

Merit-based Recruitment and Government Performance: Evidence from Random Assignment of Administrators *

Kevin Zhengcheng Liu[†] Xiaoming Zhang[‡]

Job Market Paper

November 23, 2023

[Click here for the latest version](#)

Abstract

How to recruit managers is a central question for public sector organizations. Exploiting a unique random assignment of administrators to counties in imperial China, we study how merit-based recruitment of managers affects government performance in multiple tasks. Combining personnel records and administrative archival data, we find that merit-recruited administrators through civil service exams (as opposed to other non-merit-based recruitment) performed better in public goods provision and tax collection, and were more likely to be recognized for their achievements by the local public. Evidence suggests that these administrators were associated with better management of subordinates. Examining their dual role as judges, we use criminal case data to show that cases judged by merit-recruited administrators had shorter durations, indicating enhanced judicial efficiency. Finally, merit-recruited administrators were less likely to come from elite families and have connections to incumbent top officials. Our results highlight the importance of competitive and open recruitment as a cornerstone of effective public organizations.

*We thank Yiming Cao, Yu Hao, Rongzhu Ke, James Kung, Bingjing Li, Gedeon Lim, Zhuo Nie, Cristóbal Otero, Sangyoon Park, Gary Richardson, Uta Schönberg, Jin Wang, Yanhui Wu, and seminar participants at University of Hong Kong, Peking University, Wuhan University, NUS Applied Economics Student Workshop, Chinese Political Science Speaker Series, and International Symposium on Quantitative History for helpful comments and discussion.

[†]University of Hong Kong. Email: kevin_liu@connect.hku.hk. Web: zhengcheng-liu.weebly.com

[‡]Zhejiang University. Email: zhang.xiaoming@zju.edu.cn

1 Introduction

Recruiting the right people is one of the central problems for both private and public organizations.¹ Recruitment is particularly crucial in the public sector where financial incentives are typically absent. The selection of bureaucrats is of great importance as bureaucracy plays a key role in development and state capacity (Finan et al., 2017; Besley et al., 2022). Historically, nepotism and patronage dominated the selection of bureaucrats and still pervade developing countries (Grindle, 2012).² Efforts to strengthen government effectiveness often involve introducing open and competitive recruitment, commonly through civil service exams (World Bank, 2020). Such meritocracy in selection is viewed as an essential cornerstone of effective public organizations (Weber, 1922).

The promise of such merit-based recruitment is that it could curb favoritism and select high-performing individuals. However, rule-based exams restrict the information benefit of network-based recruitment (Dustmann et al., 2016; Pallais and Sands, 2016; Voth and Xu, 2022). Exam scores may not predict on-the-job performance. Critics also argue that meritocratic selection may also create barriers for disadvantaged individuals and reproduce or even aggravate existing elite advantages (Markovits, 2019; Sandel, 2020). Yet, despite the wide use of competitive exams in recruiting government personnel, direct evidence on the causal link between exam-based recruitment and government performance remains relatively limited because it is seldom random which jobs and positions were assigned to officers.³ Bureaucrats recruited from different channels are often sorted into different job types or strategically assigned to different positions, making it difficult to establish causality between individual backgrounds and observed performance outcomes.

In this paper, we provide causal evidence on the effects of merit-based recruitment for public managers on government performance in the context of imperial China. China's imperial bureaucracy provides an ideal setting to study this question. The Chinese empire is deemed the first in the world to have developed an open and competitive recruitment system for government officers through civil service exams (Finer, 1997; Fukuyama, 2011; Kung, 2021). By curbing the dominance of aristocrats in government staffing and recruiting talent into governments in a competitive and transparent manner, the civil service exam

¹ A rich literature has studied hiring practices in the private sector. See Autor and Scarborough (2008); Horton (2017); Hoffman et al. (2018) on standardized job testing. On informal hiring by personal network, see, e.g., Burks et al. (2015); Dustmann et al. (2016); Pallais and Sands (2016); Heath (2018).

² A cross-country survey of 23,000 civil servants in ten developing countries shows that 20% of civil servants indicate that political connections are important for their recruitment (Meyer-Sahling et al., 2018).

