CENTRE FOR QUANTITATIVE HISTORY

HKU BUSINESS SCHOOL 港大經管學院

Centre for Quantitative History 量化歷史研究中心

Hong Kong University Business School:

Centre for Quantitative History Working Paper Series

No.2, Issue 1, 2025 April 2025

Awakening Latent Human Capital: The Opening-up and Entrepreneurship in 19th-Century China

By Li Duan† and Xiaoming Zhang‡

†The University of Hong Kong, ‡Zhejiang University

Abstract:

This study exploits a special historical case-openings of treaty ports in 19th-century China to examine how upper-tail human capital, quantified via book creation, impacted modernization when facing external pressures. Employing a prefecture-level panel dataset from 1840 to 1904, the study establishes book density, indicative of knowledge endowment, as a significant and positive predictor of modern firm entry following the opening of treaty ports. To understand the mechanism, a critical aspect lies in understanding the Civil Service Examination (*keju*), an indigenous institution that historically dominated talent accumulation and allocation in China. By integrating data with *keju*, we find that exposure to Western influence mobilized the segment of upper-tail human capital at the bottom or outside of the *keju* system into entrepreneurship. This paper illustrates the dynamics between indigenous institutions and external pressures.

AWAKENING LATENT HUMAN CAPITAL: THE OPENING-UP AND ENTREPRENEURSHIP IN 19TH-CENTURY CHINA*

Li Duan† Xiaoming Zhang‡

Abstract

This paper examines meritocratic recruitment and incentive issues within the context of China's initial industrialization. We analyze the impact of upper-tail human capital on entrepreneurship in China between 1840 and 1904, exploring how the Imperial Examination system influenced talent allocation between traditional sectors and modern firms. Our study leverages the treaty port system as an exogenous shock to China's industrialization, revealing heterogeneous choices made by different echelons of the upper-tail human capital. We find that traditional human capital played a diverse role in modern entrepreneurship, with some intellectuals transitioning to new opportunities while others remained entrenched in bureaucratic roles. Our findings highlight the complex interplay between civil service examinations and economic modernization, offering insights into how bureaucratic structures can influence talent reallocation and industrial development.

Keywords: Human Capital; Entrepreneurship; Bibliometrics; Treaty Ports; Civil Service Examinations.

*The authors declare that we have no relevant or material financial interests that relate to the research described in this paper. We would like to thank Zhiwu Chen, Joy Chen, Yu Hao, Sijie Hu, Ruixue Jia, James Kai-sing Kung, Jin Li, Guenther Lomas, Ruining Liu, Debin Ma ,Chicheng Ma, Albert Park, Tuan Hwee Sng, Clair Yang, Wuyue You, Yiling Zhao, Guoshun Zhang, Runzhuo Zhai, and seminar participants at HKU, LSE, NUS and Berkeley. All errors are our own.

†Faculty of Business and Economics, The University of Hong Kong; Email: dlcjy@connect.hku.hk; Address: Room 210, May Hall, The University of Hong Kong, Pokfulam Road, Hong Kong

‡Department of Sociology, Zhejiang University; Email: zhang.xiaoming@zju.edu.cn

> As often happens, however, institutions that were established originally for one purpose end up having very different and probably unanticipated consequences. The imperial service examination system eventually turned into a powerful tool to defend incumbent literati against the threat of intellectual innovators who threatened their political influence and the value of their human capital.
>
> — Joel Mokyr (2016), p.303

1 Introduction

The booming literature on bureaucracy and development identifies bureaucratic effectiveness as a core element of state capacity (e.g., Besley and Persson, 2011; Finan, Olken and Pande, 2017; Besley et al., 2022). In Max Weber's seminal analysis, one of the hallmarks of a bureaucracy in the modern sense is "prescribed special examinations as prerequisites of employment" (Weber, 1978, p. 959). Indeed, civil service examinations are central to enhancing bureaucratic effectiveness (Moreira and Pérez, 2021; Kung, Liu and Zhang, 2024; Aneja and Xu, 2024). Yet, a powerful bureaucratic recruitment mechanism also shapes the structure of human capital. When the bureaucracy becomes overly developed and consumes substantial societal resources, it can entrench intellectual elites within political sectors, thereby limiting their engagement in market competition.

This paper explores the challenges faced by late-developing countries in the process of industrialization, with a particular focus on the role of civil service examinations in shaping talent allocation and its implications for economic development. Specifically, we examine the impact of upper-tail human capital on entrepreneurship during the onset of China's industrialization from 1840 to 1904, analyzing how the the Imperial Examination system shaped talent allocation between traditional sectors and the establishment of modern firms. Our setting has two attractive features.

The first feature is the prominence of the Imperial Examination system in Chinese history. Originating during the Sui dynasty (581 CE), this system facilitated the recruitment of talent into China's elite political and bureaucratic ranks through civil service examinations. As one of the most influential systems in Chinese history, the Imperial Examination not only shaped the capacity of the bureaucracy but also influenced the circulation of ideas, the allocation of talent, and the general formation of human capital (Elman, 1991, 2000, 2013).[1] Globally, the Imperial Examination system was one of the earliest and most

[1] Although our setting is pre-modern, it does not necessarily reflect a narrative of weak state capacity or

enduring civil service examinations, even inspiring civil service reforms in Britain, as noted in the Northcote-Trevelyan Report (Northcote, Trevelyan and Jowett, 1854). Many aspects of the Examination and the challenges it faced were surprisingly modern.[2] As Woodside (2006) portrays, the Examination represents an anachronistic blend of the classical education of the aristocratic preindustrial West, reflected in its exam curricula, and the civil service recruitment norms of the postindustrial West, noted for its meritocratic and systematic approach to selecting bureaucratic officials.[3]

The second feature is the introduction of the treaty port system. The year 1840 marks a pivotal juncture in Chinese history, when the Qing dynasty of China suffered humiliating defeats in the Opium War against the British Empire. In the ensuing decades, a series of treaty ports opened China to foreign trade and markedly accelerated its industrialization (Keller and Shiue, 2021). The introduction of Western trade and legal institutions offered novel career trajectories and business paradigms, accessible to Chinese intellectuals for the first time in over a millennium (Hao, 1970*a*,*b*; van de Ven, 2014). Therefore, the competition between the entrenched Imperial Examination system and emerging Western industrial models provides a framework for investigating how bureaucracy, such as the Examination system, attenuated (or reinforced) the spread of industrialization by affecting career choices across different echelons of upper-tail human capital. We exploit the opening of these treaty ports as an exogenous shock to China's industrialization process and the reallocation of talent across the country. Therefore, we estimate a staggered difference-in-differences model, comparing prefectures with treaty ports to those without, before and after the staggered opening of these ports.

We start by estimating whether the opening of treaty ports awakened China's historical stock of upper-tail human capital to the industrialization. While the advent of treaty ports altered the career landscape for Chinese intellectuals, the extent to which Chinese intellectuals reacted positively to the forced introduction of industrialization and globalization still warrants empirical examination. Our empirical analysis is based on a panel dataset for the period of 1840 to 1904 covering 267 prefectures of Qing China. We quantify each prefecture's human capital stock using a novel bibliometric measure: the cre-

preindustrial despotism. As Besley et al. (2022) argues, autocratic regimes can build effective bureaucracies because their survival relies on either an effective police force or a social contract where the provision of public goods is essential.

[2]For instance, measures to enhance transparency included concealing candidates' names and having their answers copied in other people's handwriting. Concerns about "grade inflation" were already prevalent, and affirmative actions were implemented to ensure proper representation of disadvantaged minorities (Xue and Zhang, 2023).

[3]A helpful model for understanding the Imperial Examination is the concept of "interstitial bureaucracy" proposed by McDonnell (2017), a distinctively functioning yet embedded subsystem of a predominantly weak state (Sng and Moriguchi, 2014; Sng, 2014).

ation of new books. Our primary outcome variable of interest is the establishment of modern firms, both private and state-owned. We observe a significantly positive contribution from China's traditional human capital stock, accumulated before 1840, to entrepreneurship during the industrialization process from 1840 to 1904: a 1% increase in book density translates approximately to a 0.7% increase in firm density or a 3.5% increase in firm count. Our findings support the view that the opening up of China constituted an effective catalyst for awakening latent human capital, which translated into substantial economic outcomes in the form of innovative entrepreneurship.

Next, we try to understand what the book density metric truly captures. Does it represent the "useful knowledge" concept akin to 18th century France (Squicciarini and Voigtländer, 2015), where the pre-industrial human capital stock, rich in scientific knowledge and technological know-how, led to economic productivity with industrialization? Or does it indicate talent reallocation, suggesting intellectuals' ability to adapt to new knowledge and careers despite the fact that their prior education might be antiquated and less relevant for industrialization? Our analysis of books across different subjects, each with apparently varying industrial relevance, finds little support for the first interpretation, pointing instead towards talent reallocation.

In the remainder of this paper, we aim to understand our baseline findings through the lens of talent reallocation and the role of the Imperial Examination. We hypothesize that the opening of treaty ports had varied impacts on different intellectual echelons within the Examination system, with those more entrenched in the system being less inclined to shift towards modern entrepreneurship. Our analysis first reveals that the *Jinshi* circle's human capital played an insignificant, or possibly counterproductive, role in modern firm emergence before 1905.[4] In contrast, By using the subset of books authored by non-*Jinshi* as a proxy, we find that the positive relationship between traditional human capital and modern firm creation predominantly stems from the vast pool of intellectuals who did not participate in or receive any Examination titles. Further combining historical anecdotal evidence and new data, we provide suggestive evidence that these intellectuals mainly functioned as information bridges and conduits for promoting modernization.

Related Literature This paper makes two main contributions. First, we contribute to the literature on knowledge creation and industrialization. (e.g., Mokyr, 2002, 2005; Dittmar, 2011; Cageand Rueda, 2016; Dittmar and Seabold, 2019; Becker, Pino and Vidal-Robert, 2021). Within this body of work, there is an evolving understanding of the relation-

[4]Among the degrees obtainable from the Imperial Examination, the "Presented Scholar" (*Jìnshì*) was the most prestigious.

ship between knowledge and industrialization. Initially, this relationship is considered minor (Mitch, 1993; Sandberg, 1979; Allen, 2003; Galor, 2005). However, later research moves away from the general concept of human capital, identifying specific forms of knowledge as crucial drivers of industrialization, such as innovative technological entrepreneurship, mechanical skills, and upper-tail human capital (Mokyr, 2012; Squicciarini and Voigtländer, 2015; Kelly, Mokyr and Gráda, 2023). Our analysis further disaggregates upper-tail human capital, revealing significant heterogeneity in elites' responses to industrialization. This heterogeneity may be key to explaining the conflicting findings regarding the relationship between Chinese traditional human capital and industrialization in the current literature (e.g., Keller and Shiue, 2021; Ma, 2024).

Second, we contribute to an emerging strand within the bureaucracy and development literature that examines meritocratic recruitment and incentive issues (Acemoglu, 1995; Besley and Ghatak, 2005; Bertrand et al., 2020). As outlined at the beginning of this paper, most studies on bureaucracy and development focus on whether certain civil service reforms or merit-based recruitment systems can enhance bureaucratic effectiveness. While these studies primarily highlight short-term gains within the bureaucratic system, they often overlook how meritocratic recruitment can reshape various aspects of societal structure, potentially leading to unintended long-term consequences. As eloquently described by Besley et al. (2022), bureaucratic systems encompass not only internal interactions among various departments and hierarchical levels but also provoke external responses from the civil society in which they are embedded. In our study, the confrontation between the Imperial Examination system and treaty ports provides a context to investigate how the incentive structure derived from the Examination influenced talent reallocation towards modernization.

The rest of the paper is organized as follows. Section 2 explores the historical context relevant to our study, including book writing in pre-1912 China and bibliometric data collection, and the backdrop of treaty ports and modern firms in the late Qing period. Section 3 presents the empirical results, followed by Section 4, which offers further evidence to deepen our understanding of the talent reallocation process. Finally, Section 5 concludes the paper.

2 Historical Background and Data

In Section 2.1, we begin by elucidating China's historical book cataloging systems and then introduce the historical sources utilized in our data collection effort. We also explain in details our data generation process employed to develop our bibliometric measures,

along with essential raw data statistics. Subsequently, in Section 2.2, we delve into the institutional context of treaty ports and the various types of modern firms in the late Qing era.

2.1 Book Writing in Pre-1912 China

One major part of our contribution is the novel bibliometric data set. Bibliometric measures have already demonstrated their great potential in a few empirical studies within the contexts of Europe and the Middle East.[5] In the case of China, the application of bibliometric data is still rare, but the availability and quality of historical materials are on par with, if not better than, those of Europe.

Imperial Book Catalogs Dating back to 200 BCE, a state-sponsored book cataloging tradition started to take shape as a part of the official history of each dynasty, and this cataloging system was formalized under the Sui dynasty (581-618) and was maintained throughout the imperial period of China (Denecke, Li and Tian, 2017, p. 163). Among the Twenty-five History, the official history of Chinese dynasties, there are in total seven book catalogs for Han (202 BCE-220 CE), Sui (581-618), Tang (618-907), Song (960-1279), Ming (1368-1644) and Qing (1644-1911) dynasties. Please see Appendix Section C.2 for details. These imperial book catalogs were compiled by imperial court scholar-officials and intended as a symbol of the intellectual richness of its time by including as many books as possible from a wide variety of topics. Its limits for our research are also obvious: we have to assume a certain degree of selection bias during the book collection process.

Grand Catalog of Books in pre-1912 China To complement the primary historical materials above, we further bring in *Grand Catalog of Books in Pre-1912 China* (*Zhōngguó Gǔjí Zǒngmù*). It is a 26-volume book catalog compiled by Ancient Books Editorial Committee (2009) for books written in pre-1912 China, based on modern library holdings and private collections both domestic and overseas.Since it only collects books still in existence, it alone suffers from survival bias, but imperial catalogs address that to a certain degree. In the meantime, it complements imperial book catalogs well. It fills all the gap years and

[5] Chaney (2016) builds a catalog of books from the Islamic world primarily based on Harvard University library holding, documents the proportion of books on scientific topics, and identifies the rise and fall of Islamic science. In another project (Chaney, 2020), he further completes an ambitious book catalog of modern library holdings across Western Europe and the Middle East and empirically tests how well this book/author count measures historical city growth. Dittmar and Seabold (2019) categorizes a catalog of books into business education, Catholics, and Protestant during 1454-1600AD Europe, and empirically examines the relationship between books in Protestant or business education and city growth.

dynasties, and even for the 6 dynasties with imperial catalogs, it helps to reduce the effect of selection bias. The entire 26-volume of the Grand Catalog are digitized and extracted by AI tools with minimal manual input, such as Automated Layour Parser and Named Entity Recognition. Please refer to Appendix Section C.3 for more details on the machine learning process.

Bibliometric Measurements We need to further explain our definition of an observation unit, in comparison to what has been used in the literature. We focus on the creation of books, not the circulation of books in the book market. Therefore, our individual observation unit for a "book title" includes only one edition of each book. Later commentary books and other derivative works are counted separately. To illustrate this, the earliest version of *The Nine Chapters on the Mathematical Art* (*Jiǔ Zhāng Suànshù*) commented by Liu Hui (circa 263) is counted as an individual title, as is the later commentary work by Li Chunfeng (602-670) (*Jiǔ Zhāng Suànshù Jiǔ Juàn Yīn Yì Yī Juàn*). This definition is different from the one used by Baten and van Zanden (2008). They focus on the printing and circulation of books, so their individual units of books include both first editions and re-editions. Although we have also extracted the volume number entity in each book entry, we decide to stay with the book title count primarily for two reasons. Firstly, the content density for a single volume could be highly varied in different periods of Chinese history. Having searched the current literature, we fail to find any study that can help us to standardize the volume numbers of different time periods. Secondly, books in different topic areas could inherently require very different lengths of writing.

Our study covers the period from 1840 to 1904, but our bibliometric variable is defined as the natural logarithm of the book count per person for works written between 1644 and 1840. We are not suggesting that these authors or their books directly participated in the industrialization process post-1840. Instead, we propose that the book density accumulated between 1644 and 1840 maps the spatial distribution of upper-tail human capital from the inception of the Qing Dynasty to the eve of its opening up to the West.

We initially considered three subsamples of books: those written before 1644, between 1644 and 1840, and between 1840 and 1904. We elected to focus on the books written in the two centuries prior to 1840 for the following reasons. Firstly, Appendix Table A1 reveals a distinct spatial pattern for books written before 1644, which diverges from that of later books. This difference is consistent with the shifts in economic centers that occurred during the Northern and Southern Dynasties in the 5th century and the Song Dynasty in the 12th century. On the other hand, Appendix Table A1 shows a much higher spatial correlation between the distribution of books written from 1644 to 1840 and those written

from 1840 to 1904. This suggests that the accumulation of books in the two centuries before 1840 reflects a stable spatial distribution of intellectual elites that persisted until 1904, whereas books written significantly earlier do not. Thirdly, after 1840, numerous factors, notably Western influence, significantly altered the landscape of book writing, including their content. The prevalence of Western ideas and methodologies within these books implies that the density of books written post-1840 does not serve as an effective exogenous representation of traditional human capital.

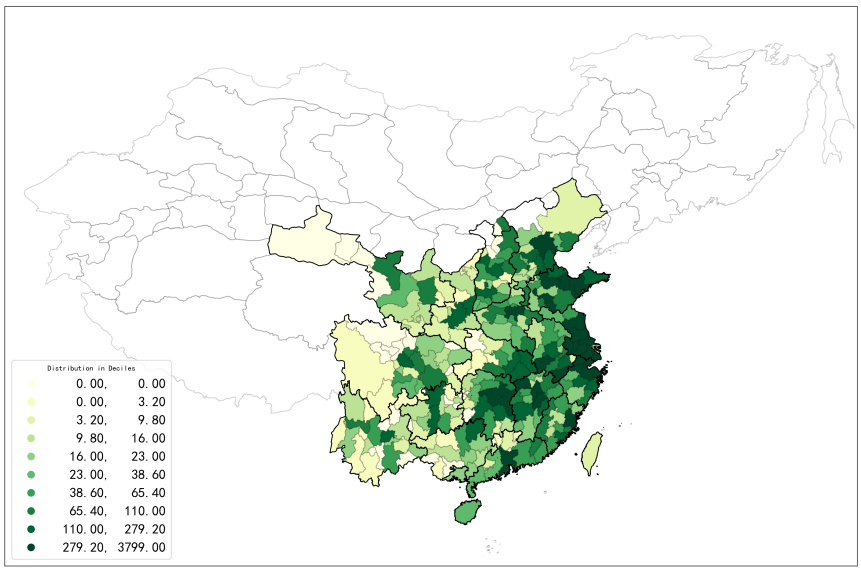


Figure 1: Spatial Distribution of Books (1644-1840)

This map plots the geographic distribution at the prefecture-level of the total number of books written between 1644 and 1840.

Matching with Biographical Data For the purpose of geo- and time-referencing our collection of books, we have linked the authors to a variety of biographical databases that catalog historical Chinese figures. The book catalogs themselves offer limited data—typically only the dynasty during which some authors lived—which does not suffice for high-frequency time-series or panel analyses. To address this shortfall, we have collated additional data on the authors' birth years and birth places from external biographical databases, thus converting our book catalog into a more robust panel dataset. The full details of our matching process are outlined in the Appendix Section C.5. In our dataset comprising 107,424 Qing dynasty books, we have successfully matched 52,435, achieving a matching rate of 48.81%. Appendix Figure C12 illustrates the distribution of these matching rates by book subjects. Figure 1 illustrates the geographical distribution of books written between 1644 and 1840.

2.2 Treaty Ports and Modern Firms in Late Qing

This section offers an overview of the history surrounding the treaty ports, aiming to elucidate the historical reasons behind their opening and identify the aspects of society and economy that their establishment perturbed.

From Isolationism to Semi-Colonialism Since the founding of the Ming dynasty in 1368, a series of foreign trade suppression policies known as "sea bans (*hǎi jìn*)" were enacted. Although the Ming government often turned a blind eye to these bans, in the 1550s, under Emperor Jiajing, a serious crackdown commenced. This effectively closed down foreign trade and reinforced China's status as an autarkic agricultural economy (Kung and Ma, 2014). When the Qing dynasty assumed control over China, it largely adopted the Ming's sea ban policies, partly to isolate the remnants of the Ming regime on the island of Taiwan. In 1684, following the re-annexation of Taiwan, Emperor Kangxi designated the coastal cities of Guangzhou, Xiamen, Songjiang, and Ningbo as ports open to foreign trade. In 1757, Emperor Qianlong restricted all foreign trade to the port of Guangzhou.[6] Between 1757 and 1842, prior to the establishment of the treaty port system, the vast majority of China's foreign trade was conducted in Guangzhou, marking the beginning of the Canton System (*Yīkǒu Tōngshāng*) (Dyke, 2007). Under this system, trade with Westerners was managed by a few officially sanctioned Chinese merchants, known as "Hongs."[7] Within Guangzhou, foreign trade was confined to a neighborhood along the Pearl River, known as the Thirteen Factories.[8]

Partly due to the growing demand for tea, porcelain, and silk in Europe, particularly in Britain where tea had become the national drink, and the poor balance of trade in goods from Europe, payments to China had to be settled in large volumes of bullion. As a result, Britain experienced a severe trade imbalance and a significant outflow of silver to China. To address the chronic trade deficit, the East India Company sold opium cultivated on its plantations to independent traders, who then smuggled it into China in exchange for silver, despite the Qing dynasty's long-established ban on opium since 1729 (Fay, 2000, p. 42). Eventually, the First Opium War broke out in 1839 between Qing China and the Royal Navy. There were numerous processes and frictions that led to this armed

[6] The motives behind this decision were complex. Notable factors include the Flint Affair, where British merchants attempted to expand their trade north, and the discovery of underground missionary activity in the late 1750s.

[7] Trade with Korea, Japan, and Russia was handled separately in other locations due to their presence on China's northern borders.

[8] The "factories" were not manufacturing facilities but rather trading posts, offices, and warehouses. The term was derived from "feitoria," which means trading post in Portuguese.

conflict, including some lesser-known factors like British territorial aggression towards Tibet through Nepal. According to Spence, the two main triggers were the crackdown on opium by the Special Imperial Commissioner, Lin Zexu, and the fact that the new British post of superintendent of foreign trade in China was a deputy of the crown, meaning any aggression towards British merchants was considered an insult to the British nation (Spence, 1999, p. 153).

The total defeat in the First Opium War resulted in the first of many unequal treaties, the Treaty of Nanjing in 1842. This treaty marked the end of the Canton System and established the concept of treaty ports, where British merchants were permitted to conduct business "without molestation or restraint" (Britannica, 2017). Five coastal locations were designated as treaty ports: Guangzhou, Xiamen (Amoy), Fuzhou, Ningbo, and Shanghai. The subsequent treaty with Britain introduced the "most-favored-nation" clause, requiring the Qing dynasty to extend any privileges granted to one country to other colonial powers. It also introduced extraterritoriality, allowing foreigners in China to be governed by the laws of their home country. A series of later treaties with other colonial powers, such as the United States and France, further expanded the network of treaty ports.[9]

In essence, China's opening-up during the latter half of the 19th century was imposed on the Qing government due to a series of military defeats and unequal treaties. The ensuing transitions were primarily driven by the interplay between Chinese social forces and foreign influence, rather than by state planning or coordination. Although some pioneering officials spearheaded industrialization efforts, known as the Self-Strengthening Movement (Kuo and Liu, 1978; Bo, Liu and Zhou, 2023), the Qing government was lingering in a near-bankruptcy state due to the substantial indemnity payments imposed by the unequal treaties, which hindered the implementation of any nationwide modernization. Additionally, the Qing imperial court remained largely divided over the merits of Westernization. Consequently, the talent reallocation examined in this article was predominantly a social response to the forced opening and foreign competition, rather than a top-down reform.

[9] Appendix Table C14 illustrates the staggered timing of the treaty ports: Wave 1 took place between 1842 and 1844, when the Treaty of Nanking, the Treaty of Wangxia, and the Treaty of Whampoa granted treaty port privileges to the United States and France; Wave 2 occurred between 1858 and 1861, following the Second Opium War and the signing of the Treaty of Tianjin, which opened 10 additional treaty ports, such as Tianjin; Wave 3 spanned from 1884 to 1887, when the Sino-French War led to the establishment of new treaty ports along the Yangtze River, including Chongqing; Wave 4 happened between 1895 and 1903, after the First Sino-Japanese War, which opened treaty ports like Suzhou and Hangzhou (Fairbank and Liu, 1980).

Selection Criteria and Potential Biases A key aspect of the history of treaty ports for our research is to understand how these ports were selected and the potential selection biases that arose from this process. In this section, we review the historical context behind the assignment of treaty ports and draw significant inspiration from the pioneering work by Jia (2014) on addressing such selection biases.

The spatial distribution of treaty ports was largely determined by accessibility via water transportation and, to a lesser degree, the proximity of certain colonial powers. Figure 2 depicts the geographic distribution of the treaty ports, accompanied by their respective years of opening. Treaty ports initially emerged along the coast, moving from south to north, and later spread inland along the Yellow River and the Yangtze River. Additionally, a few landlocked treaty ports were established along the borders. For example, Aihui in the northeast was set up at the request of Russia, which had a strong presence along the northeastern border, while Mengzi in the southwest was established upon the request of France, which exerted growing influence on neighboring South Asian countries.[10] In summary, the assignment of treaty ports was by no means a random process. Instead, the primary strategic considerations for foreign powers were to gain market access and extraterritorial rights, initially targeting coastal cities and subsequently penetrating the interior of the Qing empire via the two major river systems.

