CENTRE FOR QUANTITATIVE HISTORY

HKU Business School

HK Institute for the Humanities and Social Sciences

Centre for Quantitative History
量化歷史研究中心

CQH Working Paper Series

No. 0012 2026

Local Knowledge and State Building: Evidence from Chinese Gazetteers

Xinxian Li
The University of Hong Kong

Ningxi Liu
King's College London

Chicheng Ma
The University of Hong Kong

This paper examines the importance of local knowledge for state building. Drawing upon the quantity and quality of local gazetteers, the regional encyclopedias of imperial China, we measure the accumulation of local knowledge in 267 prefectures over a millennium. We find that the gazetteers facilitated the local penetration of state power and infrastructure between 1000 and 1820, and continued to contribute to the modernization of state infrastructure during the period of reform between 1860 and 1911. Analyzing the content of gazetteers reveals that the gazetteer effect arises from the richness, concretization, and uniqueness of records of local characteristics. These findings indicate that the state is not merely built on unified and standardized measures but also on context-based local knowledge that increases the legibility of subjects.

Local Knowledge and State Building: Evidence from Chinese Gazetteers

Xinxian Li, Ningxi Liu, and Chicheng Ma*

January 2026

This paper examines the importance of local knowledge for state building. Drawing upon the quantity and quality of local gazetteers, the regional encyclopedias of imperial China, we measure the accumulation of local knowledge in 267 prefectures over a millennium. We find that the gazetteers facilitated the local penetration of state power and infrastructure between 1000 and 1820, and continued to contribute to the modernization of state infrastructure during the period of reform between 1860 and 1911. Analyzing the content of gazetteers reveals that the gazetteer effect arises from the richness, concretization, and uniqueness of records of local characteristics. These findings indicate that the state is not merely built on unified and standardized measures but also on context-based local knowledge that increases the legibility of subjects.

* Xinxian Li, Faculty of Business and Economics, The University of Hong Kong, Pokfulam, Hong Kong. Email: felixx@connect.hku.hk. Ningxi Liu, King's Business School, King's College London, London, WC2B 4BG, United Kingdom. Email: nina.liu@kcl.ac.uk. Chicheng Ma, Faculty of Business and Economics, The University of Hong Kong, Pokfulam, Hong Kong. Email: macc@hku.hk. We thank Ying Bai, Joy Chen, Ting Chen, Zhiwu Chen, Bishnupriya Gupta, James Kung, Jin Li, Kaixiang Peng, Junichi Yamasaki, and seminar/conference participants at the New Economic History of China Workshop (Melbourne), EHS Annual Conference (Glasgow), WEHC 2025 (Lund), Second Workshop on Chinese and Comparative Economic History (Oxford), APEBH Conference 2025 (Monash), Conference on State Capacity in Comparative Perspective (CUHK), Asian Cliometrics Meeting (Seoul), AHEC 2024 (Hong Kong), 12th Australian Cliometrics Workshop (Monash), Duke Kunshan University, Shanghai University of Finance and Economics, Renmin University of China, Hong Kong University and Kyoto University for helpful comments and suggestions. We thank Xiaoyang Luan for help digitizing the gazetteer data, and Maoran Lin, Siyi Wang, and Can Yuan for research assistance. We acknowledge the Research Grants Council of Hong Kong (Grant No.: AoE/B704/22-R) for financial support. All remaining errors are our own.

1. Introduction

What is 'useful knowledge' for state building? While statecraft has relied increasingly on unified and standardized knowledge (notably statistics, registration, and more recently digitization), social scientists also perceive the value of heterogeneous local knowledge. This is especially the case for a centralized, stratified state that governs diverse localities, because local knowledge is drawn from specific regional contexts and thus complements the simplified central statistics. For instance, fiscal capacity may not be built on a formal system of tax registration alone, since the enforcement of taxation in local areas may depend on state agents' knowledge of local culture, social networks, and other idiosyncratic characteristics that are difficult to codify but fundamentally influence the cost of levying tax. The same reasoning applies to the local implementation of other central schemes. The reliance on simplified standards, argues James Scott (1998), often leads to the failure of state policies, whereas success should be built on the mastery of context-based local expertise. This resonates with Hayek's (1945) notion in "The Use of Knowledge in Society" that a state's economic system should operate upon dispersed local knowledge rather than the limited information available to the central authority. However, the effect of local knowledge on state building has yet to be examined systematically, possibly because it is challenging to measure the complex and diverse forms of local knowledge.

In this paper, we can measure the reserve of local knowledge using the local gazetteers that were compiled in imperial China. A Chinese local gazetteer is the encyclopedia of an administrative region. It not only describes the geography but also encompasses a complete set of up to several hundred local characteristics ranging from history, politics, society, culture, to miscellaneous subjects with thousands of pages of detail (see Online Appendix Figure A1 for an illustration). From the tenth century onwards, local scholars continuously compiled and updated gazetteers, by which they classified local characteristics and made them legible for reference by government (Dennis, 2015). Moreover, a gazetteer introduces the exclusive characteristics of its administrative division, which enables us to measure and compare clearly the accumulation of local knowledge across regions.

Therefore, we use the number (or volumes and pages) of local gazetteers as the primary measure of the stock of local knowledge in 267 prefectures in China proper. The gazetteers of a prefecture include both the prefectural gazetteers and the gazetteers of the counties in the prefecture. We collected most if not all the gazetteers (7,860 titles) compiled between circa 220 and 1842 based on multiple bibliographies of historical gazetteers. Further, we explored the content of gazetteers to examine the specific effect of local knowledge. We use the number of topics contained in the

gazetteer to gauge the richness of the recorded knowledge. Moreover, of the total of 10,080 topics we refine the local knowledge effect by differentiating topics that are unique to local areas from those that are routinely found in central compilations.

State building in Imperial China also varied greatly, in both time and space. Our period of analysis (1000–1911) saw sustained centralization of the imperial state. Moreover, since China opened up to the West in 1842, the last dynastic rulers endeavored to modernize state machinery until overthrown in 1911. We use this unique episode to examine the progress of modernization of the state. China's large spatial extent entailed an uneven distribution of state power (Sng, 2014). We use this variation to compare the pace and level of state penetration between regions and the extent to which it was shaped by the mastery of local knowledge.

Accordingly, we conceptualize state building as the development of state power and infrastructure in local prefectures. For traditional state building before the opening-up in 1842, we employ the proportion of tax revenue levied from each prefecture between 1077 and 1820 as a proxy for the penetration of imperial fiscal power in the local area. We measure the state's infrastructural capacity in local areas by the number of government (Confucian) schools built between 1000 and 1800 and the number of imperial courier stations erected between 1281 and 1820 in each prefecture. The government schools were built to promote state orthodoxy (Confucianism) and prepare for imperial civil examinations, and the courier stations were official communication infrastructure. During modern state-building between 1842 and 1911, we focus on the development of new state infrastructure across prefectures, including the numbers of telegraph stations and post offices that replaced the traditional courier communications and the number of new schools that replaced classical education with Western curricula.

The challenge to estimating the effect of local knowledge is that gazetteer compilation is endogenized in the process of state building. Thus, the number of gazetteers may reflect the informational or compilation capacity of the state rather than the actual effect of local knowledge. To separate the effect of knowledge from its reciprocal relationship with state building, we adopt the following empirical strategies.

First, to mitigate the effect of feedback between state building and gazetteer compilation, we use the number of gazetteers compiled from the very beginning to predict the later development of the state. Meanwhile, we control for early state development before gazetteer compilation intensified in the tenth century, for which we use the presence of pre-imperial states and the density of walled cities (770–256 BCE) and the density of county divisions of early-imperial China (221 BCE–960 CE) as proxies. In addition, we control for the density of local scholars (examination degree holders) because as local gentry they led both local public affairs and also gazetteer

compilation. We also control for geographic characteristics (number of counties, terrain, and proximity to the coast and political centers), economic conditions (population density and agricultural suitability), and the treaty ports that imported Western technologies to China. The regression analyses show a significant and positive impact of the number of gazetteers on state building even after controlling for the above confounding factors. Gazetteers increased a prefecture's share of taxation in the empire and the pace of building orthodox education and state communication stations from 1000 to 1820. Moreover, the gazetteers continued to facilitate the modernization of state infrastructures during the transitional period of 1842 to 1911.

Second, we separate the effect of knowledge from the effect of capacity for compilation by comparing the impact on state building of prefectural gazetteers with that of county gazetteers in the same prefecture. A prefecture's prefectural gazetteers and county gazetteers are subject to similar compilation capacity and other prefectural factors but vary by the concretization of records of local knowledge, to the extent that county gazetteers delineate local characteristics in a more concrete way than the prefectural gazetteers whose records are relatively synoptic. We find that county gazetteers affect state building more profoundly than prefectural gazetteers. By the same token, we compare gazetteers compiled earlier with those compiled more recently for the same prefecture, in the sense that the recent records are more up-to-date and hence contain more 'useful knowledge' for governance than their preceding editions. We find that the effect of recent gazetteers on state building is greater than that of older compilations. These findings indicate that the gazetteer effect is primarily driven by local knowledge rather than by a prefecture's capacity for compilation.

Third, to further mitigate the confounding effects of unobserved local factors, we employ the number of gazetteers compiled by a prefecture's neighboring prefectures as the instrumental variable for the number of gazetteers of this prefecture. The numbers of gazetteers in adjacent prefectures are positively correlated, because these prefectures competed with and referenced each other when compiling gazetteers. Meanwhile, given that gazetteer knowledge is restricted to the defined boundary of administrative divisions, the performance of a prefecture's governance should have little to do with the knowledge recorded in the neighboring gazetteers. To mitigate the concern that some unobserved spatial similarities (in culture and geography, for example) might cause a spurious correlation between the neighboring gazetteers and local state building, we control for the macro-regions that are divided by common geographic and cultural characteristics according to Skinner, Henderson, and Yue (2013). In addition, we control for provinces to account for the similarities of prefectures within a province. The instrumental variable estimate reaffirms that the impact of local gazetteers on state building is significant.

Next, we assess the mechanisms by which the knowledge in local gazetteers affects state building. First, if the gazetteer effect arises from records of local knowledge, the progress of state building should depend on the scale and depth of these records. To this end, we use the number of topics in the contents of gazetteers of each prefecture to measure the richness of the knowledge records and find that it affects state building positively. Second, we further investigate the value of local knowledge by dividing the gazetteer topics into two types. One pertains to routine topics, referring to those commonly codified by central authority for regular governance affairs, notably the statistical records for households, arable lands, and public goods, among others. Unlike the simplified central records for these essential items, local gazetteers contextualized them with local characteristics. The other is about locality-specific topics that are hard for central authorities to formalize and codify. We find that the context-based routine topics predict state building significantly. More interesting is that the other locality-specific topics, despite their idiosyncracies, also play a comparable role to the routine topics in state building.

In sum, these results demonstrate that state capacity was more firmly established in regions where the governors had better local knowledge. Also, the gazetteer knowledge did not become outdated in the new era of state modernization but continuously mitigated the entry costs of modern state infrastructures. This implies that the useful knowledge for state building includes not only the homogenized central statistics but also the knowledge of heterogeneous local contexts. In this connection, our study speaks to the literature on how knowledge (and information technology) makes states (e.g., Lee and Zhang, 2017; Stasavage, 2021; Beraja et al., 2023; D'Arcy, Nistotskaya, and Olsson, 2024; Garfias and Sellars, 2024; Aman-Rana and Minaudier, 2025; Cheng, Stasavage, and Wang, 2025; Dzansi et al., 2025; Wang and Yang, 2025). While previous studies focus on the state knowledge centrally collected, codified and normalized, ours documents the unique value of idiosyncratic local contexts in increasing the state's legibility over the subjects (Scott, 1998, 2017). This also coincides with Balan et al.'s (2022) finding that local chiefs who possessed local knowledge outperformed state agencies in collecting taxes in Congo.

In a broader sense, our study enriches discussion of the determinants of state building and performance (see Besley et al. (2022) and Rohner and Zhuravskaya (2025) for reviews). This literature builds on the perspectives of natural endowments (Mayshar, Moav, and Pascali, 2022; Allen, Bertazzini, and Heldring, 2023, Fernández-Villaverde et al., 2023, and Flückiger et al., 2024), wars and conflicts (e.g., Tilly, 1990; Besley and Persson, 2010; Gennaioli and Voth, 2015; Becker et al., 2025), and organizational or policy innovations (e.g., Dal Bó, Finan, and Rossi, 2013; Aneja and Xu, 2024; Angelucci, Meraglia, and Voigtländer, 2024; Blanc and Kubo, 2025; Cantoni,

Mohr, and Weigand, 2024; Carlitz et al., 2025; Mastrorocco and Teso, 2025; Chiovelli et al., 2025). Specifically to China, studies have endeavored to explain the formation and performance of the centralized empire from the perspectives of geography (Sng, 2014; Fernández-Villaverde et al., 2023), warfare (Chen and Ma, 2020; Bai, Jia, and Yang, 2023; Chen, 2024), social organizations (Wang, 2022a, 2022b), administrative innovation (Hao and Liu, 2020; Li, 2023; Chen, Wang, and Zhang, 2025; Liu and Zhang, 2025; Zhang, 2025) and in particular the rise and fall of the imperial examinations (Bai and Jia, 2016; Jia and Kung, 2025). Our study shows that compiling gazetteers also contributed to the longevity of the imperial regime by helping to maintain coherent state policies across diverse regions.

2. Historical Background

2.1. The Demand for Local Knowledge

Imperial China had a large territory and population. To govern the large-cum-heterogeneous domain effectively, the imperial authorities developed the sophisticated bureaucratic system, in which the emperors delegated their power of governance to provinces, prefectures, and counties at the bottom, and they selected qualified officials by examination. Nonetheless, the cost of governance remained evident, especially when emperors assigned local officials from outside the regions they governed under the rule of avoidance (of forming corruption and localism). As a result, local officials usually had limited knowledge of their jurisdictions.

To overcome the knowledge gap, the imperial court provided the newly-appointed local governors with basic administrative information (taxation, households, judiciary, and public infrastructures, among others) in official manuals. These standardized sources were essential for regular administration but far from sufficient to meet the needs of managing complex local affairs. To complete their tenures satisfactorily and to stand out in the competition for promotion, local governors needed to comprehend local characteristics that were usually omitted from the simplified central compilations. Knowing local history, customs, social networks and anecdotes, for instance, might be pivotal for good governance, especially when the yardstick by which their performance was evaluated was social stability rather than growth in GDP (Balan et al., 2022; Man-Cheong, 2004). [1] However, these subjects were dispersed and rooted in local

[1] This is especially true for tax collection below the county level that relied on collaboration with the local gentry. The county magistrates had to know the distribution and strength of the gentry before they could successfully complete their taxation duties (Chang, 1955). Likewise, they also needed to know the potential risk of tax resistance and conflict by looking into the gazetteer

communities. It was hard for local officials to learn them from scratch, or even within their short tenure (normally three to five years) in a region.

Thanks to local gazetteers, various types of local knowledge were compiled into an accessible form. The gazetteer records of local characteristics are not only comprehensive but of high quality, since they were written and verified by renowned scholars from careful philological research and field work (Dennis, 2015). This made gazetteers the most systematic and accessible references for use in governance. In fact, browsing local gazetteers was often the first agenda of new governors, as they could quickly gain an overview of the area before investigating specific affairs in need of governance (Dennis, 2015).

Then, to what extent did the gazetteers facilitate local governance? This can be gleaned from historical narratives and anecdotes before we turn to the metrics. For example, local officials often referred to the gazetteer records on historical geography when developing public projects (roads, waterways and irrigation systems, for instance) (Chen and Wang, 2015). Similarly, the gazetteer records of school history offered useful information for governors to renovate and expand orthodox education.[2] And the records of property rights and court verdicts were usually referenced by magistrates as 'precedents' when judging cases (Dennis, 2015). These cases attest the view of the eminent gazetteer scholar Zhang Xuecheng (1738–1801), who posits that statecraft should be rooted in concrete studies of local conditions rather than simply relying on abstract moral principles. As he submitted in the masterpiece *Wen Shi Tong Yi* (Principles of Literature and History), of numerous historical compilations, only the gazetteers could provide systematic regional knowledge that the central records could not reach (Zhang, 1985).

