at which a wife has her first son is 24.39 compared to 27.28 for a husband. Although there is no data on the actual age at the first birth for both wife and husband, the age at birth of the first son, which is highly correlated with the age at the first birth, allows us to infer that women marry earlier than men.

In terms of population data, we have both county-level population data and prefecture-level population data from two different sources. The average prefectural population figures for the three time points of 1851, 1880 and 1910 provide us with a clear picture of China's demographic changes during the last 59 years of the Qing dynasty. Between 1851 and 1880, the average population at the prefecture level fell from 3,050,700 to 2,500,200, owing to the Taiping Civil War (1851-1864), the Nian Rebellion (1851-1868), the Dungan Revolt (1862-1877), the North China Famine (1876-1879) and so on. The drop in population by 15.72% between 1851 and 1880 also reflects the dramatic decreases in opulation at the prefecture level.

With respect to education data, we present the number of current senior primary students, the number of current junior primary students, the percent of current senior primary students and the percent of current junior primary students. The average number of pupils enrolled in primary schools in the 116 counties is 5,043.38, but the average number of students enrolled in senior primary schools is just 486.24. The corresponding proportion also decreases from 1.257% to 0.103%. These suggest that, after completing their junior primary education, the vast majority of students do not choose to continue their studies at the senior primary level.

5.2 OLS results

Before presenting the instrumented results for the relationship between early industrialization and fertility patterns, OLS results, which do not take the potential reverse causality into consideration, are presented first. Table 6 shows the OLS estimates for the industry measured by textile sector product (Columns 1-3) and industry measured by product in all sectors (Columns 4-6) with or without covariates. Panel A and Panel B show the correlation between early industrialization and actual number of children, as well as age of wife at birth of the first son. Columns 1-6 in Panel A show a statistically significant negative relationship between early industrialization and number of children. The magnitude of coefficients for the industry measured by textile sector product is slightly stronger. After controlling for geographic characteristics, historical characteristics, socioeconomic characteristics, individual characteristics, province FEs and year of birth FEs, an 1% increase in (log) output per capita for the industry measured by textile sector product could reduce the number of children by 0.00234 in column (3), while an 1% increase in (log) output per capita for the industry measured by product in all sectors could reduce the number of children by 0.00219 in column (6).

In terms of the age of wife at birth of the first son, the total level of early industrialization in all sectors has a positive effect on the marriage age in column (4), although the significance of the coefficients are not robust to controlling for different covariates. The development of the textile sector, where women have a comparative advantage, shows that an 1% increase in (log) output per capita for the industry measured by textile sector product

Table 6: Early industrialization and fertility patterns: OLS results

Variable	Textile (1)	Textile (2)	Textile (3)	All (4)	All (5)	All (6)
Panel A. Dependent variable: number of children						
(log) output per capita in textile sector	-0.320^{***}	-0.245^{***}	-0.234^{***}			
	(0.056)	(0.060)	(0.056)			
(log) output per capita				-0.305^{***}	-0.233^{***}	-0.219^{***}
				(0.054)	(0.064)	(0.061)
Geographic controls	X	X	X	X	X	X
Historical controls	X	X	X	X	X	X
Socioeconomic controls		X	X		X	X
Individual controls			X			X
Province FEs	YES	YES	YES	YES	YES	YES
Year of birth FEs	YES	YES	YES	YES	YES	YES
Number of clusters	123	123	123	123	123	123
Observations	4,013	4,013	4,013	4,013	4,013	4,013
R^2	0.126	0.131	0.172	0.126	0.130	0.171
Panel B. Dependent variable: age of wife at birth of the first son						
(log) output per capita in textile sector	0.379^{**}	0.447^{**}	0.438^{**}			
	(0.148)	(0.178)	(0.179)			
(log) output per capita				0.222^{**}	0.242	0.235
				(0.110)	(0.160)	(0.160)
Geographic controls	X	X	X	X	X	X
Historical controls	X	X	X	X	X	X
Socioeconomic controls		X	X		X	X
Individual controls			X			X
Province FEs	YES	YES	YES	YES	YES	YES
Year of birth FEs	YES	YES	YES	YES	YES	YES
Number of clusters	134	133	133	134	133	133
Observations	4,139	4,119	4,119	4,139	4,119	4,119
R^2	0.064	0.067	0.073	0.061	0.063	0.070

Notes: (i) ***, **, * indicate statistical significance at the 1, 5 and 10 percent level, respectively; (ii) Standard errors in parentheses are clustered at county level; (iii) The data of cotton, linen, silk, clothing, towels, and other associated products makes up the data of the textile sector; (iv) Geographic controls include distance to coast (in km), distance to major navigable rivers (in km), rice suitability, wheat suitability, maize suitability, sweet potato suitability and economic crops suitability; (v) Historical controls include Tea center in Ming-Qing, distance to national capital in Ming-Qing (in km, logged), distance to provincial capitals in Ming-Qing (in km, logged), strength of clan in Ming-Qing (logged) and Confucian academies in Ming-Qing (logged); (vi) Socioeconomic controls include population density (per sq. li), share of child labour in population, gender ratio and urbanization rate (logged); (vii) Individual controls include number of marriages, a remarriage dummy and a first son dummy.

could delay age at first birth (measured by the age of wife at birth of the first son) for women by 0.00438 years in column (3).

As the OLS estimates do not pick up causal effects when potential reverse causality between early industrialization and fertility patterns is not considered, we are moving to instrumented results in the next subsection.

5.3 Instrumented results

This section presents the instrumented results for the causal link between early industrialization and fertility patterns in Table 7. We first show the effect of early industrialization on the number of children in Panel A, and then its effect on the age of wife at birth of the first son in Panel B. As our estimates may suffer from a weak instrument problem, in addition to presenting Kleibergen-Paap F-statistic, we also report the Anderson-Rubin test statistic in the bottom part of each panel, which is robust to weak instruments (Stock and Yogo,2005; Andrews, Stock and Sun, 2019).

From columns (1) to (6) in Panel A, there is a negative and significant relationship between early industrialization and the number of children for both when the industry is measured by product in all sectors and textile sector product. For the textile sector, where women have a comparative advantage, the coefficient of (log) output per capita in column (2) is slightly larger than the corresponding coefficient in column (1) after further controlling for a range of socioeconomic controls. The number of children born to households decreases by 0.00787 for every 1% increase in output per capita for the textile sector in column (2).