Therefore, our first robustness check is to examine whether prefectures with treaty ports differed from those without them before the opening up of China. In Appendix Table A7, we conduct linear probability regressions using the available prefecture observables in our study, following a similar approach to Jia (2014). Informed by this analysis and the historical context, we match treaty ports to a subsample of control prefectures based on a list of observables to enhance the comparability between the treatment and control groups. Most notably, we match based on the distance to the coastline and the distance to major rivers, with the linear probability results indicating these as significant differences between the treatment and control groups. To approximate the gradual northward spread of treaty ports, we also match by longitude and latitude. The linear probability results confirm that the higher the latitude, the less likely it is for a treaty port to be assigned.

Beyond geographic biases, there is another concern: foreign powers might have chosen locations that were richer or better suited for economic development, which is more challenging to address with available data. Following Jia (2014), we first select control

[10] Longzhou in Guangxi province, and Mengzi and Simao in Yunnan Province were all established in the French Treaty of 1887.

prefectures with comparable population sizes. [11] In addition, given that pre-1840 China was predominantly an agricultural economy, we also match based on terrain ruggedness and indices of grain suitability. The linear probability results indicate that both terrain ruggedness and the suitability indices for rice and wheat, China's two primary agricultural products, are significant factors. We further approximate economic development levels using data derived from social and institutional settings. First, we consider each prefecture's significance in the Qing administrative ranking system, the *chong-fan-pi-nan* grading system. Then, we include the volumes of genealogy books to proxy the establishment of Confucian clans and, finally, the presence or absence of designated market towns. The linear probability results indicate that only the latter two are significant factors.

Finally, a subsample of 91 prefectures is selected from the full sample of 267 prefectures based on the matching process. In Appendix Table A7, we observe that the differences across the aforementioned observables between the treatment and control groups are noticeably reduced in the subsample compared to those in the full sample.

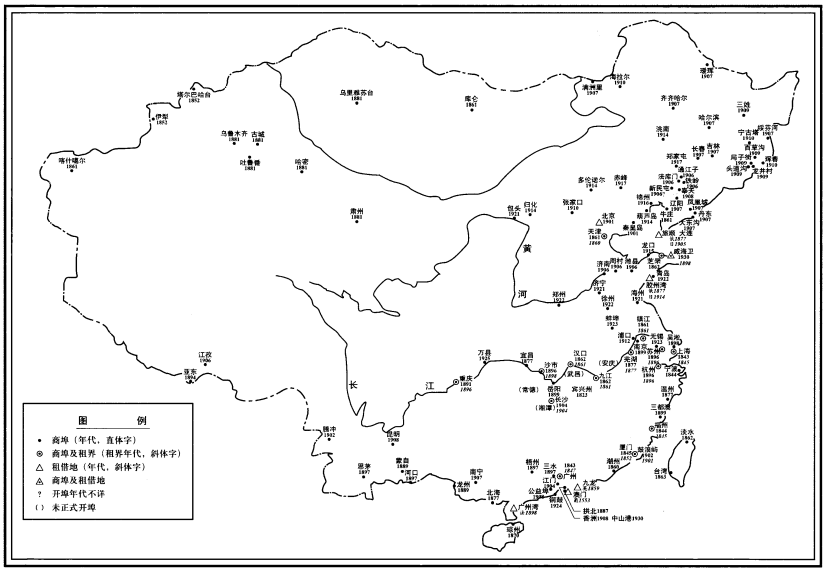


图2—1 1843—1930年间中国开辟的商埠、租界、租借地地图

Figure 2: Treaty Ports and Time of Opening

This map comes from Yan (2012) and it plots the geographic distribution of treaty ports along with their opening times. This figure includes not just treaty ports, but also concessions.

Institutional Changes and Human Capital Although the immediate purpose of treaty ports was to facilitate Western merchants' activities and protect their property rights, this

[11] In Jia (2014), population size is not significant in predicting the assignment of treaty ports, but our linear probability results indicate otherwise. This difference arises because we use the average population between 1644 and 1840, rather than the two cross-sections of population data for 1776 and 1820.

process led to significant and wide-ranging changes in China's political, economic, and social institutions. Most notably, throughout much of Chinese history, particularly under the traditional *chong-fan-pi-nan* grading system, larger and more economically developed cities were paired with more important administrative units. However, due to the treaty port system, as Murphey (1974) describes in his landmark study:

> By about the turn of the century or soon thereafter, the largest city in China was the most foreign— Shanghai, the kingpin of the treaty port system..... With the single (and significant) exception of Peking, all of the other largest cities by the 1930's—Tientsin, Canton, Wuhan, Nanking, Chungking, Mukden, Dairen, in that order—were coastal or riverine treaty ports dominated (although to different degrees) by their commercial functions as opposed to the administrative character of the largest traditional cities, and dominated by foreigners. Their populations were of course overwhelmingly Chinese, but in part a new kind of Chinese, the treaty port men, who promised to be the indigenous agents for the remaking of China along Western lines in trade, finance, transport, industry, politics, and ideology (p. 20).

We argue that the advent of the treaty port system trailblazed new business-related career paradigms for Chinese intellectuals. The nature of these new careers in business is more profound than merely increasing commercial activities and trading. Contrary to the common stereotype, 19th-century China was not a forbidden realm for merchants under a suffocating bureaucratic state. Instead, the levels of internal commerce and long-distance trade were substantial, even in per capita terms, when compared to Western Europe (e.g., Murphey, 1974; Shiue and Keller, 2007).[12] Instead, as So (2011) points out, treaty ports facilitated a "negotiated convergence" of managerial and corporate governance models. Western business practices introduced new possibilities, such as impersonal investment for a more extensive capital market, which complemented the traditional Chinese business model of network-based resource pooling (So, 2011, p. 8). Therefore, these profound changes necessitated a new type of human capital, individuals well-versed in both Chinese and Western values and institutions.

The "new kind of Chinese, indigenous agents for the remaking of China," mentioned here, are precisely the segment of human capital we are examining in this study. According to So (2011), these agents of change can be categorized into two main groups.

[12]For instance, using price correlations and cointegration analysis for internal market pairs over various distances, Shiue and Keller (2007) demonstrates that, in the late 18th century, average Chinese markets were as spatially integrated as those in Western Europe, with the gap between England's markets and those in the Yangzi Delta region of China being quite small.

First, there were the Chinese entrepreneurs. This group typically comprised compradors who assisted Western traders, migrant-returnees who brought back firsthand experience of participatory politics in Western countries, or traditional Chinese merchants endowed with business foresight. Second, there were Chinese intellectuals who advocated for new cultural movements. Their role as agents of change was primarily motivated by altruistic goals, particularly nationalism and social Darwinism. They believed that to ensure national survival, it was vital to successfully compete with foreign businesses. For Chinese intellectuals overall, So (2011) acknowledges that not all of them were receptive to these new values and institutions. Some intellectuals rejected these changes due to their adherence to traditionalist ideologies, while others were more aligned with Marxist and other leftist perspectives.

In summary, as best characterized by Murphey (1974), the treaty port system " multiplied the divisions among the traditional or reactionary gentry and official elite outside the concessions, the treaty port Chinese, and the new revolutionaries."

Treaty Port Economy and Modern Firms The treaty port was typically a concession or settlement in an urban setting, where the inhabitants were both foreigners and Chinese nationals. Foreign establishments such as clubs and churches were notable features. Treaties granted extraterritorial rights, allowing foreign nationals to avoid Chinese legal jurisdiction. The business in treaty ports were initially dominated by the foreign taipans and their Chinese compradors. The treaties also imposed a fixed low tariff on China, enforced by the Chinese Maritime Customs Service, led by expatriates with branch offices in most treaty ports. With this brief overview, what exactly defines a "treaty port economy"?

Firstly, the treaty ports' contribution to the national economy was never their primary defining feature. They accounted for no more than 10% of the nation's GNP. As Murphey (1974) aptly describes, the treaty ports were akin to "a fly on an elephant." However, it is well-documented that these ports experienced more rapid development compared to other regions in the country, and Rawski (1989) even argues that the manufacturing sector's growth rate in treaty ports was comparable to that of other industrializing nations such as Japan and the United States. Light industries, including textiles, flour milling, and matches, emerged from this "enclave industrialization" and found markets both domestically and abroad. This stable environment for the treaty port economy was protected by new institutions, such as the Chinese Maritime Customs Service and the law of extraterritoriality. These institutions effectively isolated these industrial enclaves from catastrophes like famine and civil unrest occurring in the hinterland.

Specifically, the development of Chinese modern firms can be divided into three stages

according to Zhang (1989): (i) the Infancy Stage (1848-1894), characterized by a small number of firms with modest registered capital; (ii) the Adolescence Stage (1895-1904), marked by modest growth in firm numbers and a shift towards more firms with higher registered capital; and (iii) the Adult Stage (1905-1916), during which the number of firms increased dramatically, driven mainly by firms with lower registered capital.

In the Infancy Stage, contrary to modern-day popular belief, private firms during the late Qing period experienced minimal state planning or encroachment from the Qing government (Zhang, 1989, p. 350). There were roughly four factors that shaped the distribution of private firms during this stage (Zhang, 1989, p. 315-316): (i) Natural endowment: for example, the oil pressing industry clustered in Northeast China, the textile industry clustered in Jiangsu and Zhejiang provinces, the steel industry concentrated in Hankou, and the sugar industry was located in Guangdong province; (ii) Human capital endowment: the carpet industry started in Beijing and Tianjin because a group of Tibetan carpet craftsmen relocated to Beijing; (iii) Transportation: Cities such as Harbin and Wuxi became centers of the flour industry due to their strategic locations along transportation routes for wheat; (iv) Urban areas: private firms were mainly located in urban areas.

In addition to these natural conditions, the later growth of private firms was also a spontaneous response to competition from Western businesses and imports. In 1895, following the defeat in the First Sino-Japanese War, the Qing government stepped in to encourage the establishment of private industrial firms. The top three industries, in descending order, were textile, oil pressing, and flour milling. The high concentration in the textile industry was a direct response to foreign imports, as traditional Chinese textile products faced the fiercest competition among all domestic market segments. This distribution was notably different from that of SOEs and foreign firms.

Zhang (1989) notes that the geographic distribution of private firms also differed significantly from that of SOEs and foreign firms. The top three clusters of private firms were located in Central China, South China, and North China/Northeast China. Central China was also the top choice for foreign businesses and SOEs. However, the second most popular location for foreign firms was Northeast China, owing to its proximity to Russia and Japan. For SOEs, the second most favored location was North China, as most officials responsible for state-led industrialization were part of the Beiyang Minister group and based in Zhili province, near Beijing and Tianjin. In Figure 3, our data display consistent patterns.

The data on the entry of modern firms used in this study are compiled from various sources. Foreign firm data are obtained from Zhang (1987), state-owned enterprise data from Zhang (1988*a*), and joint state-private venture data from Zhang (1988*b*). According

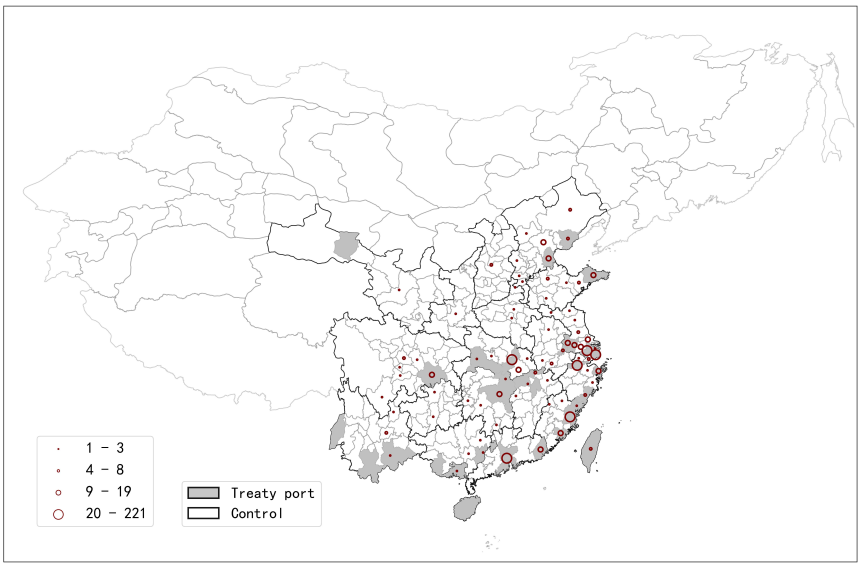


Figure 3: Spatial Distribution of Firms and Treaty Ports (1840-1904)

This map plots the prefecture locations of treaty ports and the spatial distribution of the total number of firms.

to their sources, data quality for these types of firms is generally reliable since both the Qing government and foreign businesses kept better records on firm registration information. However, data on Chinese private firms is less complete. As noted by (Zhang, 1989, p. 318), data loss mainly affected private firms in the handicraft industry. By comparing with official statistics, Zhang concludes that data loss is less of a concern for his data collection exercise since he primarily focuses on private firms with machineary. To further address data quality concerns over private firms, we included additional firm data from Du (1991).

3 Linking Book Density to Firm Creation

Before analyzing how the confrontation between the Imperial Examination system and the opening of treaty ports influenced talent reallocation in China, we must first empirically verify that traditional Chinese human capital responded positively to the globalization and industrialization introduced through the treaty ports. In other words, the treaty port system constituted a significant enough shock to challenge the monopoly of the examination-officialdom system. A review of current empirical studies on this topic reveals conflicting evidence.

On the positive side, Chinese Confucian literati exhibited a genuine interest in Western science when introduced by Jesuits (Ma, 2021); Additionally, the treaty port system

accelerated China's industrialization, with its effects extending far beyond the ports to impact much of the country (Keller and Shiue, 2021). On the negative side, while Britain directed its most creative and resourceful citizens toward entrepreneurial activities fueling industrialization, early modern China saw government service as the primary career of the most capable individuals in society (Baumol, 1990). This focus on bureaucracy and the reverence for classical studies may have acted as barriers to economic modernization when confronted with the spread of globalization and industrialization (Ma, 2024). This puzzling inconsistency in existing findings raises an important question: Could it be that China's initial industrialization happened without a positive contribution from its traditional upper-tail human capital? Consequently, establishing some baseline results on this fundamental question could not only help resolve this debate but also set the stage for our analysis into the heterogeneous roles played by different echelons of upper-tail human capital.

Section 3.1 first presents descriptive patterns from our bibliometric data and firm data. In Section 3.2, we assess the impact of books per capita on firm entry before and after the establishment of treaty ports, and investigate the heterogeneity in firm types.

3.1 Data Descriptive Patterns

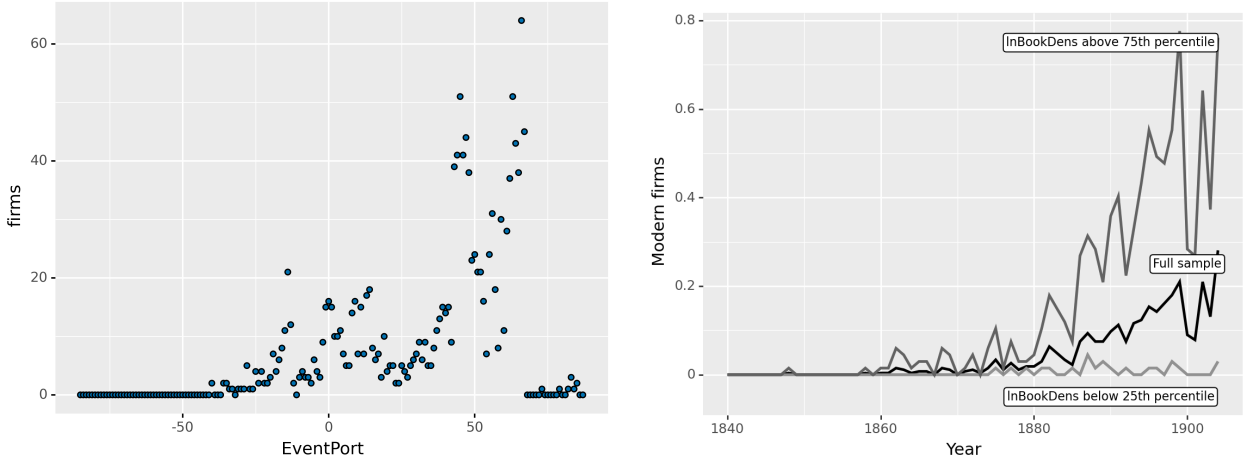
Before showing our empirical testing, we present several time series patterns that should help establish some overall picture for our historical setting. While our empirical tests are devoted to the time window from 1840 to 1904, the time series patterns presented here extend that window into 1920, providing a longer time horizon and thereby a more complete picture. In Figure 4a, we focus on the set of prefectures with treaty ports and examine the total number of new firm entries from around 50 years before treatment to around 50 years after treatment.[13] Around the relative period 0, we observe the following trends: (i) modern firm entry began to increase years prior to the opening of treaty ports; (ii) there was a sharp decline in firm entry immediately after the treaty ports opened; (iii) approximately 25 years later, modern firm entry entered a phase of steady recovery and growth.

Figure 4b compares the trends of firm entry between prefectures with book density in the first quartile and those in the fourth quartile. As the figure clearly illustrates, prior to 1880, when treaty ports were not yet widespread, these sub-samples of prefectures exhibited minimal differences. However, as more prefectures established treaty ports,

[13] As the treaty port treatment occurred in a staggered fashion and we are primarily interested in changes in firm entries due to the opening of treaty ports rather than time trends, we have converted the calendar years for each prefecture into annual periods relative to the initial time of the treaty port opening.

the discrepancy between these groups became increasingly pronounced. The sub-sample of prefectures with book density in the first quartile significantly outperformed both the group with lower book density and the full sample. Such differences clearly demonstrate the augmentation effect of book density in stimulating firm entries.

In the following section, we will introduce our empirical tests. Please refer to Appendix Section C.1 for summary statistics and data sources of the included variables.



(a) Firm Entry Relative to Treaty Ports Opening (b) Firm Creations and Book Density over Time

Figure 4: Raw Data Time Series Patterns for Modern Firm Entry

Panel (a) counts the number of new firm entries in prefectures with treaty ports in yearly periods relative to the initial treatment time. The X axis is the number of years away from the initial treatment year. For instance, -10 here means 10 years before treaty ports opening, and +10 stands for 10 years after treaty ports opening. Panel (b) illustrates the differences in firm creation time trends among prefectures with high or low book densities. The Y axis is the average number of new firms created. Besides the full sample, we split prefectures into prefectures with log book density above the 75th percentile and below the 25th percentile.

3.2 Baseline Results

To examine the link between books per capita and modern firm entry before and after the opening of treaty ports, we begin with our baseline estimation with the following specification in Equation 1:

$$\begin{aligned} lnFirm_{i,t} = &\beta \cdot lnBookDens_i \times PostPort_{i,t} + \vartheta \cdot PostPort_{i,t} + \theta \cdot \mathbf{X}_i \times PostPort_{i,t} \\ &+ \kappa \cdot \mathbf{Z}_{i,t} + \gamma_t + \lambda_i + \delta_{prov} \times \gamma_t + \lambda_i \cdot \tau + \epsilon_{i,t} \end{aligned} \tag{1}$$

where i indexes a prefecture and t indicates a year; the DiD dummy variable $PostPort_{i,t}$ is 0 before treatment and remains 1 after for the staggered entry of the absorbing treatment; $\mathbf{X}_i$ includes prefecture-level characteristics and $\mathbf{Z}_{i,t}$ represent control variables with varia-

tion in both prefectures and years; γ_t and λ_i stand for the prefecture and year fixed effects to control for all time-invariant disparities among prefectures and changes over time affecting all prefectures; we also include very flexible $\delta_{prov} \times \gamma_t$ province specific time fixed effects and $\lambda_i \cdot \tau$ prefecture specific linear time trends;[14] all standard errors are clusterd at the prefecture level.

Our key explanatory variable of interest is books per capita at the prefecture-level raised to the natural log: $lnBookDens = ln(\frac{Books}{Pop_{pre1840}} + 1)$, where *Books* is the total number of books authored by individuals born in that prefecture during the Qing dynasty up until the Opium War (between 1644 and 1840) and $Pop_{pre1840}$ is the average population at the prefecture level between 1644 and 1840. We choose not to include books written before the Qing dynasty, because the geographic distribution of intellectuals in Qing was quite different from that in earlier periods (see Appendix Table A1). *lnBookDens* is arguably one of the best measures for the local production of knowledge and, thus, for human capital stock. We also experiment with alternative constructions by dropping the normalization with population, using $lnBooks = ln(Books + 1)$ (see Appendix Table A3). Additionally, one alternative worth mentioning, $lnBookDens2 = ln(\frac{Books+1}{Pop_{pre1840}})$, is introduced in Squicciarini and Voigtländer (2015), which exploits additional variation across prefectures with zero books and different population sizes (see Appendix Table A4). Lastly, we also use alternative sources of books as the independent variable, focus on firms with high registered capital, and use the rate of urbanization as the outcome variable in Appendix Table A5.

Our baseline set of controls includes various geographic characteristics of prefectures, such as the distance to the coast, the distance to the capital, and the distance to the nearest navigable river. As noted in Section 2.2, the treaty ports were located primarily along the coast and major river systems in China. To account for disparities in prosperity, we include a set of indices on terrain ruggedness and grain suitability, along with indicators for the *chong-fan-pi-nan* grading system on prefectures.[15] Additionally, we include a panel of population density between 1840 and 1904 as a control. Finally, we include the natural log of the number of market towns in each prefecture, to control for differences in business culture, trade activities, and market structures. Please see the Baseline Controls section in Appendix Table C12 for summary statistics and data sources.

[14] This approach helps us address potential confounding factors at the provincial level. For instance, shocks such as evolving transport technology, including steamships and railways, affected different provinces differently over time.

[15] The *chong-fan-pi-nan* grading system: Prefectures were graded with a four-tier system, where 冲 (*chong*) mean places at busy highways, 繁 (*fan*) means places with numerous and complex official business, 疲 (*pi*) means places prone to tax arrears, and 难 (*nan*) means places where the population is wicked and prone to violence (Koss, 2017).

We also include a number of potential confounding factors to control for alternative explanations. Firstly, we obtain data on the number of pantheons and volumes of genealogy books in each prefecture, which serve as proxies for Confucian cultural norms. Previous research (Chen, Ma and Sinclair, 2021; Chen and Ma, 2022) has demonstrated that Confucian cultural norms had a significant impact on banking, resource pooling, and population growth, all of which were important factors behind firm creation.[16] Furthermore, we take into account the potential impact of several instances of literary inquisitions that occurred between 1711 and 1793 on both book authorship and the accumulation of social capital. As demonstrated by Xue (2021), the succession of literary inquisitions throughout the 18th century not only diminished social capital, resulting in political disengagement and diminished community involvement but also had a discernible effect on local literary output. Additionally, we control for the presence of Christianity by adding information on priest visits and church buildings. Research by Bai and Kung (2015) has shown that between 1840 and 1920, the diffusion of Protestantism affected economic prosperity in China by increasing the overall urbanization rate. Moreover, we control for hazardous shocks, such as incidents of disasters and wars, which could affect not only the normal establishment and operation of businesses, but also the effective governance of the Qing government. Finally, we include a panel of academies that were established in each prefecture between 1840 and 1904 to control for the extent to which each prefecture encouraged human capital accumulation, as the establishment of schools and the creation of firms often have a complementary relationship. Please refer to the More Controls section in Appendix Table C12 for summary statistics and data sources.

The results are presented in Table 1. Column (1) presents the results including only the prefecture and year fixed effects, and Column (2) further includes our set of baseline controls. Both show that prefectures with higher books per capita had a higher number of firm entries following the opening of treaty ports. Column (3) includes the provincial-specific time fixed effects and the prefecture-specific linear time trends, and Column (4) adds the extra set of control variables that vary at the prefecture and year level between 1840 and 1904. It should be noted that $PostPort_{i,t}$ actually had a negative and significant effect on modern firm entry, which corresponds to the short-term pattern in Figure 4a. Positive and significant coefficients of $lnBookDens_i \times PostPort_{i,t}$ imply that human capital stock, proxied by books per capita, contributed to the entry of modern firms by increasing the competitiveness of Chinese companies in the face of the sudden influx of foreign competitors. In Appendix Section A.2, when using alternative normalization methods and Poisson regression, we estimate a treatment effect where 1% increase in book density

[16] Kung and Ma (2014) also concludes that Confucianism reduced conflicts from peasant rebellions.

led to approximately a 0.7% increase in firm density or a 3.5% increase in firm count.