³ Cross-country comparison shows a strong positive correlation between the level of government performance and meritocratic recruitment, despite countries' regime type (Appendix Figure A1). Also see Evans and Rauch (1999) and Cornell et al. (2020).

played a crucial role in building an effective bureaucracy and maintaining the stability of Chinese autocratic rule (Zhao, 2015; Stasavage, 2020). Moreover, like in many developing countries today, other informal and discretionary recruitment coexisted in the entry paths into bureaucracy. We focus on county administrators, the chief managers of the lowest level of local governments, who were responsible for the provision of local public goods and tax collection. In addition, administrators multitasked by serving as both the executive and the local court judge.

We overcome the empirical challenge by exploiting a unique random assignment of administrators to counties in our context. Once selected, all to-be-appointed administrators with different recruitment backgrounds queued for vacancies together, and were then randomly assigned to counties. The key innovation of this design is that it allows us to isolate the exogenous variation in selection margin from other organizational policies, comparing officers who perform the *same task* in the *same position* under the *same incentive structure*. Frequent turnovers of administrators and the random assignment of successors generated exogenous variation in their recruitment backgrounds. We identify the causal effects of meritocratic recruitment by comparing outcomes in counties run by merit-recruited administrators through competitive exams (treatment) to those in counties under administrators recruited from other informal paths that relied more on connections and favoritism (control). In support of our research design, we conduct a series of checks showing that the two groups appear quite similar across a wide range of local characteristics.

Our study is underpinned by a novel dataset construction work. We collect 9,984 individual appointment records of county administrators from the provincial gazetteers. This dataset comprises the near universe of administrators across over 600 counties in eight provinces from 1736–1795, which allows us to identify merit-recruited administrators by their civil service exam qualifications. We link this dataset to position information from the *Roster of Government Personnel*. To assess government performance, we draw on various sources to construct a dataset across multiple local government tasks, including local public goods, fiscal performance from newly digitized tax collection records from the *Ministry of Finance Archive*, and individual records of misconduct and suspensions.

We begin by studying the impacts on public goods provision. We first focus on disaster relief as the outcome variable, a key local public good in the agricultural economy vulnerable to disaster. A sufficient and timely response to natural shocks and climate extremes is a crucial indicator of government responsiveness (Besley and Burgess, 2002). We find that counties run by merit-recruited administrators were more active in disaster relief provisions. Treated counties were 1.54 p.p. more likely to provide relief when a disaster hit the region, amounting to a 28% increase in responsiveness relative to control

counties. These results hold when controlling for officers' experience. Our findings are not driven by aid from higher levels of government.⁴ The effect is strongest during more severe disasters when the situation is more critical and an administrator's ability and effort are likely to be more important. Second, we find that treated counties were also more likely to construct schools. Given that promoting education was a non-core task with little significance in the evaluation of administrators, this result helps to mitigate the concern that administrators might reallocate effort from non-incentivized tasks to these incentivized core tasks, and thus the performance gain of one dimension might come at the expense of others (Holmstrom and Milgrom, 1991).

Next, we study performance in tax collection. Using the administrative data for tax deficits (i.e., shortfalls from the required revenue target) as measures of fiscal performance, we show that merit-recruited administrators performed better in achieving the revenue target. Moreover, we complement the results with a unique measure of subjective performance. We find that merit-recruited administrators were more likely to be listed in the *non-official* honorary record in local gazetteers in recognition of their achievements in enhancing local welfare. This implies a better perceived performance in the view of the local public.

As managers of county government agencies, administrators were responsible for supervising their subordinate officers. Consistent with the effects on government performance, we provide suggestive evidence of improved management practices. Using the misconduct records of lower level officers as a proxy of subordinate (non)performance, we find that subordinates under merit-recruited administrators were less likely to have misconduct records and be suspended, suggesting better management and monitoring. These results are in line with recent works showing that better managers can enhance subordinate outcomes (Lazear et al., 2015; Hoffman and Tadelis, 2021).