When the dynasty was modernizing the state infrastructure in the second half of the nineteenth century, the gazetteers retained their reference value, because the modernization was enforced by local governors and grounded in local circumstance (Chang, 1955). Given the persistence (and resistance) of traditional forces, the mastery of local history and features became indispensable for the elite to push the reform appropriately and reconcile the new forces with traditional incumbents.[3]

---

volumes on the chronicle of events and folk custom.

[2] For example, the rebuilding of the famous White Deer Grotto Academy in Jiujiang Prefecture of Jiangxi Province was indebted to Zhu Xi's rediscovery of this abandoned academy from the local gazetteer in 1179, when Zhu Xi served as the prefect of Nankang (Chen and Wang, 2015). He developed the academy into a top Confucian school, where he synthesized the Neo-Confucian philosophies that were promoted as state orthodoxy later.

[3] For instance, when the Qing court reformed the traditional educational system in the 1900s, they required local schools to compile a treatise on local customs (*fengtu zhi*) from the gazetteers into the curriculum (Cheng, 2003). In 1896, Chen Baozhen (1831–1900) as provincial governor of Hunan used historical maps in the gazetteer to build a telegraph line linking the provincial capital

In addition to local governance, the gazetteers also increased the transparency of local areas and exposed more local knowledge to the imperial court. This is reflected in the compilation of the Unified Chorography in the Qing dynasty, when the emperors required prefectures and counties to submit their gazetteers as source books. Statistically, we find a highly positive correlation between the number of gazetteers in a prefecture and the frequency of mentions of this prefecture in the central compilations (Online Appendix Figure A2).

2.2. The Compilation of Local Gazetteers

Before examining the impact of local gazetteers on state building, we briefly introduce the history and process of gazetteer compilation. The earliest form of gazetteers can be traced back to the Land Record (*diji*) and later the Graphical Chorography (*tujing*) between about 220 and 960. They were mostly compiled by local scholars and aristocrats to record geographic information and were also used by the government to manage household registration.

The gazetteer did not become a comprehensive regional compilation until the Song dynasty (960–1279), thanks to the rise of the literate class and their efforts to promote local cultures through publishing gazetteers (Bol, 2001). Moreover, as imperial authorities increasingly recognized the values of gazetteers, they attempted to formalize the compilation of gazetteers as part of statecraft. This explains the unprecedented growth of gazetteers following the Song dynasty. The promotion of gazetteer compilation became more intense in the Ming-Qing period (1368–1911). Consequently, most local gazetteers in Chinese history were produced in this period.

Local gazetteers were compiled in a decentralized pattern. Local scholars and officials, rather than central agencies, decided when and how to compile their gazetteers despite instructions and promotions occasionally being provided by the imperial court. Renowned scholars were the actual compilers of gazetteers under the sponsorship of local officials and other elites. They compiled and updated gazetteers for counties and prefectures. The compilation usually took years and required considerable pecuniary investment and human capital (Dennis, 2015).[4] Consequently, the number and quality

---

Changsha to Hankou. These maps allowed him to route the wire to avoid fertile farmlands and ancestral graves (Thompson, 2015). Likewise, the town gentleman Luo Shiyang (1866–1931) in Guangdong Province is remembered for his contributions to reforming local taxation and school systems. His success in reform is attributed to his comprehension of local characteristics by reading gazetteers (Liang, 2007).

[4] For example, the compilation of the County Gazetteer of Shangyuan in Jiangsu Province in 1521 involved 57 participants who included scholars, clerical staff, and other auxiliary professionals (scribes, block-carvers, and printers). The compilation of the 1640 County Gazetteer of Jiangyin in Jiangsu Province cost 298 silver taels, which was nearly ten times the annual (nominal) salary of a

of local gazetteers varied between regions depending on economic and literacy levels. For instance, a resource rich region provided not only more recordable sources but also more financial and intellectual support for the compilation. As a result, the number of gazetteers was unevenly distributed across regions throughout Chinese history (Figure 1).

3. Data

Our sample includes 267 prefectures in China proper based on the prefectural divisions as of 1820 in the China Historical GIS (CHGIS, 2016) (Figure 1). Prefectural divisions changed little throughout the Ming-Qing period (1368–1911) that covers most of our outcome variables. For the data from earlier periods with different prefectural boundaries, we match them to the 1820 prefectural divisions based on Zhou's (2017) historical evolution of Chinese administrative divisions and the CHGIS (2016), on which we will elaborate when introducing the specific data below.

3.1. Local Gazetteers

Our primary measure for the accumulation of local knowledge is the number of local gazetteers. Given the prefectural unit of analysis, the gazetteers in a prefecture include both the prefectural gazetteers (*fuzhi*) and the gazetteers of the counties (*xianzhi*) therein. Literally, the prefectural gazetteer focuses on recording characteristics of prefectural significance; it also introduces the affiliated counties but in a summary manner. The detailed county characteristics are delineated in the county gazetteer. The prefectural and county gazetteers thus provide the most complete records of local knowledge at the prefectural level.[5] Therefore, we measure a prefecture's accumulation of local knowledge using the total number (titles) of prefectural and county gazetteers that had been compiled before the times of the variables of state building. The gazetteers include both those that are newly compiled and continuations of old versions. We exclude reprints that did not produce new knowledge.

We count the gazetteers from multiple sources. The primary one is *Zhongguo Difangzhi Lianhe Mulu* (Joint Catalogue of Chinese Local Gazetteers) compiled by Beijing Astronomical Observatory, Chinese Academy of Sciences (1985), which is the most complete collection of gazetteers compiled in historical China (4,392 in our sample). Some gazetteers, especially those compiled before the seventeenth century,

county magistrate at the time (Dennis, 2015).

[5] Some towns and villages also compiled their own gazetteers, but these gazetteers are not representative and mostly not available. We do not include them in our sample.

have been lost and not included in the Joint Catalogue. To this end, we retrieved them from historians' studies of the lost gazetteers; these include Zhang's (1962) *Zhongguo Gufangzhi Kao* (An Examination of Gazetteers in Ancient China) and Liu et al.'s (2011), Gu's (2010, 2012), and Lin and Zhang's (2001) dynastic-based investigations (i.e., for the Song, Liao, Jin, Yuan and Ming dynasties, respectively). These sources record 3,896 lost gazetteers. We added them to the Joint Catalogue data and finally got a total of 8,288 gazetteers compiled between 220 and 1842. We manually identified the region of each gazetteer and matched it to the 1820 prefectural divisions, based on the region name in the gazetteer title and the evolution of prefectural and county divisions (Zhou, 2017).[6]

Next, we identified the time of compilation, primarily using the information in the prefaces of gazetteers. For those without an explicit year of compilation, we infer it from the reigns of emperors, the dates of compilers' lives, and other useful information in the gazetteer. Still, such information is missing in 428 or 5.2% of gazetteers. We excluded them and focused on the remaining 7,860 gazetteers.[7]

The data shows clearly regional and temporal variations in the number of gazetteers (Figure 1). While the gazetteers were sporadically compiled before 960 (the Song dynasty), they expanded rapidly thereafter. By 1842, the gazetteers covered nearly all of China proper, but their number still varied between prefectures on a scale from zero to 118. We take the natural logarithm of the number of gazetteers to normalize the distribution and mitigate the impact of extreme values.

3.2. State Building

Our measures of state building extend from traditional imperial governance to state modernization following the opening up in 1842. For traditional China, we focus on examining the development of the state's fiscal capacity and infrastructural capacities as defined in economic literature (e.g., Besley and Persson, 2010; Acemoglu, García-Jimeno, and Robinson, 2015). For the modernization period (1843–1911), we focus on the modernization of state machinery under Western influence. Accordingly, we define the pre-1842 development as 'traditional state building', whereas the post-1842 modernizing efforts as 'state modernization'. We introduce the specific variables below.

[6] The regional bases of some gazetteers compiled before the fourteenth century may differ from those of the 1820 prefectural divisions. Fortunately, these gazetteers account for only 23.2% of all gazetteers in the sample. We also remove these gazetteers from analyses and find that this does not alter the main results significantly.

[7] Retaining the 428 gazetteers in the cross-prefectural regressions produces similar results, suggesting that these gazetteers without compilation information does not affect our estimation of the gazetteer effect.

3.2.1. Traditional State Building

Share of Tax Revenue. We measure the imperial fiscal capacity in a prefecture using the tax revenue levied from this prefecture as a percentage of the total tax revenue of all sample prefectures. It measures the relative importance of a prefecture to the state's taxation, with a higher percentage indicating a stronger presence of imperial fiscal power. These data are available for the years 1077, 1578, 1743, and 1820, which cover nearly eight centuries and three dynasties. The long time horizon allows us to observe the development of fiscal capacity throughout four dynasties corresponding to the expansion of local gazetteers. The tax revenue from 1578 to 1820 (in the Ming and Qing dynasties) are land tax based on the estimates of Wan and Xu (2015), Guo (2022), and Liang (2008).[8] The land tax before the Ming dynasty is unavailable. Instead, we use commercial tax in 1077 based on Xu's (1957) review of the *Song Hui Yao* (Collected Essential Documents of the Song Dynasty). For consistency in measurement, we use the percentage share of the tax each prefecture contributed to the dynasty. Doing so also rules out the confounding effect of different tax policies and statistics across dynasties.

Government Schools. Our first measure of the state's infrastructural capacity is based in the context of Confucian governance in imperial China. In the absence of modern technologies, imperial China relied heavily on moral suasion. For this purpose, the authorities persistently promoted Confucianism as state orthodoxy in view of its pacifist and obedient nature. Specifically, they employed Confucian classics as the curriculum for school education and the civil examinations. By doing so, they cultivated an educated elite who served as scholar-officials in the bureaucracy or remained in their hometowns as local gentry to manage local affairs (Chang, 1955; Elman, 2000).

Confucian education was institutionalized and enforced by government schools, that is, Prefectural Schools (*fuxue*) and County Schools (*xianxue*). The schools were not simply educational institutions used to admit and finance the candidates for civil examinations but also the headquarters of local Confucian Temples that had the function of diffusing state ideology. Normally, each prefecture or county had one government school, but the time of building the school varies between regions.

[8] Wan and Xu (2015) constructed prefectural land taxes of 1578 based on the *Wanli Kuaiji Lu* (Accounting Records in Wanli Emperor Reign). The data for 1743 and 1820 are respectively obtained from Guo (2022) and Liang (2008), who reviewed the tax records in the *Daqing Yitongzhi* (Unified Chorography of the Qing Dynasty) published in 1743 and 1820. The land tax in these sources refers to the land tax quota the central authority assigned to each prefecture. The actual tax levied each year may fluctuate but is not recorded systematically.

Therefore, we explore the time variation to measure the state's capacity to develop orthodox education in local areas. Specifically, we count the government schools existing in each prefecture every 50-year from 1000 to 1800, based on the records in the *Jiaqing Chongxiu Yitongzhi* (Unified Chorography of the Qing Dynasty) published in 1820. 1,683 government schools had been built by 1800, of which 97 percent were built after 1000 CE. This is consistent with historical narratives that the official Confucian education and examination systems did not expand significantly until the Song dynasty.

Imperial Courier Stations. Besides moral suasion, the dynasties also developed communications infrastructure to facilitate central control over local areas. The communication system mainly worked through the courier routes, a traditional postal system that delivered official documents and messages between local and central governments. Given the large size of imperial China, the courier routes shaped the geography of the states' informational and governance powers (Sng, 2014). As a result, the state had relatively stronger administrative power in regions that were better integrated into the courier network. This resonates with the importance of the postal roads in sustaining the crown's governance capacity in colonial Colombia (Acemoglu, García-Jimeno, and Robinson, 2015). Therefore, we employ the number of courier stations across prefectures over three time points (1281, 1587 and 1820) depending upon data availability as the second measure of infrastructural capacity.[9] As described by Sng (2014), the courier stations were still unevenly distributed across China by 1820, and the density of the stations is positively correlated with the degree of bureaucratic sophistication and land tax quota across prefectures.

3.2.2. State Modernization (1842–1911)

China's military defeats in the two Opium Wars implied the inability of traditional Confucian statecraft to compete with the Western powers. The Qing court thus initiated the Self-strengthening (or Westernization) movement to modernize governance technologies following the Western models from the early 1860s. This state-led modernization began with the adoption of Western industrial technologies, aiming to strengthen China's industrial and military power. Later, the dynasty realized the need to reform the old institutions. The most notable in this respect is the replacement of the Confucian education-cum-examination system with a Western-style modern

[9] The data in 1281 were obtained from Xiong's (2018) study of the courier routes in the Yuan dynasty recorded in the *Xijin Zhi* (Gazetteer of Xijin). The data in 1587 are from Yang's (2006) *Mingdai Yizhan Kao* (An Examination of the Courier Stations in the Ming Dynasty). Last, the 1820 data are from Liu's (2017) *Qingdai Yizhan Kao (*An Examination of Courier Stations in the Qing Dynasty*)*.

school system from 1904 (Kung, 2022). If state failure, argues Acemoglu and Robinson (2012), can be gleaned from the failure to remake the state in response to external shocks, China's Self-strengthening reforms provide a unique context to examine the dynasty's capacity to build a modern state, and the extent to which the regional penetration of the modern governance is still shaped by the accumulation of local knowledge.

Telegraph and Post Offices. To measure the modernization progress across China, we use the numbers of telegraph stations and post offices in each prefecture to capture the modernization of the traditional communication system. Moreover, we use the number of the new schools (that adopted Western curricula and institutions) to gauge the progress of educational modernization. For each variable, we aggregate the number between 1842 and 1911 at the prefecture level and regress it against the number of gazetteers compiled beforehand in a cross-sectional setting.

For the telegraph and post offices, we obtained the data from the *Jiaotongshi Dianzhengbian* (History of Transportation Telecommunications) that were digitized by Lin et al. (2021). The telegraph lines were mostly developed by the Qing government for military and administrative use. The modern post system originated from the private postal services of the Imperial Maritime Customs (managed by the British Robert Hart) between 1878 and 1895. During this time, the postal system provided limited services between 22 cities. It was not nationalized by the Qing government until 1896 when the Great Qing Post was established. Thereafter, the Qing expanded the postal system and established 1,171 post offices in 258 prefectures.

Modern Schools. Our modern school variable includes primary and secondary schools that constituted the main body of China's new school system at the time.[10] About 90 percent of these schools were funded or sponsored by the government (Imperial Educational Ministry, 2009 [1909]). We count the schools by prefectures based on the *Di San Ci Jiaoyu Tongji Tubiao* (Tables and Figures of the Third Education Statistics) compiled by the Imperial Ministry of Education in 1909. This is the last and also the most systematic school survey during the Qing dynasty, reflecting the progress of educational modernization since the abolition of the Confucian examination system in 1904. School modernization clearly exhibits regional patterns. While nearly all the sample prefectures had established modern schools by 1909, the number varied significantly from one to 1,273.

The above variables of state building are significantly and positively correlated with the number of local gazetteers at the prefectural level (see Figure 2). Specifically, prefectures contributed higher tax revenue, developed orthodox education earlier, and

[10] There were few tertiary educational institutes (colleges and universities) in China before 1911, and most of these institutes were run by the Christian missions rather than the Qing government.

built more courier stations if these prefectures had previously compiled more gazetteers. Further, these prefectures were more capable of developing new state infrastructures (telecommunications and modern schools) when the empire started modernization. These figures indicate the potential positive role of local knowledge in building local state capacity. Next, we will examine to what extent local gazetteers contribute to (rather than simply reflect) state capacity building, and to what extent the gazetteer effect arises from the preservation of local knowledge rather than the confounding correlates of gazetteer compilation.