Further controlling for individual characteristics, the level of early industrialization in the textile sector has a slightly smaller impact on fertility in column (3). An 1% increase in (log) output per capita for the industry measured by textile sector product could reduce the family size by 0.00657. This suggests that, for those parents from different households and counties, the number of children they had is also affected by the circumstances of different counties and households. Because of these, later on we mainly refer to the coefficients in column (3) and column (6) when reporting results in Table 7.

For the industry measured by product in all sectors, we find similar changes in the magnitude of the coefficients in columns (4) to (6) compared to the results for the industry mea-

Table 7: Early industrialization and fertility patterns: instrumented results

Variable	Textile (1)	Textile (2)	Textile (3)	All (4)	All (5)	All (6)
Panel A. Dependent variable: number of children						
Dependent variable: number of children						
(log) Output per capita in textile sector	-0.764^{***}	-0.787^{***}	-0.657^{**}			
	(0.253)	(0.291)	(0.254)			
(log) Output per capita				-0.632^{***}	-0.738^{***}	-0.618^{***}
				(0.197)	(0.261)	(0.224)
Geographic controls	X	X	X	X	X	X
Historical controls	X	X	X	X	X	X
Socioeconomic controls		X	X		X	X
Individual controls			X			X
First stage						
Dependent variable: (log) output per capita						
Population change rate between 1851 and 1880 (%)	-0.021^{***}	-0.019^{***}	-0.019^{***}	-0.026^{***}	-0.020^{***}	-0.020^{***}
	(0.006)	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)
Province FEs	YES	YES	YES	YES	YES	YES
Year of birth FEs	YES	YES	YES	YES	YES	YES
Anderson-Rubin Wald test	20.83^{***}	12.69^{***}	10.09^{***}	20.83^{***}	12.69^{***}	10.09^{***}
Kleibergen-Paap F-statistic	14.43	14.90	14.78	17.59	13.23	13.15
First stage partial R^2	0.070	0.072	0.072	0.093	0.081	0.080
Number of clusters	123	123	123	123	123	123
Observations	4,013	4,013	4,013	4,013	4,013	4,013
Panel B. Dependent variable: age of wife at birth of the first son						
Dependent variable: age of wife at birth of the first son						
(log) Output per capita in textile sector	0.926	0.921	0.940			
	(0.636)	(0.726)	(0.719)			
(log) Output per capita				0.764	0.922	0.939
				(0.503)	(0.720)	(0.711)
Geographic controls	X	X	X	X	X	X
Historical controls	X	X	X	X	X	X
Socioeconomic controls		X	X		X	X
Individual controls			X			X
First stage						
Dependent variable: (log) output per capita						
Population change rate between 1851 and 1880 (%)	-0.017^{***}	-0.016^{**}	-0.016^{**}	-0.021^{***}	-0.016^{**}	-0.016^{**}
	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Province FEs	YES	YES	YES	YES	YES	YES
Year of birth FEs	YES	YES	YES	YES	YES	YES
Anderson-Rubin Wald test	2.36	1.78	1.91	2.36	1.78	1.91
Kleibergen-Paap F-statistic	7.73	6.09	6.07	9.09	5.41	5.43
First stage partial R^2	0.036	0.038	0.037	0.054	0.045	0.045
Number of clusters	134	133	133	134	133	133
Observations	4,139	4,119	4,119	4,139	4,119	4,119

Notes: (i) ***, **, * indicate statistical significance at the 1, 5 and 10 percent level, respectively; (ii) Standard errors in parentheses are clustered at county level; (iii) The data of cotton, linen, silk, clothing, towels, and other associated products makes up the data of the textile sector; (iv) Geographic controls include distance to coast (in km), distance to major navigable rivers (in km), rice suitability, wheat suitability, maize suitability, sweet potato suitability and economic crops suitability; (v) Historical controls include Tea center in Ming-Qing, distance to national capital in Ming-Qing (in km, logged), distance to provincial capitals in Ming-Qing (in km, logged), strength of clan in Ming-Qing (logged) and Confucian academies in Ming-Qing (logged); (vi) Socioeconomic controls include population density (per sq. li), share of child labour in population, gender ratio and urbanization rate (logged); (vii) Individual controls include number of marriages, a remarriage dummy and a first son dummy.

sured by textile sector product, although the magnitude of coefficients of (log) output per capita in the industry measured by product in all sectors is smaller than the corresponding coefficients for the industry measured by textile sector product. In column (6), the decrease in the number of children is 0.00618 for every 1% increase in (log) output per capita measured by product in all sectors.

In terms of the age of the wife at birth of the first son, there is no causal effect between the level of early industrialization and the age of the wife at birth of the first son, which is different from the OLS results in Table 6. This insignificant relationship indicates Chinese families had fewer children in the period of early industrialization is more likely due to women opting for intra-marital birth control rather than marry later to postpone the first birth[10]. By contrast, some scholars suggest the main reason for demographic transition in the Europe is the female decision to marry later as a result of economic growth. However, it is also worth noting that age of wife at the first birth is not precisely measured in the data, and this may also lead to a lack of significance.

All the results discussed above are confirmed when we use weak-instrument-robust Anderson-Rubin test to analyze significance of the effect of industrialization. By comparing the OLS and 2SLS estimates for the number of children, in Table 6 and Table 7 relevant results show an upward bias of the OLS estimates. The bias with reverse causality is usually not straightforward to sign, especially with multiple control variables. However, in a simple simultaneous equations two-variable model, the positive bias in OLS would be consistent with a negative causal effect of industrialization on fertility and a positive causal effect of fertility on industrialization.[11]

In addition, from both Panel A and Panel B of Table 7, the first-stage results also present a significantly negative relationship between the rate of population change between 1851 and 1880 and output per capita, which is in line with our argument for the validity of our instrumental variable.

[10]Lee and Feng (1999) found evidence of not only artificial birth control such as contraception and abortion but also of enormous infanticides, especially of female infanticides, in traditional Chinese society.

[11]Note that if $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ and $X_i = \alpha_0 + \alpha_1 Y_i + u_i$, then $cov(X_i, \epsilon_i) = \frac{\alpha_1 Var(\epsilon_i)}{1-\alpha_1\beta_1}$ and the sign of $cov(X_i, \epsilon_i)$ determines the sign of the OLS bias. Now, if $\alpha_1 > 0$ and $\beta_1 < 0$, then we have a positive bias.