Table 1: Baseline Firm and BookCount

	(1) lnfirms	(2) lnfirms	(3) lnfirms	(4) lnfirms
lnBookDens × PostPort	0.906***	0.696***	0.839***	0.826***
	(0.185)	(0.104)	(0.233)	(0.241)
PostPort	0.024	-1.052***	-1.263**	-1.186**
	(0.024)	(0.384)	(0.546)	(0.546)
Prefecture FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Baseline Controls		✓	✓	✓
Prefecture × Year			✓	✓
Province × Year FE			✓	✓
More Controls				✓
Adj R-square	0.293	0.301	0.549	0.550
Observations	17355	17355	17355	17355

Note. This table reports the impact of log book density on log firm after the opening of treaty ports, compared with that before the treaty ports. Column (1) controls for both the prefecture and year-fixed effects. Column (2) also controls for our baseline set of control variables. Column (3) adds the province-specific time-fixed effects and the prefecture-specific linear time trends. Column (4) includes a set of additional confounding factors that control for alternative explanations. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We also examine the link between books per capita and modern firm entry dynamically using an event study design, in addition to the static DiD estimates above. The purpose is to confirm two things: (i) whether there were already different trends in the firm entry for prefectures with different books per capita before the opening of treaty ports; (ii) Whether the positive and significant effect captured above was actually short-lived. The specification is as follows in Equation 2:

$$lnFirm_{i,t} = \sum_{j \in J} \beta_j \cdot \mathbf{1}\{t - E_i \in j\} \times lnBookDens_i + \theta \cdot \mathbf{X}_i \times t + \kappa \cdot \mathbf{Z}_{i,t} + \gamma_t + \lambda_i + \epsilon_{i,t} \quad (2)$$

where E_i is the time when prefecture i initially receives the binary absorbing treatment, treaty ports; $t - E_i$ stands for the relative time periods based on the initial treatment timing; $\mathbf{1}\{t - E_i \in j\}$ collects the set of dummy variables corresponding to the relative periods; again γ_t and λ_i stand for the prefecture and year fixed effects.

Our variable of interest in this event study design is the interaction between books per capita and the relative periods, $\sum_{j \in J} \mathbf{1}\{t - E_i \in j\} \times lnBookDens_i$. We include all relative periods before and after treatment and bin them by decades, and we also bin remote pe-

riods that are more than 50 years away from treatment at either side of the tails. Figure 5 presents the results. We can confirm that, before the opening of treaty ports, prefectures with different books per capita showed no significant disparity in their trends of firm entry. We also observe that the positive impact of books per capita occurred after the opening of treaty ports and that its magnitude not only persisted but also increased with time. The effect of the initial treatment time on firm entry was about 0.53 and increased steadily over time, reaching a peak of approximately 4.4 after 50 years. Overall, the dynamic results suggest that human capital stock, as measured by books per capita, played a significant role in facilitating the industrialization of China during the late Qing dynasty. Moreover, the positive impact of books per capita on firm entry not only endured but also grew stronger over time, indicating its increasing importance as a driver of economic development.

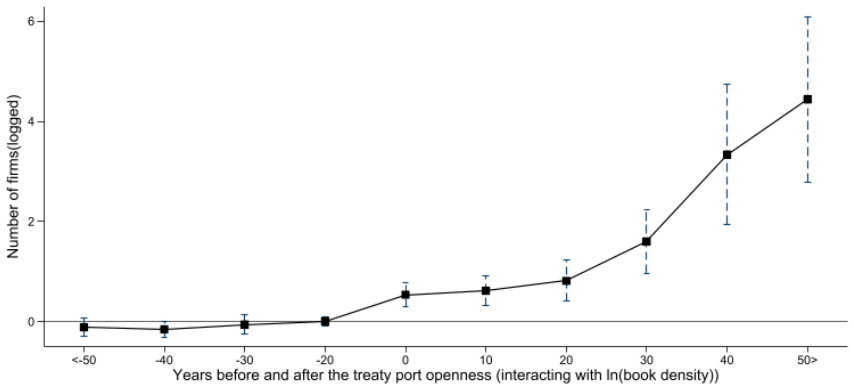


Figure 5: The Dynamic Effects of Books Density on Modern Firms Creation
This figure uses 10 years before treatment as the reference, where the solid line connects the coefficients and the dashed lines indicate the 95% confidence intervals.

We further examine potential heterogeneity in the impact of books per capita on different types of firms. Because human capital, as measured by books per capita, was mostly a spontaneous formation due to local economic and social endowment, we expect it to be a first-order causal factor in the establishment of private firms. In contrast, the state's top-down plan of the Self-strengthening Movement prioritized political and military considerations over local human capital stock. Thus, we anticipate that the impact of books per capita on private firms will be more significant than on state-owned firms. Table 2 presents the results for the three main types of firms: private firms, state-owned firms, and joint government-merchant ventures. Column (1) shows that the positive impact of books per capita on private firm entry remained significant and had the largest magnitude compared to the other two types of firms. In Columns (2) and (3), local human

capital endowment had no significant effect on the entry of state-owned firms or joint government-merchant ventures. Furthermore, we observe that the negative impact of foreign competition was significant only for private firms and joint state-merchant ventures, but not for state-owned enterprises. This is not surprising, as state-owned enterprises were concentrated in heavy industries where foreign capital had little or no access. Overall, our results demonstrate the crucial role played by local human capital endowment in facilitating bottom-up industrialization during the late Qing dynasty.

Table 2: Heterogeneity: Firm Types

	(1) lnprivate_firms	(2) lnstate_firms	(3) lnstate_private_firms
lnBookDens × PostPort	0.879***	-0.018	0.050
	(0.191)	(0.031)	(0.075)
PostPort	-0.854**	0.052	-0.552***
	(0.416)	(0.082)	(0.136)
Prefecture FE	✓	✓	✓
Year FE	✓	✓	✓
Baseline Controls	✓	✓	✓
More Controls	✓	✓	✓
Prefecture × Year	✓	✓	✓
Province × Year FE	✓	✓	✓
Adj R-square	0.548	0.077	0.307
Observations	17355	17355	17355

Note. This table reports the impact of log book density on the natural log of 3 different types of firms after the opening of treaty ports, compared with that before the treaty ports. All regressions are run with the same full set of control variables and fixed effects as Column (4) of Table 1. Column (1) has the natural log of private firms as the dependent variable. Column (2) has the natural log of state-owned enterprises as the dependent variable. Column (3) has the natural log of joint government-merchant ventures as the dependent variable. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

As indicated in Appendix Table C14, the openings of treaty ports occurred in 18 distinct time periods. Our estimates, derived from Equations 1 and 2, reflect average effects across these staggered events. Recent advancements in the econometric literature on staggered DiD highlight potential biases from heterogeneous treatment timing (e.g., Goodman-Bacon, 2021; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021). To address this concern, we apply two popular methods from this literature. Firstly, from the host of new estimators proposed in the literature, we adopt the flexible interaction weighted estimator by Sun and Abraham (2021), suitable for continuous treatment with staggered timing. An alternative approach to circumvent this issue with TWFE DiD estimator is the stacked regression method, as summarized by Baker, Larcker and Wang (2022). We first demarcate different treatment cohorts based on treatment timing and an-

alyze the treatment effect for each cohort against a "clean" control sample. Here, "clean" control units imply not only the inclusion of propensity score matching but also the restriction to later-treated cohorts and never-treated units as delineated by Goodman-Bacon (2021). We perform standard TWFE DiD estimations across event-specific datasets for specific estimates and stack these datasets for an aggregate estimate. For a more in-depth discussion, please refer to Appendix Section A.6.

For further robustness checks on our baseline findings, please refer to the following sections in the Appendix. In Appendix Section A.7, we perform a series of subsample estimations to address biases in treatment selection, including propensity score matching on a wide variety of covariates. In Appendix Section A.8, we utilize an instrumental variable on our bibliometric measure to further mitigate any omitted variable bias. To account for the spillover effect of treaty ports as identified by Keller and Shiue (2021), in Appendix Table A10, we replace the post indicator variable with a dynamic distance to the closest treaty port.

4 Interpretation and Mechanisms

In the baseline analysis, we evaluated the overarching influence of book writing in Qing China by aggregating works of all subjects, aiming to quantify the geographic distribution of educated elites prior to the country's opening. This approach yields two plausible interpretations of our results: (i) "Useful Knowledge": A subset of the traditional Chinese human capital may have harbored useful scientific knowledge that, although dormant, was activated by the jolt of establishing treaty ports, thereby becoming economically productive; (ii) "Re-allocation to Modern Learning": The traditional educational system and knowledge, while antiquated, may have sufficiently equipped Chinese intellectuals with the foundational skills and cognitive abilities necessary to rapidly absorb new knowledge and transition into burgeoning careers in the productive sectors upon the opening up of China.

In this section, we present multiple pieces of evidence for or against each hypothesis. In Section 4.1, we disaggregate the book catalog by subject matter to test the "Useful Knowledge" hypothesis, finding no empirical patterns that support this hypothesis. Consequently, Section 4.2 focuses on the reallocation hypothesis and investigates the diverse career trajectories pursued by intellectuals. Specifically, we disaggregate the book catalog based on the authors' positions in relation to the spectrum of upper-tail human capital defined by the Imperial Examination.

4.1 Useful Knowledge

Should the former interpretation hold true, our positive and significant baseline results would suggest that a subset of the pre-1840 publications likely contained knowledge relevant to industrialization. This would lead us to anticipate distinct treatment effects among books of different subjects; for example, we would expect books related to STEM fields to be pivotal, while those pertaining to Arts & Literature might prove to be non-essential (e.g., Curtis and de la Croix, 2023). To test this hypothesis, we disaggregate our baseline model by subject matter, applying it to distinct categories of books to discern the subject-specific contributions to the observed economic outcome.

Book Content Classification The imperial cataloging system not only records key information, namely the book title and the authors, but also classifies books into different themes based on their content. The Han dynasty (202 BCE–220 CE) catalog established the "Six-group Classification" (*Liù Bù Fēnlèi Fǎ*), which divides books based on their content into six major groups. Each of the six groups further contains a finer level of classification, with a total of 38 smaller groups. Starting with the Sui dynasty (581-618) catalog, this system evolved into the "Four-group Classification" (*Sì Bù Fēnlèi Fǎ*) and was followed by all later dynasties. Books are divided into four major groups and a total of 60 smaller groups. The entire 26-volume Grand Catalog also follows the classic "Four-group Classification". Please refer to Appendix Section C.4 for details. We also explored the use of a classifier model based on book titles to perform a robustness check, drawing inspiration from approaches used by Gentzkow and Shapiro (2010) and Dittmar and Seabold (2019). However, given that our classifier had to discern among more than two categories and considering the abstract nature of traditional Chinese book titles, the model yielded unsatisfactory accuracy scores. Furthermore, it failed to demonstrate any improvement in performance, even with the expansion of the training dataset. Therefore, for the content classification within this paper, we rely on the readily available traditional content classification systems, and align them with contemporary subject categories. To summarize, our books are categorized into 28 modern subjects, which are further consolidated into broader topic areas, including STEM, Philosophy, Religion, Social Sciences, and Arts&Literature.

The Role of Book Content We acknowledge the anachronistic nature of these subject names when applied to pre-industrial books; for instance, subjects within the 'STEM' topic area—like "Engineering", "Physics", and "Mathematics"—may misleadingly suggest direct relevance to industrialization. For clarity in our exposition, we have opted for

these familiar modern terminologies, using "STEM" in lieu of the historically accurate "natural philosophy" and "medicine" in substitution for "herbology". Despite potential limitations, we posit that our five-category schema offers a reasonable approximation of the quintessential domains engaged by Chinese intellectuals, which include: (i) STEM, addressing challenges in the natural world; (ii) Social Sciences, dealing with human organizational issues; (iii) Arts & Literature, indulging in aesthetic pursuits; (iv) Philosophy, examining human relationships; and (v) Religion, contemplating transcendental questions.

Table 3: Firms and Books in Different Topic Areas

	(1) lnfirms	(2) lnfirms	(3) lnfirms	(4) lnfirms	(5) lnfirms	(6) lnfirms
lnBookDens_AL × PostPort	1.299***					
	(0.368)					
lnBookDens_SocialSci × PostPort		2.690***				
		(0.848)				
lnBookDens_Phil × PostPort			4.959***			
			(1.707)			
lnBookDens_Relig × PostPort				14.593		
				(18.015)		
lnBookDens_STEM × PostPort					4.871***	
					(1.663)	
lnBookDens_nonSTEM × PostPort						0.905***
						(0.261)
PostPort	-1.069**	-1.436**	-1.126**	-1.477**	-1.033*	-1.196**
	(0.536)	(0.587)	(0.540)	(0.594)	(0.570)	(0.544)
Prefecture FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓	✓	✓	✓
Prefecture Year	✓	✓	✓	✓	✓	✓
Province Year FE	✓	✓	✓	✓	✓	✓
More Controls	✓	✓	✓	✓	✓	✓
Adj R-square	0.550	0.550	0.550	0.548	0.549	0.550
Observations	17355	17355	17355	17355	17355	17355

Note. This table reports the impact of log book density on log firm across books in different topic areas. Column (1) focuses on books in Arts & Literature. Column (2) focuses on books in Social Sciences. Column (3) focuses on books in Philosophy. Column (4) focuses on books in Religion. Column (5) focuses on books in STEM. Column (6) groups the 4 non-STEM topic areas into one broader category. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3 presents the disaggregated regression outcomes by book topic areas. Contrary to expectations, we find that books across all categories, with the exception of Religion, exert a significant and positive influence on the inception of modern firms. These surprising

findings suggest that the specific content of books may not be as crucial as hypothesized, thus challenging the notion that STEM books exclusively capture "useful" scientific and technological knowledge pertinent to industrialization. Overall, the findings here suggest that the true value of traditional education may lie not in the content per se.

4.2 Re-allocation to Modern Learning

Should the reallocation hypothesis hold, our baseline findings suggest that talent reallocation towards modernization indeed occurred on a significant scale, indicating that the establishment of treaty ports provided a sufficient impetus to dislodge a sizable segment of intellectuals from the entrenched Examination-officialdom system. Consequently, we would anticipate distinct treatment effects among books authored by different groups of intellectuals, reflecting their varied career choices. For instance, given that the Imperial Examination previously dictated the incentive structure for Chinese intellectuals, we would expect those who stood to benefit more from the established Examination system to be less inclined to transition into modernization. To test this hypothesis, we disaggregate the books by the authors' qualifications within the Examination system and aim to identify which segments of Chinese intellectuals were incentivized to pursue modern knowledge and business.

4.2.1 The Imperial Examination

To better understand what constituted a typical career path for a Chinese intellectual prior to the year 1840 and the complex composition of upper-tail human capital, we first need to outline the historical background and structure of the exam system.

The Structure of the Examination The exam consisted of three major levels, which are widely known today. The first level was the triennial provincial-level exam (*Xiāng Shì*) held at provincial capitals. Successful candidates were granted the title of "recommended man" (*Jǔ Rén*) and were qualified to take the next level of the exam. The second level was the triennial national exam (*Huì Shì*) held at the national capital. Those who passed this level were granted the title of "Passed Scholars" (*Gòng Shì*) and were qualified for the final level of the exam. The third and final level was held at the imperial court in front of the emperor (*Diàn Shì*), and served to rank all the "Passed Scholars." All candidates who passed the final exam were granted the prestigious title of "Presented Scholars" (*Jìn Shì*). *Jinshi* degree holders were qualified for mid to high official positions in the national capital or provinces other than their home provinces.

To qualify for the provincial-level *Xiangshi*, they first needed to pass three levels of licensing tests, which were held twice every three years. All prospective candidates at the beginning were referred to as "Apprentice Students" (*Tóng Shēng*), who were preparing for the licensing tests (*Tóng Shì*). Candidates who successfully passed these licensing tests were officially registered as "Government Students" (*Shēng Yuán*) at one of their regional government schools.[17] Appendix Section B.1 provides more historical details on the distribution of Chinese intellectuals across these Examination titles and demonstrates that the vast majority of Chinese upper-tail human capital was positioned at the lower tiers or entirely outside the Examination system.

The Examination after 1840 The emergence of treaty ports signified a profound transformation in China's economic framework, with differential impacts across the intellectual spectrum. Notably, the career trajectories of incumbent bureaucrats and top-tier examination laureates, particularly the *Jinshi*, appeared largely insulated from these shifts. Scholarly contributions by Bai (2019) and Bai and Jia (2016) suggest that this echelon of human capital pivoted toward engagement in modern business and political reforms primarily after the momentous decree on September 2, 1905, which dismantled the imperial examination system. In stark contrast, for the literati who occupied the lower echelons of the examination hierarchy or those positioned entirely outside of it, the treaty ports arguably ushered in unprecedented vocational alternatives. These newly available opportunities potentially matched or exceeded the attractiveness of their conventional roles in low-tier government service or teaching, marking a significant departure from a millennium-long status quo.

Hence, we hypothesize that the *Jinshi* population or other scholars close to the center of the Imperial Examination were unlikely to reposition towards industrialization following the opening of treaty ports, but for literati at the bottom or outside of the Examination system, the opening-up brought sufficient incentives to shift towards industrialization. The subsequent sections are dedicated to the quantitative assessment of this hypothesis, aiming to shed light on the complex dynamics between the opening-up and the Examination at shaping the observed talent re-allocation.

4.2.2 Dissecting Human Capital and Contrasting Contributions to Firm Entry

In this section, we integrate various measures of human capital to systematically dissect and quantify the complex composition of upper-tail human capital during late Qing China. Through a clear dissection, we first assess whether the *Jinshi* circle contributed to

[17] "Government Students" were also commonly known as "the literati" (*Xiù Cái*).

the emergence of modern firms in China between 1840 and 1904. Second, we explore the intricate human capital composition within the non-*Jinshi* population and elucidate the contributions of its various subgroups.

The existing literature commonly employs two measures as proxies for human capital, and coincidentally, both of them are related to the Imperial Examination. The most commonly used and easily accessible measure is the prefecture-level exam quotas for *Shengyuan* from the Imperially Established Institutes and Laws of the Great Qing Dynasty (Kun et al., 1899). Quota density serves as a useful indicator of human capital at the entry level of the examination system and closely approximates the total number of exam takers in each prefecture. However, it is important to acknowledge the limitations of this measure. Firstly, the quota assignments remained largely unchanged throughout most of the Qing Dynasty, partly due to the limitations in the state's capacity (Chang, 1955). This raises doubts about the accuracy of quotas in reflecting the actual human capital landscape of each locality. Additionally, the original purpose of the quota system was to "confine and regulate the power of elites" and shape the size and composition of successful candidates in the licensing tests (Elman, 2000, p. 140). Considering the political motivations underlying this system, it is prudent to exercise caution regarding the potential endogeneity associated with quota density.

The other measure is *Jinshi*. The use of prefecture-level *Jinshi* count or density as a measure of human capital provides better variation across both space and time compared to quotas. Furthermore, with the availability of the complete biographical database on *Jinshi* from the book *Ming-Qing Jinshi Timing Beilu Suoyin* (Official Directory of Ming-Qing Civil Exam Graduates) by Zhu and Xie (1979), we not only have the count but also access to basic biographical information for each *Jinshi*. However, it is important to acknowledge the inherent disadvantages of using *Jinshi* as a measure of human capital. In terms of representativeness, *Jinshi* is an even narrower proxy compared to quotas. As acknowledged by Chen, Kung and Ma (2020), there are also endogeneity concerns associated with the variation in *Jinshi* density. The Qing Dynasty governed the number of *Jinshi* through a province-level quota system, and as noted by Xue and Zhang (2023), this quota system was influenced by political considerations similar to the motivations behind contemporary Affirmative Action policies in the United States. These factors raise concerns about the representativeness and the potential endogeneity embedded within *Jinshi* density as a measure of human capital.

In contrast, our bibliometric measure not only encompasses a much wider range of human capital but also mitigates the influence of political motivations or institutional biases. The cohort of authors in our book catalog consists of individuals who participated in

the examination system, spanning from those at the highest echelons to those at the lowest, as well as intellectuals who operated outside of the Examination system. Although there appears to be a hierarchy among these authors based on their examination achievements, we contend that all books are indicators of upper-tail human capital. Given the exclusivity of the Examination system, even entry-level qualifications represent a form of upper-tail human capital. [18] Hence, our decomposition of human capital does not revolve around upper-tail versus average human capital, but rather human capital within versus outside the Examination system.

Table 4 below summarizes the correspondence between the four human capital measures utlized in this study and the actual population of intellectuals within and outside the Examination system. Although there are two relatively robust measures for the *Jinshi* population—the density of *Jinshi* and the density of books authored by *Jinshi*—accurately gauging the population in the intermediate levels of the examination system and those without any title is considerably more challenging.

Table 4: Keju Population and Measures

Measures / Groups	**Jinshi**	**Juren to Shengyuan**	**Intellectuals w/o exam titles**
Jinshi Density	✓		
Book Density (Jinshi)	✓		
Quota Density	✓	✓	
Book Density (Non-Jinshi)		✓	✓

Note. This table categorizes intellectuals from historical China into three groups and identifies which of these are captured by the four human capital measures utilized in this research.

Jinshi vs. Non-Jinshi Now we first focus on the *Jinshi* population and examine whether the *Jinshi* population was successfully motivated into modern entrepreneurship during the period 1840-1904. The results from our static Difference-in-Differences (DiD) estimations are presented in Table 5. In Columns (1) to (4), our focus is on the interaction between *Jinshi* density and treaty port, and we find that the significance of *Jinshi*-related human capital diminishes after accounting for alternative explanatory factors. We also conduct a dynamic event study on *Jinshi* density using the same specification as Equation 2. Appendix Figure B3 illustrates that the impact of Jinshi density did not exhibit significance until 30 years after the treatment. These results underscore the necessity of carefully selecting appropriate measures of human capital when examining its role in the

[18] Numerous renowned historical figures in China achieved only entry-level titles in their attempts at the Examination: the famous herbalist 李时珍(Li Shizhen) held the title of *Shengyuan*; the esteemed scientist 宋应星(Song Yingxing) obtained the title of *Juren*.

emergence of modern firms in historical China, while also highlighting the importance of our data collection efforts.

Table 5: Books vs. *jinshi*-related Human Capital

	(1) lnfirms	(2) lnfirms	(3) lnfirms	(4) lnfirms	(5) lnfirms
lnJinshiDens × PostPort	1.753***	1.152***	1.227***	1.024*	-2.057
	(0.540)	(0.432)	(0.420)	(0.613)	(1.355)
lnBookDens × PostPort					1.502***
					(0.542)
PostPort	0.020	-1.356**	-1.584**	-1.529**	-0.873*
	(0.026)	(0.530)	(0.700)	(0.637)	(0.455)
Prefecture FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Baseline Controls		✓	✓	✓	✓
Prefecture × Year			✓	✓	✓
Province × Year FE			✓	✓	✓
More Controls				✓	✓
Adj R-square	0.284	0.299	0.547	0.548	0.550
Observations	17355	17355	17355	17355	17355

Note. This table reports the impact of log *jinshi* density on log firm after the opening of treaty ports, compared with that before the treaty ports. Column (1) controls for both the prefecture and year-fixed effects. Column (2) also controls for our baseline set of control variables. Column (3) adds the province-specific time-fixed effects and the prefecture-specific linear time trends. Column (4) includes a set of additional confounding factors that control for alternative explanations. Column (5) reports a horse race regression between log book density and log *jinshi* density. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Column (5), we conduct horse race regressions by comparing *Jinshi*-related human capital with our bibliometric measure. The results indicate that, after controlling for book density, the previously observed positive but unstable effects of *Jinshi* become insignificant. These findings not only serve as an additional test of the robustness of our main result but also suggest that the previously identified link between *Jinshi*-related human capital and firm entries was spurious and driven by the correlation between *Jinshi*-related human capital and bibliometric human capital.

In order to further strengthen this finding, we also measure *Jinshi* human capital with our bibliometric measures, the density of books authored by *Jinshi*. The availability of a complete biographical database on *Jinshi* enables us to identify authors who held the prestigious *Jinshi* title in our sample, thereby facilitating the decomposition of our bibliometric measure.[19] We conduct horse race regressions on books written by *Jinshi* and books written by non-*Jinshi* to determine which one better explains the variations in firm

[19]For a more thorough examination of raw data patterns, please consult the Appendix Section B.3.

Table 6: Books by *jinshi* vs. Books by Others

	(1) lnfirms	(2) lnfirms	(3) lnfirms
lnBookDens_*jinshi* × PostPort	-1.531		
	(1.254)		
lnBookDens_not*jinshi* × PostPort	2.331**		
	(0.920)		
lnAuthorDens_*jinshi* × PostPort		-11.161**	
		(4.527)	
lnAuthorDens_not*jinshi* × PostPort		11.165***	
		(3.076)	
lnBookDens_*jinshi*_nonSTEM × PostPort			-1.267
			(1.341)
lnBookDens_not*jinshi*_nonSTEM × PostPort			2.311**
			(0.976)
PostPort	-0.652	-0.633	-0.819*
	(0.492)	(0.422)	(0.477)
Prefecture FE	✓	✓	✓
Year FE	✓	✓	✓
Baseline Controls	✓	✓	✓
More Controls	✓	✓	✓
Prefecture Year	✓	✓	✓
Province Year FE	✓	✓	✓
Adj R-square	0.550	0.551	0.550
Observations	17355	17355	17355

Note. This table reports horse race regressions between bibliometric proxy of *jinshi* related human capital and that of non-*jinshi* related. Column (1) compares books written by *jinshi* and books written by other non-*jinshi* authors. Column (2) compares *jinshi* authors and non-*jinshi* authors. Column (3) compare the non-STEM books written by *jinshi* and others. All regressions are run with the full set of fixed effects and control variables. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

creation. As shown in Table 6, we group the bibliometric measures at both the book level and the author level. Column (1) reveals that only the subsample of books written by non-*Jinshi* positively contributed to firm creation, while the subsample of books written by *Jinshi* had a non-significant negative effect. In Column (2), the pattern is similar when we examine the author level, except that the negative effect from *Jinshi* authors is now significant at the 5% level. Please refer to Appendix Figure B7 for the correlations among these measures.