3.3. Control Variables

Number of Counties. We control for the number of counties in each prefecture, simply because it directly determines the number of county gazetteers therein. Meanwhile, the number of counties reflects a prefecture's bureaucratic sophistication and hence the penetration of state capacity. We obtained the number of counties in each prefecture from the county atlases in CHGIS between 960 and 1911.[11]

Economic Conditions. The prosperous areas were more likely to be able to afford to compile more and higher-quality gazetteers. To mitigate the concern that the effect of gazetteers is driven by economic prosperity, we control for the latter using population density as the proxy.[12] In addition, we control for the average suitability index for planting major crops (wheat, rice, sweet potato, and maize) based on the Global Agro-Ecological Zones (GAEZ) database of the Food and Agriculture Organization of the United Nations (2002), reasoning that the suitability reflects potential agricultural productivity in each prefecture. When examining the gazetteer effect during the modernization period (1842–1911), we additionally control for a dummy variable indicating whether a prefecture had treaty ports, in the light of the leading role of these ports in diffusing Western technologies in China (Jia, 2014).

Geographical Factors. First, we control for the terrain ruggedness index using data from U.S. Geological Survey (2012), since geographic barriers affect the costs of governance. By the same token we control for the shortest distance from a prefecture's centroid to the coastline. Second, we control for the shortest distance to the imperial

[11] In the following panel data regressions, the number of counties varies by the year of the outcome variables. In the cross-sectional regressions, we take the average number of counties from 960 to 1820 (or 1842).

[12] The population data were obtained from Le (1936) and Cao (2024). Le provides the prefectural population in 976 and 1102, and Cao provides the prefectural population in 1393, 1589, 1630, 1650, 1680, 1776, and 1820. In our panel regressions, we control for the population figures of the nearest time to that of the dependent variable. For cross-sectional regressions, we control for the average population between 976 and 1820 in each prefecture.

capital and the shortest distance to the provincial capital (i.e. political centers for the sake of clarity), since the state control over local areas could be reduced by geographic distance from the power centers (Besley and Persson, 2010; Sng, 2014).[13] Online Appendix Table A1 summarizes the statistics for the main variables used in our analysis.

4. The Impact of Local Gazetteers on State Building

In this section, we assess the effect of local gazetteers on state building. To mitigate reverse causality, our primary strategy is to use the historical accumulation of gazetteers to predict the subsequent state building. Meanwhile, we control for the gamut of geographic and economic factors as introduced in Section 3 and in addition control for the level of early state development before the gazetteer compilation intensified in the 10th century, and the number of local scholars who dominated both gazetteer compilation and local public affairs. Second, to further disentangle the effect of local knowledge from the confounding effect of compilation capacity, we compare the gazetteers that are subject to the similar compilation capacity of the prefecture but vary by the level of concretization in knowledge records. Lastly, since we cannot rule out all confounding factors, we employ an instrumental variable approach to approximate the exogenous effect of local gazetteers.

Given the long time span of traditional state building, we examine the gazetteer effect on it in a panel data setting specified in Equation (1):

$$Y_{it} = \alpha + \beta \sum_{220CE}^{t} \text{Gazetteers}_{it} + \Gamma \mathbf{X}_{it} + prefecture_i + year_t + \epsilon_{it} \qquad (1)$$

where Y_{it} denotes the variables of state building in prefecture i at time t. They include the share of tax revenue in 1077, 1578, 1743, and 1820, the number of government schools every 50 years between 1000 and 1800, and the number of imperial courier stations in 1281, 1587, and 1820. The main explanatory variable, $Gazetteers_{it}$, is the number of prefectural and county gazetteers that had been compiled before each time of the dependent variable. It measures the accumulation of local knowledge in each prefecture. We control for both prefectural and year effects. $\mathbf{X}_{it}$ is a vector of the

[13] The distance to political centers varied over time because of changes in the imperial and provincial capitals. Based on the data in CHGIS, we compute the distances on a dynastic basis (i.e., Song (1080), Yuan (1330), Ming (1578), and Qing (1820)) for the panel data regression. For cross-sectional estimates, we employ a time-invariant measure by using the average of the distances across the four dynasties for each prefecture.

control variables introduced above. For the time-invariant controls, we interact each of them with the time dummies τ_t.

On the variables of state modernization between 1842 and 1911, we examine the effect of local gazetteers compiled before 1842 in a cross-sectional setting, based on the specification:

$$Y_i = \alpha + \beta \sum_{220CE}^{1842} \text{Gazetteers}_i + \Gamma \boldsymbol{X}_i + \epsilon_i \qquad (2)$$

where Y_i measures state modernization at the prefectural level using the numbers of telegraph stations (1881–1911), post offices (1896–1911), and modern schools (1909). Gazetteers_i is now measured by the number of prefectural and county gazetteers compiled between 220 and 1842 in each prefecture. The control variables in $\mathbf{X}_i$ are the same as those in Equation (1) except they are organized at the cross-prefectural level and include also the treaty port dummy. In addition, we control for the provincial effects, in the sense that Qing's modernization efforts were largely undertaken by provincial governors and thus have a provincial pattern (Feuerwerker, 1958; Spence, 1990).

4.1. The Main Results

Table 1 reports the OLS estimates of the gazetteer effect conditional on other prefectural factors. The number of local gazetteers significantly predicts the subsequent state building. Columns 1 to 3 show a positive effect of local gazetteers on the development of the dynasty's fiscal and infrastructural capacities at the prefectural level. Approximately, doubling the number of gazetteers increases the share (percentage points) of a prefecture's tax revenue in the total revenue (of all 266 prefectures) by 29 percent when evaluated by the sample mean (0.404 × 0.725). Similarly, doubling the number of gazetteers increases the number of government schools by about 1.16 or 35 percent against the sample mean and the number of courier stations by 1.82 (or 43 percent).

The gazetteers continuously predict the progress of state modernization after 1842 (columns 4 to 6 of Table 1). In prefectures that had previously compiled more gazetteers, telegraphs, post offices, and modern schools were more likely to be established between 1842 and 1911. Doubling the number of gazetteers is associated with 0.6 more telegraph stations and 1.52 more post offices, which translate into 42 percent and 35 percent increases against their sample means, respectively. The effect on the modernization of Confucian education becomes even greater; the number of

modern schools increased by 71 percent as a result of doubling the number of gazetteers in each prefecture.

In addition to the number of gazetteers, we use the number of volumes and pages of all gazetteers in each prefecture as alternative measures of local knowledge accumulation. The volumes and pages can measure the size of gazetteer records although there is limited data. The volume information is only available for the gazetteers compiled in the Qing dynasty (1644–1842) based on the Joint Catalogue of Chinese Local Gazetteers. We count the volumes by prefectures for all the 3,543 gazetteers compiled in this period (which account for 45 percent of all gazetteers before 1842). We do the same for the pages, which are available in the 2,085 gazetteers which were compiled between 1368 and 1842 and digitized in the Diaolong Full-text Database of Chinese and Japanese Ancient Books. The volumes and pages cannot cover the full gazetteers but form representative sub-samples, in the light of their high correlation with the number of gazetteers of the full sample across prefectures.[14] We find that both the volumes and pages of gazetteers have a significant and positive impact on the variables of state building (Online Appendix Table A2).

4.2. Controlling for Early State Development and Local Scholars

In this subsection, we further control for the early state development (before gazetteer compilation), since it may affect a prefecture's capacity for gazetteer compilation and also persistently affect state-building in the long run. Another concern is the number of local scholars, who led the compilation of gazetteers and meanwhile affected local state capacity by actively engaging in various public affairs under the county level.

We define early state development as that before 960, i.e., the beginning of the Song dynasty when gazetteer compilation intensified (see Online Appendix Figure A3). We employ three proxies for early state development at the prefectural level. The first pertains to bureaucratic sophistication, which is measured by the average density of the county-level administrative units from the first empire (Qin, 221–206 BCE) to 960 CE based on the CHGIS. This measure resembles Acemoglu et al. (2015) in using the presence of state functionaries to capture state capacity in local areas. The other two variables can be traced back to the pre-imperial origins of Chinese states based on archaeological records, since the geography of pre-imperial state development could persistently shape the regional disparities in political-economic development in the long

[14] The correlation coefficient between the number of gazetteers of the volume sample and the number of gazetteers of the full sample across prefectures is 0.9 and significant at the 1% level. The correlation between the Diaolong digitized gazetteers and the full sample of gazetteers is 0.66 and significant at the 1% level.

run (Chen and Wang, 2024). Specifically, we constructed a dummy indicator for pre-imperial state presence using the state borders in circa 350 BCE according to Herrmann (1935). To further capture the level of early state development, we use the density of 'cities' (settlements with city walls) from circa 770–256 BCE based on Xu (2017). Given the high correlations between the three measures, we use their first principal component as a proxy for the level of early state development.

Local scholars in imperial China refer to civil examination degree holders. As a cultural and political elite, they were promoted as local gentry to lead various public affairs (typically taxation, conflict resolution, and provision of public goods) in towns and villages to which the state bureaucracy rarely reached. We measure the size of this elite using the number of the highest exam degree holders (*jinshi*) produced in each prefecture. Although many *jinshi* holders were appointed out of town, their numbers reflect the local educational level and thus approximately the number of local scholars. Specifically, we use the cumulative number of *jinshi* holders conferred from 960 to the times of our dependent variables to measure the strength of local scholars, and we normalize the number by prefectural population.[15]

Table 2 shows the results after controlling for the early state development and the size of local scholars. Both have a positive effect on the development of traditional state capacity (columns 1–3), an effect consistent with historical accounts of the importance of the early states and the civil examination elite in imperial governance (Chang, 1955). However, their effects fade out and even become negative during the period of state modernization (columns 4–6), possibly because the traditional bureaucratic and educational systems could not adapt to (and thus were replaced by) the new governance technologies under Western influence. However, the effect of local gazetteers on state building remains significantly positive from the traditional to the modernizing periods. These results indicate that gazetteer knowledge may not be as outdated as the other traditional governance tools for state modernization and, more importantly, the gazetteer effect is unlikely to be driven by early state development and the size of the educated elite, albeit their measurement is still rough.

4.3. Comparing Different Gazetteers in the Same Prefecture

To further separate the knowledge effect from the confounding effect of the capacity for compilation, we compare different types of gazetteer that varied in the usefulness

[15] The data on the *jinshi* degree holders in the Ming-Qing period (1371–1840) were obtained from Zhu and Xie (1980), and the data for the previous Yuan, Liao, Jin, and Song dynasties from Shen (2019); Gao (2015); Xue (2004); Gong and Zu (2014), respectively. The population data were obtained from Le (1936) and Cao (2024). We take the average population (in millions) of available years between 960 and the time of the dependent variable.

of knowledge records but were compiled in the same prefecture and hence were subject to similar prefectural factors including compilation capacity.

We first compare the prefectural gazetteers and the county gazetteers in terms of their effects on state building. In a prefecture, the compilation of the county and prefectural gazetteers is shaped by similar local factors. As shown in Online Appendix Table A3, the number of county gazetteers and the number of prefectural gazetteers are consistently correlated with the same set of prefectural variables, in particular economic conditions, proximity to political centers, and the size of local elite that illuminate local recording capacity.

However, they recorded local knowledge at different levels of concretization (i.e., quality, depth and informativeness of the text). The county gazetteers usually record the most systematic and specific local characteristics below the county level, whereas the prefectural gazetteers focus on the prefecture-level characteristics and omit much information about its sub-regions. For instance, while a county can be represented by 501 pages in its gazetteer, this number is reduced to only 151 pages in the prefectural gazetteer (calculated based on the sample mean). If our gazetteer measure captures the effect of local knowledge rather than the prefecture's capacity for compilation, the county gazetteers should outperform the prefectural gazetteers in affecting state building. Otherwise, if the capacity for compilation matters, the prefectural and county gazetteers should have comparable effects on state building.

To test this, we horse race the number of county gazetteers and the number of prefectural gazetteers in each prefecture on their impact on state building (Table 3). The county gazetteers have a significantly positive impact on state building. However, the effect of prefectural gazetteers is much smaller and is statistically insignificant. These results indicate that the effect of gazetteers arises from the record of knowledge rather than local state capacity for gazetteer compilation. They also indicate that concrete local knowledge is more valuable than summary records for building state capacity.

In a similar vein, we also compare the gazetteers compiled earlier to those compiled more recently in the same prefecture, reasoning that the recent compilations contain updated and more useful knowledge for governance. However, given the high cost of a compilation project, the compilation and updating of gazetteers was never a regular practice but was undertaken in a nearly haphazard way at the will and preference of the local elite. On average, the update of prefectural and county gazetteers happened every 87 years during the Ming-Qing period. The new gazetteers are indeed more informative than the old. For instance, the average number of topics and pages of the gazetteers compiled in the Qing period (1644–1842) is 1.36 and 1.25 times those of Ming (1368–1643).

To test the advantage in knowledge of the new gazetteers, we divide the gazetteers by their period of compilation. For the panel regressions of traditional state building, we compare the number of gazetteers compiled in the most recent century to those of the previous centuries. On cross-sectional regressions of state modernization after 1842, we compare the effects of the gazetteers compiled in the recent mid-to-late Qing period (1744–1842) to those of the longer past, i.e., early Qing (1644–1743), Ming (1368–1643), and pre-Ming periods. Figure 3 shows that the gazetteers compiled more recently have a greater impact on state building than the earlier compilations, demonstrating the value of local knowledge for contemporaneous governance.[16]

4.4. Neighboring Gazetteers as the Instrument

Our last resort to isolate the effect of gazetteer knowledge from unobserved prefectural factors is to use an instrumental variable (IV) to predict the exogenous variation in the number of gazetteers. We employ the total number of gazetteers compiled in a prefecture's neighboring (adjacent) prefectures as the instrumental variable for the number of gazetteers of this prefecture (see Panel A of Figure 4 for an illustration). We justify the validity of this instrumental variable as follows.

First-stage Relationship. The number of gazetteers in neighboring prefectures is positively correlated (see Panel B of Figure 4). The correlation may arise from inter-prefectural competition for cultural and political status as well as cross-referencing during gazetteer compilation. First, given the state's promotion of compiling gazetteers, local governors would feel pressure and have an incentive to compile or update their gazetteers in accordance with their neighboring prefectures' performance in this respect. Given the centralized personnel system, falling behind in the gazetteer undertaking would risk a local governor's performance evaluation and promotion prospects. Consequently, the gazetteer was not only a reference book but also a competitive arena for local officials (Man-Cheong, 2004).[17]

Besides officials, local scholars also competed for regional cultural hegemony by compiling gazetteers. The early rise of local gazetteers in the Song dynasty, for instance, was allegedly rooted in the rise of 'localism', during which scholars sought regional

[16] That said, this interpretation is premised upon a constant capacity for complication of a prefecture. It is likely that the continuous compilations and updates of gazetteers reflect the growth of compilation capacity in a prefecture. We interpret the results of the temporal comparison with caution.

[17] Such a peer effect can be gleaned from a case of imperial inspection in Jiangxi Province in the sixteenth century. When an imperial commissioner pointed out that the Prefectural Gazetteer of Nan'an was incomplete, governors of the neighboring Ganzhou prefecture quickly responded to this news by revising their own prefectural gazetteer (Prefectural Gazetteer of Nan'an, 1536, Preface).

cultural identity and reputation by compiling their own gazetteers (Bol, 2001). This inter-regional cultural competition persisted to the Ming-Qing period. By systematically recording a region's history and achievements, the gazetteer was promoted as a cultural symbol of the region and used to project an image of local standing to the audience of the empire (Dennis, 2015).

Second, the gazetteers of neighboring prefectures, if any, provided valuable references for compiling new gazetteers. As a comprehensive cultural project, gazetteer compilation relied on rich reference materials including existing gazetteers as templates. Given the limited communications in pre-industrial society, neighboring gazetteers provide one of the most accessible sources of reference. In this sense, a prefecture was more likely to publish more and higher-quality gazetteers if there had been more reference sources available (Dennis, 2015).

Exclusion Restriction. Gazetteer knowledge is arguably orthogonal to the state building of neighboring prefectures, simply because gazetteer knowledge has clear 'boundaries' based on the administrative divisions. A prefecture's gazetteers almost exclusively introduce characteristics specific to this prefecture rather than the neighboring ones. Consequently, state capacity in a prefecture should have little to do with the quantity and quality of gazetteers compiled in the neighboring prefectures.