In the second part of the paper, we examine the dual role of county administrators as judges to study the impacts of merit-based recruitment on judicial performance. As the judge of the local court, administrators were in charge of civil dispute resolutions and criminal case judgments. We link unique data on criminal cases to our administrator data. Following the law and economics literature, we use case duration as a measure of efficiency performance (Ponticelli and Alencar, 2016; Kondylis and Stein, 2021). We begin by showing that individual officers can account for a significant amount of variation in case duration. By gradually including case characteristics, county, time, and administrator fixed effects in a regression on duration, administrator effects increase the R-squared by a

⁴In addition, we find no evidence of the differential occurrence of disaster. The incidence of disasters or extreme precipitations does not covary with administrator backgrounds.

sizeable magnitude (from 0.195 to 0.303). Examining the effect on judicial efficiency, we find that cases judged by merit-recruited administrators had a 0.6 months shorter duration, a 5% reduction relative to the mean. Moreover, the effects are not driven by the differential composition of cases or workload. Lastly, using the likelihood of being reversed by the “Supreme Court” (i.e., the Ministry of Justice) as a proxy for decision quality, we show that the improvement in efficiency does not come at the expense of worsened decision quality.

Why did meritocratic recruitment of administrators by civil service exams lead to better government performance in multiple tasks? In the last part of this paper, we discuss underlying mechanisms. We first consider whether the performance improvement may be due to their advantaged family backgrounds or family connections. Linking data on high-ranking officials, we identify family ties between previous top officials and administrators using individuals’ surnames and hometowns, and construct an elite family indicator. In contrast to the prediction of elite reproduction, we find that merit-recruited administrators were actually less likely to come from elite families and had fewer family connections to top officials.⁵ These results imply that competitive and open recruitment by exams helped to curtail nepotism in the selection into government. In addition, we also show that the exam-based connections (“quasi-classmate”) in superior officials played a limited role in explaining the effects on performance.

Second, we find no evidence that the link between promotion and performance indicators or experience changes markedly with officers’ recruitment backgrounds, implying that the performance gap is unlikely to be the result of differential promotion incentives.

Finally, merit-based recruitment could improve on-the-job performance by selecting higher-quality officers. This is plausible, given the extreme competitiveness of the exams that tested general cognitive ability in our context. We provide evidence consistent with the selection for quality by exploring the intensive margin of competitive selection in civil service exams. We find that administrators with higher-ranking exam qualification appear to have greater effects on performance. This is consistent with [Dahis et al. \(2020\)](#) who find that judges’ rank in entry exams is positively associated with their performance, and also in line with recent findings that cognitive ability measured by exam scores can predict entrepreneurial success ([Levine and Rubinstein, 2017](#); [Bai et al., 2021](#)).

Related literature. This paper contributes to several strands of the literature. First, we contribute to the literature on the personnel economics of the state and the impacts of bureaucracy for development ([Finan et al., 2017](#); [Besley et al., 2022](#)). Studies in this literature examine the impacts of financial and career incentives on performance ([Khan et al., 2016, 2019](#); [Bertrand et al., 2020](#); [Leaver et al., 2021](#); [De Janvry et al., 2023](#)) the role of

⁵ These patterns hold true within the same surname and when using a subsample of rare surnames.

public workforce selection (Dal Bó et al., 2013; Deserranno, 2019; Ashraf et al., 2020; Otero and Muñoz, 2022), and the internal allocation of bureaucrats (Limodio, 2021; Xu et al., 2021; Bergeron et al., 2022; Liu and Zhang, 2023). We contribute to this literature by studying a prevalent form of recruitment: competitive exams. This is a widely-used personnel policy in public sector organizations but has received less attention in the literature (Dahis et al., 2020; Moreira and Pérez, 2021). Our paper provides quasi-experimental evidence that open and competitive recruitment by civil service exam improved government performance in multiple tasks.⁶ While existing literature mainly studies frontline workers with specialized tasks, we instead focus on *managers* of the general local government agency. More broadly, we also speak to the literature on standardized job testing and screening tools versus human discretion in the hiring process (Autor and Scarborough, 2008; Hoffman et al., 2018; Li et al., 2020; Moreira and Pérez, 2022).

Second, we add to the nascent literature on the impact of managers and management in public sector organizations. A growing body of work has documented that management practices matter for organizational performance (Bloom et al., 2015; Rasul and Rogger, 2018; Bloom et al., 2020; Rasul et al., 2021). Another strand of work studies the impact of individual managers on organizational performance, such as hospitals (Janke et al., 2019; Otero and Muñoz, 2022), procurement offices (Best et al., 2023), social security agencies (Fenzia et al., 2022), and schools (Muñoz and Prem, 2022). We contribute to this literature by providing causal evidence on the impacts of managers of general administrative agencies. By showing the performance implications of competitive recruitment, we complement recent research on the recruitment reform of public managers (Muñoz and Prem, 2022; Otero and Muñoz, 2022).