5.4 Robustness checks

In this section, we assess the robustness of our estimates by clustering standard errors at the prefecture level and making use of alternative independent and dependent variables . Given that there is no causal effect between the level of early industrialization and age of wife at birth of the first son in the main regression, we mainly focus on the robustness of causal effect between early industrialization and number of children in this section.

Table 8: Alternative independent variable: capital per capita

	Textile			All		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Dependent variable: number of children						
Dependent variable: number of children						
(log) Capital per capita in textile sector	−1.247**	−1.314**	−1.090**			
	(0.505)	(0.590)	(0.494)			
(log) Capital per capita				−0.626***	−0.698**	−0.584**
				(0.224)	(0.276)	(0.239)
Geographic controls	X	X	X	X	X	X
Historical controls	X	X	X	X	X	X
Socioeconomic controls		X	X		X	X
Individual controls			X			X
First stage						
Dependent variable: (log) capital per capita						
Population change rate between 1851 and 1880 (%)	−0.013***	−0.011***	−0.011***	−0.026***	−0.021***	−0.021***
	(0.004)	(0.004)	(0.004)	(0.006)	(0.005)	(0.005)
Province FEs	YES	YES	YES	YES	YES	YES
Year of birth FEs	YES	YES	YES	YES	YES	YES
Anderson-Rubin Wald test	20.83***	12.69***	10.09***	20.83***	12.69***	10.09***
Kleibergen-Paap F-statistic	9.35	8.93	9.05	16.42	16.05	16.06
First stage partial R^2	0.054	0.053	0.053	0.127	0.115	0.114
Number of clusters	123	123	123	123	123	123
Observations	4,013	4,013	4,013	4,013	4,013	4,013

Notes: (i) ***, **, * indicate statistical significance at the 1, 5 and 10 percent level, respectively; (ii) Standard errors in parentheses are clustered at county level; (iii) The data of cotton, linen, silk, clothing, towels, and other associated products makes up the data of the textile sector; (iv) Geographic controls include distance to coast (in km), distance to major navigable rivers (in km), rice suitability, wheat suitability, maize suitability, sweet potato suitability and economic crops suitability; (v) Historical controls include Tea center in Ming-Qing, distance to national capital in Ming-Qing (in km, logged), distance to provincial capitals in Ming-Qing (in km, logged), strength of clan in Ming-Qing (logged) and Confucian academies in Ming-Qing (logged); (vi) Socioeconomic controls include population density (per sq. li), share of child labour in population, gender ratio and urbanization rate (logged); (vii) Individual controls include number of marriages, a remarriage dummy and a first son dummy.

First, we use capital per capita at the county level instead of output per capita to measure the level of early industrialization in Chinese society. We get similar findings from Table 8 as in Table 7, although the magnitude of capital per capita coefficients is larger for industries measured by textile sector capital and is slightly smaller for capital in all sectors.

In terms of the alternative dependent variable, we use the number of children per woman to measure fertility, which is equal to number of children per household divided by number

Table 9: Alternative dependent variable: number of children per woman

	Textile			All		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Dependent variable: number of children per woman						
Dependent variable: number of children per woman						
(log) Output per capita in textile sector	-0.629^{***}	-0.633^{**}	-0.617^{**}			
	(0.210)	(0.242)	(0.243)			
(log) Output per capita				-0.524^{***}	-0.597^{***}	-0.580^{***}
				(0.161)	(0.214)	(0.213)
Geographic controls	X	X	X	X	X	X
Historical controls	X	X	X	X	X	X
Socioeconomic controls		X	X		X	X
Individual controls			X			X
First stage						
Dependent variable: (log) output per capita						
Population change rate between 1851 and 1880 (%)	-0.022^{***}	-0.019^{***}	-0.019^{***}	-0.026^{***}	-0.020^{***}	-0.020^{***}
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Province FEs	YES	YES	YES	YES	YES	YES
Year of birth FEs	YES	YES	YES	YES	YES	YES
Anderson-Rubin Wald test	19.74^{***}	11.69^{***}	9.77^{***}	19.74^{***}	11.69^{***}	9.77^{***}
Kleibergen-Paap F-statistic	14.10	14.54	14.51	16.98	12.89	12.95
First stage partial R^2	0.070	0.072	0.071	0.093	0.080	0.080
Number of clusters	123	123	123	123	123	123
Observations	3,902	3,902	3,902	3,902	3,902	3,902

Notes: (i) ***, **, * indicate statistical significance at the 1, 5 and 10 percent level, respectively; (ii) Standard errors in parentheses are clustered at county level; (iii) The data of cotton, linen, silk, clothing, towels, and other associated products makes up the data of the textile sector; (iv) Geographic controls include distance to coast (in km), distance to major navigable rivers (in km), rice suitability, wheat suitability, maize suitability, sweet potato suitability and economic crops suitability; (v) Historical controls include Tea center in Ming-Qing, distance to national capital in Ming-Qing (in km, logged), distance to provincial capitals in Ming-Qing (in km, logged), strength of clan in Ming-Qing (logged) and Confucian academies in Ming-Qing (logged); (vi) Socioeconomic controls include population density (per sq. li), share of child labour in population, gender ratio and urbanization rate (logged); (vii) Individual controls include number of marriages, a remarriage dummy and a first son dummy.

of wives in the corresponding household. Comparing Table 9 to Table 7, we do not see any stark differences in the relationship between the level of early industrialization and the number of children.

Finally, as the instrumental variable (population change rate) varies at the prefecture level while the independent variable (output per capita) varies at the county level, we present estimates with standard errors clustered at the prefecture level where the variation of the instrumental variable occurs in order to account for within prefecture correlation and heteroskedasticity. Our conclusions are unaffected by findings presented in Table 10.