The exogeneity of the number of neighboring gazetteers is supported by a horse-race with the number of local prefecture's gazetteers in affecting state building (see Online Appendix Table A4). To provide a benchmark, we first examine the reduced-form relationship between the number of neighboring gazetteers and the variables of state building in local prefectures, which is significant and positive. Then, after we include the number of local prefecture's gazetteers into regressions, the effect of the neighboring gazetteers decreases greatly and becomes statistically insignificant on most state building outcomes. The effect of local gazetteers, however, remains significant and positive. These findings suggest that the effect of the instrumental variable on state building operates primarily through affecting the number of local gazetteers. Certainly, these results are suggestive rather than conclusive since both the local and neighboring gazetteers are still endogenous in the regressions. Nonetheless, the horse race results lessen the concern that neighboring gazetteers spuriously affect local state building.

The Instrumented Results. We report in Table 4 the two-stage least squares (2SLS) estimates, in which the number of neighboring gazetteers predicts the number of local gazetteers in the first stage. We control for the same set of prefectural factors as those previously included in the OLS estimation. The statistical significance of the coefficient and the F-statistics of the first-stage results show a strong correlation between the IV and the number of local gazetteers. The second-stage results show a

significantly positive impact of the instrumented number of local gazetteers on state building. And the magnitude of the coefficient is somewhat greater than the corresponding OLS ones.[18] The instrumented results remain robust to controlling for early state capacity and local scholars (Online Appendix Table A5).

To further mitigate the possibility that the neighboring gazetteers are correlated with unobserved factors of state building in the local prefecture, we use the number of neighboring gazetteers compiled in earlier periods (arbitrarily 50 to 200 years before the times of state building). We assume that the earlier gazetteers in neighboring prefectures could still provide incentives and references for the compilation of gazetteers in the local prefecture, but they should be less likely to affect the contemporary state building of other prefectures because of the time gap. We find that the effect of local gazetteers remains robust to using these lagged instrumental variables (Online Appendix Table A6).

Another concern arises from changes in the prefectural boundary over time. This may lead to overlaps in gazetteer knowledge between prefectures and hence erode the exclusion restriction of the IV.[19] Fortunately, 76% of gazetteers in our sample were compiled during the Ming-Qing period when prefectural boundaries changed little. For robustness, we restrict the analysis to the Ming-Qing period and obtain consistent results with those of the full sample (Online Appendix Table A7). This implies that boundary changes did not bias the gazetteer effect significantly.

Last but not the least is that the neighboring gazetteers may affect a prefecture's state capacity through unobserved spatial correlations. For instance, adjacent prefectures usually share some common geographic and cultural characteristics, and these common characteristics may induce a spurious correlation between the number of neighboring gazetteers and local state building. To address this concern, we control for spatial similarities using the macro-region fixed effects and cluster standard errors at the macro-region level. The macro-regions are defined by Skinner, Henderson, and Yue (2013), who divides China proper into 18 macro-regions based on similarities in physiographic, cultural, and socioeconomic fundamentals. Accounting for the spatial similarities does not change the significant effect of local gazetteers (see Online Appendix Table A8).[20] Moreover, we also examine the spatial autocorrelation of

[18] The only exception is that the effect of gazetteers on the telegraph is positive but statistically insignificant. A possible reason is the small regional variations in the number of telegraph stations before 1911 (only 1.44 stations per prefecture). The telegraph network expanded to more prefectures and counties for administrative purposes after 1911 (Lin et al., 2021). We find that the instrumented gazetteer effect becomes significant when we extend the sample period to 1927 (not reported).

[19] For example, if some counties of a prefecture were previously affiliated with a neighboring prefecture, the gazetteer of this neighboring prefecture compiled at the earlier time would cover part of the local prefecture.

[20] Our sample prefectures cover 16 macro-regions. The results are based on cross-prefectural

standard errors between neighboring prefectures using the method of Colella et al. (2023). We cluster standard errors within 131 kilometers of the spatial centroid of each prefecture, a radius that roughly covers the neighboring prefectures. The spatially-clustered standard errors do not mitigate the significant effect of local gazetteers effect (Online Appendix Table A9).

5. Testing the Knowledge Effect of Local Gazetteers

In this section, we investigate the content of the gazetteers to substantiate the knowledge mechanism by which the gazetteers facilitated state building. We begin by examining the effect of the richness of the gazetteer knowledge based on the diversity of contents in the gazetteers. Further, we provide evidence for the value of local knowledge by distinguishing the local-specific topics from the common or routine topics recorded in the gazetteers.

5.1. The Richness of Local Knowledge

We measure the richness using the total number of unique topics covered by the gazetteers. Unlike the number, volumes and pages of gazetteers which capture the magnitude of knowledge, the number of topics reflects the diversity and comprehensiveness of it. We define topics according to the chapter titles given in the gazetteer content lists. Content lists are available for 2,085 prefectural and county gazetteers compiled during the Ming-Qing period (1368–1842); and these gazetteers are digitized in the Diaolong Full-text Database of Chinese and Japanese Ancient Books. We identified a total of 10,080 unique topics with an average of 296 per prefecture.[21] Given the limited temporal variation in topics during Ming-Qing period, we organize the data in a cross-sectional setting and regress state building in or after 1820 against the number of topics in the gazetteers compiled between 1368 and 1820.

Since some topics with different titles describe the same or similar subject matter, simply counting the raw topics may overestimate the degree of richness of knowledge. To address this concern, we refine the definition of unique topics by clustering

estimates for the nineteenth century, because the macro-regions are defined based on the situation in nineteenth-century China. And the macro-region fixed effects do not apply econometrically to the panel data regressions given that the prefectural fixed effects have already been included.

[21] The dataset can be accessed via http://hunteq.com/ancientc/ancientkm. As introduced in subsection 4.1, the Diaolong digitized gazetteers cover 67% of the 267 prefectures of China proper. The cross-prefectural distribution of the number of Diaolong gazetteers is highly correlated with that of the full sample of gazetteers (correlation coefficient is 0.7 and significant at 1% level). This indicates that the selection of the Diaolong gazetteers is not biased. We drop 88 prefectures that are not covered by the Diaolong database.

synonymous topics. Specifically, we calculate the cosine similarity between topics using OpenAI's text embedding method, i.e., translating all topic texts into numerical vectors and compare their semantic meaning mathematically (in the context of Chinese historical gazetteers). We cluster topics whose mutual cosine similarity scores are above 0.7 as synonyms and hence count them as one clustered topic. We obtained 5,664 clustered topics from the original 10,080 raw topics. We use the number of clustered topics as an alternative measure for the richness of local knowledge.[22]

Table 5 presents the OLS estimates for the effect of topics conditional on the same set of controls as that of Table 1. The effect of the topics is consistently positive and significant in explaining tax revenue, educational and communication infrastructures, and the modernization of state's infrastructure after 1842. In terms of magnitude, a one-standard-deviation increase (about 182) in the number of raw topics is associated with an increase of 15 percent in the share of land tax, nearly 0.67 government schools and 1 courier stations by 1820, and further 0.38 telegraph stations, 0.66 post offices, and 11 percent of modern schools in the modernization period. We omit the IV estimate since the topic data are drawn from a subsample of (digitized) gazetteers and hence cannot be predicted precisely by the number of neighboring gazetteers.

5.2. Routine and Non-routine Knowledge

The value of local gazetteer lies not only in the encyclopedic content but also in the unique local knowledge that is usually overlooked in other official compendia. In this section, we further discern the unique local knowledge from the topics of gazetteers and examine their values for state-building.

For the sake of clarification, all topics in the gazetteers are about local knowledge. But these topics differ in the degree of localization and uniqueness. We divide the gazetteer topics into approximately two groups: one pertains to the routine knowledge—to borrow the concept of organizational economics (Nelson and Winter, 1982), which refers to subjects standardized and codified for government use, such as statistics on households, taxation, personnel records, arable land, public facilities, and other administrative routines. These routine subjects can also be accessed from central compilations (for example, unified tax books and household records), but these records are highly summarized and uniform. The gazetteers contextualize these items with rich

[22] We set the threshold at 0.7 to balance dimensional space and semantic precision. A lower threshold, for instance, 0.5, would group topics that have loose thematic associations rather than real synonyms. Setting a higher threshold, however, would yield more fragmented clusters that exclude many synonymous raw topics.

local information. For example, the chapters of taxation in the gazetteers usually elaborate on land tax by introducing land endowments and types, and they also detail a complete set of other locality-specific tax sources (e.g., taxes on salt, tea, fishing, and miscellaneous levies).

The other group includes topics of non-routine knowledge, which are entrenched in local contexts and introduce idiosyncratic and nuanced material that is hard to normalize in the central compendium.[23] Therefore, the non-routine topics constitute more localized knowledge than the routine topics; and more importantly, the gazetteers systematically compiled these idiosyncratic topics and made them accessible.

To distinguish routine from non-routine topics, we refer to the content of the *Jiaqing Chongxiu Yitongzhi* (Unified Chorography of the Qing Dynasty) compiled by the imperial court in 1820. By 'Unified' it means that it records the standardized common information across all administrative divisions for reference by central authority. In this sense, the topics in the Chorography largely covered the routine knowledge of interest to the emperor at the time. There are a total of 25 topics listed in the Chorography. We employ these topics to distinguish routine from non-routine topics in local gazetteers.[24]

From the 2,085 digitized gazetteers in the Diaolong database, we find 4,298 raw topics corresponding to the 25 routine topics defined by the Chorography. The remaining 5,782 raw topics in local gazetteers, which differ from any topics in the Chorography, are counted as non-routine knowledge (see Online Appendix Table A10 for the 25 routine topics and examples of non-routine topics). We use the number of routine topics and the number of non-routine topics in all gazetteers of a prefecture to measure the accumulation of the two types of knowledge, respectively. Given the high correlation between the numbers of the routine and non-routine topics, we cannot directly compare their effects in the same regression. Instead, we examine their effects separately.[25]

---

[23] Most non-routine knowledge is about local characteristics and specialties. For example, some coastal regions introduce their harbors, sea transport and communications with foreign cultures in the gazetteers; southwestern frontier regions introduce their chieftains as traditional local elite; and regions rich in herbal medicine, minerals and other endowments usually detail these characteristics in the gazetteers.

[24] Alternatively, we also used the items in another central compilation, namely the *Daoren Xuzhi* (Handbooks for Newly Appointed Officials) collected in the *Ming Huidian* (Statutes of the Ming Dynasty) as the reference to define the routine knowledge topics. The results (not reported) are consistent with the Unified Chorography.

[25] We check the robustness of the manual identification of non-routine topics using the text-embedding method. We compute the cosine similarity between each gazetteer topic and each of the 25 routine topics in the Chorography. We define a gazetteer topic as non-routine if the maximum of the 25 similarity scores of this topic is less than the mean of the maximum score of all the 10,080 gazetteer topics. This approach identifies 6,176 non-routine topics. Their prefectural distribution is

The results in Table 6 (Panel A) show the significant and positive impact of the routine knowledge on state building. This is straightforward since most routine items address essential governance needs and provide rich local context. More interesting is that the number of non-routine topics also positively and significantly affects state building (Panel B). Moreover, the magnitude of the effect of non-routine knowledge is greater than that of routine knowledge. These findings suggest that the various local-specific records in the gazetteers were not dispensable miscellanies but provide useful references for state building. This adds more credence to the value of local knowledge through which the gazetteers contributed to state building.

6. Conclusion

This paper shows that the local development of state capacity in a region depends on the extent to which the state has knowledge about this region. Thanks to the encyclopedic gazetteers the local elite compiled to specify regional characteristics in historical China, we were able to use the number of gazetteers to quantify the accumulation of local knowledge. We find that prefectures that compiled more gazetteers contributed higher tax revenue and built more Confucian schools and courier stations in traditional China. When China endeavored to modernize the state machinery, the regional progress of modernization was still shaped by the stock of local gazetteers.

We document that the gazetteer effect arose from the local knowledge recorded in them, in light of the findings that the pace of state building in a region is positively impacted by the richness, concretization, and uniqueness of knowledge records. We find the knowledge effect robust after we control for the factors underlying the local capacity for compilation and explore the exogenous variation in the accumulation of local knowledge using the number of neighboring gazetteers as the instrument.

States are built on the acquisition of useful knowledge. While the acquisition has increasingly relied on simplification and codification thanks to the development of information technology, we show that the local expertise is still indispensable for building state capacity, since it contextualized the simplified codes and statistics into local specifics and thus mitigated the information costs between the state and the heterogeneous localities. Our findings also shed light on why imperial China persistently invested in compiling and recording gazetteers, especially when the empire became more unified and centralized following the fourteenth century.

highly correlated with that of the manual division (with a correlation of 0.99). And they also predict variables of state building positively (see Online Appendix Table A11).

References

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *American Economic Review*, 105 (8): 2364–2409.

Acemoglu, Daron, and James A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty.* New York: Crown Publisher.

Allen, Robert C., Mattia C. Bertazzini, and Leander Heldring. 2023. "The Economic Origins of Government." *American Economic Review*, 113 (10): 2507–2545.

Aman-Rana, Shan, and Clement Minaudier. 2025. "Spillovers in State Capacity Building: Evidence from the Digitization of Land Records in Pakistan." Working Paper.

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.

Angelucci, Charles, Simone Meraglia, and Nico Voigtländer. 2024. "Organizing a Kingdom." NBER Working Paper No. 32542.

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.

Bai, Ying, Ruixue Jia, and Jiaojiao Yang. 2023. "Web of Power: How Elite Networks Shaped War and Politics in China." *Quarterly Journal of Economics*, 138 (2): 1067–1108.

Balan, Pablo, Augustin Bergeron, Gabriel Tourek, and Jonathan Weigel. 2022. "Local Elites as State Capacity: How City Chiefs Use Local Information to Increase Tax Compliance in the Democratic Republic of the Congo." *American Economic Review*, 112 (3): 762–797.

Becker, Sascha O., Andreas Ferrara, Eric Melander, and Luigi Pascali. 2025. "Wars, Taxation and Representation: Evidence from Five Centuries of German History." *Journal of the European Economic Association*, 23 (6): 2241–2288.

Beijing Astronomical Observatory, Chinese Academy of Sciences. 1985. *Zhongguo Difangzhi Lianhe Mulu (Joint Catalogue of Chinese Local Gazetteers).* Beijing: Zhonghua Shuju.

Beraja, Martin, Andrew Kao, David Y. Yang, and Noam Yuchtman. 2023. "AI-tocracy." *Quarterly Journal of Economics*, 138 (3): 1349–1402.

Besley, Timothy, Robin Burgess, Adnan Khan, and Guo Xu. 2022. "Bureaucracy and Development." *Annual Review of Economics*, 14 (1): 397–424.

Besley, Timothy, and Torsten Persson. 2010. "State Capacity, Conflict, and Development." *Econometrica*, 78 (1): 1–34.

Blanc, Guillaume, and Masahiro Kubo. 2025. "The Making of France." Working Paper.

Bol, Peter K. 2001. "The Rise of Local History: History, Geography, and Culture in Southern Song and Yuan Wuzhou." *Harvard Journal of Asiatic Studies*, 61 (1): 37–76.

Cantoni, Davide, Cathrin Mohr, and Matthias Weigand. 2024. "The Rise of Fiscal Capacity: Administration and State Consolidation in the Holy Roman Empire." *Econometrica*, 92 (5): 1439–1472.

Cao, Shuji. 2024. *The Population History of China (1368–1953)*. Leiden: Brill.

Carlitz, Ruth, Ameet Morjaria, Joris Mueller, and Philip Osafo-Kwaako. 2025. "State Building in a Diverse Society." *Review of Economic Studies*, 92(6): 3704–3740.

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

Chen, Joy. 2024. "State Formation and Bureaucratization: Evidence from Pre-Imperial China." *Journal of Economic History*, 84 (3): 690–726.

Chen, Joy, Erik H. Wang, and Xiaoming Zhang. 2025. "From Powerholders to Stakeholders: State Building with Elite Compensation in Early Medieval China." *American Journal of Political Science*, 69 (2): 607–623.