We contend that the differential impacts observed between *Jinshi* and non-*Jinshi* populations do not stem from variations in the content of the books they authored. Instead, these differences are primarily attributed to the distinct career prospects of the *Jinshi* population, for whom the rewards within the exam-officialdom system were sufficiently sub-

stantial to overshadow the incentives presented by nascent modern business opportunities. In Appendix Figure B6, we examine books by subject matter, illustrating their composition based on authorship by *Jinshi* and non-*Jinshi* writers. We find minimal disparity in the range of subjects covered by both groups, with the possible exception of STEM fields, where non-*Jinshi* authors appear to have made greater contributions. Therefore, in Column (3), after excluding STEM-related works from the analysis for both cohorts, the results reaffirm our initial findings: when controlling for content similarity in the books, we continue to discern a positive and significant contribution solely from the non-*Jinshi* demographic.

We also adopt the approach used by Becker and Woessmann (2009) in their study on the human capital theory of Protestant economic history, where the authors compare the effects of two variables, Protestant share and literacy rate. To carry out this check, we add the nearest distance to bamboo as an instrumental variable to our horse race specification.

$$\widehat{lnFirm_{i,t}} = lnFirm_{i,t} - \psi \cdot lnBookDens\_notJinshi_i \times PostPort \tag{3}$$

$$\widetilde{lnFirm_{i,t}} = lnFirm_{i,t} - \beta \cdot lnBookDens\_Jinshi_i \times PostPort_{i,t} \tag{4}$$

First, we examine the variation in firm entries that cannot be attributed to books authored by non-*Jinshi*, as specified in Equation 3. Here, ψ represents the coefficient of lnBookDens_not*Jinshi* in Column (1) of Table 6. We then test the effect of books written by *Jinshi* using the nearest distance to bamboo as an instrumental variable. As shown in Column (1) of Table 7, after accounting for the firm entry variation due to books authored by non-*Jinshi*, we find that books authored by *Jinshi* still had no significant effect on the creation of modern firms.

Next, we focus on the variation in firm entries not explained by books authored by *Jinshi*, as specified in Equation 4, where β is the coefficient of lnBookDens_*Jinshi* in Column (1) of Table 6. We investigate the effect of books written by non-*Jinshi*, again using the nearest distance to bamboo as an instrumental variable. Column (2) of Table 7 shows that even after removing the firm entry variation due to books authored by *Jinshi*, books authored by non-*Jinshi* continued to have a significant and positive effect.[20]

In summary, we consistently find that *Jinshi*-related human capital did not make a positive contribution to modern firm creation before 1905, and the decomposition of human capital precisely identifies that the portion of human capital outside of the *Jinshi* circle was mobilized into industrial entrepreneurship.

[20]This robustness check is conducted only at the individual book level, as the nearest distance to bamboo is a weak instrument for author distribution.

Table 7: IV Approach on Books by *jinshi* vs. by Others

	(1) $\widehat{lnFirm_{i,t}}$	(2) $\widetilde{lnFirm_{i,t}}$
	IV	IV
lnBookDens_*jinshi* × PostPort	-1.065	
	(0.739)	
lnBookDens_not*jinshi* × PostPort		2.687***
		(0.560)
PostPort	-0.616	-0.504
	(0.468)	(0.437)
	1st Stage	
	lnBookDens_*jinshi* × PostPort	lnBookDens_not*jinshi* × PostPort
lnBambooDist × PostPort	-0.041***	-0.053***
	(0.007)	(0.009)
PostPort	0.207**	-0.044
	(0.099)	(0.124)
Prefecture FE	✓	✓
Year FE	✓	✓
Baseline Controls	✓	✓
More Controls	✓	✓
Prefecture × Year	✓	✓
Province × Year FE	✓	✓
Kleibergen-Paap F-stat	31.745	37.147
Observations	17355	17355

Notes. This table examines whether the finding about *jinshi* vs. non-*jinshi* books are robust to using distance to the nearest bamboo site as the instrument variable. Columns (1)-(2) present the second stage result at the top and the first stage result at the bottom. Column (1) presents the result for the log density of books written by *jinshi* when instrumented by bamboo distance. Column (2) presents the result for the log density of books written by non-*jinshi* when instrumented by bamboo distance. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Zoom into the Non-Jinshi Population Our analysis now shifts towards pinpointing the particular segment of human capital outside the *Jinshi* circle that aligns with the positive outcomes observed in the overall human capital stock. As delineated in Table 4, conceptually, the non-*Jinshi* intellectuals can be categorized into two principal groups. The first group encompasses those Examination participants who attained any title, ranging from *Shengyuan* to *Juren*. The second group consists of the numerous intellectuals who did not secure any official exam title. In an ideal research scenario, these groups would be measured separately, or our bibliometric indicators would be decomposed to reflect the contributions from each category of authorship. Regrettably, comprehensive rosters and biographical data for individuals with lower to mid-level titles are scarce, and data

on the second group is nearly non-existent, posing a substantial obstacle to our analytical efforts.

Owing to these constraints, scholars have largely focused on quota density when examining the broader intellectual spectrum beyond the *Jinshi* population. As discussed in Appendix Section B.1, this approach fails to account for the 98.5% of examinees who did not secure the *Shengyuan* title in each examination cycle, as well as intellectuals who did not participate in the competitive examinations. In an effort to bridge this gap, our research makes a novel contribution by identifying books penned by non-*Jinshi* authors, thereby capturing the entirety of the intellectual cohort lacking any examination titles. For the purposes of our research, it is necessary to acknowledge the intrinsic overlaps between quota density and *Jinshi* density, as well as between non-*Jinshi* authorship and quota density metrics (see Table 4).

Bearing this in mind, we first aim to ascertain whether the human capital as represented by quota density played any role in the establishment of modern firms. In Appendix Figure B4, the dynamic event study result is similar to that of *Jinshi* density. In static Difference-in-Differences (DiD) estimations, Columns (1) to (4) of Table 8 confirm that the segment of human capital represented by quota density was not instrumental to the emergence of modern firms between 1840 and 1904.

Subsequently, we attempt to delineate which specific subgroup within the non-*Jinshi* category was instrumental in this economic transformation. In Column (5), we conduct a horse-race regression among quota density, *Jinshi* density, and the density of books authored by non-*Jinshi* intellectuals. This analysis is designed to disentangle the confounding overlaps and identify the specific subgroup that significantly influenced the emergence of modern firms. According to Appendix Figure B7, the correlations among these three variables are not high, with the highest at 0.55 between *Jinshi* density and the density of books authored by non-*Jinshi*. In Column (6), we replicate this horse-race regression, except that we use books authored by *Jinshi* as the proxy for *Jinshi* human capital.

The coefficient for quota density is intended to isolate the impact of human capital represented by individuals who attained titles from *Shengyuan* to *Juren*. The results suggest that these intellectuals, possessing lower to mid-level exam titles, were not significantly inclined towards modern entrepreneurial ventures. Furthermore, the coefficient for *Jinshi* density reiterates that higher concentrations of *Jinshi* were likely associated with a hesitancy towards the establishment of modern firms. In contrast, the positive and statistically significant coefficient for the density of books by non-*Jinshi* authors indicates that the positive relationship between the general human capital stock and modern firm creation, as observed in our baseline analysis, predominantly originates from the segment of

intellectuals without any examination titles.

Table 8: Firm and Quota Density

	(1) lnfirms	(2) lnfirms	(3) lnfirms	(4) lnfirms	(5) lnfirms	(6) lnfirms
lnQuotaDens × PostPort	-1.508**	0.347	1.509***	0.777	0.577	0.331
	(0.595)	(0.645)	(0.560)	(0.520)	(0.608)	(0.618)
lnBookDens × PostPort				0.751***		
				(0.241)		
lnBookDens_not*jinshi* × PostPort					2.206***	2.051*
					(0.728)	(1.073)
lnJinshiDens × PostPort					-2.237*	
					(1.221)	
lnBookDens_*jinshi* × PostPort						-1.246
						(1.417)
PostPort	0.269***	-1.283**	-1.893***	-1.419***	-0.675	-0.826
	(0.082)	(0.578)	(0.570)	(0.511)	(0.495)	(0.531)
Prefecture FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Baseline Controls		✓	✓	✓	✓	✓
Prefecture × Year			✓	✓	✓	✓
Province × Year FE			✓	✓	✓	✓
More Controls			✓	✓	✓	✓
Adj R-square	0.280	0.297	0.548	0.550	0.551	0.550
Observations	17355	17355	17355	17355	17355	17355

Note. This table reports the impact of quota density on firm before and after the opening of treaty ports. Column (1) controls for both the prefecture and year-fixed effects. Column (2) also controls for our baseline set of control variables. Column (3) adds the province-specific time-fixed effects, the prefecture-specific linear time trends, and a set of additional confounding factors that control for alternative explanations. Column (4) reports a horse race regression between log book density and quota density. Column (5) reports a horse race regression among quota density, jinshi density, and the density of books by non-*jinshi*. Column (6) reports the same horse race regression, except that it uses books written by *Jinshi* authors as the proxy for the *Jinshi* huamn capital. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

4.2.3 Harbingers of Modern Learning

So far, our quantitative evidence has highlighted the contributions from the segment of intellectuals without examination titles. However, we have yet to determine precisely where these intellectuals relocated during the industrialization process. In Section 2.2, our historical survey indicates that these "indigenous agents for the remaking of China" primarily operated in two domains: as entrepreneurs and as intellectuals disseminating new ideas. In Appendix Section B.5, historical anecdotal evidence suggests that most industrial entrepreneurs indeed lacked examination titles, yet it also indicates that many of

these entrepreneurs might not have been intellectuals. Given the relatively small number of entrepreneurs compared to the large number of intellectuals without examination titles, we hypothesize that the majority of these intellectuals served as conduits for promoting modernization. To quantify the effort at spreading new ideas, we collect information on the establishment of scientific journals and societies from 1840 to 1904.

According to the compilation by Yao, Wang and Yao (2008), there is a discernible temporal trend in the establishment of scientific journals from 1792 to 1949. Initially, journals bore classical Chinese titles and predominantly served to introduce Western novelties, exemplified by titles such as *Records of Observations in China and the West* (*Zhōng Xī Jiànwén Lù*) and *Enlightenment and Knowledge Newspaper* (*Qǐméng Gézhì Bào*). Over time, however, a shift toward specialization is evident, with journals adopting more professional monikers, indicative of their focus on particular fields and industries—examples include *Journal of Chemical Pharmaceutical Industry* (*Huàxué Yàoyè Zázhì*) and *Chinese Fisheries Magazine* (*Zhōnghuá Shuǐchǎn Zázhì*). Although the actual content of these journals remains beyond our reach, their titles and brief histories offer substantial insight into the gradual assimilation of Western scientific knowledge by Chinese intellectuals, and the distinct trajectory from generalist to specialist publications. This transition is poignantly encapsulated in the prelude poetry of a medical journal:

> Apollo, the deity of medicine, drives a train long like a serpent, charging forward with lightning speed. He races through history, sweeping across the globe, darting past those ancient, dim stations, never stopping, never stopping... Listen! The roars of Shennong and Huangdi from behind us: "You unworthy descendants, why do you cling to the dead remains of us? Hurry forward! You have your era; we have our graves, quickly catch up, your era will never pause for those who fall behind!"
>
> — (Yao, Wang and Yao, 2008, p. 4)

This prelude captures the intense yearning and anguish of Chinese intellectuals to promote Western medical knowledge. The invocation of legendary figures such as Shennong and Huangdi underscores the profound disparity between traditional knowledge and modern medical understanding.

Moreover, China Kexie Development Research Center (2014) collates data on the founding of science and technology societies from 1840 to 1949. The initial wave was largely dedicated to enlightenment and basic dissemination of Western science, integral to the "The Spread of Western Learning to the East" (*Xīxué Dōngjiàn*) movement. Collectively,

these enlightenment objectives during the nascent stages of these journals and societies suggest that, in the face of burgeoning industrialization, Chinese intellectuals—armed with traditional human capital—were not as immediately productive as their French contemporaries referenced in Squicciarini and Voigtländer (2015). Nevertheless, they demonstrated agility in adapting their scholarly pursuits to embrace modern knowledge streams.

Table 9: Modern Journals and Societies and Books

	(1) lnjnsoc	(2) lnjnsoc	(3) lnjnsoc	(4) lnjnsoc
lnBookDens × PostPort	0.101***	0.131**		
	(0.027)	(0.052)		
lnBookDens_not*jinshi* × PostPort			0.217***	0.514*
			(0.074)	(0.283)
lnBookDens_*jinshi* × PostPort				-0.436
				(0.356)
lnQuotaDens × PostPort				0.018
				(0.211)
PostPort	-0.005	-0.092	-0.052	0.036
	(0.004)	(0.095)	(0.094)	(0.157)
Prefecture FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Baseline Controls		✓	✓	✓
Prefecture × Year		✓	✓	✓
Province × Year FE		✓	✓	✓
More Controls		✓	✓	✓
Adj R-square	0.115	0.292	0.292	0.292
Observations	17355	17355	17355	17355

Note. This table reports the impact of book density on the establishment of modern journals and societies for the learning of modern knowledge. Columns (1) and (2) include all books while varying the specifications. Column (3) only includes books written by non-*Jinshi* authors. Column (4) conducts a horse race among books written by non-*Jinshi* authors, books written by *Jinshi* authors, and quota density. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In Table 9, our analysis seeks to determine whether Chinese intellectuals, particularly those without examination titles, played a role in the dissemination of modern learning and industrialization. Columns (1) and (2) first confirm a significant and positive relationship between the reservoir of traditional human capital and the proliferation of modern science and technology journals and societies. Column (3) focuses on books authored by non-*Jinshi* individuals and finds that this group made a significant contribution of much larger magnitude. In Column (4), we perform a horse race regression among the same three groups of intellectuals as in Table 8, highlighting that intellectuals without exami-

nation titles were the only group actively promoting the spread of modernization.

5 Conclusion

To the best of our knowledge, our research is the inaugural systematic effort to dissect and quantify the complex composition of upper-tail human capital in historical China. Our disaggregated analysis of various segments of human capital unveils a nuanced landscape of the diverse career trajectories pursued by intellectuals, each positioned differently in relation to the Examination spectrum. Numerous studies have delineated the misallocation of talent attributable to the Imperial Examination (Baumol, 1990; Elman, 1991; Lin, 1995; Elman, 2000; Lowery and Baumol, 2013; Elman, 2013). Consistent with this body of work, our findings suggest that the Examination system confined *Jinshi*—China's crème de la crème—within the bureaucratic apparatus, impeding their engagement in innovative entrepreneurial activities. Furthermore, our research is the first to illuminate a significant talent re-allocation into modern entrepreneurship from the extensive pool of human capital positioned at the lower tiers or entirely external to the Examination system.

Although our study examines a historical setting and the Examination system no longer exists, the implications of our research may still be relevant for more recent periods and other institutional contexts. For example, modern China continues to select talent through the Nationwide Unified Examination for Admissions to General Universities and Colleges (*Gāo kǎo*). When Deng Xiaoping opened China to globalization in the 1970s, a significant number of *Gaokao* graduates transitioned towards entrepreneurship, while those with higher exam scores tended to remain within the bureaucratic system (Bai et al., 2024). Thus, the patterns identified in our study persist in contemporary contexts.

References

Acemoglu, Daron. 1995. "Reward structures and the allocation of talent." *European Economic Review*, 39: 17–33.

Allen, Robert C. 2003. "Progress and Poverty in Early Modern Europe." *Economic History Review*, 56: 403–443.

Ancient Books Editorial Committee. 2009. *Zhongguo Guji Zongmu (Grand Catalog of Books in pre-1912 China).* Shanghai: Shanghai Classical Publishing House.

Aneja, Abhay, and Guo Xu. 2024. "Strengthening State Capacity: Civil Service Reform and Public Sector Performance during the Gilded Age." *American Economic Review*, 114(8): 2352–2387.

Bai, Chong-En, Ruixue Jia, Hongbin Li, and Xin Wang. 2024. "Entrepreneurial Reluctance: Talent and Firm Creation in China." *The Economic Journal*.

Bai, Ying. 2019. "Farewell to confucianism: The modernizing effect of dismantling China's imperial examination system." *Journal of Development Economics*, 141.

Bai, Ying, and James Kai-sing Kung. 2015. "Diffusing Knowledge While Spreading God's Message: Protestantism and Economic Prosperity in China, 1840–1920." *Journal of the European Economic Association*, 13: 669–698.

Bai, Ying, and Ruixue Jia. 2016. "Elite Recruitment and Political Stability: The Impact of the Abolition of China's Civil Service Exam." *Econometrica*, 84(2): 677–733.

Baker, Andrew, David F. Larcker, and Charles C.Y. Wang. 2022. "How much should we trust staggered difference-in-differences estimates?" *Journal of Financial Economics*, 144: 370–395.

Baten, Joerg, and Jan Luiten van Zanden. 2008. "Book production and the onset of modern economic growth." *Journal of Economic Growth*, 13(3).

Baumol, William J. 1990. "Entrepreneurship: Productive, Unproductive, and Destructive." *Journal of Political Economy*, 98(5).

Becker, Sascha O., and Ludger Woessmann. 2009. "Was Weber Wrong? A Human Capital Theory of Protestant Economic History." *The Quarterly Journal of Economics*, 124(2): 531–596.

Becker, Sascha O., Francisco J. Pino, and Jordi Vidal-Robert. 2021. "Freedom of the Press? Catholic Censorship During the Counter-Reformation." CEPR Discussion Paper No. DP16092.

Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu. 2020. "The Glittering Prizes: Career Incentives and Bureaucrat Performance." *The Review of Economic Studies*, 87(2): 626–655.

Besley, Timothy, and Maitreesh Ghatak. 2005. "Competition and Incentives with Motivated Agents." *American Economic Review*, 95(3): 616–636.

Besley, Timothy, and Torsten Persson. 2011. *Pillars of Prosperity: The Political Economics of Development Clusters.* Princeton University Press.

Besley, Timothy, Robin Burgess, Adnan Khan, and Guo Xu. 2022. "Bureaucracy and Development." *Annual Review of Economics*, 14: 397–424.

Bo, Shiyu, Cong Liu, and Yan Zhou. 2023. "Military investment and the rise of industrial clusters: Evidence from China's self-strengthening movement." *Journal of Development Economics*, 161: 103015.

Britannica, Encyclopedia. 2017. "Treaty of Nanjing — China-United Kingdom [1842]." Accessed 17 March 2017.

Cage, Julia, and Valeria Rueda. 2016. "The Long-Term Effects of the Printing Press in Sub-Saharan Africa." *American Economic Journal: Applied Economics*, 8(3): 69–99.

Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics*, 225: 200–230.
Chaney, Eric. 2016. "Religion and the Rise and Fall of Islamic Science." Working paper.
Chaney, Eric. 2020. "Modern Library Holdings and Historic City Growth." Working paper.
Chang, Chung-Li. 1955. *The Chinese Gentry: Studies on Their Role in Nineteenth-Century Chinese Society.* Seattle: University of Washington Press.
Chen, Ting, James Kai-sing Kung, and Chicheng Ma. 2020. "Long Live Keju! The Persistent Effects of China's Civil Examination System." *The Economic Journal*, 130(631): 2030–2064.
Chen, Zhiwu, and Chicheng Ma. 2022. "Prioritizing Stability over Productivity: How Confucianism Facilitated Population Growth in Preindustrial China." Working Paper.
Chen, Zhiwu, Chicheng Ma, and Andrew J Sinclair. 2021. "Banking on the Confucian Clan: Why China Developed Financial Markets So Late." *The Economic Journal*.
China Kexie Development Research Center. 2014. *Jindai Zhongguo Keji Shetuan (Modern Chinese Science and Technology Societies).* China Science and Technology Press.
Curtis, Matthew, and David de la Croix. 2023. "DP18321 Seeds of Knowledge: Premodern Scholarship, Academic Fields, and European Growth." CEPR Press CEPR Discussion Paper No. 18321, Paris & London.
Denecke, Wiebke, Wai-yee Li, and Xiaofei Tian. 2017. *The Oxford Handbook of Classical Chinese Literature (1000 BCE-900 CE).* Oxford University Press.
Dittmar, Jeremiah, and Skipper Seabold. 2019. "New Media and Competition: Printing and Europe's Transformation after Gutenberg." Centre for Economic Performance, LSE CEP Discussion Papers dp1600.
Dittmar, Jeremiah E. 2011. " Information Technology and Economic Change: The Impact of The Printing Press *." *The Quarterly Journal of Economics*, 126(3): 1133–1172.
Du, Xuncheng. 1991. *Minzu Ziben Zhuyi yu JiuZhongguo Zhengfu 1840-1937 (national capitalism and old Chinese government 1840-1937).* Shanghai Academy of Social Sciences Press.
Dyke, Paul A. Van. 2007. *The Canton Trade: Life and Enterprise on the China Coast, 1700-1845.* Hong Kong University Press.
Elman, Benjamin A. 1991. "Political, Social, and Cultural Reproduction via Civil Service Examinations in Late Imperial China." *The Journal of Asian Studies*, 50(1): 7–28.
Elman, Benjamin A. 2000. *A Cultural History of Civil Examinations in Late Imperial China.* Berkeley: University of California Press.
Elman, Benjamin A. 2013. *Civil Examinations and Meritocracy in Late Imperial China.* Cambridge: Harvard University Press.
Fairbank, John King, and Kwang-Ching Liu. 1980. *The Cambridge History of China...: Late Ch'ing, 1800-1911.* Vol. 11, Cambridge University Press.
Fay, Peter Ward. 2000. *The Opium War, 1840–1842: Barbarians in the Celestial Empire in the Early Part of the Nineteenth Century and the Way by Which They Forced the Gates Ajar.* Chapel Hill, North Carolina:University of North Carolina Press.
Finan, F., B.A. Olken, and R. Pande. 2017. "Chapter 6 - The Personnel Economics of the Developing State." In *Handbook of Economic Field Experiments*. Vol. 2 of *Handbook of Economic Field Experiments*, , ed. Abhijit Vinayak Banerjee and Esther Duflo, 467–514. North-Holland.
Galor, Oded. 2005. "From Stagnation to Growth: Unified Growth Theory." In *Handbook of Economic Growth*. Vol. 1, , ed. Philippe Aghion and Steven N. Durlauf. Amsterdam:Elsevier.

Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica*, 78(1): 35–71.
Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225: 254–277.
Hao, Yen-p'ing. 1970*a*. *The Comprador in Nineteenth Century China: Bridge Between East and West. Harvard East Asian series*, Harvard University Press.
Hao, Yen-p'ing. 1970*b*. "A "New Class" in China's Treaty Ports: The Rise of the Comprador-Merchants." *The Business History Review*, 44(4): 446–459.
Jia, Ruixue. 2014. "The Legacies of Forced Freedom: China's Treaty Ports." *The Review of Economics and Statistics*, 96(4).
Keller, Wolfgang, and Carol H. Shiue. 2021. "The Economic Consequences of the Opium War." National Bureau of Economic Research Working Paper 29404.
Kelly, Morgan, Joel Mokyr, and Cormac Ó Gráda. 2023. "The Mechanics of the Industrial Revolution." *Journal of Political Economy*, 131(1): 59–94.
Koss, Daniel. 2017. "Political Geography of Empire: Chinese Varieties of Local Government." *The Journal of Asian Studies*, 76(1): 159–184.
Kun, Gang, et al. 1899. *Qinding da Qing huidian shili (Imperially Established Institutes and Laws of the Great Qing Dynasty)*. Beijing: Zhonghua shu ju.
Kung, James Kai-sing, and Chicheng Ma. 2014. "Can cultural norms reduce conflicts? Confucianism and peasant rebellions in Qing China." *Journal of Development Economics*, 111: 132–149.
Kung, James, Kevin Zhengcheng Liu, and Xiaoming Zhang. 2024. "Merit-based Recruitment and Government Performance: Evidence from Random Assignment of Administrators." Working paper.
Kuo, Ting-yee, and Kwang-Ching Liu. 1978. "Self-strengthening: the pursuit of Western technology." *The Cambridge History of China*, 10(part 1): 491–542.
Lin, Justin Yifu. 1995. "The Needham Puzzle: Why the Industrial Revolution Did Not Originate in China." *Economic Development and Cultural Change*, 43: 269–292.
Lowery, Ying, and William J. Baumol. 2013. "Rapid Invention, Slow Industrialization, and the Absent Innovative Entrepreneur in Medieval China." *Proceedings of the American Philosophical Society*, 157(1): 1–21.
Ma, Chicheng. 2021. "Knowledge Diffusion and Intellectual Change: When Chinese Literati Met European Jesuits." *The Journal of Economic History*, 81(4): 1052–1097.
Ma, Chicheng. 2024. "Classicism and Modern Growth: The Shadow of the Sages." *The Journal of Economic History*, 84(2): 395–431.
McDonnell, E. M. 2017. "Patchwork Leviathan: How Pockets of Bureaucratic Governance Flourish within Institutionally Diverse Developing States." *American Sociological Review*, 82(3): 476–510.
Mitch, D. 1993. "The Role of Human Capital in the First Industrial Revolution." In *The British Industrial Revolution: An Economic Perspective*. , ed. J. Mokyr. Boulder, CO:Westview.
Mokyr, Joel. 2002. *The Gifts of Athena: Historical Origins of the Knowledge Economy*. Princeton, NJ:Princeton University Press.
Mokyr, Joel. 2005. "The Intellectual Origins of Modern Economic Growth." *Journal of Economic History*, 65(2): 285–351.
Mokyr, Joel. 2012. "Entrepreneurship and the Industrial Revolution in Britain." *The Invention of Enterprise: Entrepreneurship from Ancient Mesopotamia to Modern Times*, , ed. David S. Landes, Joel Mokyr and William J. Baumol, 183–210. Princeton: Princeton University Press.