Third, we contribute to the literature on meritocracy and civil service rules. In his seminal work, Weber (1922) highlights the importance of impersonal and meritocratic recruitment for bureaucratic organizations. Early cross-country comparison works suggest that meritocratic recruitment positively correlated with economic growth and bureaucratic performance (e.g., Evans and Rauch, 1999; Rauch and Evans, 2000). Several studies document the impacts of civil service reforms on election and fiscal outcomes, focusing on the transition of US bureaucracy from the “spoils system” to the “merit system” in the 19th and 20th centuries.⁷ There are also a few recent works showing that protection from politicians’

⁶ By providing evidence that merit-based recruitment reduced the share of officers with elite backgrounds and family ties, we also relate to studies on the cost of favoritism and patronage in public organizations (Fisman et al., 2018; Xu, 2018; Colonnelli et al., 2020; Li, 2020; Brassiolo et al., 2021; Riaño, 2022).

⁷ For instance, Folke et al. (2011) show that civil service reform reduced incumbent parties’ control of patronage and their reelection rate. Ujhelyi (2014) studies the introduction of the merit system on state governments’ spending. Bostashvili and Ujhelyi (2019) show that the reforms reduced political budget cycles.

influence improves bureaucrat performance (Ornaghi, 2019; Aneja and Xu, 2023).⁸

Despite the widespread use of open and competitive exams in recruiting civil servants, direct evidence on how they shape government performance remains relatively scarce. Two recent papers study the impact of entry exams on the performance of bureaucrats with specialized tasks and find mixed results (Moreira and Pérez, 2021; Dahis et al., 2020).⁹ Our paper provides causal evidence that civil service exams improved the performance of general government organizations in *multiple tasks*. We build on and differ from existing work in two ways. First, as Besley et al. (2022) observe: “The research that exists relies on variation in test scores *conditional* on being selected, thus leaving the extensive margin unexplored.” Comparing officers selected from exams versus ones from non-merit-based channels, we identify the causal effects of meritocratic recruitment using quasi-experimental variation in the *extensive margin*. Second, unlike previous work studying civil service reforms which usually involve bundles of policy changes, our setting enables us to clearly separate the exogenous variation only in recruitment margin from other organizational practices, comparing officers who perform the *same task* in the *same position* under the *same evaluation*. Our work also relates to recent work studying the effects of merit-based organizational practices such as promotion (Deserranno et al., 2021) and its broad socio-economic impacts beyond organization (Bai and Jia, 2016; Chen et al., 2020b; Moreira and Pérez, 2022; Xu and Adhvaryu, 2023).

2 Institutional background

2.1 Civil service exams in imperial China

China has a long history of centralized bureaucratic rule. The origins of China’s civil service examination (*Keju*) can be traced back to the seventh century. Later in the Song Dynasty (960–1279), the civil service exam was formally consolidated in the recruitment paths of bureaucrats and then kept in operation for nearly one thousand years (Elman, 2000). It was the world’s earliest meritocracy institution designed to recruit learned individuals into the government by competitive and open exams, nearly 800 years earlier than similar practices in European states (Finer, 1997; Stasavage, 2020; Kung, 2021). However, similar to

⁸ Ornaghi (2019) shows that the merit system in US municipal police department decreased crime rates. Aneja and Xu (2023) show that the Pendleton Act improved the delivery quality of the US Post Office, by the insulation of bureaucrats from party politics.

⁹ Moreira and Pérez (2021) study 1883 Pendleton Act with a focus on the US Customs Office. Comparing custom districts adopting exam-based hiring with those not, they find that the competitive exam improved the quality of new hires but had no impact on revenue collections. Dahis et al. (2020) estimate the individual fixed effects of Brazilian state judges on judicial efficiency and show document a positive association between entry exam rankings and performance.

many developing countries today, other informal recruitments like patronage, nepotism, and the sale of offices were also widespread and kept coexisting with civil service exams in the entry paths. In the 18th and 19th centuries, while the civil service exam had been well institutionalized in recruitment channels, less than 25% of the total officers in the formal bureaucracy were recruited via merit civil exam qualifications (Chen et al., 2020a).