Chen, Shuo, and Debin Ma. 2020. "States and Wars: China's Long March Towards Unity and Its Consequences, 221 BC–1911 AD." CEPR Discussion Paper No. DP15187.

Chen, Xi, and Zhongjing Wang. 2015. "Nansong Difangzhi Yu Difang Zhengwu (Local Gazetteers and Local Governance in the Southern Song Dynasty)." *Zhongnan Minzu Daxue Xuebao (Renwen Shehui Kexue Ban)* , 35 (4): 176–180.

Chen, Zhiwu, and Wanda Wang. 2024. "Unraveling Seven Millennia of Change: What Reversed China's Enduring Neolithic Economic Geography?" SSRN Working Paper No. 4767886.

Cheng, Cheng, David Stasavage, and Yuhua Wang. 2025. "The Written Word and the Development of the State in China and Europe." SSRN Working Paper No. 5241327.

Cheng, Meibao. 2003. "You Aixiang Er Aiguo: Qingmo Guangdong Xiangtu Jiaocai De Guojia Huayu (To Love My Native Place, to Love My Country: The National Discourse on Native Place Textbooks in Late Qing)." *Lishi Yanjiu*, 4: 68–84.

CHGIS, Version 6. 2016. Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University. Available online at https://dataverse.harvard.edu/dataverse/chgis_v6.

Chiovelli, Giorgio, Leopoldo Fergusson, Luis Martinez, Juan David Torres, and Felipe Valencia Caicedo. 2025. "Bourbon Reforms, State Capacity, and Revolution in the Spanish Empire." SSRN Working Paper No. 4545407.

Colella, Fabrizio, Rafael Lalive, Seyhun Orcan Sakalli, and Mathias Thoenig. 2023. "acreg: Arbitrary Correlation Regression." *The Stata Journal*, 23(1): 119-147.

D'Arcy, Michelle, Marina Nistotskaya, and Ola Olsson. 2024. "Cadasters and Economic Growth: A Long-Run Cross-Country Panel." *Journal of Political Economy*, 132 (11): 3785–3826.

Dal Bó, Ernestom, Frederico Finan, and Martín A. Rossi. 2013. "Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service." *Quarterly Journal of Economics*, 128 (3): 1169–1218.

Dennis, Joseph R. 2015. *Writing, Publishing, and Reading Local Gazetteers in Imperial China, 1100–1700*. Leiden: Brill.

Dzansi, James, Anders Jensen, David Lagakos, and Henry Telli. 2025. "Technology and Tax Capacity: Evidence from Local Governments in Ghana." NBER Working Paper No. 29923.

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

Fernández-Villaverde, Jesús, Mark Koyama, Youhong Lin, and Tuan-Hwee Sng. 2023. "The Fractured-Land Hypothesis." *Quarterly Journal of Economics*, 138 (2): 1173–1231.

Feuerwerker, Albert. 1958. *China's Early Industrialization: Sheng Hsuan-Huai (1844–1916) and Mandarin Enterprise*. Cambridge, MA: Harvard University Press.

Flückiger, Matthias, Mario Larch, Markus Ludwig, and Luigi Pascali. 2024. "The Dawn of Civilisation: Metal Trade and the Rise of Hierarchy." CESifo Working Paper No. 10929.

Food and Agriculture Organization of the United Nations. 2002. Global Agro-Ecological Zones (GAEZ).

Gao, Shunfu. 2015. *Keju Yu Liaodai Shehui (The Imperial Examination and Society in Liao Dynasty)*. Beijing: Zhongguo Shehui Kexue Chubanshe.

Garfias, Francisco, and Emily A. Sellars. 2024. "Fiscal Legibility and State Development: Theory and Evidence from Colonial Mexico." *American Journal of Political Science*, 00: 1–16.

Gennaioli, Nicola, and Hans-Joachim Voth. 2015. "State Capacity and Military Conflict." *Review of Economic Studies* 82, (4): 1409–1448.

Gong, Yanming, and Hui Zu. 2014. *Songdai Dengke Zonglu (Official Directory of Song Civil Exam Graduates)*. Guilin: Guangxi Shifan Daxue Chubanshe.

Gu, Hongyi. 2010. *Songchao Fangzhi Kao (An Examination of Gazetteers in the Song Dynasty)*. Shanghai: Shanghai Guji Chubanshe.

Gu, Hongyi. 2012. *Jin Yuan Fangzhi Kao (An Examination of Gazetteers in the Jin and Yuan Dynasties)*. Shanghai: Shanghai Guji Chubanshe.

Guo, Yongqin. 2022. *Land and Labor Tax in Imperial Qing China (1644–1912)*. Leiden: Brill.

Hao, Yu, and Kevin Z. Liu. 2020. "Taxation, Fiscal Capacity, and Credible Commitment in Eighteenth-Century China: The Effects of the Formalization and Centralization of Informal Surtaxes." *Economic History Review*, 73 (4): 914–939.

Hayek, Friedrich A. 1945. "The Use of Knowledge in Society." *American Economic Review*, 35(4): 519–530.

Herrmann, Albert. 1935. "The Contending States—Boundaries of 350 B.C.." In *Norman B. Leventhal Map and Education Center Collection*. Cambridge, MA: Harvard University Press.

Imperial Educational Ministry. 2009 [1909]. "Xuantong Yuannian Fen Di San Ci Jiaoyu Tongji Tubiao (The Third Educational Statistical Charts for the First Year of the Xuantong Reign)." In *Guojia Tushuguan Cang Jindai Tongji Ziliao Congkan*, vols. 32–33, edited by Guojia Tushuguan Gujiguan. Beijing: Yanshan Chubanshe.

Jia, Ruixue. 2014. "The Legacies of Forced Freedom: China's Treaty Ports." *Review of Economics and Statistics*, 96 (4): 596–608.

Jia, Ruixue, and James Kai-sing Kung. 2025. "The Culture and Institutions of Confucianism." NBER Working Paper No. 33883.

Kung, James Kai-sing. 2022. "The Economic Impact of the West: A Reappraisal." In *The Cambridge Economic History of China*, edited by Debin Ma and Richard von Glahn, pp. 354–413. Cambridge: Cambridge University Press.

Le, Shi. 1936 [930–1007]. *Taiping Huanyu Ji* (Universal Geography of the Taiping Era). Beijing: The Commercial Press.

Lee, Melissa M., and Nan Zhang. 2017. "Legibility and the Informational Foundations of State Capacity." *Journal of Politics*, 79 (1): 118–132.

Li, Weijia. 2023. "Dual Leadership and Bureaucratic Capacity." Working Paper.

Liang, Fangzhong. 2008. *Zhongguo Lidai Hukou Tiandi Tianfu Tongji (Statistics of Household, Land and Tax Registration in Chinese History)*. Beijing: Zhonghua Shuju.

Liang, Minling. 2007. "Qingji Minchu Yige Yuedong Difang Dushuren De Licheng—Yi Luo Shiyang (1866–1931) Wei Ge'an (The Journey of a Local Scholar in Eastern

Guangdong During the Late Qing and Early Republican Periods: A Case Study of Luo Shiyang (1866–1931)).” *Lishi Renleixue Xuekan* 5 (2): 39–84.

Lin, Chen, Chicheng Ma, Yuchen Sun, and Yuchen Xu. 2021. “The Telegraph and Modern Banking Development, 1881–1936.” *Journal of Financial Economics*, 141 (2): 730–749.

Lin, Ping, and Jiliang Zhang. 2001. *Mingdai Fangzhi Kao (An Examination of Gazetteers in the Ming Dynasty)*. Chengdu: Sichuan Daxue Chubanshe.

Liu, Kevin Z., and Xiaoming Zhang. 2025. “Discretion, Talent Allocation, and Governance Performance: Evidence from China’s Imperial Bureaucracy.” *Journal of Development Economics*, 172: 103391.

Liu, Weiyi, Chaohua Wang, Shuting Zhao, and Meiling Zheng. 2011. *Song Liao Jin Yuan Gufangzhi Jiyi (An Examination of Lost Gazetteers in the Song, Liao, Jin, and Yuan Dynasties)*. Shanghai: Shanghai Guji Chubanshe.

Liu, Wenpeng. 2017. *Qingdai Yizhan Kao (An Examination of Courier Stations in the Qing Dynasty)*. Beijing: Renmin Chubanshe.

Man-Cheong, Iona. 2004. *The Class of 1761: Examinations, State, and Elites in Eighteenth-Century China*. Stanford, CA: Stanford University Press.

Mastrorocco, Nicola, and Edoardo Teso. 2025. “State Capacity as an Organizational Problem.” *Review of Economic Studies*, forthcoming.

Mayshar, Joram, Omer Moav, and Luigi Pascali. 2022. “The Origin of the State: Land Productivity or Appropriability?” *Journal of Political Economy*, 130 (4): 1091–1144.

Nelson, Richard R., and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, MA: Belknap Press.

Rohner, Dominic, and Ekaterina Zhuravskaya. 2025. “The Economics of Nation-Building: Methodological Tool Kit and Policy Lessons.” *Annual Review of Economics*, 17: 453–478.

Scott, James C. 1998. *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*. New Haven, CT: Yale University Press.

Scott, James C. 2017. *Against the Grain: A Deep History of the Earliest States*. New Haven, CT: Yale University Press.

Shen, Wanli. 2019. *Yuandai Keju Xintan (A New Study of the Civil Service Examinations in the Yuan Dynasty)*. Beijing: Renmin Chubanshe.

Skinner, G. William, Mark Henderson, and Zumou Yue, 2013, “Physiographic Macror egions of China.” Harvard Dataverse, V1. Available online at https://doi.org/10.7910/DVN/4BEKKX, Harvard Dataverse, V1.

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.

Spence, Jonathan D. 1990. *The Search for Modern China*. New York: WW Norton & Company.

Stasavage, David. 2021. "Biogeography, Writing, and the Origins of the State." In *The Handbook of Historical Economics*, edited by Alberto Bisin and Giovanni Federico, 881–902. Cambridge, MA: Academic Press.

Thompson, Roger R. 2015. "The Wire: Progress, Paradox, and Disaster in the Strategic Networking of China, 1881–1901." *Frontiers of History in China* 10 (3): 395–427.

Tilly, Charles. 1990. *Coercion, Capital, and European States, AD 990–1990*. Cambridge, MA: Blackwell.

U.S. Geological Survey, Community for Data Integration. 2012. U.S. Geological Survey.

Wan, Ming, and Yingkai Xu. 2015. *Mingdai "Wanli Kuaiji Lu" Zhengli Yu Yanjiu (A Collation and Study of the Ming "Wanli Kuaiji Lu")*. Beijing: Zhongguo Shehui Kexue Chubanshe.

Wang, Shaoda, and David Y. Yang. 2025. "Policy Experimentation in China: The Political Economy of Policy Learning." *Journal of Political Economy*, 133 (7): 2180–2228.

Wang, Yuhua. 2022a. "Blood Is Thicker than Water: Elite Kinship Networks and State Building in Imperial China." *American Political Science Review*, 116 (3): 896–910.

Wang, Yuhua. 2022b. *The Rise and Fall of Imperial China: The Social Origins of State Development*. Princeton, NJ: Princeton University Press.

Xu, Hong. 2017. *Xianqin Chengyi Kaogu (Enclosures in Pre-Qin China)*. Beijing: Jincheng Chubanshe and Shuyuan Chubanshe.

Xu, Song, ed. 1957. *Song Hui Yao Jigao (Recovered Draft of the Collected Essential Documents of Song)*. Beijing: Zhonghua Shuju.

Xue, Ruizhao. 2004. *Jindai Keju (Civil Service Examinations in the Jin Dynasty)*. Beijing: Zhongguo Shehui Kexue Chubanshe.

Yang, Zhengtai. 2006. *Mingdai Yizhan Kao (An Examination of Courier Stations in the Ming Dynasty)*. Shanghai: Shanghai Guji Chubanshe.

Zhang, Xiaoming. 2025. "Information Foundations of State Capacity: Evidence from the Imperial Bureaucracy of China." *Social Forces*, forthcoming.

Zhang, Xuecheng. 1985 [1832]. *Wen Shi Tong Yi (Principles of Literature and History)*. Beijing: Zhonghua Shuju.

Zhou, Zhenhe, ed. 2017. *Zhongguo Xingzheng Quhua Tongshi (A General History of Chinese Administrative Divisions)*. Vols. 8–13. Shanghai: Fudan Daxue Chubanshe.

Figures

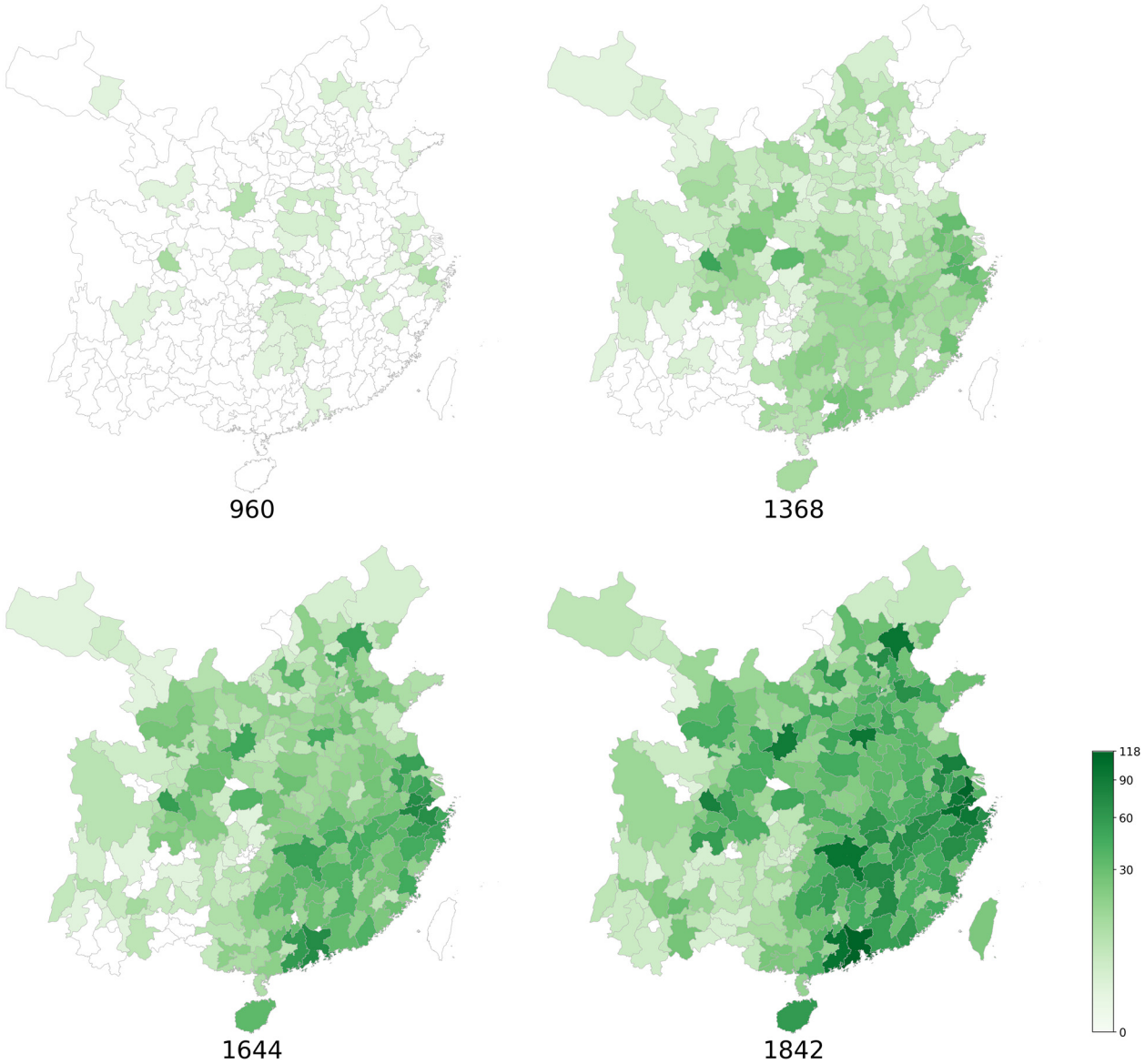
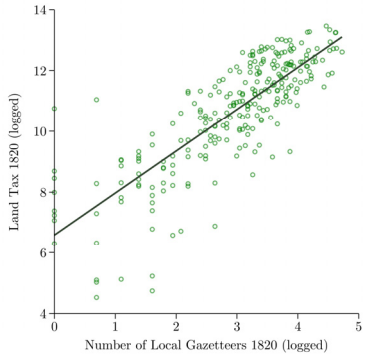
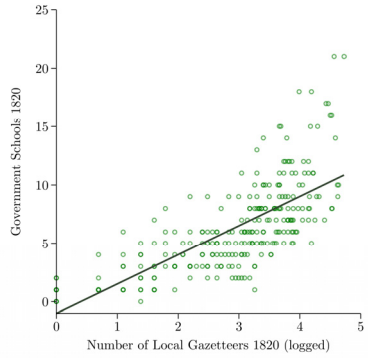


Figure 1: Distribution of Local Gazetteers in the Sample Prefectures (960–1842)

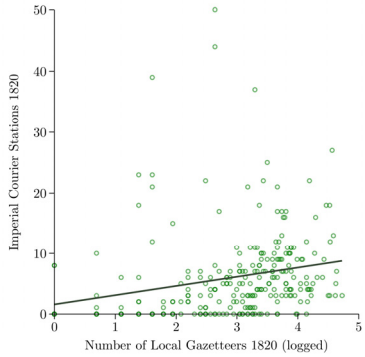
Notes: The figures show the cumulative number of the county and prefectural gazetteers compiled at the beginning of the Song dynasty (960), Ming dynasty (1368), Qing dynasty (1644) and the year of opening-up of the Qing (1842). The sample includes 269 prefectures of China proper based on the 1820 prefectural division in CHGIS. See the text for the sources of data for gazetteers.