Mokyr, Joel. 2016. *A Culture of Growth: The Origins of the Modern Economy.* Princeton University Press.
Moreira, Diana, and Santiago Pérez. 2021. "Civil Service Exams and Organizational Performance: Evidence from the Pendleton Act." National Bureau of Economic Research Working Paper 28665.
Murphey, Rhoads. 1974. "The Treaty Ports and China's Modernization." In *The Chinese City Between Two Worlds*. , ed. Mark Elvin and G. William Skinner, 17–72. Redwood City:Stanford University Press.
Northcote, SH, SCE Trevelyan, and B Jowett. 1854. *Report on the Organisation of the Permanent Civil Service.* London:Eyre & Spottiswoode.
Rawski, Thomas G. 1989. *Economic Growth in Prewar China.* Berkeley, CA:University of California Press.
Sandberg, Lars G. 1979. "The Case of the Impoverished Sophisticate: Human Capital and Swedish Economic Growth before World War I." *Journal of Economic History*, 39: 225–241.
Shiue, Carol H., and Wolfgang Keller. 2007. "Markets in China and Europe on the Eve of the Industrial Revolution." *American Economic Review*, 97(4): 1189–1216.
Sng, Tuan-Hwee. 2014. "Size and Dynastic Decline: The Principal-Agent Problem in Late Imperial China, 1700–1850." *Explorations in Economic History*, 54: 107–127.
Sng, Tuan-Hwee, and Chiaki Moriguchi. 2014. "Asia's Little Divergence: State Capacity in China and Japan before 1850." *Journal of Economic Growth*, 19(4): 439–470.
So, Billy K.L. 2011. "Modern China's Treaty Port Economy in Institutional Perspective: An Introductory Essay." In *Treaty-port Economy in Modern China: Empirical Studies of Institutional Change and Economic Performance*. , ed. Billy K.L. So and Ramon H. Myers, 1–27. Berkeley, CA:Institute of East Asian Studies.
Spence, Jonathan D. 1999. *The Search for Modern China.* . 2nd ed., New York:W. W. Norton.
Squicciarini, Mara P., and Nico Voigtländer. 2015. "Human Capital and Industrialization: Evidence from the Age of Enlightenment." *The Quarterly Journal of Economics*, 130(4).
Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2).
van de Ven, Hans. 2014. *Breaking with the Past: The Maritime Customs Service and the Global Origins of Modernity in China.* Columbia University Press.
Weber, Max. 1978. *Economy and Society: An Outline of Interpretive Sociology.* Berkeley:University of California Press.
Woodside, Alexander. 2006. *Lost Modernities: China, Vietnam, Korea, and the Hazards of World History.* Harvard University Press.
Xue, Melanie Meng. 2021. "Autocratic Rule and Social Capital: Evidence from Imperial China." SSRN Working Paper No. 2856803.
Xue, Melanie Meng, and Boxiao Zhang. 2023. "The Short- and Long-Run Effects of Affirmative Action: Evidence from Imperial China." Working Paper.
Yan, Zhongping. 2012. *Zhongguo Jindai Jingjishi Tongji Ziliao Xuanji(Early Modern China Economic History Statistics Book). Institute of Economics Chinese Academy of Social Sciences, bulletin on early modern Chinese economic history*, China Social Sciences Press.
Yao, Yuan, Rui Wang, and Shufeng Yao. 2008. *Zhongguo Jindai Keji Qikan Yuanliu 1792-1949 (he Origin and Development of Modern Chinese Science and Technology Journals).* Shandong Education Publishing House.
Zhang, Yufa. 1987. *Qingmo Minchu de Waizi Gongye (Foreign Ownership Firms in Early Modern China). Bulletin of Institute of Modern History, Academia Sinica*, Institute of Modern

History, Academia Sinica.
Zhang, Yufa. 1988*a*. *Qingmo Minchu de Guanban Gongye(1860-1896)(Public Ownership Firms in Early Modern China).* 清季自强运动研讨会论文集, Institute of Modern History, Academia Sinica.
Zhang, Yufa. 1988*b*. *Qingmo Minchu de Guandu Shangban Gongye(State-Private Ownership Firms in Early Modern China). Bulletin of Institute of Modern History, Academia Sinica*, Institute of Modern History, Academia Sinica.
Zhang, Yufa. 1989. *Qingmo Minchu de Minying Gongye(Private Ownership Firms in Early Modern China). Bulletin of Institute of Modern History, Academia Sinica*, Institute of Modern History, Academia Sinica.
Zhu, Baojiong, and Peilin Xie. 1979. *Ming Qing Jinshi Timing Beilu Suoyin (Ming and Qing dynasty Jinshi Name Roster).* Shanghai: Shanghai Guji Chuban She (Shanghai Classics Publishing House).

Online Appendix

AWAKENING LATENT HUMAN CAPITAL:
THE OPENING-UP AND ENTREPRENEURSHIP IN 19TH-CENTURY CHINA

Li Duan Xiaoming Zhang

A Additional Results for Baseline

A.1 The Persistence in the Spatial Distribution of Books

In this section, we present spatial correlations among books written in different time periods. As discussed in Section 2.1 of the main text, we have chosen books written between 1644 and 1840 as our primary explanatory variable. Here, we offer empirical support for this choice.

We perform spatial correlation analyses at the prefecture level among three subsamples of books: those written before 1644, between 1644 and 1840, and between 1840 and 1904. In addition, we separately consider the correlations for books written by *Jinshi*-degree holders and those written by authors without a *Jinshi* degree. Our objective is to provide further justification for our choice of book measure in Section 4.

Table A1: Spatial Correlations: Books 1644-1840, Books 1840-1904 and Books Pre 1644

Pair	Correlation	Significance
lnBookDens1644-1840 vs. lnBookDens1840-1904	0.761	***
lnBookDens1644-1840_*jinshi* vs. lnBookDens1840-1904_*jinshi*	0.651	***
lnBookDens1644-1840_non*jinshi* vs. lnBookDens1840-1904_non*jinshi*	0.667	***
lnBookDens1644-1840 vs. lnBookDensPre1644	0.519	***
lnBookDens1644-1840_*jinshi* vs. lnBookDensPre1644_*jinshi*	0.385	***
lnBookDens1644-1840_non*jinshi* vs. lnBookDensPre1644_non*jinshi*	0.612	***

Note. This table presents the spatial correlations among three groups of books. Rows 1 to 3 compare books written during 1644 and 1840 and books written during 1840 and 1904. The first row shows the correlation among all books. The second row shows the correlation between the log density of books written by *jinshi*, and the third row shows that of books written by non-*jinshi*. Rows 4 to 6 compares books written during 1644 and 1840 and books written before 1644.

A.2 Concerns with the Log Transformation

In this section, we confront the intensifying scrutiny surrounding "log-like" transformations. Transformations such as the logarithmic adjustment, $\log(1+Y)$, and the inverse hyperbolic sine, $\text{arcsinh}(Y)$, have been traditionally employed to normalize outcomes Y that are predominantly non-negative, with a notable frequency at zero. These techniques facilitate the interpretation of treatment effects in relative "percentage" terms, a necessity in many research contexts.

Recent literature has begun to highlight the limitations of these transformations, advising caution with their application, particularly in the interpretation of treatment effects (e.g., Aihounton and Henningsen, 2021; Chen and Roth, 2023; Mullahy and Norton, 2023). Through theoretical exposition and simulation, it has been illustrated that the estimated marginal effects derived from such transformations are sensitive to the scaling on the units of Y. Specifically, Chen and Roth (2023) articulate that scale dependence is an inescapable characteristic of any Average Treatment Effect (ATE) that remains well-defined when Y includes zero values. No transformation or alternative estimation strategy offers a panacea. Furthermore, they elucidate a fundamental trilemma: any parameter that is well-defined with zero-valued Y cannot simultaneously satisfy three conditions—being an average of individual-level treatment effects, being invariant to the rescaling of Y, and being point identified. As such, they posit that the selection of a target parameter and the prioritization among these conditions should be informed by the specific economic question and the underlying motivation of the research.

In this study, we adopt the transformation $\log(1+Y)$, prioritizing the attainment of a point-identified Average Treatment Effect (ATE) over scale invariance, for several compelling reasons. Primarily, our analytical goal is to ascertain the influence of traditional human capital on the emergence of modern firms, with a particular focus on the direction and statistical significance of the effect. Therefore, identification is more important than functional form fit. We acknowledge, however, that the log transformation imparts a degree of uncertainty to the magnitude of our coefficients, rendering them less straightforward to interpret. Consequently, as proposed by Chen and Roth (2023), we implement alternative methodologies to furnish a more precise estimate of book density's impact, which we present in Table A2.

To begin with, we eschew the intensive margins of treatment effects in favor of the extensive margins. We transform our outcome variable into a binary indicator reflecting above-zero firm entry. The results, displayed in Columns (1) and (2) of a linear probability model, reveal that a doubling of book density correlates with a 65% increase in the likeli-

hood of modern firm entry. Next, we employ an alternative normalization technique, as recommended by Chen and Roth (2023), wherein we normalize the number of modern firms by the population. Columns (3) and (4) detail the outcomes of this approach, indicating an association where a 1% increase in book density yields a 0.7% rise in the number of firms per capita. Lastly, we utilize Poisson regression to estimate a scale-invariant ATE expressed as a percentage change. Given the excess of zeros in our firm count data, we opt for the Poisson Pseudo-Maximum Likelihood (PPML) estimator to accommodate overdispersion. The choice of PPML, as opposed to the Zero-Inflated Negative Binomial (ZINB) estimator, is due to its efficiency in handling models with high-dimensional fixed effects, which would otherwise lead to computational issues with ZINB. The findings from Column (5) suggest that a 1% augmentation in book density is associated with a 3.5% increment in the count of modern firms.

Table A2: Alternatives to the Log Transformation

	(1)	(2)	(3)	(4)	(5)
	$I_{Firm>0}$	$I_{Firm>0}$	firm density	firm density	firms
lnBookDens × PostPort	0.690***	0.645***	0.684***	0.660***	3.460*
	(0.159)	(0.242)	(0.185)	(0.233)	(1.868)
PostPort	0.048*	-0.720	0.047*	-1.715***	3.403
	(0.026)	(0.466)	(0.027)	(0.542)	(6.411)
Prefecture FE	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓
Baseline Controls		✓		✓	✓
Prefecture × Year		✓		✓	
Province × Year FE		✓		✓	
More Controls		✓		✓	
Observations	17355	17355	17355	17355	3493

Note. This table reports the results alternative to our baseline $log(Y+1)$ transformation. In Columns (1) and (2), we calibrate the value of Y and separately estimate the effects for the extensive margins by using a dummy Y variable indicating the entry of a new firm. In Columns (3) and (4), we adopt an alternative normalization method, expressing the outcome in per million people. In Column (5), we adopt Poisson regression, except that given the mass of zero in our outcome variable, we use the Poisson Pseudo-Maximum Likelihood estimator to allow for overdispersion and zero-inflated data. We omit the additional set of control variables and time trends because they would lead to computational issues. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.3 Natural Logarithm of Book Count

In our main text, we choose to first normalize books by population and then take the natural logarithm for two main reasons. Firstly, for the same consideration as Squicciarini and Voigtländer (2015), more populous prefectures will naturally have more book

writers. On the supply side, similar to the process of premodern technological inventions theorized by Lin (1995), we posit that the premodern creation of new books also follows a Poisson process, where the mean parameter is closely correlated with population. On the demand side, although this paper does not delve into the demand and circulation of books due to data limitations, population is an obvious first-order determinant. Secondly, it is well-documented that the Chinese population experienced a significant surge during the Ming and Qing dynasties (Cao, 2000). Our time series patterns for book data also reflect a surge in book counts during these periods. Therefore, in the Chinese context, it is crucial to directly normalize for population in the book measure, in addition to controlling for population in the regression equation.

In this section, we demonstrate that our baseline findings remain robust even when using the natural logarithm of books without normalizing them by population. In Table A3, we present results from the same four specifications as in Table 1, confirming that the core relationships observed in our analysis persist regardless of whether we adjust for population.

Table A3: Firm and Alternative BookCount

	(1) lnfirms	(2) lnfirms	(3) lnfirms	(4) lnfirms
lnBooks × PostPort	0.064***	0.068***	0.056***	0.044*
	(0.016)	(0.018)	(0.018)	(0.025)
PostPort	-0.133**	-1.256***	-1.447**	-1.419**
	(0.054)	(0.425)	(0.635)	(0.594)
Prefecture FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Baseline Controls		✓	✓	✓
Prefecture × Year			✓	✓
Province × Year FE			✓	✓
More Controls				✓
Adj R-square	0.290	0.301	0.547	0.548
Observations	17355	17355	17355	17355

Note. This table reports the impact of log book density on log firms after the opening of treaty ports, compared with that before the treaty ports. Unlike Table 1, the log book measure here is not normalized by population. Column (1) controls for both the prefecture and year-fixed effects. Column (2) also controls for our baseline set of control variables. Column (3) adds the province-specific time-fixed effects and the prefecture-specific linear time trends. Column (4) includes a set of additional confounding factors that control for alternative explanations. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.4 Alternative Density Measure

For prefectures without any book, our baseline measure of log book density does not take full advantage of available variation: population difference. Intuitively, zero books for a prefecture with 1,000,000 inhabitants should reflect a much lower human capital density than zero books for a prefecture with 50,000 inhabitants. To exploit this additional variation in the population, we borrow the alternative density calculation from Squicciarini and Voigtländer (2015):

$$lnBookDens2 = ln(\frac{Books + 1}{Pop_{pre1840}}) \tag{5}$$

where *Books* is the total number of books authored by individuals born in that prefecture during the Qing dynasty up until the Opium War (between 1644 and 1840). Table A4 presents the results using this alternative density calculation, which is very similar to our baseline pattern.

Table A4: Baseline Firm and BookCount

	(1) lnfirms	(2) lnfirms	(3) lnfirms	(4) lnfirms
lnBookDens2 × PostPort	0.077***	0.073***	0.087***	0.076***
	(0.022)	(0.020)	(0.022)	(0.023)
PostPort	0.361***	-1.095**	-1.282**	-1.254**
	(0.087)	(0.425)	(0.621)	(0.595)
Prefecture FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Baseline Controls		✓	✓	✓
Prefecture × Year			✓	✓
Province × Year FE			✓	✓
More Controls				✓
Adj R-square	0.288	0.300	0.548	0.549
Observations	17355	17355	17355	17355

Note. This table reports the impact of log book density on log firms after the opening of treaty ports, compared with that before the treaty ports. Unlike Table 1, log book density here is calculated as shown in Equation 5. Column (1) controls for both the prefecture and year-fixed effects. Column (2) also controls for our baseline set of control variables. Column (3) adds the province-specific time-fixed effects and the prefecture-specific linear time trends. Column (4) includes a set of additional confounding factors that control for alternative explanations. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.5 Alternative measures of books and firms

Since our analysis above hinges on the novel books per capita explanatory variable, we begin by testing the robustness of our results using alternative measures of books. Furthermore, we demonstrate the robustness of our findings using alternative outcome variables to address any concerns over the data quality of modern firms.

Using Authors per Capita When constructing human capital proxies out of our book catalog data, we have considered aggregating the catalogs not only at the individual book level but also at the author level. However, we ultimately choose books per capita as our baseline measure because author-level aggregation omits the productivity disparity across different authors. Aggregating at the author level weighs authors with high or low book writing productivity equally, and the best weight we can apply to authors is, in fact, just aggregation at the book level. From a conceptual standpoint, we do not discern any clear advantages or disadvantages between the two measures when used as a proxy for human capital stock. In Chaney (2016), the unit of analysis is at the book level because the author aims at measuring knowledge production in science, whereas in Chaney (2020), the unit of analysis becomes the authors because the goal is an alternative proxy for the urban population. Nevertheless, the result for using authors per capita is presented in Column (1) of Table A5.

Using Siku Book Catalog Although our book catalog aims to be comprehensive by combining various historical primary sources and modern library holdings, we test the robustness of our results using a more familiar source of book catalogs to dispel concerns over data quality. Given our Qing dynasty setting, we focus on the *Siku Quanshu*, a monumental book catalog commissioned by the Qianlong Emperor. Between 1773 and 1784, over 3,600 scholars, supervised by officials handpicked by the emperor himself, collected over 3,000 works and 36,000 volumes, making the *Siku Quanshu* an essential reference work for scholars and officials during the Qing dynasty. However, being an imperial project, the *Siku Quanshu* is potentially plagued with selection biases, and when compared to our catalog dataset, it is more of a highly selective mini-catalog. Nonetheless, we construct an alternative measure of books per capita based on the *Siku Quanshu*, along with its later extension and supplementary works (please see Table C12 for more details). The positive and significant finding on the impact of books from *Siku Quanshu*, presented in Column (2) of Table A5, helps to boost confidence in our book catalog data quality.

Table A5: Robustness Check: Alternative Measures

	(1) lnfirms	(2) lnfirms	(3) lnfirms_big	(4) urbanrate	(5) urbanrate	(6) lnurban
lnAuthorDens × PostPort	2.638***					
	(0.778)					
lnBookDens_Siku × PostPort		4.245***				
		(1.289)				
lnBookDens × PostPort			0.762***	18.375**	17.929*	1.382*
			(0.158)	(8.489)	(10.733)	(0.806)
PostPort	-1.102*	-1.020*	0.283	0.610	20.065	1.825
	(0.602)	(0.564)	(0.410)	(1.299)	(27.642)	(1.988)
Prefecture FE	✓	✓	✓	✓	✓	✓
Year FE	✓	✓	✓	✓	✓	✓
Baseline Controls	✓	✓	✓		✓	✓
More Controls	✓	✓	✓		✓	✓
Prefecture Year	✓	✓	✓		✓	✓
Province Year FE	✓	✓	✓		✓	✓
Adj R-square	0.549	0.550	0.548	0.498	0.541	0.647
Observations	17355	17355	17355	534	534	534

Notes. This table examines whether our findings are robust to alternative sources of the explanatory bibliometric variable and alternative outcome variables. Column (1) replaces our book density with log density of the authors. Column (2) uses the alternative book catalog, *Siku Quanshu*. Column (3) only keeps firms with registered capital above a threshold of 10,000 silver coins. Columns (4) to (6) use the urbanization rate as the outcome variable, where Columns (4) and (5) use the percentage point and Column (6) uses the natural logarithm of that percentage point. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Using Firms with High Registered Capital We further address the following concerns with our firm measure: (i) even though Zhang and Du's firm data collection primarily focuses on modern firms, some of the firms may not be exactly modern; (ii) if human capital stock predominantly gave rise to low-tech entrants, excessive competition and a saturated market could displace high-quality firms, thus negatively affecting the industrialization of late Qing China. To address these potential issues, we narrow our focus to firms with registered capital exceeding 10,000 taels of silver. This strategy enables us to concentrate on the most significant and well-capitalized firms, which are less likely to be affected by the issues raised. Column (3) of Table A5 shows that our baseline bibliometric variable had a positive and significant effect on these well-resourced firms.

Using Urbanization Rate Since we are linking upper-tail human capital to the overall industrialization process, we now demonstrate its relevance to an alternative outcome variable with broader implications for industrialization: the rate of urbanization. Before 1840, China was predominantly an agricultural state, but the forceful opening of its mar-

ket and the onset of industrialization initiated the urbanization process. Therefore, the urbanization rate serves as a clear indicator of industrialization. However, due to data constraints, we only have prefecture-level urbanization rates for 1776 and 1910 from Cao (2000). To address this, we collapse our panel dataset into a two-period panel, treating 1776 as the pre-treatment period and 1910 as the post-treatment period, and apply the same difference-in-differences specification as Equation 1. Columns (4) and (5) of Table A5 present results using the percentage points of urbanization rate as the outcome variable, while Column (6) uses the natural logarithm of that percentage point. All three specifications indicate that upper-tail human capital had a positive and significant effect on the urbanization process.

Collectively, these results confirm that our baseline findings are robust to alternative measures. Furthermore, the positive and significant impact of books per capita on high-caliber firms and urbanization provides additional evidence of the substantial contribution upper-tail human capital stock made to industrialization in late Qing China.

A.6 Staggered Difference-in-Differences Designs

A review of this literature reveals a host of novel estimators engineered to be robust against the heterogeneous timing of treatment (e.g., De Chaisemartin and D'Haultfœuille, 2020, 2022; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021; Wooldridge, 2021). However, our context diverges from the standard scenario of staggered binary treatments typically examined in this body of work. We are dealing with two staggered treatments: (i) the binary variable of staggered treaty port openings, and (ii) the interaction between the staggered treaty port openings and the time-invariant and continuous log book density measure. This complexity is not catered to by standard estimators. To navigate these intricacies, we employ a dual-pronged approach to address the issue of heterogeneous treatment timing. In this section, we first conduct an analysis on an event-by-event basis and then employ a novel estimator that is robust to the heterogeneous timing of treatment.

Biases in TWFE Staggered DiD Estimation First of all, we must determine whether these robustness checks are necessary and if our baseline estimates are likely impacted by biases due to heterogeneous treatment timing. There are two primary types of heterogeneity when examining a treatment: heterogeneity in the treatment effect and heterogeneity in treatment timing. The former involves variations in the magnitude of the treatment effect, which can fluctuate across groups, over time, or both. The latter occurs when the treatment is introduced to different groups in a staggered timeline. Due to the

scope of this paper, we will only briefly discuss how the TWFE DiD estimate can yield biased results in the presence of staggered treatment timing. Baker, Larcker and Wang (2022) employs COMPUSTAT sample simulations to effectively illustrate the creation of such biases. According to their findings, TWFE DiD estimates are unbiased in situations with staggered treatment timing only if the treatment effect is homogenous across both groups and time. In many natural experiments, including ours, there is no compelling reason to assume that the effect of treaty port openings on modern firm entry is homogeneous across prefectures and years. Given the geographical variation of these treaty ports and the rapidly evolving socio-economic circumstances, we should be particularly concerned about treatment effect heterogeneity in our context. Furthermore, according to Sun and Abraham (2021), unlike the static TWFE staggered DiD, event study estimates can remain contaminated by heterogeneity in treatment timing even with treatment effect homogeneity. In summary, our context combines staggered treatment timing with treatment effect heterogeneity, either across groups or over time. Consequently, biases are likely to plague our TWFE staggered DiD estimates and event study estimates.

Our research is particularly concerned with the potential for biases in TWFE staggered DiD estimates to reverse the sign of the estimates. To illustrate this, we employ the decomposition method utilized in Goodman-Bacon (2021). A TWFE staggered DiD with G different treatment-timing cohorts can invariably be decomposed into G^2 standard DiD constituents, each with two groups and two periods. In conducting a TWFE staggered DiD, we are contrasting not only various treatment cohorts with the control group (G of the 2×2 constituents) but also earlier treated cohorts with later treated cohorts ($G^2 - G$ of the 2×2 constituents). Biases originate from the problematic $G^2 - G/2$ of the 2×2 constituents, where the early-treated cohorts serve as controls for the later-treated cohorts. In these problematic comparisons, the difference in the outcome for the later treated cohort (the effective treatment) is subtracted by the difference in the outcome for the earlier treated cohort (the effective controls). While such difference-in-differences remains a valid estimate of the treatment effect when the treatment effect is homogeneous, the presence of treatment effect heterogeneity can yield negative estimates even when all ATTs are genuinely positive. For instance, if the change in the outcome for the earlier treated cohort (the effective control) surpasses the change in the outcome for the later treated cohort (the effective treatment), the problematic comparison would yield a negative estimated treatment effect. In our case, we have reasons to anticipate that the change in modern firm entry for early treaty ports is greater than the change for later treaty ports: (i) earlier treaty ports were more strategically located for international trade; (ii) later treaty ports faced heightened competition from earlier treaty ports and other access points to Western

markets.

Improper weighting can lead to the reversal of the overall TWFE staggered DiD estimate's sign due to the opposing signs from problematic 2×2 constituents. The TWFE staggered DiD assigns weights to each 2×2 constituent based on the cohort sample size and treatment variance. According to Goodman-Bacon (2021), assuming all else remains equal, constituents where treatment groups receive treatment closer to the center of the time window are allocated more significant weights due to their larger treatment variance. As stated by Baker, Larcker and Wang (2022), even a slight degree of treatment effect heterogeneity can result in a high propensity for Type I and Type II errors in TWFE staggered DiD.