5.5 Placebo tests

In this subsection, we conduct a placebo test on three representative sectors, including printing, grain processing and metal smelting, where women do not have a comparative advan-

Table 10: Standard errors clustered at prefecture level

Variable	Textile (1)	(2)	(3)	All (4)	(5)	(6)
Panel A. Dependent variable: number of children						
Dependent variable: number of children						
(log) Output per capita in textile sector	-0.764^{***}	-0.787^{**}	-0.657^{**}			
	(0.260)	(0.305)	(0.274)			
(log) Output per capita				-0.632^{***}	-0.738^{***}	-0.618^{***}
				(0.188)	(0.253)	(0.229)
Geographic controls	X	X	X	X	X	X
Historical controls	X	X	X	X	X	X
Socioeconomic controls		X	X		X	X
Individual controls			X			X
First stage						
Dependent variable: (log) output per capita						
Population change rate between 1851 and 1880 (%)	-0.021^{***}	-0.019^{***}	-0.019^{***}	-0.026^{***}	-0.020^{***}	-0.020^{***}
	(0.006)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
Province FEs	YES	YES	YES	YES	YES	YES
Year of birth FEs	YES	YES	YES	YES	YES	YES
Anderson-Rubin Wald test	23.57^{***}	16.80^{***}	12.17^{***}	23.57^{***}	16.80^{***}	12.17^{***}
Kleibergen-Paap F-statistic	15.18	16.51	16.46	23.66	20.91	20.99
First stage partial R^2	0.070	0.072	0.072	0.093	0.081	0.080
Number of clusters	82	82	82	82	82	82
Observations	4,013	4,013	4,013	4,013	4,013	4,013

Notes: (i) ***, **, * indicate statistical significance at the 1, 5 and 10 percent level, respectively; (ii) Standard errors in parentheses are clustered at prefecture level; (iii) The data of cotton, linen, silk, clothing, towels, and other associated products makes up the data of the textile sector; (iv) Geographic controls include distance to coast (in km), distance to major navigable rivers (in km), rice suitability, wheat suitability, maize suitability, sweet potato suitability and economic crops suitability; (v) Historical controls include Tea center in Ming-Qing, distance to national capital in Ming-Qing (in km, logged), distance to provincial capitals in Ming-Qing (in km, logged), strength of clan in Ming-Qing (logged) and Confucian academies in Ming-Qing (logged); (vi) Socioeconomic controls include population density (per sq. li), share of child labour in population, gender ratio and urbanization rate (logged); (vii) Individual controls include number of marriages, a remarriage dummy and a first son dummy.

tage. Because women bear the majority of childcare costs, the expansion of sectors in which women have a comparative advantage, as well as an increase in the level of industrialization of the society will affect women's fertility or childbearing decisions. The criterion by which we identify whether there are sectors in which women do not have a comparative advantage is the share of female workers in Table 3. For example, the percentage of female employees in the textile sector and in all sectors is 40.191% and 26.916%, respectively, while the percentage of female employees in printing, grain processing and metal smelting is just 0.720%, 0.808% and 1.010%, respectively.

We found above that industrialization in the textile sector affects fertility. By contrast, the growth of printing, grain processing and metal smelting should not significantly affect the number of children. We present OLS estimates of the relationship between output per capita and the number of children for printing, grain processing and metal smelting in panel A of Table 11. For grain processing, printing and metal smelting, which rarely employ

Table 11: Placebo test: sectors in which women have no comparative advantage

	Grain proc.		Printing		Metal smelting	
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Dependent variable: number of children						
(log) Output per capita in corresponding sector	−0.100	−0.091	0.220	0.140	0.027	−0.025
	(0.065)	(0.060)	(0.136)	(0.135)	(0.155)	(0.147)
Geographic controls	X	X	X	X	X	X
Historical controls	X	X	X	X	X	X
Socioeconomic controls	X	X	X	X	X	X
Individual controls		X		X		X
Province FEs	YES	YES	YES	YES	YES	YES
Year of birth FEs	YES	YES	YES	YES	YES	YES
Number of clusters	123	123	123	123	123	123
Observations	4,013	4,013	4,013	4,013	4,013	4,013
R^2	0.123	0.164	0.123	0.164	0.122	0.163
Panel B. Dependent variable: age of wife at birth of the first son						
(log) Output per capita in corresponding sector	0.099	0.110	−0.278	−0.301	0.253	0.229
	(0.198)	(0.195)	(0.433)	(0.418)	(0.615)	(0.606)
Geographic controls	X	X	X	X	X	X
Historical controls	X	X	X	X	X	X
Socioeconomic controls	X	X	X	X	X	X
Individual controls		X		X		X
Province FEs	YES	YES	YES	YES	YES	YES
Year of birth FEs	YES	YES	YES	YES	YES	YES
Number of clusters	133	133	133	133	133	133
Observations	4,119	4,119	4,119	4,119	4,119	4,119
R^2	0.062	0.069	0.062	0.069	0.062	0.069

Notes: (i) ***, **, * indicate statistical significance at the 1, 5 and 10 percent level, respectively; (ii) Standard errors in parentheses are clustered at county level; (iii) Geographic controls include distance to coast (in km), distance to major navigable rivers (in km), rice suitability, wheat suitability, maize suitability, sweet potato suitability and economic crops suitability; (iv) Historical controls include Tea center in Ming-Qing, distance to national capital in Ming-Qing (in km, logged), distance to provincial capitals in Ming-Qing (in km, logged), strength of clan in Ming-Qing (logged) and Confucian academies in Ming-Qing (logged); (v) Socioeconomic controls include population density (per sq. li), share of child labour in population, gender ratio and urbanization rate (logged); (vi) Individual controls include number of marriages, a remarriage dummy and a first son dummy.

female workers, our finding of insignificant results is strong evidence for the soundness of the research design.

In terms of the relationship between output per capita and age of women at birth of the first son across sectors (panel B of Table 11), we do not find a significant effect between these two either. This is in line with our findings in Table 6 and Table 7.

It is worth noting that our argument for the validity of demographic shocks as the instrument for industrialization stops holding in these sectors. In particular, an increased abundance of land and a switch from food crops to cash crops as a result of population decline has little effect on these industries. This implies that, population change will be weakly, if at all, correlated with industrialization in these sectors, a claim confirmed by very low values of the first stage F statistics (0.98, 0.18 and 0.10 for grain processing, printing and metal smelting, respectively) obtained when we indeed run 2SLS. As there is not enough proof of a reverse causal link between childbearing patterns and the development of sectors where women have no comparative advantage, our instrument is not used.

5.6 Mechanism analysis

Early industrialization could reduce the number of children a couple has but does not postpone age of the first birth for women in the 19th and early 20th century China, as we already know in the last section. Therefore, in this subsection, we investigate two main channels through which early industrialization may have affected the number of children in Chinese households.