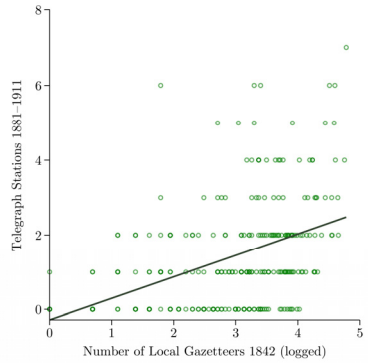
Land Tax 1820



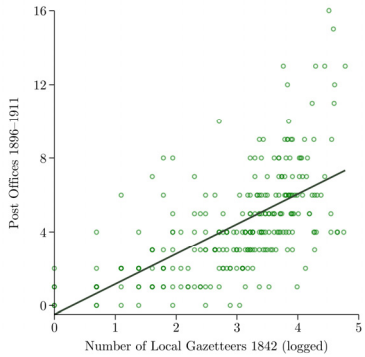
Number of Government Schools 1820



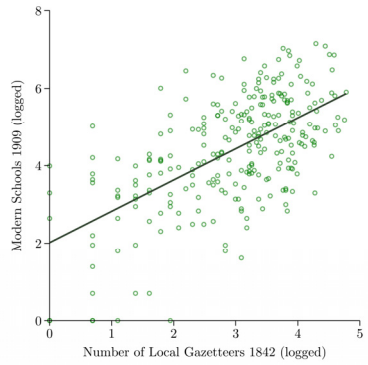
Number of Imperial Courier Stations 1820



Number of Telegraph Stations 1881–1911



Number of Post Offices 1896–1911



Number of Modern Schools 1909

Figure 2: Correlations between Local Gazetteers and Measures of State Building

Notes: All are cross-prefectural correlations between state building status on or after 1820 and the total number of gazetteers previously compiled. See the text for data sources.

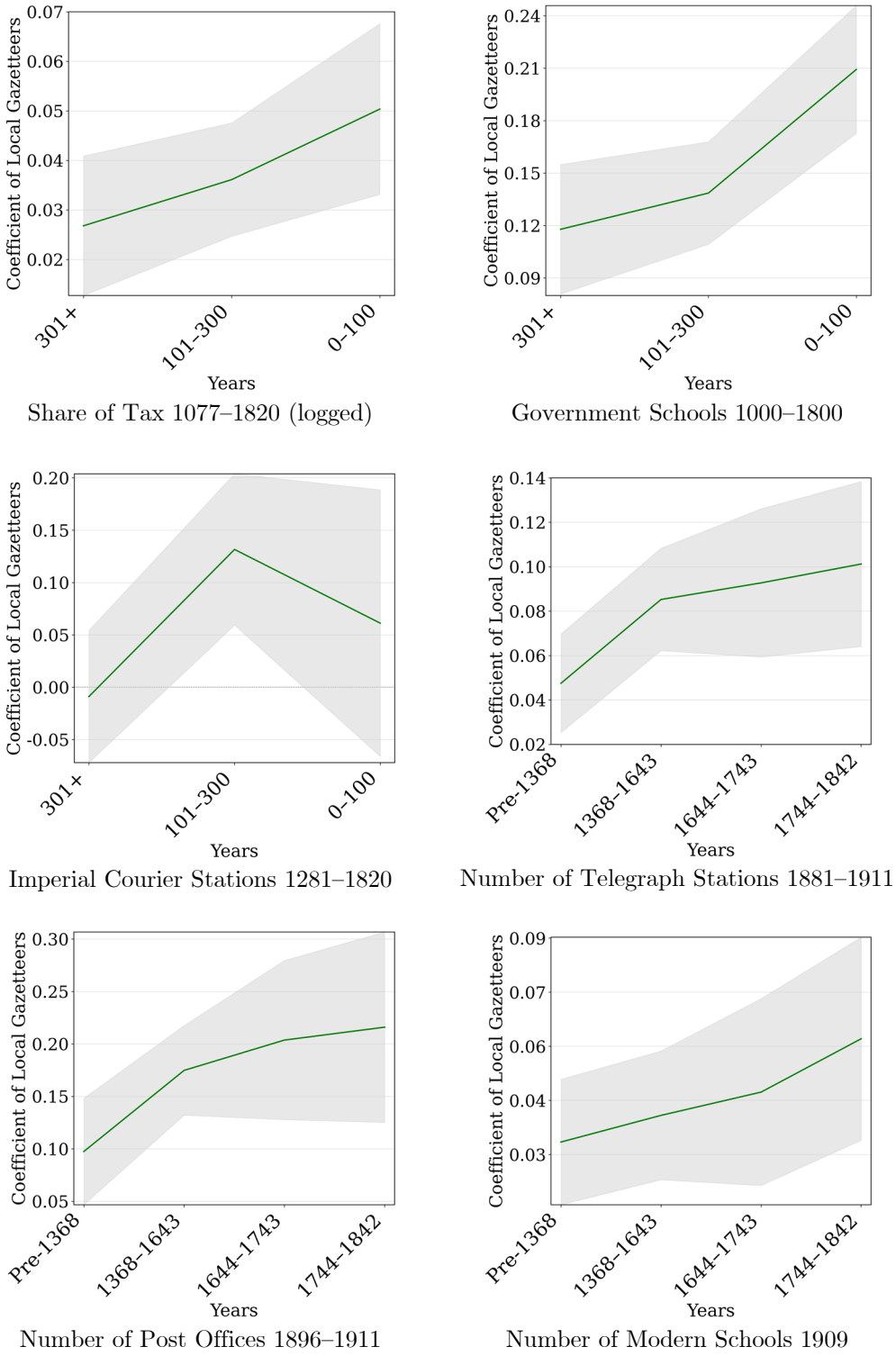
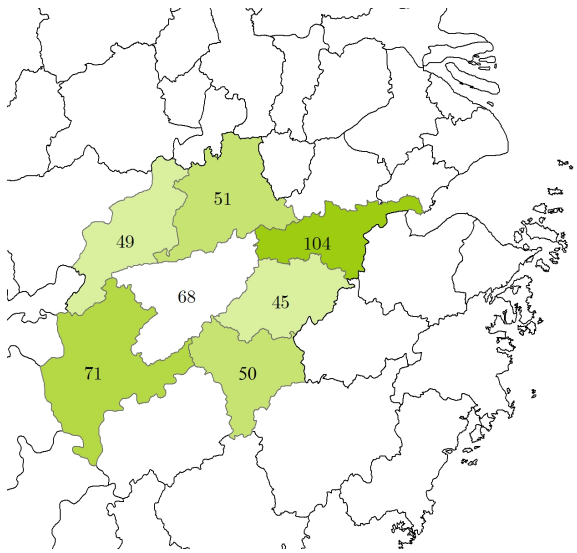
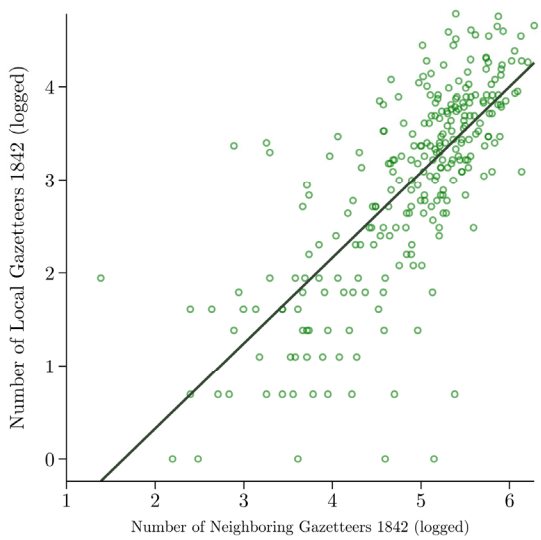


Figure 3: Comparing the Effects of Local Gazetteers Compiled at Different Periods within the Same Prefecture

Notes: The figures display the coefficients (and 90% confidence intervals) of the number of gazetteers compiled in different periods on measures of state building. For panel regressions of traditional state building, gazetteers are divided by their compilation years in the last century, previous two centuries and earlier. For the cross-sectional regressions of state modernization, gazetteers are grouped by mid-to-late Qing (1744–1842), early-to-mid-Qing (1644–1743), Ming (1368–1643), and pre-Ming. All regressions are based on the same specifications used for Table 1 with a full set of controls.



Panel A: The number of gazetteers in Huizhou (Anhui Province) and its neighboring prefectures (in green shade) by 1842



Panel B: Correlation between the numbers of local and neighboring gazetteers

Figure 4: Number of Neighboring Gazetteers as the Instrumental Variable

Notes: The map in Panel A illustrates the construction of the instrumental variable using the number of gazetteers in a prefecture's neighboring (adjacent) prefectures based on the reasoning that neighboring prefectures usually competed with and referenced each other when compiling gazetteers, whereas the gazetteers are mutually exclusive records of local knowledge due to prefectural boundaries. For Huizhou Prefecture in the map, the number of 68 gazetteers is instrumented by the total number of gazetteers of the neighboring prefectures around (i.e., 370 gazetteers). Panel B shows the positive correlation between the number of local gazetteers and the number of neighboring gazetteers compiled by 1842 for all 267 prefectures. See the text for data sources.

Tables

Table 1. Impact of Local Gazetteers on State Building: OLS Estimates

	Traditional State Building			State Modernization (Since 1842)		
	Share of Tax 1077–1820 (logged)	Government Schools 1000–1800	Imperial Courier Stations 1281–1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Local Gazetteers (logged)	0.725	1.162	1.819	0.601	1.518	0.714
	(0.109)***	(0.133)***	(0.478)***	(0.127)***	(0.265)***	(0.108)***
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Prefecture and Time FEs	Yes	Yes	Yes			
Province FEs				Yes	Yes	Yes
Observations	989	4,539	801	267	267	267
Number of Prefectures	266	267	267	267	267	267
Adjusted R^2	0.804	0.811	0.484	0.400	0.630	0.669

Notes: Columns 1–3 are panel data regressions of traditional state building against the number of local gazetteers. The share of tax refers to the percentage (×100) of a prefecture's tax revenue in total tax revenue of all sample prefectures in a given year. The data are available for four times (1077, 1578, 1743, and 1820), in which the commercial tax is used for 1077 and the land tax is for the remaining years. Government schools denote the number of existing government schools in each prefecture every 50 years between 1000 and 1800. The number of imperial courier stations spans three times: 1281, 1587, and 1820. Standard errors are clustered at the prefectural level (in parentheses). Columns 4–6 are cross-prefectural regressions of the post-1842 state modernization (1842–1911) on the total number of local gazetteers compiled before 1842. We use the number of telegraph stations established in 1881 to 1911, the number of post offices established in 1896 to 1911, and the number of modern (primary and secondary) schools in 1909 to respectively measure the modernization of state infrastructures in each prefecture. Robust standard errors are reported in parentheses. *, ** and *** indicate the level of significance at 10%, 5% and 1%, respectively.

Table 2. Impact of Local Gazetteers on State Building: OLS Estimates with Additional Controls

	Traditional State Building			State Modernization (Since 1842)		
	Share of Tax 1077–1820 (logged)	Government Schools 1000–1800	Imperial Courier Stations 1281–1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Local Gazetteers (logged)	0.680	1.034	1.592	0.755	1.777	0.539
	(0.102)***	(0.125)***	(0.526)***	(0.151)***	(0.324)***	(0.117)***
Early State Development	[0.000]***	[0.000]***	[0.510]	-0.030	-0.251	-0.131
				(0.113)	(0.159)	(0.068)*
Local scholars (logged)	0.055	0.142	0.133	-0.118	-0.194	0.136
	(0.046)	(0.061)**	(0.234)	(0.073)	(0.099)*	(0.053)**
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Prefecture and Time FEs	Yes	Yes	Yes			
Province FEs				Yes	Yes	Yes
Observations	989	4,539	801	267	267	267
Number of Prefectures	266	267	267	267	267	267
Adjusted R^2	0.815	0.829	0.484	0.403	0.636	0.679

Notes: This table replicates Table 1 except for additionally controlling for early state building and the number of local scholars. Early state development is measured by the first principal component of the density of counties from 221 BCE to 960 CE (i.e., before gazetteer compilation intensified in the Song dynasty), the density of walled cities in the Eastern Zhou period (circa 771–256 BCE), and a dummy for the presence of the Sinic state around 350 BCE. We interact the principal component with the time dummies in columns 1–3 and report their joint significance (p-value) of F-statistics in brackets accordingly. The number of local scholars is measured by the number of the highest (*jinshi*) exam degree holders per million people from 960 to the time point of the dependent variable.

Table 3. Comparing the Effects of County Gazetteers and Prefectural Gazetteers of the Same Prefecture, OLS Estimates

	Traditional State Building			State Modernization (Since 1842)		
	Share of Tax 1077–1820 (logged)	Government Schools 1000–1800	Imperial Courier Stations 1281–1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Number of County Gazetteers (logged)	0.606	1.313	1.746	0.535	1.410	0.674
	(0.093)***	(0.142)***	(0.481)***	(0.107)***	(0.211)***	(0.088)***
Number of Prefectural Gazetteers (logged)	0.209	0.041	0.006	-0.006	0.034	-0.059
	(0.128)	(0.200)	(0.659)	(0.128)	(0.185)	(0.108)
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Prefecture and Time FEs	Yes	Yes	Yes			
Province FEs				Yes	Yes	Yes
Observations	989	4,539	801	267	267	267
Number of Prefectures	266	267	267	267	267	267
Adjusted R^2	0.806	0.820	0.485	0.403	0.650	0.685

Notes: All are the same as Table 1 except we separately count the county and prefectural gazetteers and compare their relative importance in explaining state building. County gazetteers contain richer local knowledge than the prefectural gazetteers but are subject to the same set of prefectural factors (including the compilation capacity) as the prefectural gazetteers. The greater (and more significant) effect of the county gazetteers indicates that the gazetteer effect observed in Table 1 mainly arises from local knowledge rather than the correlates of local compilation capacity.