Therefore, eliminating problematic comparisons and discovering an appropriate weighting method for the remaining constituents are pivotal to rectifying biases attributable to heterogeneous treatment timing. We address this issue in this paper with a two-fold strategy: (i) we perform an event-by-event analysis that scrutinizes the treatment effect of each treatment cohort against the correct control group, and (ii) we employ the robust weighting method proposed by Sun and Abraham (2021) to yield robust overall estimates that apply to both our static DiD estimate and event study estimates.

Event-by-Event Analysis The challenging part is that our setting differs from the standard scenario of staggered binary treatments typically examined in the recent literature on staggered difference-in-differences designs. Papers in this literature address staggered introduction of treatment where the treatment is binary and absorbing (e.g., Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Baker, Larcker and Wang, 2022). In Equation 1, $PostPort_{i,t}$ represents treaty port openings, the standard binary staggered absorbing treatment. However, our main explanatory variable of interest, $lnBookDens_i \times PostPort_{i,t}$, is the interaction between staggered treaty port openings and time-invariant upper-tail human capital. The econometric theories and newly proposed estimators in the current literature do not specifically address this type of staggered treatment. Therefore, to maintain rigor in this robustness check exercise, instead of applying any readily available estimators directly, we follow Goodman-Bacon (2021) and decompose our TWFE staggered DiD into many constituent standard DiD, each with uniform treatment timing. To achieve this, we first create 18 event-specific panel datasets corresponding to our 18 staggered events. We then generate event-specific estimates using separate regressions for each event.

Each event e-specific panel dataset includes the prefectures in treatment cohort e ($e \in E = \{2, 3, 4, 5..., 18, 19]\}$, where E corresponds to the 18 cohorts in Table C14) and other

clean control prefectures for a window of relative event time extending from 30 years prior to, and up to 40 years following, event e. Clean controls encompass never-treated prefectures as well as later-treated prefectures whose treatment times fall outside the time window of the event-specific (e-specific) panel dataset. This inclusion eliminates contamination from the problematic comparisons. For each event-specific panel dataset, we execute a regression similar to Equation 1:

$$lnFirm_{i,t} = \beta \cdot lnBookDens_i \times PostPort_t + \gamma_t + \lambda_i + \epsilon_{i,t} \tag{6}$$

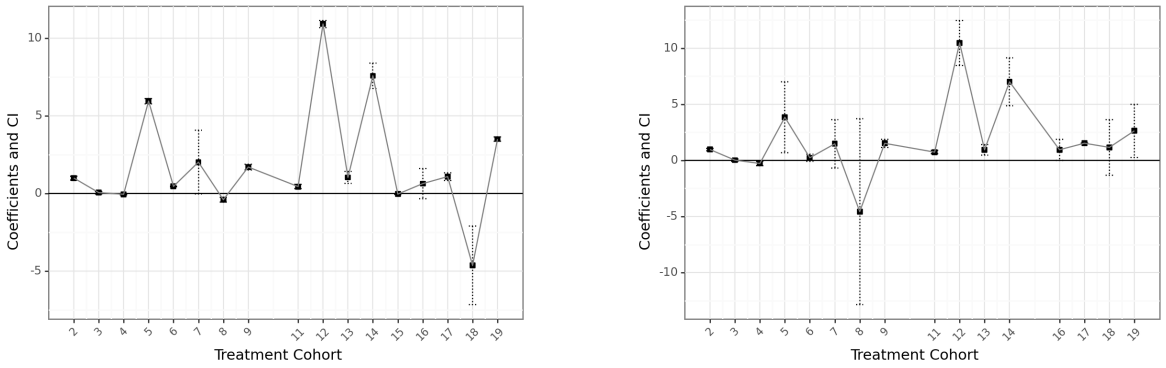
except where the treaty port opening $PostPort_t$ is now uniform across all prefectures in treatment cohort e. Due to the small sample size and minimal variation in our control variables, which led to a significant number being dropped due to multicollinearity, we are omitting them from this event-by-event exercise. All standard errors are clustered at the prefecture level.[1] Figure A1a shows the result for each event-specific analysis, along with 95% confidence intervals. There is clear evidence of positive but heterogeneous effects of book density on modern firm entry across events: 13 (76.47%) of the events are positive, of which 10 are significant at the 5 percent level.

Besides the omission of control variables, the small sample sizes also raise another concern: inference in difference-in-differences with few treated groups (e.g., Bertrand, Duflo and Mullainathan, 2004; Ferman and Pinto, 2019). Within each event-specific dataset, although we have removed earlier-treated cohorts as control groups, the number of prefectures within each treatment cohort remains substantially smaller than in the control group (as detailed in Table C14). Given that our DiD specifications cluster standard errors at the prefecture level—thereby avoiding assumptions of homoskedasticity within the group $\times$ time aggregate model—our standard error estimation may be compromised due to the small number of treated groups. However, our situation, characterized by heteroskedasticity not attributable to variation in group sizes, precludes the use of the cluster residual bootstrap method suggested by Ferman and Pinto (2019). Instead, we approach this issue of too few treated groups by further shrinking down our control groups. We retain the same set of later-treated cohorts as controls, considering that these prefectures, which have not yet undergone treatment, theoretically provide high-quality control under the assumptions of no anticipation effects and adherence to parallel trends. For the numerous never-treated prefectures, we confine our focus to the three nearest neighbors of each treated prefecture, as determined by the propensity score matching process outlined in Table A7. This method of selecting control prefectures also provides some justification for

[1] Cohort 10 are omitted because it only has 1 treatment prefecture that has 0 firm and 0 book density.

our omission of control variables since the matching incorporates all relevant covariates. The results, depicted in Figure A1b,[2] demonstrate a positive and significant effect of book density on modern firm entry, reinforcing the robustness of our findings: 13 (81.25%) of the event estimates are positive, with 9 attaining significance at the 5 percent level.

In sum, these event-specific analyses lend further credence to our baseline results, confirming the overarching positive impact of treaty port openings for the majority of the 18 cohorts studied and indicating that the aggregate positive estimate is not disproportionately influenced by a few extreme individual events.



(a) Without Matching (b) With Matching

Figure A1: Event-specific Estimates

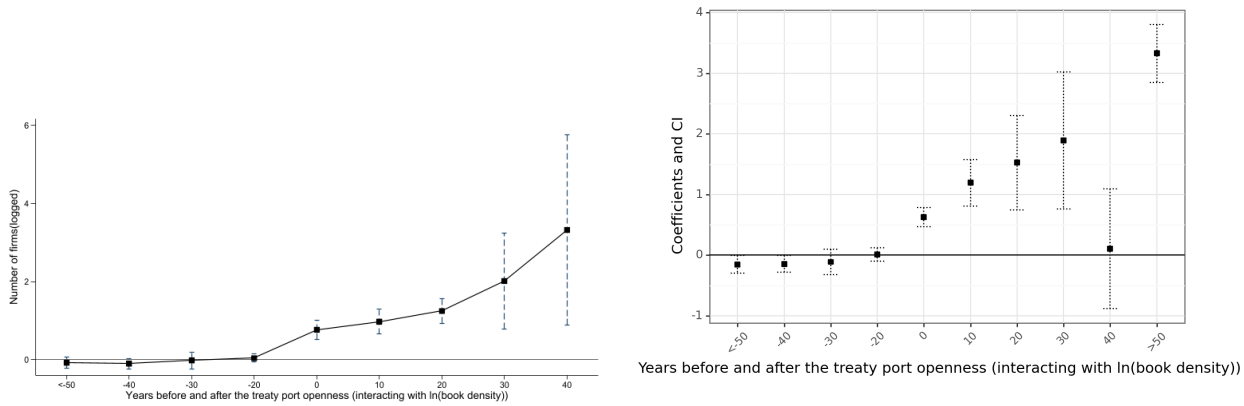
The figure shows the event-specific DiD estimates across all treaty port openings. The control groups included are different in these two panels. For each event e-specific analysis, Panel (a) incorporates all never-treated prefectures along with later-treated prefectures whose treatment times fall outside the time window of the e-specific panel dataset. Panel (b) includes the same set of later-treated prefectures but restricts the never-treated prefectures to those successfully matched to prefectures in cohort e. The dashed lines indicate the 95% confidence intervals.

Alternative Aggregate Estimations In addition to the detailed constituent DiD estimates at the micro level, we aim to provide an alternative to our baseline aggregate estimate that is more robust to the issue of staggered treatment timing in the presence of heterogeneous treatment effects. As outlined above, the key is to apply a proper weighting method to all the cohort-specific treatment effects. To this end, we first use the stacked regression estimation where we stack all the event-specific datasets above and run a standard TWFE DiD regression (e.g., Cengiz et al., 2019; Deshpande and Li, 2019). We also adopt the flexible interaction-weighted estimation method proposed by Sun and Abra-

[2]Cohort 15 is omitted due to lack of common support.

ham (2021).[3] In particular, we include the same set of control variables as in Equations 1 and 2. The event study estimates displayed in Figure A2 and the static DiD results in Table A6 further substantiate our baseline findings from Table 1 and Figure 5, confirming a statistically significant positive effect.

Overall, these results are reassuring, as they show that our baseline findings are robust to the issue of heterogeneity in both treatment timing and effects. The congruence between estimates from the micro-level event-by-event analysis and the properly weighted aggregate effect lends further credibility to the significant and persistent influence of book density on the entry of modern firms.



(a) Stacked Regression (b) Interaction-weighted Estimators

Figure A2: The Dynamic Effects of Books Density on Modern Firms Creation

These figures uses 10 years before treatment as the reference and apply alternative estimators that are robust to the staggered timing of treatment. Panel (a) uses the stacked regression estimator, and Panel (b) applies the interaction-weighted estimator proposed by Sun and Abraham (2021). The solid squares mark the coefficients, and the dashed lines indicate the 95% confidence intervals.

[3]The estimator by Sun and Abraham (2021) is preferred over other robust estimators, such as those by Goodman-Bacon (2021) and Callaway and Sant'Anna (2021), primarily due to the non-binary nature of our interaction term $lnBookDens_i \times PostPort_{i,t}$. The associated Stata packages for these papers only require inputs that define treatment cohorts or timing and do not accommodate non-binary treatment variables. Conversely, the *eventstudyinteract* package by Sun and Abraham (2021) allows for custom coding of staggered treatment variables and the inclusion of covariates.

Table A6: Firm and BookCount: Stacked Regression and Alternative Estimator

	Stacked TWFE	Interaction-Weighted Estimator		
	(1) lnfirms	(2) lnfirms	(3) lnfirms	(4) lnfirms
lnBookDens × PostPort	1.018***	1.171***	1.041***	0.834***
	(0.263)	(0.164)	(0.071)	(0.084)
PostPort	-0.499	0.046**	-0.017	-0.911***
	(1.490)	(0.022)	(0.201)	(0.170)
Prefecture FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Baseline Controls	✓		✓	✓
More Controls	✓			✓
Prefecture × Year	✓			
Province × Year FE	✓			
Adj R-square	0.331	0.307	0.310	0.311
Observations	16344	17355	17355	17355

Note: This table provides alternatives to our baseline aggregate estimate that are robust to the staggered timing of treatment. In Column (1), we stack all the event-specific datasets and run regular TWFE DiD with the same specification as our baseline results. Columns (2) to (4) employ the interaction-weighted estimator proposed by Sun and Abraham (2021). We cannot include the prefecture-specific time trend or the province-specific time fixed effects, because the *eventstudyinteract* package does not allow such input. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.7 Subsample Estimations

As discussed in Section 2.2, the selection of treaty port locations took into consideration factors such as water transportation accessibility, proximity of colonial powers, and local economic prosperity. Therefore, it is critical to address biases in treatment selection using subsample estimations.

A.7.1 Propensity Score Matching

Matching Variables Selection Considering that various factors, ranging from geographic to economic conditions, influenced the opening of treaty ports, we consider all the time-invariant prefecture characteristics among our control variables in this propensity score matching exercise. In Panel A of Table A7, we regress the treatment status variable on these prefecture characteristics individually and narrow our matching criteria down to the sub-list of controls that are significantly correlated with the opening of treaty ports. Column (1) lists all the time-invariant prefecture characteristics included in the initial one-on-one regressions. Column (2) reports the coefficients and standard errors, and we select only the variables with a significance level of 10% or higher.

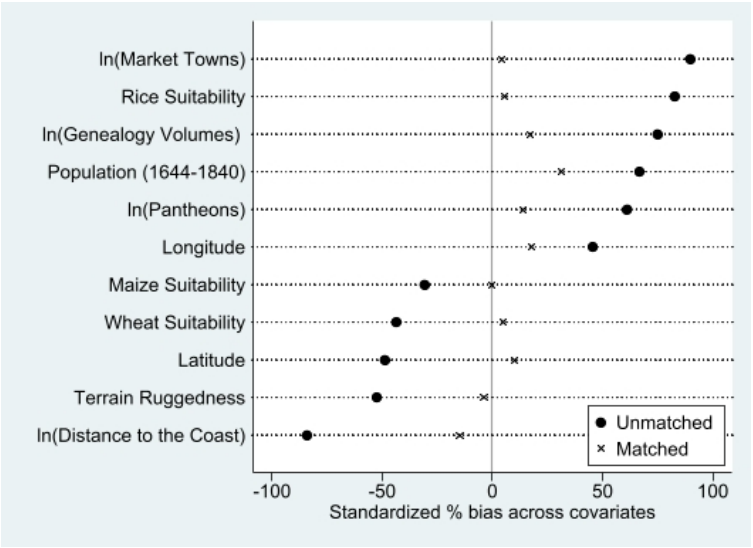
Matching and Subsample Analysis During the propensity score matching, we keep the three nearest neighbors of each treatment prefecture as control prefectures with replacement. Consequently, 33 of the 35 prefectures with treaty ports are on common support,

and 58 of the 232 prefectures without treaty ports are on common support and selected as the control sample, together forming the subsample that we use to conduct our analysis. The figure in Table A7 compares the standardized bias across the covariates between our unmatched full sample and the matched subsample. We can see that the matching subsampling has effectively reduced the differences between treatment prefectures and control prefectures along important prefecture characteristics. In Panel B of Table A7, we run two separate regressions within this matched subsample. In Column (4), we directly use this subsample and find the result stable and very close to our baseline outcome. In Column (5), we further drop all prefectures that were provincial capitals, and the result remains significant and grows in magnitude.

Overall, these findings provide further evidence of the strong positive relationship between human capital and industrialization in late Qing China, even when controlling for potential biases in treatment selection.

Table A7: Robustness Check: Propensity Score Matching (Dependent Var: $lnfirms$)

Panel A: Correlations with Treaty Ports		Panel B: Subsample Estimations		
(1)	(2)	(3)	(4)	(5)
Controls	Correlation		All	No Provincial Capitals
Population (1644-1840)	0.248***	lnBookDens × PostPort	0.796***	1.263***
	(0.060)		(0.195)	(0.318)
ln(Distance to the Capital)	0.078	PostPort	-1.962**	-1.292**
	(0.061)		(0.746)	(0.527)
ln(Distance to the Coast)	-0.327***	Prefecture FE	✓	✓
	(0.058)	Year FE	✓	✓
Terrain Ruggedness	-0.161***	Baseline Controls	✓	✓
	(0.061)	More Controls	✓	✓
Wheat Suitability	-0.140**	Prefecture × Year	✓	✓
	(0.061)	Province × Year FE	✓	✓
Rice Suitability	0.289***	Adj R-square	0.616	0.639
	(0.059)	Observations	5850	5200
Sweet Suitability	0.048	**Matching Variables**		
	(0.061)			
Maize Suitability	-0.101*			
	(0.061)			
ln(Distance to River)	0.014			
	(0.061)			
chong	0.068			
	(0.061)			
fan	0.018			
	(0.061)			
pi	0.061			
	(0.061)			
nan	-0.007			
	(0.061)			
ln(Lit Inq)	0.095			
	(0.061)			
ln(Market Towns)	0.269***			
	(0.059)			
ln(Pantheons)	0.196***			
	(0.060)			
ln(Genealogy Volumes)	0.256***			
	(0.059)			
Longitude	0.166***			
	(0.061)			
Latitude	-0.166***			
	(0.061)			



Notes. This table examines whether our findings are robust to a subsample of treatment and control prefectures that are balanced on an expansive suit of characteristics. Panel A conducts one-on-one correlations between the treatment status and all the control variables we have and filters out the less relevant characteristics. Panel B presents two regressions run by propensity score matching. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.7.2 Dropping all Prefectures in Jiangnan

Here we show that our main finding is not driven by a few more prosperous prefectures. During the late Qing, the Jiangnan region (Zhejiang, Jiangsu, and Anhui provinces) were among the most prosperous and also saw a high concentration of treaty ports. Therefore, we drop all prefectures in the Jiangnan region in this robustness check exercise. Table A8 presents the results using this subsample. The positive and significant effects of book density remain largely stable, with the exception of the addition of the prefecture-specific time linear trend.

Table A8: Firm and BookCount with Jiangnan region dropped

	(1) lnfirms	(2) lnfirms	(3) lnfirms	(4) lnfirms
lnBookDens × PostPort	0.912***	1.236**	1.790**	-0.597
	(0.342)	(0.485)	(0.785)	(0.545)
PostPort	0.039	-1.753***	-1.628***	-0.242
	(0.027)	(0.480)	(0.474)	(0.561)
Prefecture FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Baseline Controls		✓	✓	✓
More Controls			✓	✓
Province × Year FE			✓	✓
Prefecture × Year				✓
Adj R-square	0.231	0.244	0.254	0.474
Observations	15015	15015	15015	15015

Note. This table reports the impact of log book density on log firm after the opening of treaty ports, compared with that before the treaty ports. In comparison to Table 1, all regressions here are run with prefectures in the Jiangnan region dropped. Column (1) controls for both the prefecture and year-fixed effects. Column (2) also controls for our baseline set of control variables. Column (3) adds the province-specific time-fixed effects and the additional set of additional confounding factors that control for alternative explanations. Column (4) includes the prefecture-specific linear time trends. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.8 Instrumental Variable Estimations

Although our specifications so far have controlled for a wide range of covariates and the difference-in-differences strategy has helped to rule out biases from omitted variables whose effects remained constant before and after the opening of treaty ports, we remain concerned about potential omitted variables whose effects may have changed with treatment. To address this concern, we introduce an instrument for books per capita. Results from both stages, along with the reduced form estimates, are presented in Table A9.

We adopt the instrument variable from Chen, Kung and Ma (2020). In their study, the authors use the nearest distance to bamboo as one of the instruments for *Jinshi* density by arguing that bamboo, a key ingredient for printing paper, dictated the location of printing centers and the availability of books, which in turn affected the preparation for the Imperial Examination. According to Figure 3A and 3B in Chen, Kung and Ma (2020), the 19 printing centers during the Qing dynasty were located in close proximity to bamboo habitats. Therefore, the nearest distance to bamboo is a natural instrument for our bibliometric measure.

Based on the GIS Point data of bamboo habitats, we calculate for each of our 267 prefectures its distance to the nearest bamboo site. Column (1) of Table A9 reports the reduced-form result and shows that this instrument variable is significantly correlated with modern firm entry, after including our full suite of controls. The first stage results in columns (2) and (3) show that our instrument has a significant impact on book density, serving as basic relevance tests. Considering that none of our settings assumes i.i.d. errors and all cluster standard errors at the prefecture level, we report Kleibergen-Paap rk Wald F statistic for the weak identification test. We follow the approach suggested by Baum, Schaffer and Stillman (2007) and apply the Stock and Yogo (2005) critical value table based on TSLS size with a significance level at 5%.[4] For both specifications in Columns (2) and (3), we can safely reject the weak identification null hypothesis.[5]

The unique advantage of using bamboo locations as the instrument is that we find no firm in logging or bamboo-related industries in our firm sample. Upon close inspection, none of the industries present in our firm data had noticeable usage of bamboo either in the setup process or in the production process. Into the second half of the 19th century and beyond, bamboo availability no longer had any role in the distribution of books, because the new paper-making technology switched to use rags, asphalt and wood as the raw materials and printing technology also moved away from the traditional woodblock printing (Reed, 2004). Furthermore, according to Chen, Kung and Ma (2020), bamboo was not intentionally planted in historical China as an economic crop, not even for commercial printing, so its geographic distribution can be considered exogenously determined.

[4]Critical values do not yet exist for Kleibergen-Paap rk Wald F statistic, and Baum, Schaffer and Stillman (2007) suggests to apply with caution the critical values compiled by Stock and Yogo (2005) for the i.i.d. case Cragg-Donald Wald F statistic or apply the old rule of thumb of Staiger and Stock (1997) that the Cragg-Donald Wald F statistic should be at least 10.

[5]According to the table, for the case of one endogenous variable and one instrument, the critical values are between 5.53 and 16.38. In the first stage result of Column (2), a Kleibergen-Paap rk Wald F statistic of 11.07 is above the rule-of-thumb threshold value of 10 and indicates that we have a rejection rate of 15% of a 5% Wald test, when we only control for two-way fixed effects. In Column (3), the first stage Kleibergen-Paap rk Wald F statistic of 26.029 allows us to even more safely reject the null, when including our full suite of controls

They further show that the instrument variable is uncorrelated with a set of prefecture characteristics ranging from population density to urbanization rate both during and after the Ming-Qing period. Therefore, we have confidence in this instrumental variable for satisfying exclusion restrictions.

Column (2) reports the 2SLS estimates that include only the two-way fixed effects for simplicity, and Column (3) adds the full suite of control variables and time trends. The results remain stable and significant throughout. These findings provide further evidence of the strong positive relationship between human capital and industrialization in late Qing China, even when controlling for potential biases from omitted variables that may have changed with treatment.

Table A9: Robustness Check: Instrumental Variable Results (Dependent Var: $lnfirms$)

	(1)	(2)	(3)
	Reduced Form	IV	IV
lnBookDens × PostPort		1.143***	1.004**
		(0.225)	(0.393)
lnBambooDist × PostPort	-0.081**		
	(0.031)		
PostPort	-0.940	-0.0005	-1.118**
	(0.508)	(0.021)	(0.497)
		1st Stage	
		lnBookDens × PostPort	lnBookDens × PostPort
lnBambooDist × PostPort		-0.063***	-0.081***
		(0.020)	(0.013)
PostPort		0.372***	0.177
		(0.102)	(0.177)
Prefecture FE	✓	✓	✓
Year FE	✓	✓	✓
Baseline Controls	✓		✓
More Controls	✓		✓
Prefecture × Year	✓		✓
Province × Year FE	✓		✓
Kleibergen-Paap F-stat		10.107	38.315
Observations	17355	17355	17355

Notes. This table examines whether our findings are robust to using distance to the nearest bamboo site as the instrument variable. Column (1) presents the reduced form regression between the IV to log firms. Columns (2)-(3) present the second stage result at the top and the first stage result at the bottom. Column (2) only controls for the two-way fixed effects. Column (3) further adds our full set of fixed effects and control variables. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

A.9 Distance to the Nearest Treaty Port

We then address the concern that treaty ports potentially had spillover effects on neighboring prefectures. To capture spatial spillover effects, we calculate the distance between each prefecture and its nearest treaty port. However, since treaty ports opened over time in a staggered fashion, the closest treaty port for each prefecture changed dynamically. Consequently, we construct a panel of time-varying distances for all 267 prefectures as follows: (i) Before 1843 when the first treaty port opened, we calculate for each prefecture the distance to Guangzhou prefecture, the only port open to foreign trade designated by the Qing court; (ii) After 1843, we calculate the closest treaty port distance year by year based on the available treaty ports each year; (iii) For the 48 prefectures that received their own treaty ports by 1904, we consider them as control prefectures in the distance calculation when they were not yet treated. Accordingly, Guangzhou has a distance of 0 throughout, and Ningbo has a distance of 1,114 km to Guangzhou between 1840 and 1842. The distance then changes to 142 km to Songjiang in 1843 after Songjiang opened a treaty port. Eventually, the distance becomes 0 between 1844 and 1904 after Ningbo itself became a treaty port in 1844.

To examine the link between books per capita and modern firm entry as the distance to the nearest treaty port varies, we use the following specification:

$$\begin{aligned} lnFirm_{i,t} =& \beta \cdot lnBookDens_i \times lnDistance_{i,t} + \vartheta \cdot lnDistance_{i,t} + \theta \cdot \mathbf{X}_i \times PostPort_{i,t} \\ &+ \kappa \cdot \mathbf{Z}_{i,t} + \gamma_t + \lambda_i + \delta_{prov} \times \gamma_t + \lambda_i \cdot \tau + \epsilon_{i,t} \end{aligned} \tag{7}$$

where $lnDistance_{i,t}$ is the natural log of the closest treaty port distance. Table A10 presents the results. Across columns (1) to (4), the interaction term $lnBookDens_i \times lnDistance_{i,t}$ has coefficients that are negative and significant, and the coefficients for $lnDistance_{i,t}$ are positive but mostly insignificant except in Column (2). Since the interaction term remains significant, it indicates that the relationship between books per capita and modern firm entry changes as the distance changes, even though the main effect of the distance is not significant. The interpretation here is that the positive effect of a 1% increase in books per capita on the percentage change in the number of firms becomes 0.08% stronger as the distance to the nearest treaty port decreases by 1%. Overall, this alternative specification provides evidence that traditional knowledge stock, as proxied by book density, had a stronger effect on firm entry in locations closer to treaty ports during the emergence of modern firms in China. Together with our baseline findings above, it points to the critical role that connections to global trade networks and opening-up played in awakening the relationship between upper-tail human capital and economic development during this

transformative stage of initial industrialization.