The civil service exam was open to all men regardless of wealth or social status. A large sample estimation shows that 45% of 2nd-tier qualification holders and 37% of 1st-tier qualification holders in the Qing Dynasty came from a commoner background (Ho, 1962). The exam resulted in an impersonal and merit-based selection with a standardized format for testing and evaluation criteria (Elman, 2013). It ensured equality of opportunity: selection was only by the performance of the exam. Jiang and Kung (2021) show that a candidate's measurable ability significantly predicts success in the exam, while direct proxies for wealth do not. The government enforced strict measures to prevent cheating, corruption, and manipulation.¹⁰

The exams were extremely competitive. Appendix Figure A2 illustrates the levels of the exam and the corresponding passing rate. The county and prefecture level exams were the starting point with a passing rate of 1.5% by estimation (Elman, 2000, p. 143). Successful candidates would obtain the licentiate (*Shengyuan*) title, which granted some social privilege as members of the gentry class. However, this title did not qualify them for official positions in the government.

The provincial exam was the crucial step at the next level. The overall passing rate of this stage was as low as 6%. Those who passed the exam were awarded the recommended man (*Juren*) title, the 2nd-tier exam qualification. Normally all 2nd-tier qualification holder would continue to attempt the 1st-tier exam qualification (*Jinshi* or advanced scholar) in the national exam, the top layer of the exam system, with an overall passing rate of 17.7% (Kung, 2021). Under the personnel regulation of the Qing government, *Juren* and *Jinshi* were combinedly classified as one category of qualification (*Kejia*), which granted recruitment into the bureaucracy. We refer to these two exam qualifications as “merit exam qualifications.” Newly minted officers with 1st-tier qualification enjoyed priority in the wait list for assignment and more opportunities for central government positions.

The basic format of the civil service exam was essay-writing tests composed of three sessions. In session one, candidates were required to write logical argumentation essays on Confucian classics with specified and elaborate stylistic forms (the 8-legged essay).¹¹ From

¹⁰ For example, examiners read anonymous papers during assessments. To prevent examiners from recognizing a particular candidate through their handwriting, all exam scripts were hand-copied first, and examiners only graded the transcript of exam papers (Miyazaki, 1981).

¹¹ The eight-legged essay (*Baguwen*) was a style of essay for argumentation and reasoning writing with

a modern perspective, the classical curriculum encompasses philosophy, ethics, literature, statecraft, social sciences, and history (He, 1998). Session two tested administrative writings, such as proclamations and ordinances. Session three was policy questions on administrative practices and economic and political issues.¹²

What did the exam actually screen on? First, the exam was a test of refined literacy and clerical ability. This made a pragmatic sense, given the crucial importance of clear written communication in daily administration (Rowe, 2009, p. 46). Second, it was essentially a test of *general cognitive ability*. Being distinguished in the test required (1) logical reasoning and abstract thinking ability, (2) literary talent, and (3) memory ability (He, 1998). Overall, the civil service exam “based on nontechnical moral and statecraft theory was as suitable for selection of elites in China to serve the imperial state at its highest echelons of power as humanism and a classical education served elites in the nation-states of early-modern Europe” (Elman, 2013).

2.2 County governments and administrators

The local administration in imperial China had three layers: eighteen provinces, approximately 250 prefectures, and over 1,600 counties. The county was the lowest level of administration unit. The average county population was around 100 thousand. As the chief managers of county governments, administrators were the foremost authority within its jurisdictions who actually carried out policies and “governed” the country. They were in charge of the management of their offices. Their duties, among others, include the day-to-day operation of the county, the management and supervision of their subordinate officers, personnel management of staff, and the management of financial resources. In the sense that all these functions were under the administrator’s charge, he represented a “one-man government” (Ch’u, 1962).

County governments were responsible for various aspects of state functions, including local public goods provision, tax collection, policing, agricultural production, education, and judicial matters. The overall performance of county administrators was evaluated on a three-year basis. There were three major tasks. The first task was public goods provision. Among them, the most important local public good was disaster relief. Late imperial China was an agricultural economy where livelihoods were tightly linked with the weather. Frequent natural disasters endangered people’s lives and triggered social unrest. The gov-

a requirement of balanced clauses, phrases, and characters. It was a rhetorical style of argumentation in traditional China comparable to the Aristotelian syllogism (Elman, 2013, pp. 58–64).