Table 4. Impact of Local Gazetteers on State Building: Instrumented Results

	Traditional State Building			State Modernization (Since 1842)		
	Share of Tax 1077–1820 (logged)	Government Schools 1000–1800	Imperial Courier Stations 1281–1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Local Gazetteers (logged)	1.251	1.065	3.616	0.275	1.805	0.619
	(0.366)***	(0.407)***	(0.899)***	(0.256)	(0.375)***	(0.211)***
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Prefecture and Time FEs	Yes	Yes	Yes			
Province FEs				Yes	Yes	Yes
Observations	989	4,539	801	267	267	267
Number of Prefectures	266	267	267	267	267	267
Adjusted R^2	0.786	0.811	0.464	0.382	0.627	0.667
K-P F-statistic	17.36	96.98	37.65	50.78	52.82	53.21
First-Stage for the Number of	0.297	0.353	0.414	0.579	0.573	0.571
Neighboring Gazetteers (logged)	(0.071)***	(0.036)***	(0.067)***	(0.081)***	(0.079)***	(0.078)***

Notes: All are the same as Table 1 except we instrument the number of local gazetteers using the number of gazetteers in the neighboring (adjacent) prefectures. The number of neighboring gazetteers is positively correlated with the number of local prefecture's gazetteers as shown in the bottom line for the first-stage results but is arguably orthogonal to local prefecture's state building due to the knowledge boundary of gazetteers along administrative division.

Table 5. Impact of the Richness of Gazetteer Knowledge on State Building

	Traditional State Building			State Modernization (Since 1842)		
	Land Tax 1820 (logged)	Government Schools 1820	Imperial Courier Stations 1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Number of Raw Topics	0.0008	0.0037	0.0056	0.0021	0.0036	0.0006
	(0.0003)***	(0.0014)***	(0.0029)*	(0.0008)***	(0.0011)***	(0.0004)*
Adjusted R^2	0.724	0.598	0.209	0.316	0.514	0.618
Panel B						
Number of Clustered Topics	0.0009	0.0041	0.0062	0.0023	0.0041	0.0007
	(0.0003)***	(0.0016)***	(0.0033)*	(0.0009)***	(0.0013)***	(0.0004)*
Adjusted R^2	0.724	0.597	0.209	0.314	0.513	0.618
Controls in Both Panels						
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Province FEs				Yes	Yes	Yes
Observations	176	178	178	180	180	180

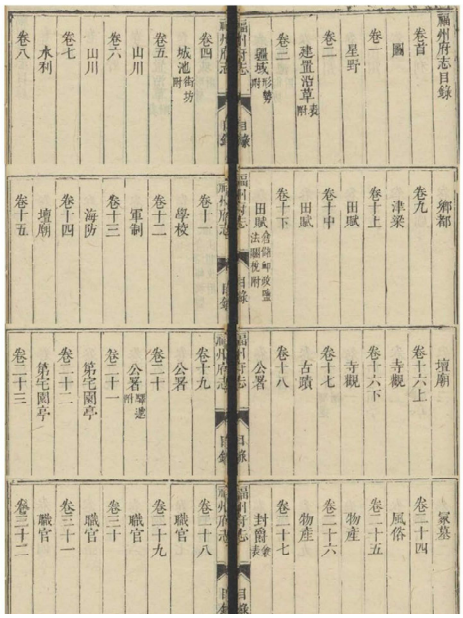
Notes: All are the same as Table 1 except that we are now examining the effect of the richness of the gazetteer knowledge in a limited sample of 180 prefectures. The richness is measured by the number of topics introduced in all prefectural and county gazetteers in each prefecture. In Panel A, the topics count all the raw topics (chapter titles) in the contents of gazetteers compiled between 1368 and 1820 (or 1842 for columns 4–6), which are available in the Diaolong database for digitized gazetteers (see the text for details). In Panel B, the raw topics are aggregated into synonymous clusters based on OpenAI's embedding vector similarity. Robust standard errors are reported in parentheses. *, ** and *** indicate the level of significance at 10%, 5% and 1%, respectively.

Table 6. Impacts of the Routine and Non-Routine Knowledge on State Building

	Traditional State Building			State Modernization (Since 1842)		
	Land Tax 1820 (logged)	Government Schools 1820	Imperial Courier Stations 1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Number of Routine Topics	0.0004	0.0021	0.0033	0.0010	0.0016	0.0002
	(0.0002)***	(0.0008)**	(0.0015)**	(0.0004)**	(0.0006)**	(0.0002)
Adjusted R^2	0.723	0.596	0.209	0.310	0.504	0.615
Panel B						
Number of Non-routine Topics	0.0009	0.0041	0.0067	0.0019	0.0033	0.0004
	(0.0003)***	(0.0016)***	(0.0029)**	(0.0008)**	(0.0011)***	(0.0003)
Adjusted R^2	0.725	0.601	0.213	0.315	0.511	0.615
Controls in All Panels						
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Province FEs				Yes	Yes	Yes
Observations	176	178	178	180	180	180

Notes: All are the same as Table 5 except we further divide the topics into routine and non-routine. The routine topics refer to those more standardized and codified forms of knowledge commonly used for regular governance. They are defined according to the routine topics in the Unified Chorography of the Qing Dynasty compiled by central authority. The non-routine topics are those not included in the central chorography but unique to local contexts. Robust standard errors are reported in parentheses. *, ** and *** indicate the level of significance at 10%, 5% and 1%, respectively.

Online Appendix

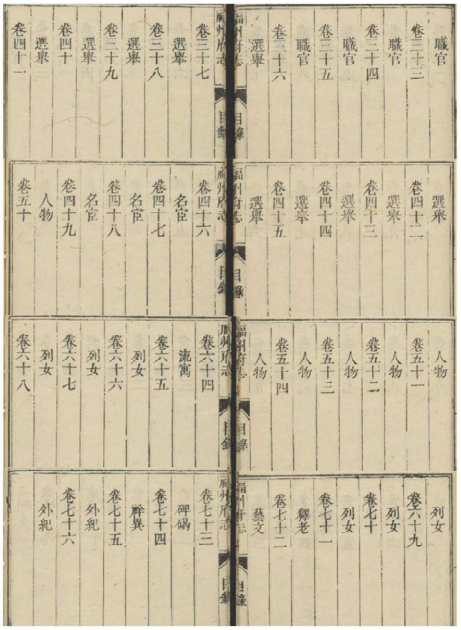


Chapter 1 Astrology (星野)
Chapter 2 Administrative History (建制沿革)
Chapter 3 Territories (疆域)
Chapter 4 City Walls (城池)
Chapter 5–6 Mountains and Rivers (山川)
Chapter 7 Irrigation (水利)
Chapter 8 Township (乡都)

Chapter 9 Bridges (津梁)
Chapter 10 Land Tax (田赋)
Chapter 11 Schools (学校)
Chapter 12 Military Affairs (军制)
Chapter 13 Coastal Defense (海防)
Chapter 14–15 Shrines (坛庙)

Chapter 16 Buddhist/Taoist Temples (寺观)
Chapter 17 Historical Sites (古迹)
Chapter 18–20 Government Offices (公署)
Chapter 21–22 Houses and Gardens (第宅园亭)

Chapter 23 Graves (冢墓)
Chapter 24 Customs (风俗)
Chapter 25–26 Crops and Products (物产)
Chapter 27 Nobility (封爵)
Chapter 28–35 Government Officials (职官)



Chapter 36–45 Civil Examinations (选举)

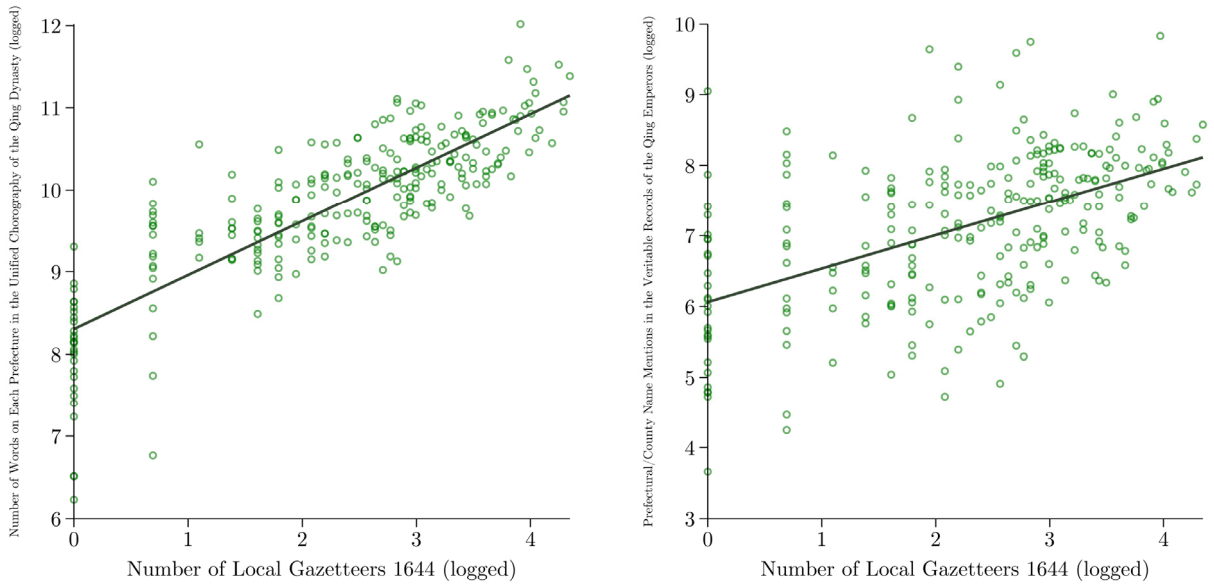
Chapter 46–48 Notable Officials (名宦)
Chaptcr 50–63 Local Cclcbritics (人物)

Chapter 64 Notable Migrants (流寓)
Chapter 65–70 Chaste Women (列女)

Chapter 71 Buddhists and Lao Tzu (释老)
Chapter 72 Arts and Literature (艺文)
Chapter 73 Stele (碑碣)
Chapter 74 Abnormal Events (祥异)
Chapter 75–76 Miscellaneous Records (外纪)

Figure A1: Contents in the Prefectural Gazetteer of Fuzhou (Fujian), 1754

Source: Chinese Text Project (ctext.org).



Panel A. *Unified Chorography of the Qing Dynasty* 1820

Panel B. *Veritable Records of the Qing Emperors* 1644–1911

Figure A2. Local Gazetteers and Regional Exposure in Central Compilations

Notes: Panel A shows the correlation between the text length for a prefecture (and its counties) in the *Unified Chorography of the Qing Dynasty* and the number of gazetteers of this prefecture compiled before the Qing dynasty. Panel B shows the number of mentions of a prefecture (and its counties) in another central chronicle, the *Veritable Records of the Qing Emperors*, and the number of gazetteers of this prefecture compiled before the Qing.

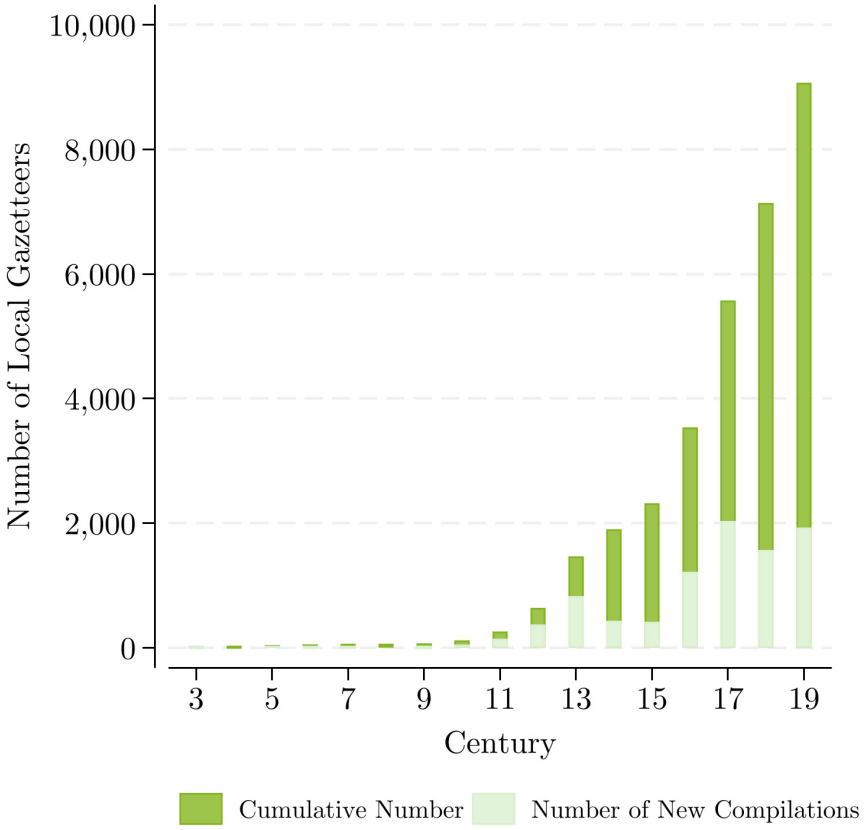


Figure A3: Number of Local Gazetteers in China Proper by Century

Notes: This figure shows the total number of county and prefectural gazetteers compiled for the 267 prefectures of China proper by century. See the text for data sources.

Table A1. Descriptive Statistics

	Structure	N	Mean	SD	Min	Max
Variables on Local Knowledge						
Number of Local Gazetteers	Panel	4,539	9.09	14.17	0	107
Number of Local Gazetteers	Cross-sectional	267	29.13	24.58	0	118
Number of County Gazetteers	Panel	4,539	6.76	11.73	0	92
Number of Prefectural Gazetteers	Panel	4,539	2.32	3.30	0	22
Number of Raw Topics	Cross-sectional	180	295.9	181.65	12	826
Number of Clustered Topics	Cross-sectional	180	271.8	159.83	12	720
Number of Volumes	Cross-sectional	267	187.49	184.75	0	1,350
Number of Pages	Cross-sectional	180	5,657.66	5,145.17	32	29,779
Number of Routine Topics	Cross-sectional	180	389.5	310.23	7	1,639
Number of Non-routine Topics	Cross-sectional	180	223.24	177.21	5	841
Variables on State Building						
Share of Tax 1281–1800 (percent × 100)	Panel	989	0.4	0.49	0.0001	5.3
Land Tax 1820 (sliver *liang*)	Cross-sectional	261	114,519.38	133,206.43	90.32	701,787.75
Government Schools, 1000–1800	Panel	4,539	3.31	3.64	0	21
Government Schools 1820	Cross-sectional	267	6.27	3.89	0	21
Imperial Courier Stations, 1281–1820	Panel	801	4.21	5.21	0	50
Imperial Courier Stations 1820	Cross-sectional	267	5.96	7.20	0	50
Telegraph Stations 1881–1911	Cross-sectional	267	1.44	1.47	0	7
Post Offices 1896–1911	Cross-sectional	267	4.39	2.87	0	16
Modern Schools 1909	Cross-sectional	267	176.47	223.46	0	1,273
Control Variables						
Number of Counties	Panel	4,539	5.01	4.29	0	62
Number of Counties	Cross-sectional	267	5.01	3.74	0	20.13
Population Density (people/km^2)	Panel	4,539	47.14	74.48	0.0004	1,039.25
Population Density (people/km^2)	Cross-sectional	267	62.84	70.36	0.19	604.18
Distance to Imperial Capital (km)	Panel	4,539	1,059.12	554.06	0	2,370.28
Distance to Imperial Capital (km)	Cross-sectional	267	989.18	455.34	370.7	2,126.11
Distance to Provincial Capital (km)	Panel	4,539	265.55	206.07	0	1,695.81
Distance to Provincial Capital (km)	Cross-sectional	267	291.99	144.93	0	1,069.03
Distance to the Coast (km)	Cross-sectional	267	515.23	369.51	2.16	1,925.56
Agricultural Suitability Index	Cross-sectional	267	16.32	6.05	0.0005	29.91
Terrain Ruggedness Index	Cross-sectional	267	27.59	18.20	0.68	91.36
Treaty Port 1904	Cross-sectional	267	0.09	0.29	0	1
Early State Development (first principal component)	Cross-sectional	267	0	1.43	-1.41	6.13
Local Scholars (people/million)	Panel	4,539	456.56	1691.83	0	42,603.44
Local Scholars (people/million)	Cross-sectional	267	341.01	569.05	0	5,836.65
Instrumental Variables						
Number of Neighboring Gazetteers	Panel	4,539	54.59	70.02	0	486
Number of Neighboring Gazetteers	Cross-sectional	267	174.03	109.33	4	532

Table A2. Impact of the Volumes and Pages of Gazetteers on State Building

	Traditional State Building			State Modernization (Since 1842)		
	Land Tax 1820 (logged)	Government Schools 1820	Imperial Courier Stations 1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Number of Volumes (logged)	0.538	0.861	1.490	0.375	0.821	0.407
	(0.067)***	(0.144)***	(0.311)***	(0.088)***	(0.198)***	(0.082)***
Observations	261	267	267	267	267	267
Adjusted R^2	0.732	0.700	0.287	0.396	0.600	0.649
Panel B						
Number of Pages (logged)	0.133	0.556	0.314	0.369	0.516	0.158
	(0.073)*	(0.224)**	(0.678)	(0.139)***	(0.224)**	(0.063)**
Observations	176	178	178	180	180	180
Adjusted R^2	0.724	0.595	0.200	0.320	0.506	0.626
Controls in Both Panels						
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Province FEs				Yes	Yes	Yes

Notes: All are the same as Table 1 except that we use the volumes and pages as alternative measures of the accumulation of local knowledge in a limited sample of 180 prefectures. Number of volumes refers to the total number of volumes of all county and prefectural gazetteers compiled in each prefecture from 1644 to 1820 (or 1842 for columns 4–6). Number of pages denotes the total number of pages in all county and prefectural gazetteers in each prefecture compiled 1368 to 1820 (or 1842 for columns 4–6). See the text for data sources. Robust standard errors are reported in parentheses. *, ** and *** indicate the level of significance at 10%, 5% and 1%, respectively.