First, we look into whether early industrialization reduces fertility by increasing the opportunity cost of childcare for women. The opportunity cost of childbearing is borne primarily by women as females often face a trade-off between child-rearing and work. Women have more possibilities to participate in the labour market as a result of more jobs for women created by early industrialization. Still, childbearing could cause women to be temporarily absent from work and thus lose their current wage or even affect the lifetime wage. As a result, wage loss may cause more women to have fewer children or delay timing of first birth. Through results shown in Table 7, we may suggest that especially for Chinese society in the 19th and early 20th centuries, women usually reduce their fertility through intra-marital birth control rather than delay timing of birth by late marriage. Therefore, early industrialization mainly reduced the number of children rather than delayed age of the first birth for women in China in the 19th and early 20th centuries.

Table 12: Mechanism analysis: opportunity cost of childbearing

	Textile			All		
Variable	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable: share of female workers in female population (%)						
(log) output per capita in textile sector	0.814^{***}	0.851^{***}	0.721^{***}			
	(0.218)	(0.229)	(0.180)			
(log) output per capita				0.797^{***}	0.888^{***}	0.783^{***}
				(0.184)	(0.214)	(0.183)
Geographic controls	X	X	X	X	X	X
Historical controls		X	X		X	X
Socioeconomic controls			X			X
Province FEs	YES	YES	YES	YES	YES	YES
Observations	133	133	132	133	133	132
R^2	0.730	0.747	0.773	0.742	0.776	0.792

Notes: Notes: (i) ***, **, * indicate statistical significance at the 1, 5 and 10 percent level, respectively; (ii) The data of cotton, linen, silk, clothing, towels, and other associated products makes up the data of the textile sector; (iii) Robust standard errors are shown in parentheses; (iv) Geographic controls include distance to coast (in km), distance to major navigable rivers (in km), rice suitability, wheat suitability, maize suitability, sweet potato suitability and economic crops suitability; (v) Historical controls include Tea center in Ming-Qing, distance to national capital in Ming-Qing (in km, logged), distance to provincial capitals in Ming-Qing (in km, logged), strength of clan in Ming-Qing (logged) and Confucian academies in Ming-Qing (logged); (vi) Socioeconomic controls include population density (per sq. li), share of child labour in population, gender ratio and urbanization rate (logged).

We use female employment rate at the county level as an outcome to capture the opportunity cost of female childcare because of the trade-off between child-rearing and work for women. In Table 12, there are statistically significant coefficients in columns (1) to (6). We include geographic, historical and socioeconomic controls when presenting results in column (3) and column (6). An 1% increase in (log) output per capita for all sectors would increase the proportion of female workers to female population by 0.00783 pp. (around 1% of the mean value) at the county level. Compared with all sectors, the effect of early industrialization on the proportion of female workers is slightly smaller in the textile sector, at 0.00721 pp. These results mean that higher levels of industrialization, particularly in the textile industry where women have a comparative advantage, are associated with more women choosing work rather than child-rearing when facing the trade-off between these two.

In addition, we focus on the trade-off between quality and quantity of children through which early industrialization may lead to a decrease in fertility. Due to the lack of child labour legislation during the early industrialization phase, when the use of technology or

Table 13: Mechanism analysis: quantity and quality trade-off

	Developing		Undeveloped	
Variable	(1)	(2)	(3)	(4)
Panel A. Senior primary students				
Dependent variable: number of children				
Percentage of current senior primary students in the population (%)	-3.371^{**}	-4.381^{**}	0.294	-1.447
	(1.584)	(1.407)	(2.052)	(2.414)
Individual controls	X	X	X	X
Geographic controls	X	X	X	X
Historical controls		X		X
Province FEs	YES	YES	YES	YES
Year of birth FEs	YES	YES	YES	YES
Number of clusters	51	51	54	54
Observations	747	747	625	625
R^2	0.267	0.275	0.178	0.206
Panel B. Junior primary students				
Dependent variable: number of children				
Percentage of current junior primary students in the population (%)	-0.293^{*}	-0.357^{*}	-0.060	-0.145
	(0.165)	(0.181)	(0.171)	(0.161)
Individual controls	X	X	X	X
Geographic controls	X	X	X	X
Historical controls		X		X
Province FEs	YES	YES	YES	YES
Year of birth FEs	YES	YES	YES	YES
Number of clusters	51	51	54	54
Observations	747	747	625	625
R^2	0.265	0.271	0.178	0.207

Notes: (i) ***, **, * indicate statistical significance at the 1, 5 and 10 percent level, respectively; (ii) Standard errors in brackets are clustered at county level; (iii) Geographic controls include distance to coast (in km), distance to major navigable rivers (in km), rice suitability, wheat suitability, maize suitability, sweet potato suitability and economic crops suitability; (iv) Historical controls include Tea center in Ming-Qing, distance to national capital in Ming-Qing (in km, logged), distance to provincial capitals in Ming-Qing (in km, logged), strength of clan in Ming-Qing (logged) and Confucian academies in Ming-Qing (logged); (v) Individual controls include number of marriages, a remarriage dummy and a first son dummy.

machines was not as widespread, child labour was more common in all sectors. However, as technology advances, the wage income of adult labour will be higher than the wage income of child labour, and the wage gap could raise the cost of child-rearing for parents while lowering the cost of schooling for children (Ager, Herz and Brueckner, 2020). Hence, child labour could be less valuable in manufacturing, and the wage gap between adults and child labour causes parents to prioritise the quality of children over the quantity of children and replace child labour with child education, leading to lowered fertility.

China went through a challenging and tortuous early stage of industrialization after the First Opium War in 1842, with factories using machines to produce products on a large scale gradually replacing the manual workshops, thus providing an ideal context for such quantity-quality trade-off to emerge. To explore whether the investment in child human capital based on quantity-quality trade-off theory as a channel between the early industrialization and the number of children, we regress the number of children on the percentage of current senior primary students in the population (%) and on the percentage of current junior primary students in the population (%). We focus on individuals born between 1875 and 1889 because the year of the survey for current primary school students in the Republic of China is 1918. This allows us to ensure that their children were around primary school age in 1918, and we could gain a relatively complete fertility history. In addition, as quantity-quality trade-off emerged in the wake of technological advances and more widespread employment of machines, we divided the counties of the Republic of China into two groups based on whether county-level output per capita is greater than the median value (1.676).