Table A10: Firm, BookCount and Distance to the Nearest Treaty Port

	(1) lnfirms	(2) lnfirms	(3) lnfirms	(4) lnfirms
lnBookDens × lnDistance	-0.177***	-0.144***	-0.098***	-0.081***
	(0.029)	(0.018)	(0.027)	(0.029)
lnDistance	0.006	0.042***	0.008	0.004
	(0.006)	(0.013)	(0.014)	(0.013)
Prefecture FE	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Baseline Controls		✓	✓	✓
Prefecture × Year			✓	✓
Province × Year FE			✓	✓
More Controls				✓
Adj R-square	0.290	0.307	0.547	0.547
Observations	17355	17355	17355	17355

Note. This table examines whether our findings are robust to including the potential spillover effects of the treaty ports. All regressions are run with an interaction term between log book density and log distance to the nearest treaty port, instead of that between log book density and an indicator for treaty ports. Standard errors in parentheses are clustered at the prefecture level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B Mechanism

B.1 The Imperial Examination

As early as 220 BCE, the Han dynasty established the "Nine-Rank System" (*Jiǔpǐn Zhōngzhèng Zhì*) to recruit elites into the bureaucratic system. It was a system with meritocracy in design but rampant nepotism in practice. The system was eventually replaced by the Imperial Examination in 605 under the Sui dynasty (581–618). The Imperial Examination, also known as *Kējǔ*, was adopted and held regularly by later dynasties until its abolition in 1905, with the exception of the Yuan dynasty (1276–1368), where the exam was sporadically held 16 times. Starting with the Song dynasty (960-1279), the Examination was open to men from almost all socio-economic backgrounds, with only a few exceptions.[6]

The Odds of becoming Jinshi During the Ming dynasty (1368–1644), the Imperial Examination produced a total of 24,866 *Jinshi* in 89 triennial exam cycles, with an average

[6]such as craftsmen, businessmen, criminals, and religious personnel.

of 279 per cycle (Wu, 2009). In the 112 cycles during the Qing dynasty (1644–1911), the Examination produced a total of 27,009 *Jinshi*, with an average of 241 per cycle (Zhu and Xie, 1979). In 1394, with a population of approximately 65 million, the number of *Jinshi* titles granted was only 100, resulting in a ratio of one *Jinshi* per 650 thousand persons. The competition for the *Jinshi* title became even more exclusive during the Qing dynasty against the backdrop of a population boom. In 1850, out of a pool of two million participants (from a total population of 410 million), 1.5% achieved the *Shengyuan* status, further 5% of the latter passed the provincial-level exam, and less than 20% of the remaining contenders eventually pass the final level of the Examination, meaning that only one in 6,000 candidates or one per two million population would become *Jinshi* (Elman, 1991).

The lower degrees, unlike the top prizes, carried no qualification in most official positions, but they covered a much broader population of literati in China. For example, the club of *Shengyuan* was much more inclusive and was governed by a quota system during the Qing Dynasty. In 1400, Ming China cumulatively had 30,000 *Shengyuan* out of a population of 65 million, resulting in a ratio of 1 per 2,200 persons. By 1700, the number of *Shengyuan* rose to 500,000 in a population of 150 million, resulting in a ratio of 1 per 300 persons. During the Qing, the sum of quotas for each exam cycle across all prefectures was around 30,000.

Scholar Gentry in the Local Society Therefore, the true allure of the Examination in shaping knowledge production across China lies in the entry to intermediate exam titles, offering low-hanging rewards for the majority. Unlike their *Jinshi* counterparts, who were often appointed to mid or high official positions outside their home regions, intellectuals with entry to intermediate exam titles remained in their localities. This stratum of intellectuals acted as intermediaries between government officials and the populace and assumed a pivotal role in the governance of local communities, including the construction and upkeep of public infrastructure. As their numbers grew, these titleholders' state-recognized privileges evolved into markers of social status, forming the "scholar gentry" as a distinct social class.

Beyond their administrative contributions, the gentry were instrumental in founding private academies, which stood in complement to the state-run schools overseen by educational officials. These private institutions, managed by the local scholar gentry, emerged as the epicenters of scholarly activity and intellectual exchange (Chang, 1955, p. 63). Since the Examination system and its conferred degrees were key to the membership of the scholar gentry group, Chang contends that the gentry served as custodians, advocates, and embodiments of the knowledge fields within the Examination curriculum. Quantita-

tively, the scholar gentry class represented a significant reservoir of human capital. In the 19th century, they numbered approximately 5.5 million, or 1.3% of the population (Chang, 1955, p. 139). As Ji (2006) posits, this influence extended beyond degree holders to their kin, with average household sizes of 5.68 and 5.33 in the Ming and Qing dynasties, respectively (Ho, 1962). Their cultural influence permeated other social strata, as evidenced by merchants aspiring to adopt the scholar-gentry lifestyle (Mote, 1999, pp. 764-765). In summary, both the quantity and social influence of human capital beyond the *Jinshi* are crucial for understanding development in Chinese history.

B.2 Jinshi and Quota Density

Figure B3 illustrates that the impact of *Jinshi* density did not exhibit significance until 30 years after the treatment. As our analysis employs a staggered event study with relative time indicators, it is challenging to pinpoint the exact calendar year for the relative period 30. However, we can provide some insight into the timing. Based on Appendix Table C14, most treaty ports were established between 1860 and 1900. Therefore, the approximately 30-year delay in the significance of *Jinshi*-related human capital aligns roughly with the abolition of the *keju* examination system. Combined with Bai (2019)'s findings, it is relatively safe to conclude that the *Jinshi* population did not actively participate in modern entrepreneurship until the abolition of the *keju* exam.

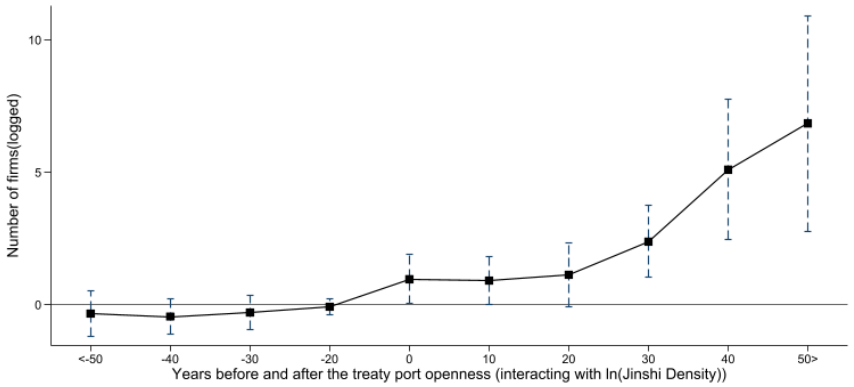


Figure B3: The dynamic effects of log *Jinshi* density on modern firms creation

These figures use 10 years before treatment as the reference, where the solid line connects the coefficients and the dashed lines indicate the 95% confidence intervals.

In Figure B4, we apply the same dynamic event study on quota density as the specification in Equation 2. Similar to the result for *Jinshi* density, there is a 20-year delay in significance for quota density. Notably, in the case of quota density, there is no obvious break in the trend from period -20 to period 0, and the first significant break occurs only

30 years after the shock. This observation complements the work of Bai (2019). When examining quota density, Bai (2019) demonstrates that $ln(QuotaDens + 1) \times Post1904$ has a positive and significant effect on firm creations. We further show that if the *Post* indicator is moved to any event prior to the abolition of the *keju* system, the effect likely disappears.

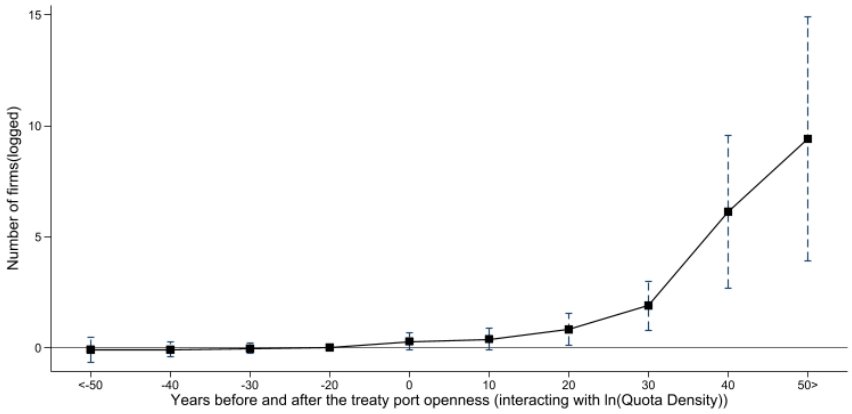


Figure B4: The dynamic effects of log quota density on modern firms creation

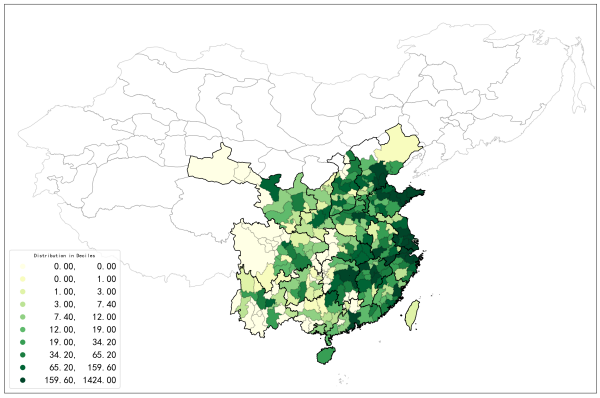
These figures use 10 years before treatment as the reference, where the solid line connects the coefficients and the dashed lines indicate the 95% confidence intervals.

B.3 Books by Jinshi vs. Books by Others

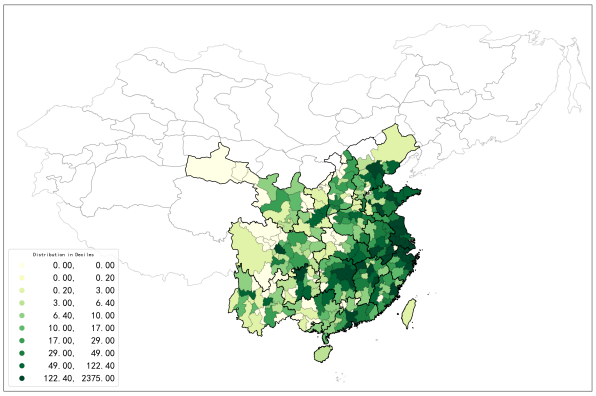
In this section, we present additional analysis results and raw data patterns on the comparison between human capital inside and outside the *Jinshi* circle, to help understand our analysis results on them in the main text. Figure B5 compares the spatial distribution of these two groups of books between 1644 and 1840.

Figure B6 illustrates a comparative analysis of book topics authored by *Jinshi* and non-*Jinshi* writers.

Figure B7 presents the spatial correlations among the different human capital measures used in the analysis. It provides support to our horse-race results.



(a) Total No. of Books Written by *Jinshi* at Prefecture Level (1644-1840)



(b) Total No. of Books Written not by *Jinshi* at Prefecture Level (1644-1840)

Figure B5: Spatial Distribution of Books by *Jinshi* and by Others

These GIS figures visualize the different spatial distributions of books written by *Jinshi* and books written by non-*Jinshi* authors.

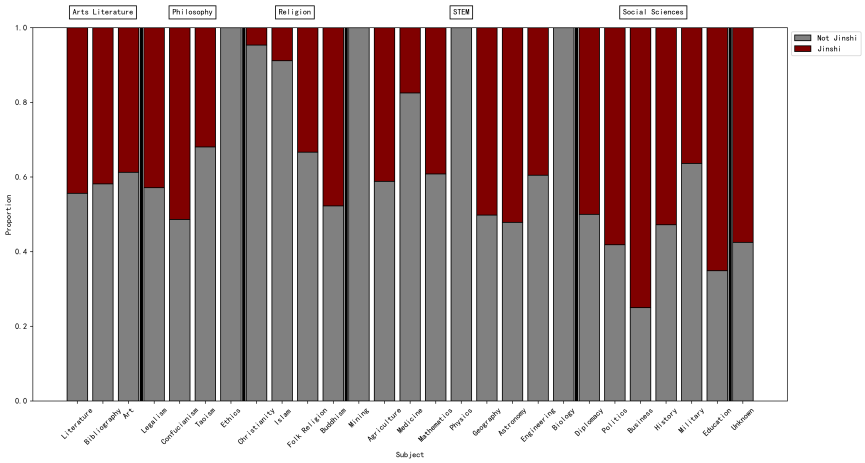


Figure B6: Book Proportions across Subjects (*Jinshi* vs. Others)

For each subject, the stacked bar shows the proportions written by *Jinshi* and by non-*Jinshi*. The top maroon portion represents the books written by *Jinshi*, and the bottom grey portion represents the books written by non-*Jinshi*.

	QuotaDens	BookDens_NotJinshi	JinshiDens	BookDens_Jinshi	AuthorDens_Jinshi	AuthorDens_NotJinshi
QuotaDens	1.00	-0.16	0.09	-0.08	0.01	-0.14
BookDens_NotJinshi	-0.16	1.00	0.55	0.80	0.71	0.96
JinshiDens	0.09	0.55	1.00	0.70	0.82	0.53
BookDens_Jinshi	-0.08	0.80	0.70	1.00	0.87	0.80
AuthorDens_Jinshi	0.01	0.71	0.82	0.87	1.00	0.72
AuthorDens_NotJinshi	-0.14	0.96	0.53	0.80	0.72	1.00

Figure B7: Correlations among Human Capital Measures

This table presents the spatial correlations among the bibliometric proxies and *keju*-related proxies of human capital. Darker red colors indicate higher correlations.

B.4 Jinshi in Entrepreneurship

We examine the participation of *Jinshi* in modern firm creation between 1840 and 1927. First, we conduct a comparative analysis of the number of firms founded by individuals holding the *Jinshi* title before and after the abolition of the *keju* system in 1905. As illustrated in Figure B8, firm creation by *Jinshi* was a sporadic phenomenon before 1905, with the few instances primarily concentrated between 1900 and 1905. In contrast, after 1905, firm creation by *Jinshi* occurred in substantially larger quantities almost every year. Additionally, we investigate the involvement of *Jinshi* in entrepreneurship by examining the number of *Jinshi* who joined entrepreneurial ventures before and after 1905. Our analysis reveals that 12 *Jinshi* participated in entrepreneurship between 1840 and 1904, while 63 *Jinshi* entrepreneurs emerged between 1905 and 1927. These findings suggest that the *keju* system indeed constrained the *Jinshi* human capital until its abolishment.

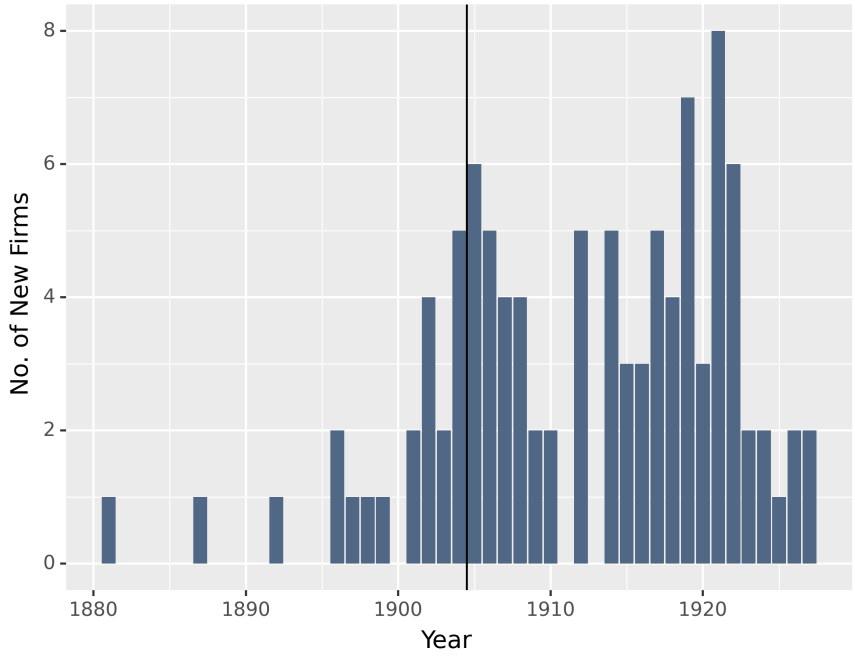


Figure B8: Number of Firms founded by *Jinshi* over Years

This figure plots the yearly count of new firms with *Jinshi* founders. The vertical line marks the abolition of *keju* in 1904.

B.5 History Behind: Industrial Entrepreneurs

To portray a more vivid picture of what happened, we meticulously collected and examined the life stories of 197 founders in our firm sample. Eventually, we are able to find biographical records for 76 of them. The objective is to discern their educational back-

grounds and career paths and compare these anecdotal patterns with our quantitative insights. In this section, we will discuss the three main types of entrepreneurs present in Table 4, each category introduced by an illustrative personal story.

Jinshi: Zhang Jian Having achieved the *Shengyuan* title, Zhang Jian made five unsuccessful attempts at the provincial-level exam at the ages of 18, 21, 23, 24, and 27. However, in an extraordinary turn of events in 1894, the 41-year-old not only obtained the *Jinshi* degree but also topped the Imperial Examination, gaining the coveted title of "Optimus" (*Zhuangyuan*).[7] Zhang's endeavors primarily revolved around business and industrialization. He championed the establishment and management of both state-owned and private firms such as the Dasheng Cotton Mill, Nantong Textile School, and Nantong Normal School. Subsequently, he earned a reputation as one of the top three businessmen of late Qing China, sharing this distinction with Sheng. This is just part of the story of Zhang Jian, the most famous *Zhuangyuan* entrepreneur in Chinese history.

For *Jinshi*, most of their firms were founded after they became *Jinshi* and officials, which is to be expected given the competitiveness of the exam. As per regulations, the Qing government did not prohibit officials from getting involved with private firms. Among the opposition voices against this policy, most were grounded on Confucian moral values rather than anti-corruption concerns. Still, most of the firms founded by *Jinshi* were either directly part of the state-initiated Self-Strengthening Movement or in response to the widespread call for import substitution amid the surge in nationalism. In other words, before 1904, the motivation for *Jinshi* to join entrepreneurship was still mostly political in nature, and such cases remained isolated phenomena and were not yet widespread. Please refer to Appendix Section B.4 for the *Jinshi* population's entrepreneurial activities over time.

Shengyuan to Juren: Sheng Xuanhuai In 1867, 1873, and 1876, Sheng Xuanhuai attempted the provincial-level exam (*xiangshi*) three times, but was unsuccessful on each occasion. Hailing from a lineage of officials and *keju* laureates, Sheng remained at the *Shengyuan* level before ultimately diverting his ambitions toward industrial entrepreneurship, foregoing further attempts at the *Juren* title. As a key figure in the Self-Strengthening Movement, Sheng played a pivotal role in the establishment and management of several significant enterprises, including the Beiyang Shipping Company, the Tianjin Machinery Bureau, and the Hanyang Iron Works. Sheng's extraordinary accomplishments in business and industry earned him recognition as one of the top three most successful and

[7] From 622 to 1904, the Examination crowned a scant total of 654 *Zhuangyuan* in its 1282-year existence.

revered businessmen of late Qing China (Feuerwerker, 1958). Nonetheless, in his personal memoir, he expressed the following sentiments about the top three regrets of his life:

> 吾祖吾父以科第起家，吾少壮时锐欲继绳，而卒屡蹶于秋驾；家有治谱，常以理繁治剧自许，而未尝假乎一州一邑为亲民之官；保使才，办洋务，日与友邦人士相周旋，而足迹未履欧美一步。此生平三憾事也！(My ancestors made their names in *keju*, and in my youth, I aspired to do the same. However, I failed in all my attempts. Despite my family's reputation for producing competent officials, I never held a governing position. I engaged in business and industrialization projects, interacting with foreign friends. Yet, I never set foot in any Western country. These are the three regrets of my life.)
>
> — Sheng (1916)

This quote encapsulates the thought process behind his career transition from an aspiring *keju* scholar to a modern entrepreneur. Initially, Sheng aspired to the *Jinshi* title and a subsequent bureaucratic career, and he only embraced entrepreneurship after repeated exam failures. In other words, despite the burgeoning opportunities in modern business, the allure of the *keju* and officialdom remained compelling in the minds of Chinese intellectuals. Even among the few exam titleholders who ventured into modern entrepreneurship, such as Sheng, they were predominantly affiliated with state officials who were proponents of the Self-Strengthening Movement. Their commercial endeavors were frequently safeguarded by these officials. For example, Sheng Xuanhuai received substantial support from the eminent minister Li Hongzhang.

Founders without Exam Titles: Zhu Dachun Zhu Dachun was born into a poor family and faced the hardship of losing his father at a young age. Instead of going to private Confucian academies like some kids did, Zhu took on various odd jobs as an errand boy and apprentice in different businesses at the age of 16. Through perseverance and honesty, Zhu grew up to be a successful comprador himself and worked very closely with Jardine Matheson in Shanghai. He also established many well-known companies such as Yuan Chang Machinery and Hardware Factory and Heng Chang Yuan Yarn Factory.

This group of entrepreneurs is not only the largest, but also is the group with the most diverse backgrounds. One prominent commonality is that many of them started their careers as errand boys or apprentices in Chinese or foreign businesses at a young age due to hardships. Similar to Zhu, some of these apprentices ultimately became compradors. For instance, Zhang Bishi, who also came from poverty, had to leave his hometown due to

a famine to find apprentice jobs in Jakarta. However, he later founded the famous winery company Changyu Pioneer Wine in 1892, which is the oldest and largest winery in China today. For others, becoming an apprentice first and then a comprador was a natural career path because they attended schools founded by Christian missionaries from the start. For example, Tong Mow Chee attended the Morrison School in Macau. He was one of the earliest groups of students to be educated in Western-style schools and also worked as a comprador for Jardine Matheson in Tianjin.

Table B11 presents the distribution of the 76 entrepreneurs across the three groups introduced above. In Column (1), we can identify a total of 12 founders who had achieved the *Jinshi* degree by matching with the *Jinshi* data. Column (2) lists another 12 entrepreneurs with documented examination titles; within this cohort, 7 attained the *Juren* title and 5 the *Shengyuan* title. Column (3) highlights that the most substantial segment of our entrepreneur sample—52 individuals—have no recorded examination titles.

We approach these statistics with prudence, given the absence of information for 121 entrepreneurs. However, this caveat should not detract from the observation that during this period of modern firm creation, the primary contributors were intellectuals without any examination titles. The exhaustive nature of the *Jinshi* biographical database ensures that there are no gaps in our data regarding entrepreneurs with a *Jinshi* degree. Among the 121 unrecorded entrepreneurs, it is reasonable to surmise that exam titleholders constitute a minority for two reasons: (i) historically, exam titleholders represented a relatively small fraction within the literati population, and (ii) the records of exam titleholders are presumably more systematically maintained and have a higher likelihood of persisting into contemporary times compared to those of non-titleholding entrepreneurs from diverse backgrounds.

Table B11: Distribution of Entrepreneur Types

	(1) *jinshi*	(2) *Shengyuan* to *Juren*	(3) Intellectuals w/o exam titles
Count	12	12	52
Percentages	15.79%	15.79%	68.42%
Average No. of Firms	2.5	2.58	2.10
Top 3 Industry	Shipping	Silk reeling	Silk reeling
	Spinning	Fuel and other mining	Shipping
	Fuel and other mining	Flour industry	Machinery manufacturing

Note. This table accounts for the distribution of the three types of entrepreneurs present in our data. Column (1) represents entrepreneurs who received the *jinshi* title. Column (2) represents entrepreneurs who achieved lower to mid-level exam titles. Column (3) represents entrepreneurs who had no record of any exam titles.

In summary, our historical anecdotal patterns align with the results of our regression

analysis: the reallocation of traditional human capital into industrial entrepreneurship predominantly originated from individuals beyond the *Jinshi* circle, particularly from those within the non-*Jinshi* demographic who lacked any examination titles.