Table A3. Prefectural Correlates of the Number of Prefectural and County Gazetteers

	Number of Prefectural Gazetteers 1842	Number of County Gazetteers 1842
	(1)	(2)
Population Density (logged)	1.325	7.955
	(0.312)***	(1.762)***
Agricultural Suitability (logged)	-0.390	-3.562
	(0.153)**	(0.856)***
Terrain Ruggedness Index (logged)	0.057	2.056
	(0.243)	(1.336)
Distance to Imperial Capital (logged)	0.375	-2.084
	(0.702)	(4.008)
Distance to Provincial Capital (logged)	-0.786	-7.590
	(0.353)**	(0.912)***
Distance to the Coast (logged)	-0.387	-2.483
	(0.237)	(1.399)*
Early State Development	-0.201	-0.452
	(0.193)	(1.148)
Local Scholars (logged)	0.807	2.976
	(0.104)***	(0.542)***
Observations	267	267
Adjusted R^2	0.417	0.480

Notes: Both columns report the OLS estimates of the correlations between prefectural characteristics and the number of prefectural and county gazetteers compiled in each prefecture. Robust standard errors are reported in parentheses. *, ** and *** indicate the level of significance at 10%, 5% and 1%, respectively.

Table A4. Exclusion Restriction Tests

	Share of Tax 1077–1820 (logged)	Share of Tax 1077–1820 (logged)	Government Schools 1000–1800	Government Schools 1000–1800	Imperial Courier Stations 1281–1820	Imperial Courier Stations 1281–1820
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Neighboring Gazetteers (logged)	0.372	0.167	0.376	-0.038	1.497	0.900
	(0.126)***	(0.120)	(0.160)**	(0.159)	(0.464)***	(0.482)*
Number of Local Gazetteers (logged)		0.691		1.173		1.443
		(0.109)***		(0.142)***		(0.511)***
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	989	989	4,539	4,539	801	801
Adjusted R^2	0.776	0.805	0.792	0.811	0.477	0.488
	Telegraph Stations 1881–1911	Telegraph Stations 1881–1911	Post Offices 1896–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)	Modern Schools 1909 (logged)
	(7)	(8)	(9)	(10)	(11)	(12)
Number of Neighboring Gazetteers (logged)	0.159	-0.245	1.035	0.212	0.353	-0.070
	(0.165)	(0.156)	(0.206)***	(0.273)	(0.137)**	(0.143)
Number of Local Gazetteers (logged)		0.698		1.435		0.742
		(0.134)***		(0.319)***		(0.121)***
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	267	267	267	267	267	267
Adjusted R^2	0.343	0.402	0.562	0.630	0.595	0.668

Notes: All are the same as Table 1 except we include the number of neighboring gazetteers to compare its effect with the number of local gazetteers. The controls include the same set of prefectural factors as used in Table 1.

Table A5. Impact of Local Gazetteers on State Building: Instrumented Results Controlling for Early State Development and Local Scholars' Influence.

	Traditional State Building			State Modernization (Since 1842)		
	Share of Tax 1077–1820 (logged)	Government Schools 1000–1800	Imperial Courier Stations 1281–1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Local Gazetteers (logged)	1.202	0.915	3.908	0.241	2.229	0.316
	(0.380)***	(0.450)**	(1.092)***	(0.362)	(0.605)***	(0.315)
Early State Development	[0.000]***	[0.000]***	[0.804]	-0.043	-0.239	-0.138
				(0.110)	(0.150)	(0.068)**
Local scholars (logged)	0.036	0.155	-0.169	0.019	-0.315	0.196
	(0.043)	(0.079)*	(0.223)	(0.101)	(0.178)*	(0.093)**
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Prefecture and Time FEs	Yes	Yes	Yes			
Province FEs				Yes	Yes	Yes
Observations	989	4,539	801	267	267	267
Adjusted R^2	0.798	0.829	0.454	0.374	0.631	0.673
K-P F-statistic	15.16	71.68	24.62	24.54	25.05	25.43

Notes: All are the same as Table 4 except that we additionally control for early state development and the size of local scholars. Please refer to the text for the definition and sources of the two additional controls.

Table A6. Impact of Local Gazetteers on State Building: The Effect of Lagged Instrumental Variables

	Traditional State Building			State Modernization (Since 1842)		
	Share of Tax 1077–1820 (logged)	Government Schools 1000–1800	Imperial Courier Stations 1281–1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: IV with 50-year Lag						
Number of Local Gazetteers (logged)	1.172	1.128	4.667	0.273	1.712	0.590
	(0.278)***	(0.351)***	(1.033)***	(0.257)	(0.356)***	(0.206)***
Adjusted R^2	0.791	0.811	0.432	0.382	0.628	0.666
K-P F-statistic	36.88	128.7	40.62	49.72	51.64	51.81
Panel B: IV with 100-year Lag						
Number of Local Gazetteers (logged)	0.993	1.191	4.733	0.219	1.659	0.498
	(0.230)***	(0.343)***	(1.248)***	(0.254)	(0.353)***	(0.203)**
Adjusted R^2	0.799	0.811	0.430	0.376	0.629	0.661
K-P F-statistic	39.56	122.2	33.05	48.16	49.81	49.99
Panel C: IV with 200-year Lag						
Number of Local Gazetteers (logged)	1.438	1.508	8.531	0.132	1.609	0.487
	(0.268)***	(0.425)***	(2.336)***	(0.265)	(0.394)***	(0.226)**
Adjusted R^2	0.771	0.809	0.196	0.363	0.630	0.660
K-P F-statistic	23.10	67.11	14.22	38.17	39.16	39.06
Controls in All Panels						
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Center	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Prefecture and Time FEs	Yes	Yes	Yes			
Province FEs				Yes	Yes	Yes
Observations	989	4,539	801	267	267	267

Notes: All are the same with Table 4 except we instrument the number of local gazetteers using the number of gazetteers of the neighboring (adjacent) prefectures lagged by 50, 100 and 200 years, respectively.

Table A7. Impact of Local Gazetteers on State Building: Instrumented Results in the Ming-Qing Period (1368–1911)

	Traditional State Building			State Modernization (Since 1842)		
	Land Tax 1820 (logged)	Government Schools 1820	Imperial Courier Stations 1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Local Gazetteers (logged)	1.182	1.036	0.706	0.404	2.106	0.734
	(0.207)***	(0.324)***	(1.038)	(0.273)	(0.402)***	(0.226)***
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Province FEs				Yes	Yes	Yes
Observations	261	267	267	267	267	267
Adjusted R^2	0.767	0.718	0.266	0.398	0.623	0.661
K-P F-statistic	54.42	66.93	66.93	48.83	50.69	51.33

Notes: All are the same as Table 4 except we use the number of local gazetteers in the Ming and Qing dynasties to predict state building in a cross-sectional setting. Robust standard errors are reported in parentheses. *, ** and *** indicate the level of significance at 10%, 5% and 1%, respectively.

Table A8. Impact of Local Gazetteers on State Building: Instrumented Results Controlling for the Macro-region Effect

	Traditional State Building			State Modernization (Since 1842)		
	Land Tax 1820 (logged)	Government School 1820	Imperial Courier Stations 1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Local Gazetteers (logged)	1.047	1.398	2.608	0.232	1.840	0.575
	(0.224)***	(0.296)***	(0.881)***	(0.391)	(0.479)***	(0.289)**
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Province FEs				Yes	Yes	Yes
Macro-region FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	261	267	267	267	267	267
Adjusted R^2	0.785	0.750	0.532	0.403	0.641	0.690
K-P F-statistic	30.59	35.96	35.96	19.28	19.65	19.97

Notes: All are the same as Table 4 except we additionally control for the macro-region fixed effects to account for the potential confounding effect of similarities between neighboring prefectures. The sample consists of 16 macro-regions based on William Skinner, Henderson, and Yue (2013)'s division of nineteenth-century China upon biogeographic similarities. Standard errors are clustered at the macro-region level (in parentheses). *, ** and *** indicate the level of significance at 10%, 5% and 1%, respectively.

Table A9. Impact of Local Gazetteers on State Building: Instrumented Results Accounting for Spatial Correlation between Neighboring Prefectures

	Traditional State Building			State Modernization (Since 1842)		
	Land Tax 1820 (logged)	Government School 1820	Imperial Courier Stations 1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Number of Local Gazetteers (logged)	1.251	1.065	3.616	0.275	1.805	0.619
	(0.464)***	(0.470)**	(1.068)***	(0.261)	(0.375)***	(0.216)***
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Prefecture and Time FEs	Yes	Yes	Yes			
Province FEs				Yes	Yes	Yes
Observations	989	4,539	801	267	267	267
R^2	0.847	0.825	0.650	0.443	0.663	0.700
K-P F-statistic	12.13	87.49	15.46	40.64	41.50	41.81

Notes: All are the same as Table 4 except we cluster the standard error within a radius of 131 km to account for possible spatial correlation among neighboring prefectures based on Colella et al. (2023). *, ** and *** indicate the level of significance at 10%, 5% and 1%, respectively.

Table A10. The Routine Topics in the Unified Chorography of the Qing Dynasty and Selected Non-routine Topics in Local Gazetteers

Routine Topics (Chapters) in the Unified Chorography of the Qing Dynasty
Territories (疆域). Astrology (分野). Topographies (形势). Customs (风俗). City Walls (城池). Schools (学校). Administrative History (建置沿革). Households (户口). Land Tax (田赋). Taxation (税课). Official Posts (职官). Mountains and Rivers (山川). Historical Sites (古迹). Passes (关隘). Bridges (津梁). Dikes (隄堰). Graves (陵墓). Shrines (祠庙). Buddhist/Taoist Temples (寺观). Eminent Officials (名宦). Personage (人物). Sojourners (流寓). Chaste Women (列女). Daoist and Buddhist Immortals (仙释). Local Specialties (土产)
Non-Routine Topics in Local Gazetteers (Examples of the total 5,782 topics)
Maritime Related: Tides (潮汐), Seaports (海港), Maritime Shipping (海运), Sea Routes (海道), Sea Flood (海汛), Coastlines (海岸), Coastal Defense (戍海), Fishing (捕海鱼), Maritime Hazards (海险), Typhoons (海飓), Maritime Customs (海关), Ports (口岸). Medical Related: Medicine (医学), Medical Officers (医官), Medical Societies (医社), Renowned Physicians (名医), Hereditary Physicians (世医), Diseases (疾病), Pharmacies (药局), Medicinal Materials (药材), Miracle Physicians (神医). Architecture Related: Pavilions (亭), Pagodas (塔), Towers (楼阁), Halls (堂), Residential Houses (民居), Palaces (宫室). Military and Public Security: Military Affairs (军事), Frontier Defense (边防), Gunpowder Bureau (火药局), Bandits (盗贼), Fire Control Matters (救火事宜). Mining Related: Gold (金), Mining Administration (矿务), Mine Security (矿防), Mining and Metallurgy (矿冶), Mining Areas (矿山), Mining Hazards (矿害), Coal Mines (煤矿), Goldworks (金厂). Stores and Manufacturing: Warehouses (仓厂), Horse Yards (马厂), Forage Yards (草厂), Government Workshops (官厂), Imperial Kilns (官窑), Tile Kilns (瓦窑). Frontiers and Foreign Affairs: Chieftains (土司), Tributary States (藩封), Western Regions (西域), Southern Barbarians (南蛮), Barbarian Tribes (夷族), Barbarian Languages (夷语), Barbarian Affairs (夷情), Foreign Countries (外国), Southeast Asia (南洋). Rites and Music: Music (音乐), Musical Instruments (乐器), Musical Institutions (乐制), Temperament (乐律), Elegant Music (雅乐), Song and Dance (歌舞), , Sacrificial Music (释奠乐), Music of Great Achievement (大成乐章), Music Director (司乐). Economy and Statecraft: Economy (经济), Administrative Strategy (经略), Principles of Tao (道理), Discourses (论议), Natural History (博物), Anecdotal Records (稗谈), Corvée Discussions (役议), Statecraft (经国用), Principles of Things (物理), Discourse on Taxes and Corvée (赋役公议说). Anomalies and Heresy: Anecdotes (异闻), Rumors (传疑), Eccentrics (异人), Auspicious and Anomalous Signs (瑞异), Celestial Anomalies (天时变异), Heterodox Sects (孽教), Supernatural Anomalies (妖异), Monstrous Portents (妖孽).

Table A11: Impacts of the Routine and Non-Routine Knowledge on State Building: Alternative Classification using Text Embedding

	Traditional State Building			State Modernization (Since 1842)		
	Land Tax 1820 (logged)	Government Schools 1820	Imperial Courier Stations 1820	Telegraph Stations 1881–1911	Post Offices 1896–1911	Modern Schools 1909 (logged)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A						
Number of Routine Topics	0.0005	0.0027	0.0042	0.0012	0.0020	0.0002
	(0.0002)**	(0.0011)**	(0.0019)**	(0.0005)**	(0.0008)**	(0.0002)
Adjusted R^2	0.723	0.598	0.210	0.310	0.506	0.615
Panel B						
Number of Non-routine Topics	0.0007	0.0029	0.0049	0.0015	0.0024	0.0003
	(0.0002)***	(0.0012)**	(0.0022)**	(0.0006)**	(0.0008)***	(0.0003)
Adjusted R^2	0.725	0.597	0.211	0.314	0.508	0.616
Controls in Both Panels						
Number of Counties	Yes	Yes	Yes	Yes	Yes	Yes
Population Density	Yes	Yes	Yes	Yes	Yes	Yes
Agricultural Suitability	Yes	Yes	Yes	Yes	Yes	Yes
Distance to Political Centers	Yes	Yes	Yes	Yes	Yes	Yes
Distance to the Coast	Yes	Yes	Yes	Yes	Yes	Yes
Terrain Ruggedness	Yes	Yes	Yes	Yes	Yes	Yes
Treaty Port				Yes	Yes	Yes
Province FEs				Yes	Yes	Yes
Observations	176	178	178	180	180	180

Notes: All are the same with Table 6 except we distinguish the routine and non-routine topics using the cosine similarity between each gazetteer topic and the 25 routine topics from the Chorography (see the text for details). Robust standard errors are reported in parentheses. *, ** and *** indicate the level of significance at 10%, 5% and 1%, respectively.