In Panel A of Table 13, our findings demonstrate an inverse correlation between the proportion of current senior primary school students and the number of children born in developing regions where machines were already used in some industrial activities. Although only suggestive, these findings imply that when machinery was widely used, there was a trade-off between the quantity and quality of children in the 19th and early 20th century China. Due to the increasing cost of child-rearing and the widening wage gap between adults and children caused by technological advancements, parents are inclined to allow children who have completed junior primary school to continue their studies in senior primary or even secondary schools. Fewer children are born due to increased human capital investment based on the quantity-quality trade-off theory, but a larger dataset might be needed to further strengthen the support for this hypothesis.

It is also worth noting that, for developing regions, there is a much smaller effect detected in Panel B compared to Panel A. As the primary goal of the junior primary school was to eradicate illiteracy, its coefficients explain the quantity and quality trade-off theory weakly. By contrast, parents' decision to let their children complete junior primary school and proceed to senior primary school could better suggest the importance they attached to the quality of their children.

6 Conclusion

In this paper, we investigate the relationship between early industrialization and fertility patterns by matching genealogical records from the Chinese Genealogy Database with industrial data from the China Industrial Survey Report and population data from both the Compilation of Historical Population in the Republic of China and the Population History of China (Qing Dynasty) to create a unique dataset. To address the reverse causality between early industrialization and fertility patterns, we employ the demographic shocks between 1851 and 1880 as the instrumental variable. In addition, we also analyze three channels linking early industrialization to fertility patterns.

Our findings present a significantly negative relationship between early industrialization and the number of children but no significant causal link between early industrialization and age of wife at the first birth in the 19th and early 20th century China. While this shows that industrial development could reduce fertility, it may also reveal that, in traditional Chinese society, parents are more inclined to adopt intra-marital birth control methods like contraception, abortion and even infanticides rather than delaying timing of the first birth by late marriage to control the number of children in each household, as Lee and Feng (1999) discovered. Our findings, on the other hand, differ from those of studies that use the Western world as a backdrop. Hajnal (2017) found, for example, that women delaying timing of the first birth by late marriage are a direct cause of fertility decline in the West. Furthermore, when examining the heterogeneous effects of early industrialization across sectors, the development of those sectors in which women do not have a comparative advantage is less likely to limit fertility.

Finally, from our mechanism analyses, the increasing opportunity cost of childcare for women caused by early industrialization contributes to fewer children in a Chinese household in the 19th and early 20th centuries. There is no evidence to support the household income generated by child labour is an important motivator for parents to have more children. By contrast, due to the wake of technological advances and more widespread employment of machines leading to a wider wage gap between adults and children, parents tend to replace child labour with child education, which could reveal the existence of quality-quantity trade-off allowing Chinese couples to have fewer children.

References

Abramitzky, R., Delavande, A., and Vasconcelos, L. (2011). Marrying up: the role of sex ratio in assortative matching. *American Economic Journal: Applied Economics*, 3(3):124–57.

Ager, P., Herz, B., and Brueckner, M. (2020). Structural change and the fertility transition. *The Review of Economics and Statistics*, 102(4):806–822.

Andrews, I., Stock, J. H., and Sun, L. (2019). Weak instruments in IV regression: Theory and practice. *Annual Review of Economics*.

Angrist, J., Lavy, V., and Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, 28(4):773–824.

Bai, Y., Jia, R., and Yang, J. (2023). Web of power: How elite networks shaped war and politics in China. *The Quarterly Journal of Economics*, 138(2):1067–1108.

Basu, K. and Van, P. H. (1998). The economics of child labor. *American Economic Review*, pages 412–427.

Battistin, E., Becker, S. O., and Nunziata, L. (2022). More choice for men? Marriage patterns after World War II in Italy. *Journal of Demographic Economics*, 88(3):447–472.

Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political Economy*, 81(4):813–846.

Bellou, A. (2015). The impact of internet diffusion on marriage rates: Evidence from the broadband market. *Journal of Population Economics*, 28(2):265–297.

Bo, F. (1990). *Zhonghua renmin gongheguo jingjishi [Economic history of the People's Republic of China]*. Harbin: Heilongjiang Education Press.

Bradby, B. (1975). The destruction of natural economy. *Economy and Society*, 4(2):127–161.

Budig, M. J. and England, P. (2001). The wage penalty for motherhood. *American Sociological Review*, pages 204–225.

Cao, S. (2001). *Zhongguo renkou shi: Qing shiqi [The population history in China (Qing dynasty)]*. Shanghai: Fudan University Press.

Chan, K. S. (2014). The late Qing Dynasty to the Early Republic of China: A period of great institutional transformation 1. In *Routledge Handbook of the Chinese Economy*, pages 21–40. Routledge.

Chen, S. and Kung, J. K.-s. (2016). Of maize and men: the effect of a new world crop on population and economic growth in China. *Journal of Economic Growth*, 21:71–99.

Chen, Z., Ma, C., and Sinclair, A. J. (2022). Banking on the Confucian clan: Why China developed financial markets so late. *The Economic Journal*, 132(644):1378–1413.

China Historical Geographic Information System (2016). CHGIS Version 6 dataverse, Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University. `https://dataverse.harvard.edu/dataverse/chgis_v6`.

Ciyu, C. (1982). Jindai zhongguo chaye de fazhan yu shijie shichang [Development of the modern Chinese tea business and the world market]. *Taipei: Academia Sinica*.

De Moor, T. and Van Zanden, J. L. (2010). Girl power: The European marriage pattern and labour markets in the North Sea region in the late medieval and early modern period. *The Economic History Review*, 63(1):1–33.

Dennison, T. and Ogilvie, S. (2014). Does the European marriage pattern explain economic growth? *The Journal of Economic History*, 74(3):651–693.

Do, Q.-T., Levchenko, A. A., and Raddatz, C. (2016). Comparative advantage, international trade, and fertility. *Journal of Development Economics*, 119:48–66.

Fan, H., Li, C., Xue, C., and Yu, M. (2023). Clan culture and patterns of industrial specialization in China. *Journal of Economic Behavior & Organization*, 207:457–478.

Fan, X. and Wu, L. (2023). The shaping of a gender norm: Marriage, labor, and foot-binding in historical China. *International Economic Review, Forthcoming*.

Foreman-Peck, J. and Zhou, P. (2018). Late marriage as a contributor to the industrial revolution in England. *The Economic History Review*, 71(4):1073–1099.

Global AgroEcological Zones (2012). GAEZ Datasets, The Food and Agriculture Organization of the United Nations (FAO). `http://fao.org/Ag/AGL/agll/gaez/index.htm`.