C Data

C.1 Summary Statistics

Table C12: Summary Statistics for the Main Variables

	(1) VARIABLES	(2) N	(3) mean	(4) sd	(5) Sources
Modern Firms	firms	17,355	0.045	0.503	1,2,3,4,5
	ln(firms+1)	17,355	0.020	0.157	1,2,3,4,5
	private firms	17,355	0.029	0.391	1,2,3,4,5
	state-owned firms	17,355	0.004	0.073	1,2,3,4,5
	state-private joint firms	17,355	0.010	0.164	1,2,3,4,5
	big firms	17,355	0.028	0.374	1,2,3,4,5
Treaty Port	Post Port	17,355	0.056	0.230	17
Sci Socities & Journals	ModernJournalSociety	17,355	0.00490	0.135	27,28
Bibliometric Measures	Book Count	267	211.491	618.300	6,23
	Book by Jinshi	267	98.558	262.353	6,11,23
	Book not by Jinshi	267	112.933	369.691	6,11,23
	Author	267	35.172	86.995	6,23
	Author Jinshi	267	16.150	31.52	6,11,23
	Author not Jinshi	267	19.022	57.841	6,11,23
	Book from siku	267	15.93	47.96	7,8,9,10
Baseline Controls	Average Population 1644-1840	267	1,080	982.8	14
	Population Density	17,355	0.127	0.123	14
	coastal distance_km	267	511.9	370.60	13
	Terrain ruggedness	267	234.0	182.4	13
	Wheat suitability	267	23.60	11.21	13
	Rice suitability	267	10.62	10.60	13
	Sweet suitability	267	8.981	9.631	13
	Maize suitability	267	22.11	11.51	13
	Distance to navigable river	267	26.97	36.52	13
	Distance to Capital	267	1,224	573.0	13
	chong	267	0.685	0.652	24
	fan	267	0.948	0.509	24

Table continues next page

	VARIABLES	N	mean	sd	Sources
	pi	267	0.408	0.672	24
	nan	267	0.839	0.587	24
	Market towns	267	27.086	22.311	25
More Controls	Disasters	17,355	0.409	1.159	15
	Wars	17,355	0.026	0.182	16
	Academies	17,355	0.096	0.400	18
	Church	17,355	0.280	3.100	20
	Priest	17,355	0.467	5.048	20
	Pantheons	267	12.06	10.35	22
	Genealogies volumes 1644-1840	267	372.5	1,516	21
	Literary inquisitions	267	0.352	0.837	26
Instrumental Variable	Min distance to bamboo	267	160.8	162.8	13
	ln(min distance to bamboo)	267	4.572	1.124	13
keju Human Capital	Jinshi 1644-1840	267	70.15	119.3	11
	ln(jinshi density)	267	0.0487	0.048	11
	Avg quota 1644-1840	267	109.1	74.46	12
	ln(quota density)	267	0.118	0.054	12

1: Zhang (1988*a*);
2: Du (1991);
3: Zhang (1989);
4: Zhang (1988*b*);
5: Zhang (1987);
6: Ancient Books Editorial Committee (2009);
7: Yong (1792);
8: Wu, Zhang and Wen (2000);
9: 《续修四库全书》编委会 (2002);
10: 《四库未收书辑刊》编纂委员会 (1997);
11: Gong (2021);
12: Kun et al. (1899);
13: Chen, Kung and Ma (2020);
14: Cao (2000);
15: Zhang et al. (2004);
16: 《中国军事史》编写组 (2003);
17: Yan (2012).
18: Ji (1996)
19: Chen, Xiao and Xiong (2012)
20: Stauffer (1922)
21: Shanghai Library (2009)
22: Zhange and Xien (2008)
23: Zhonghua Shu Ju (2000)
24: Liu (1993)
25: Harvard Yenching Institute and ICHGS at Fudan University (2016)
26: 北平故宫博物院文献馆 (2011)
27: China Kexie Development Research Center (2014)
28: Yao, Wang and Yao (2008)

C.2 Book Catalogs in the Twenty-Four Histories

Primary Source Book Catalogs Among the history literature in the Chinese language, to the best of our knowledge, all studies on Chinese book catalogs and bibliometric measures rely on the book catalog chapters in the Twenty-Four Histories. "Twenty-Four Historie" is a collection of 24 dynastic history books for the dynasties prior to Qing and was officially recognized by Emperor Qianlong of Qing. Among 7 of the 24 histories, each of them has chapters dedicated to cataloging books of its time. Please see the following list:

- 《汉书艺文志》(Book Catalog for Han Dynasty)
- 《隋书经籍志》(Book Catalog for Sui Dynasty)
- 《旧唐书经籍志》(Old Book Catalog for Tang Dynasty)
- 《新唐书艺文志》(New Book Catalog for Tang Dynasty)
- 《宋史艺文志》(Book Catalog for Song Dynasty)
- 《明史艺文志》(Book Catalog for Ming Dynasty)
- 《清史稿艺文志》(Book Catalog for Qing Dynasty)

Avoding Duplicates One important detail worth pointing out is that: one cannot simply count and concatenate all the entries from these 7 book catalogs. All of the catalogs, except the catalog for the Ming dynasty, recorded the stocks of books at their times, which means that they also included books written by authors from previous dynasties. Consequently, duplicate records are common across catalogs of different dynasties. In contrast, the book catalog for the Ming only recorded books written during its time. For our purpose, considering that we only want to add book entries that are not recorded in the *Guji Zongmu*, we need to be extra careful at avoiding duplicated entries.

Thankfully, all of the Twenty-Four Historie are already digitized, so we can directly apply our NER model to identify entities like authors and book titles. Then, we rely on the combination of book titles and authors as the unique identifier for each book and filter our duplicated books before adding them in.

C.3 Data Extraction on *Zhongguo Guji Zongmu*

Here we explain how we have extracted all the individual book entries from the *Zhongguo Guji Zongmu* (Grand Catalog of Books in pre-1912 China).

Extraction on Book Entries The entire 26 volumes of books are organized in a bullet-point style, with clear clues of indentations and spacing. Thanks to the hard work of the Ancient Books Editorial Committee (2009), we can train a basic image recognition algorithm to crop each page into individual book entries based on these clear indentation rules. Figure C9b is an example page cropped by the bot. The slightly messy part of this step is the fact that the last book entry of many pages goes into the next page. We have to put in manual effort to fix and examine these broken book entries. After that, we push these individual book entry images through OCR tools and end up with individual book entries in text. To get more information about each book, we next need to extract key entities within each book entry. Thankfully, Ancient Books Editorial Committee (2009) also maintained a very unified structure within each book entry and the book preface contains a chapter outlining the formatting rules. Basically, each book entry recorded key entities in a set order and with clear spacing and indentations. Combining the format clues, we also trained a Named Entity Recognition algorithm, where we targeted key entities like authors, dynasties, authorship types, and version history. Figure C10 is an example flash-card from our training set, where we manually defined these key entities, and Figure C11 is a robustness check on our NER algorithm by comparing its answer with that of our manual input. Interestingly, our manual input on the top actually overlooked the volume number entity, but our NER algorithm was able to correctly separate the volume entity from the book title entity.

(a) A Sample Table of Content

(b) A sample page cropped by AI

Figure C9: Sample Pages from *Guji Zongmu*

Panel (a) is a page of the table of content from group Zi of the Four-Group Classification. Panel (b) is a sample of how we use image recognition to crop each page into individual book entries.

Figure C10: Manual Entity Tagging for the Training Set

This is the web system that we use for manually tagging a training set for our machine learning.

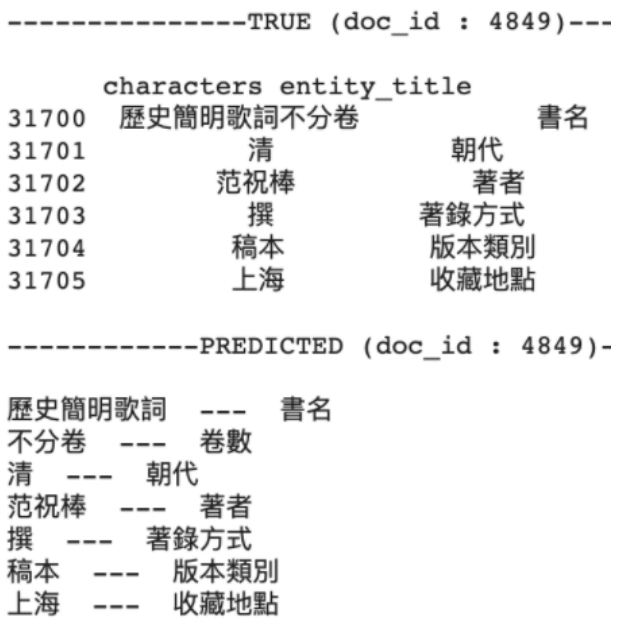


Figure C11: An Example of Manual vs. Machine Tagging

This is a flashcard from the training set, showing human tagging on the top and machine tagging on the bottom. In this particular case, the human tagger made a mistake but the machine was able to correct it.

C.4 Content Classification

Four-group Classification In the main text, we have explained briefly that the traditional book content classification method started during the Han dynasty as the Six-group Classification and gradually evolved into the Four-group Classification system during the Sui dynasty. Please see the list below for details:

六部分类法**(Six-group Classification)**	四部分类法**(Four-group Classification)**
六艺(six arts)	经(classics)
诸子(various masters)	史(history)
诗赋(poetry)	子(miscellaneous)
兵书(warfare)	集(literature)
数术(math and astrology)	
方技(medicine and miscellaneous)	

Table C13: Comparative table of the Six-group and Four-group classification methods.

As mentioned in the main text, although at the top level, it reduced from six to four groups, the classification system actually evolved to be more granular. The Six-group

Classification has around 38 categories at its lowest level, and the Four-group Classification has around 60 categories at its most granular level. Therefore, when coding the content of each book entry, we rely on its most granular level of category in the Four-group Classification. For instance, in some cases, people count all books under the 子 (miscellaneous) category as STEM-related. However, our review of the finer subcategories under 子 (miscellaneous) reveals that a significant proportion of this group comprises miscellaneous studies in non-mainstream philosophy.

The table of contents of *Zhongguo Guji Zongmu* follows the traditional "四部分类法" (four-group classification). 26 volumes of books are grouped into 4 and within each group there are further finer levels of classifications. Figure C9a is an example page from the table of contents of the 子 (miscellaneous) group. On this page, we can see five major categories: Astronomy/Astrology, Mathematics/Arithmetics, Arts, Encyclopedia, and Miscellaneous. Within each category, there are even lower levels of classifications. We manually coded the table of contents into 28 modern subjects and then further grouped these subjects into five major topic areas.

C.5 Historical Biography Query

CBDB The primary source of biography information for the authors is the China Biographical Database (CBDB) (Harvard University, Academia Sinica and Peking University, 2022). CBDB is a freely accessible database with biographical information about approximately 500,000 individuals primarily from the 7th through 19th centuries. Among the entities we have extracted from the book catalog, we have the name and dynasty of each author, and together they serve as the identifier we can use to query for biography in CBDB. First of all, we have to filter out all the figures in CBDB with duplicates in both dynasty and name. For the rest of the figures, we match them with the authors' roster based on the unique identifier in "name" + "dynasty". We mainly look for birthplace and birth year of each author from CBDB.

The Jinshi Database Our next main source is the biographical database on *Jinshi* by Gong (2021). Since many *Jinshi* were also active writers and scholars, naturally we want to match our authors to this database. Unfortunately, we also need to drop all the *Jinshi* with identical names and dynasties. The matching procedure is identical to that of CBDB case above. Additionally, we also use similar approach to identify which of our authors were *Jinshi*, except that we do not need to drop *Jinshi* with identical names and dynasties in the *Jinshi* database. For authors with multiple matches from the *Jinshi* sides, although

we cannot pinpoint the exact one, we can at least safely conclude that these authors were highly likely to be *Jinshi*.

Wikipedia and Baidu Baike To further increase our coverage rate, we loop the rest of the authors through the two search engines: Wikipedia and Baidu Baike. Fortunately, many of the biography pages on these two sites contain tables that list the basics such as birthplace and birth year. It enables us to precisely extract these two pieces of information from the HTML file without the need to do parsing and cleaning on the biography articles themselves. For all three processes listed here, for books with multiple authors, we treat such book as separate authorships and match them to their respective birthplace and birth year.

Appendix Figure C12 shows that while most subjects have matching rates ranging from 40% to 60%, certain areas such as Ethics, Folk Religion, Buddhism, and Biology present more substantial challenges. However, the relatively lower matching rates in these four subjects are unlikely to impact our study adversely. These fields were not predominant in the landscape of knowledge production in historical China, and their influence on the emergence of modern firms is presumed to be minimal.

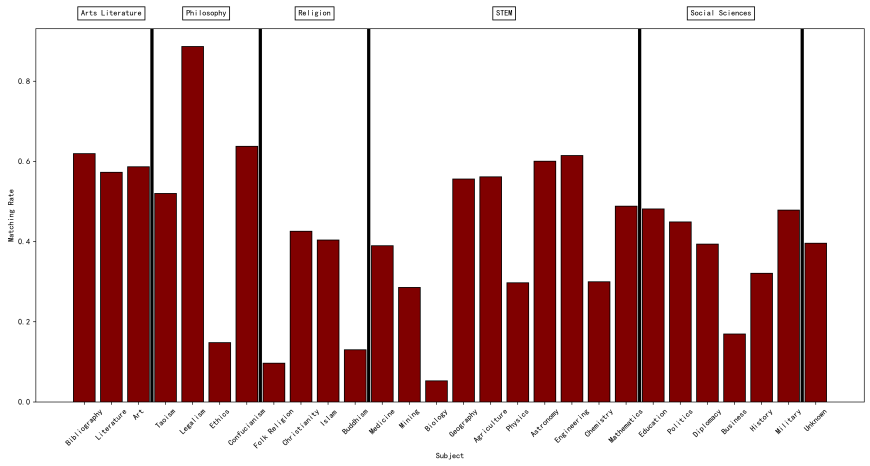


Figure C12: Book Matching Rate across Subjects (Qing)

This figure plots the matching rates for books in different subjects. Each bar represents the proportion of books in that subject with successful matching for the information on the authors.

C.6 Treaty Ports

In this section, we explicitly list the prefectures in our sample that we identified as treatment prefectures between 1840 and 1904. Table C14 lists the treatment prefectures and their initial treatment time, and Column "PortCohort" is an identifier for different

treatment cohorts who shared the same initial treatment time. Among the 34 prefectures with treaty ports, there were 18 cohorts who received their initial treatment in 18 different years. The cohort identifier is also used in the robust estimation of staggered DiD.

Table C14: Prefectures with Treaty Ports

Prefecture	Prefecture(Eng)	PortYear	PortCohort
广州府	Guangzhou	1843.0	2
松江府	Songjiang	1843.0	2
宁波府	Ningbo	1844.0	3
福州府	Fuzhou	1844.0	3
泉州府	Quanzhou	1845.0	4
潮州府	Chaozhou	1860.0	5
天津府	Tianjin	1861.0	6
镇江府	Zhenjiang	1861.0	6
九江府	Jiujiang	1862.0	7
台湾府	Taiwan	1862.0	7
汉阳府	Hanyang	1862.0	7
登州府	Dengzhou	1862.0	7
琼州府	Qiongzhou	1870.0	8
太平府	Taiping	1877.0	9
宜昌府	Yichang	1877.0	9
廉州府	Lianzhou	1877.0	9
温州府	Wenzhou	1877.0	9
肃州	Suzhou	1881.0	10
临安府	Lin'an	1889.0	11
太平府(桂)	Taiping (Gui)	1889.0	11
重庆府	Chongqing	1891.0	12
杭州府	Hangzhou	1896.0	13
苏州府	Suzhou	1896.0	13
荆州府	Jingzhou	1896.0	13
开化府	Kaihua	1897.0	14
普洱府	Puer	1897.0	14
梧州府	Wuzhou	1897.0	14
太仓州	Taicang	1898.0	15
岳州府	Yuezhou	1899.0	16
江宁府	Jiangning	1899.0	16
福宁府	Funing	1899.0	16
永平府	Yongping	1901.0	17
腾越厅	Tengyue	1902.0	18
长沙府	Changsha	1904.0	19

References

Aihounton, Ghislain B D, and Arne Henningsen. 2021. "Units of measurement and the inverse hyperbolic sine transformation." *The Econometrics Journal*, 24(2): 334–351.

Ancient Books Editorial Committee. 2009. *Zhongguo Guji Zongmu (Grand Catalog of Books in pre-1912 China).* Shanghai: Shanghai Classical Publishing House.

Bai, Ying. 2019. "Farewell to confucianism: The modernizing effect of dismantling China's imperial examination system." *Journal of Development Economics*, 141.

Baker, Andrew, David F. Larcker, and Charles C.Y. Wang. 2022. "How much should we trust staggered difference-in-differences estimates?" *Journal of Financial Economics*, 144: 370–395.

Baum, Christopher F., Mark E. Schaffer, and Stephen Stillman. 2007. "Enhanced Routines for Instrumental Variables/GMM Estimation and Testing." *Stata Journal*, 7(4): 465–506.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics*, 119(1): 249–275.

Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with multiple time periods." *Journal of Econometrics*, 225: 200–230.

Cao, Shuji. 2000. *Zhongguo Renkou Shi, Diwu Juan (History of Population in China,Volume 5).* Shanghai: Fudan University Press.

Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs." *The Quarterly Journal of Economics*, 134(3): 1405–1454.

Chaney, Eric. 2016. "Religion and the Rise and Fall of Islamic Science." Working paper.

Chaney, Eric. 2020. "Modern Library Holdings and Historic City Growth." Working paper.

Chang, Chung-Li. 1955. *The Chinese Gentry: Studies on Their Role in Nineteenth-Century Chinese Society.* Seattle: University of Washington Press.

Chen, Jiafeng, and Jonathan Roth. 2023. "Logs with Zeros? Some Problems and Solutions." *The Quarterly Journal of Economics*.

Chen, Ting, James Kai-sing Kung, and Chicheng Ma. 2020. "Long Live Keju! The Persistent Effects of China's Civil Examination System." *The Economic Journal*, 130(631): 2030–2064.

Chen, Zhenhan, Guoliang Xiao, and Zhengwen Xiong. 2012. *Qingshilu jingjishi ziliao (Economic history materials from the Veritable Records).* Beijing: Peking University Press.

China Kexie Development Research Center. 2014. *Jindai Zhongguo Keji Shetuan (Modern Chinese Science and Technology Societies).* China Science and Technology Press.

De Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–96.

De Chaisemartin, Clément, and Xavier D'Haultfœuille. 2022. "Difference-in-Differences Estimators of Intertemporal Treatment Effects." National Bureau of Economic Research Working Paper 29873.

Deshpande, Manasi, and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs." *American Economic Journal: Economic Policy*, 11(4): 213–248.

Du, Xuncheng. 1991. *Minzu Ziben Zhuyi yu JiuZhongguo Zhengfu 1840-1937 (national capitalism and old Chinese government 1840-1937).* Shanghai Academy of Social Sciences

Press.
Elman, Benjamin A. 1991. "Political, Social, and Cultural Reproduction via Civil Service Examinations in Late Imperial China." *The Journal of Asian Studies*, 50(1): 7–28.
Ferman, Bruno, and Cristine Pinto. 2019. "Inference in Differences-in-Differences with Few Treated Groups and Heteroskedasticity." *The Review of Economics and Statistics*, 101(3): 452–467.
Feuerwerker, Albert. 1958. *China's Early Industrialization: Sheng Hsuan-huai (1844-1916) and Mandarin Enterprise.* Harvard University Press.
Gong, Yanming. 2021. "Lidai Jinshi Dengke Shuju Ku (Biographical Database of Jinshi across dynasties)."
Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225: 254–277.
Harvard University, Academia Sinica, and Peking University. 2022. "China Biographical Database (CBDB)."
Harvard Yenching Institute, and ICHGS at Fudan University. 2016. "CHGIS, version 6."
Ho, Ping-ti. 1962. "Studies on the Population of China, 1368–1953." *Comparative Studies in Society and History*, 5(1): 112–114.
Ji, Xiaofeng. 1996. *Zhongguo Shuyuan Cidian (A Compendium on the Chinese Academies).* Hangzhou: Zhejiang Jiaoyu Chubanshe.
Ji, Yingying. 2006. "Mingqing kejuzhi de shehui zhenghe gongneng: yi shehui liudong wei shijiao (On the Social Integrative Function of the Chinese Imperial Examination System in Ming-Ch'ing Period: From the Perspective of Social Mobility)." *Chinese Journal of Sociology*, 26(6): 197–215.
Kun, Gang, et al. 1899. *Qinding da Qing huidian shili (Imperially Established Institutes and Laws of the Great Qing Dynasty).* Beijing: Zhonghua shu ju.
Lin, Justin Yifu. 1995. "The Needham Puzzle: Why the Industrial Revolution Did Not Originate in China." *Economic Development and Cultural Change*, 43: 269–292.
Liu, Cheng-yun. 1993. "Chong, Fan, Pi, and Nan: An Exploration of the Ranking of Qing Administrative Units." *Bulletin of Institute of History and Philology, Academia Sinica*, 64(1).
Mote, F.W. 1999. *Imperial China: 900-1800.* Cambridge, MA: Harvard University Press.
Mullahy, John, and Edward C. Norton. 2023. "Why Transform Y? The Pitfalls of Transformed Regressions with a Mass at Zero*." *Oxford Bulletin of Economics and Statistics*.
Reed, Christopher A. 2004. *Gutenberg in Shanghai: Chinese Print Capitalism, 1876–1937.* Vancouver:UBC Press.
Shanghai Library. 2009. *Zhongguo Jiapu Zongmu (A Comprehensive Catalogue on the Chinese Genealogy).* Shanghai: Shanghai Guji Chuban She (Shanghai Classics Publishing House).
Sheng, Chongyin. 1916. *Sheng Xuanhuai Xingshu(a brief biography of Xuanhuai Sheng).* Shanghai People's Press. reprinted in 1961.
Squicciarini, Mara P., and Nico Voigtländer. 2015. "Human Capital and Industrialization: Evidence from the Age of Enlightenment." *The Quarterly Journal of Economics*, 130(4).
Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica*, 65(3): 557–586.
Stauffer, Milton Theobald. 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 Comm. on Survey and Occupation, China Continuation Comm., 1918-1921.* Shanghai: China Continuation Committee.
Stock, James H., and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear

IV Regression." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. , ed. Donald W. K. Andrews and James H. Stock, 80–108. Cambridge University Press.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2).

Wooldridge, Jeffrey M. 2021. "Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators." *SSRN*. SSRN Working Paper No. 3906345.

Wu, Sugong, Sui Zhang, and Xiangfeng Wen. 2000. *Siku Jinhuishu Congkan(Books banned from Complete Library of the Four Treasure).* The Beijing Publishing Group.

Wu, Xuande. 2009. *Mingdai Jinshi De Dili Fenbu (Geographic Distribution of Jinshi during the Ming dynasty).* Hong Kong: The Chinese University of Hong Kong.

Yan, Zhongping. 2012. *Zhongguo Jindai Jingjishi Tongji Ziliao Xuanji(Early Modern China Economic History Statistics Book). Institute of Economics Chinese Academy of Social Sciences, bulletin on early modern Chinese economic history*, China Social Sciences Press.

Yao, Yuan, Rui Wang, and Shufeng Yao. 2008. *Zhongguo Jindai Keji Qikan Yuanliu 1792-1949 (he Origin and Development of Modern Chinese Science and Technology Journals).* Shandong Education Publishing House.

Yong, Rong. 1792. *Siku Quanshu(Complete Library of the Four Treasure). photocopy in 1986*, The Commercial Press Taiwan.

Zhang, D., et al. 2004. *Zhongguo Sanqiannian Qixiang Jilu Zongji (A Compendium of Chinese Meteorological Records of the Last 3000 years).* Phoenix Publishing & Media, Inc.

Zhange, Mu, and Pan Xien. 2008. *Uniform Sourcebook of the Great Qing (Daqing Yitong Zhi).* Shanghai: Shanghai Guji Chuban She (Shanghai Classics Publishing House).

Zhang, Yufa. 1987. *Qingmo Minchu de Waizi Gongye (Foreign Ownership Firms in Early Modern China). Bulletin of Institute of Modern History, Academia Sinica*, Institute of Modern History, Academia Sinica.

Zhang, Yufa. 1988*a*. *Qingmo Minchu de Guanban Gongye(1860-1896)(Public Ownership Firms in Early Modern China).* 清季自强运动研讨会论文集, Institute of Modern History, Academia Sinica.

Zhang, Yufa. 1988*b*. *Qingmo Minchu de Guandu Shangban Gongye(State-Private Ownership Firms in Early Modern China). Bulletin of Institute of Modern History, Academia Sinica*, Institute of Modern History, Academia Sinica.

Zhang, Yufa. 1989. *Qingmo Minchu de Minying Gongye(Private Ownership Firms in Early Modern China). Bulletin of Institute of Modern History, Academia Sinica*, Institute of Modern History, Academia Sinica.

Zhonghua Shu Ju. 2000. *Er Shi Si Shi (Twenty-Four Histories).* Beijing: Zhonghua Shu Ju (Zhonghua Book Company).

Zhu, Baojiong, and Peilin Xie. 1979. *Ming Qing Jinshi Timing Beilu Suoyin (Ming and Qing dynasty Jinshi Name Roster).* Shanghai: Shanghai Guji Chuban She (Shanghai Classics Publishing House).

《中国军事史》编写组**.** 2003. *Zhongguo Lidai Zhanzheng Nianbiao(Chronology of Wars in Chinese History).* People's Liberation Army Publishing House.

《四库未收书辑刊》编纂委员会**.** 1997. *Siku Weishou Shu Jikan(Books not collected by Complete Library of the Four Treasure).* The Beijing Publishing Group.

《续修四库全书》编委会**.** 2002. *Xuxiu Siku Quanshu(Extension on Complete Library of the Four Treasure).* Shanghai Chinese Classics publishing house.

北平故宫博物院文献馆**.** 2011. *Qingdai Wenziyu Dang(Archives of the Literary Inquisition during the Qing Dynasty).* Shanghai Bookstore Publishing House.