Greif, A. and Tabellini, G. (2010). Cultural and institutional bifurcation: China and Europe compared. *American Economic Review*, 100(2):135–40.

Hajnal, J. (2017). *European marriage patterns in perspective*. Routledge.

Hatton, T. J. and Martin, R. M. (2010). The effects on stature of poverty, family size, and birth order: British children in the 1930s. *Oxford Economic Papers*, 62(1):157–184.

Hazan, M. and Berdugo, B. (2002). Child labour, fertility, and economic growth. *The Economic Journal*, 112(482):810–828.

Ho, P.-t. (2013). *Studies on the population of China, 1368–1953*. Harvard University Press.

Hornung, E. (2014). Immigration and the diffusion of technology: The huguenot diaspora in Prussia. *American Economic Review*, 104(1):84–122.

Ji, X. (1996). *Zhongguo shuyuan cidian [A Compendium on the Chinese Academies]*. Hangzhou: Zhejiang Education Press.

Ko, D., Haboush, J. K., and Piggott, J. (2003). *Women and Confucian cultures in premodern China, Korea, and Japan*. University of California Press.

Kulu, H., Hannemann, T., Pailhé, A., Neels, K., Krapf, S., González-Ferrer, A., and Andersson, G. (2017). Fertility by birth order among the descendants of immigrants in selected European countries. *Population and Development Review*, pages 31–60.

Landry, A. (1934). La révolution démographique: Études sur les problèmes de la population. *Paris: Librairie Sirey*.

Lee, J. Z. and Feng, W. (1999). *One quarter of humanity: Malthusian mythology and Chinese realities, 1700–2000*. Harvard University Press.

Li, N. (2015). Legacy of war: The long-term effect of taiping rebellion on economic development in modern China. *Available at SSRN 3220756*.

Li, N. and Lin, C. (2015). Taiping tianguo zhanzheng dui jindai renkou yingxiang de zaiguji [Re-estimation of the impact of the Taiping War on China's modern population]. *China Economic Quarterly*, 14:1325–1346.

Li, W., Cheng, X., Liu, Y., and Xia, M. (1994). *Zhongguo jindai shida zaihuang [Top ten modern Chinese disasters]*. Shanghai Renmin Press.

Liang, H. (2011). *Waiguo zaihua zhijie touzi yu jindai zhongguo jingji fazhan [Foreign direct investment in China and the economic development of early modern China]*. Beijing: China Social Science Press.

Loughran, D. S. and Zissimopoulos, J. M. (2009). Why wait? The effect of marriage and childbearing on the wages of men and women. *Journal of Human Resources*, 44(2):326–349.

McElderry, A. (1986). Confucian capitalism? Corporate values in Republican banking. *Modern China*, 12(3):401–416.

Miguel, E., Satyanath, S., and Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of Political Economy*, 112(4):725–753.

Murphy, M. and Knudsen, L. B. (2002). The intergenerational transmission of fertility in contemporary Denmark: The effects of number of siblings (full and half), birth order, and whether male or female. *Population Studies*, 56(3):235–248.

Notestein, F. W. (1945). Population-the long view. *Food for the World.*, pages 36–57.

Perkins, D. H. (2013). *Agricultural development in China, 1368-1968*. AldineTransaction.

Qian, N. (2008). Missing women and the price of tea in China: The effect of sex-specific earnings on sex imbalance. *The Quarterly Journal of Economics*, 123(3):1251–1285.

Rosenzweig, M. R. and Zhang, J. (2009). Do population control policies induce more human capital investment? Twins, birth weight and China's "one-child" policy. *The Review of Economic Studies*, 76(3):1149–1174.

Sato, Y. and Yamamoto, K. (2005). Population concentration, urbanization, and demographic transition. *Journal of Urban Economics*, 58(1):45–61.

Shanghai Library (2009). *Zhongguo zupu zongmu [The general catalog of Chinese genealogy]*. Shanghai: Shanghai Ancient Books Publishing House.

Shen, Z. (1999). *Jindai zhongguo qiye: Zhidu yu fazhan [Modern Chinese enterprise: Institutions and development]*. Shanghai: Shanghai Academy of Social Sciences Press.

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

Spagnoli, P. G. (1983). Industrialization, proletarianization, and marriage: A reconsideration. *Journal of Family History*, 8(3):230–247.

Stasson, A. H. (2021). Revolutionary christian attitudes toward women and family in late Qing and republican-era China. *Review of Religion and Chinese Society*, 8(2):173–193.

Stauffer, M. T. (1922). *The Christian Occupation of China: A General Survey of the Numerical Strength and Geographical Distribution of the Christian Forces in China Made by the Special Committee on Survey and Occupation 1918–1921*. Shanghai: China Continuation Committee.

Stock, J. H. and Yogo, M. (2005). *Testing for Weak Instruments in Linear IV Regression*, page 80–108. Cambridge University Press.

Telford, T. A. (1992). Covariates of men's age at first marriage: The historical demography of Chinese lineages. *Population Studies*, 46(1):19–35.

Voigtländer, N. and Voth, H.-J. (2013). How the west "invented" fertility restriction. *American Economic Review*, 103(6):2227–64.

Wakeman, F. (1997). *Strangers at the gate: Social disorder in South China, 1839-1861*. University of California Press.

Wanamaker, M. H. (2012). Industrialization and fertility in the nineteenth century: Evidence from South Carolina. *The Journal of Economic History*, 72(1):168–196.

Wang, Y. (2010). *Minguo jiaoyu tongji ziliao huibian [Education statistics of the Republic of China]*. The National Library Press.

Watkins, S. C. and Menken, J. (1985). Famines in historical perspective. *Population and Development Review*, pages 647–675.

Yin, M. and Tian, Q. (2009). *Minguo renkou huji shiliao huibian [Compilation of historical population in the Republic of China]*. The National Library Press.

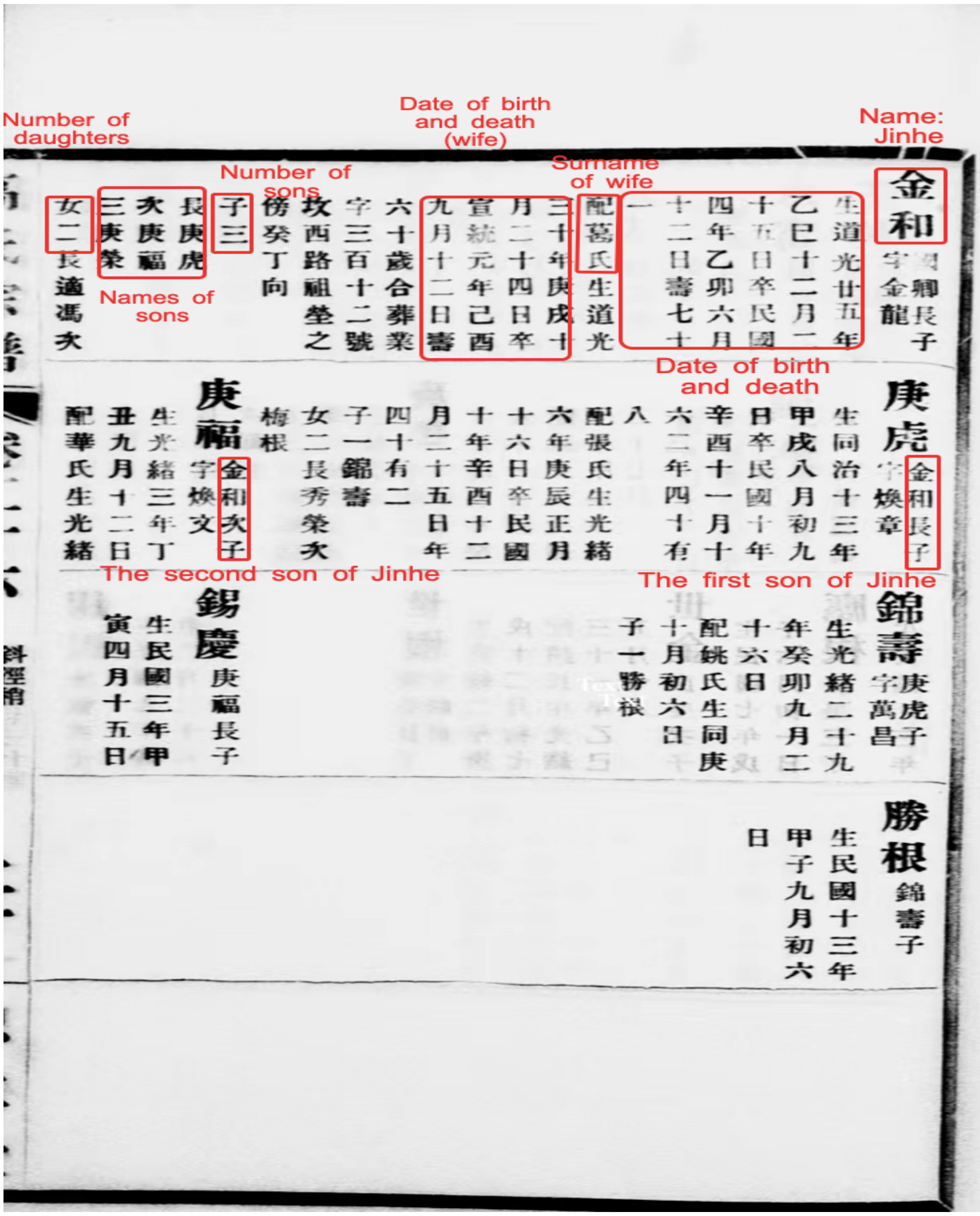
Yoon, W. (2015). Dashed expectations: Limitations of the telegraphic service in the late Qing. *Modern Asian Studies*, 49(3):832–857.

Zhang, C. (2019). Family support or social support? The role of clan culture. *Journal of Population Economics*, 32:529–549.

Zhao, Z. (1997). Long-term mortality patterns in Chinese history: Evidence from a recorded clan population. *Population Studies*, 51(2):117–127.

A Sample pages for the Chinese genealogy

Figure 2: Sample pages for the Chinese genealogy



Notes: Data source: Chinese Genealogy Database

B Other control variables

Table 14: Summary statistics for other control variables

Variable	Mean	St. dev.	Min	Max	Observations
Prefecture level:					
Distance to coast (in km)	304.18	276.9	18.62	1079.35	85
Distance to major navigable rivers (in km)	23.93	49.65	1.08	342.4	85
Rice suitability	13.29	11.57	0.074	45.21	85
Wheat suitability	27.65	9.65	8.89	54.64	85
Maize suitability	26.70	10.43	8.18	53.65	85
Sweet potato suitability	5.52	7.97	0	27.32	85
Economic crops suitability	9.58	3.94	1.84	23.64	85
Tea center in Ming-Qing (1=yea,0=no)	0.2248	0.419	0	1	85
Distance to national capital in Ming-Qing (in km, logged)	13.45	1.61	0.00004	14.52	85
Distance to provincial capitals in Ming-Qing (in km, logged)	11.79	0.732	10.14	13.02	85
Strength of clan in Ming-Qing (logged)	0.178	0.605	-0.36	2.97	85
Confucian academies in Ming-Qing (logged)	5.06	1.29	1.61	7.76	85
Urbanization rate in 1920 (logged)	-4.14	1.47	-6.91	-0.64	85

Table 15: Data source for other control variables

Variable	Level	Source
Distance to coast (in km)	Prefecture	CHGIS (2016)
Distance to major navigable rivers (in km)	Prefecture	CHGIS (2016)
Rice suitability	Prefecture	Global Agro-Ecological Zones Database (2012)
Wheat suitability	Prefecture	Global Agro-Ecological Zones Database (2012)
Maize suitability	Prefecture	Global Agro-Ecological Zones Database (2012)
Sweet potato suitability	Prefecture	Global Agro-Ecological Zones Database (2012)
Economic crops suitability	Prefecture	Global Agro-Ecological Zones Database (2012)
Distance to national capital in Ming-Qing (in km, logged)	Prefecture	CHGIS (2016)
Distance to provincial capitals in Ming-Qing (in km, logged)	Prefecture	CHGIS (2016)
Strength of clan in Ming-Qing (logged)	Prefecture	The General Catalog of Chinese Genealogy compiled by Shanghai Library (2009)
Confucian academies in Ming-Qing (logged)	Prefecture	Ji (1996)
Tea center in Ming-Qing (1=yea, 0=no)	Prefecture	Chen (1982)
Urbanization rate in 1920 (logged)	Prefecture	Stauffer (1922)