¹² The top five themes of the policy question in the Zhejiang Provincial exam during the mid-eighteenth century were (1) political economy, (2) classic intellectual history, (3) literature, (4) geography, and (5) history (Elman, 2013, p. 254).

ernment played an active role in relief actions throughout the 18th century and operated a nationwide granary system for food distribution (Will, 1990). Local governments managed “ever-normal granaries” in every county and supervised charity granaries in major towns and rural community granaries (Will and Wong, 1991). Monitoring and reporting disasters, assessing economic losses, and delivering relief were core tasks of local governments. There were also non-core tasks in public goods provision, including promoting education and charity, encouraging farming, and maintaining other infrastructural works like irrigation and dams. These tasks were subject to limited monitoring and evaluation (Ch’u, 1962).

The second was tax collection. Each county was required to complete a predetermined tax collection task for land tax each year. The central government designated explicit and nearly fixed tax revenue targets for each county (Wang, 1973). There was also grain tax in kind with a similar fixed quota in some southern provinces. The collection of land tax and grain tax was subject to strict performance evaluation. Failure to meet the revenue target (i.e., tax deficit) was a veto criterion for career advancement and would lead to disciplinary action according to the deficit rate (Ch’u, 1962).

Third, county governments were responsible for the judiciary. In imperial China, the judiciary was not an independent system and was instead regarded as a part of administrative functions. The office of county administration was also the local court and the county administrator also served as the judge of the local court. Administrators adjudicated civil dispute cases and made initial decisions on the conviction and sentencing of criminal cases. For serious criminal cases, they were required to finish trial and make a judgment within a stipulated time limit. Failure to meet deadlines would lead to disciplinary punishment. All decisions on serious criminal cases involving death sentences were subject to recheck by higher judicial officers at the province level, and then submitted to the “Supreme Court”, i.e., the Ministry of Justice (MOJ). Finally, the MOJ would certify or overrule the decision on the case after careful review (Bodde and Morris, 1967). The judgment could be overruled by the MOJ due to wrongful conviction and sentencing, incorrect legal citations, and insufficient evidence (Zheng, 2022). County administrators, as judges of local courts, would face punishment if the MOJ reversed the case they judged.

2.3 The assignment rule

There were several staffing channels for county administrators. Newly minted bureaucrats with merit civil exam qualifications were important origins of administrators. Meanwhile, other informal and opaque recruitment were also prevalent, which relied more on personal connections and corruption. For instance, one could obtain appointment qualification by a discretionary recommendation from high-ranking officials (*Baoju*), by the sale of

office(*Juanna*), or if he was a son of a top official (legacy qualification, or *Yinsheng*) ([Ch’u, 1962](#)).

The Ministry of Personnel (MOP) used a rule-based process to randomly assign administrators to county vacancies. The rationale of this random assignment is to minimize rent-seeking and favoritism in the allocation of positions. ([Will, 2002](#)). All to-be-appointed officers already qualified for the administrator position would be grouped by qualification types and queued for vacancies together in a certain sequence.¹³ When new vacancies were created periodically due to idiosyncratic factors (e.g., transfer, promotion, suspension, illness leave, or retirement), they were allocated to officers in a principle of randomization. The assignment was held monthly, and the process can be divided into three steps ([Liu, 1988; Zhang, 2010](#)).¹⁴ Appendix Figure A3 illustrates the assignment process.

Step 1. Every month, the MOP counted the number of new vacancies and determined the size of the allocation bracket for this month.

Step 2. Waiting officers filled in the allocation bracket following a stipulated sequence of qualification categories. The resulting allocation bracket was a list of officers with certain weights of qualification types.

Step 3. Officers of this allocation bracket were randomly matched to vacant counties by a lottery.

This centralized assignment rule under the MOP applied to 75% of administrator positions in normal counties (*Xian/Zhou*) (1,066 out of 1,440 normal counties). Figure 1 shows the location of all counties by assignment methods.¹⁵ The assignment method remained fixed for the majority of positions with a few adjustments. Historian studies suggest that the random assignment process was strictly implemented in the Qing dynasty ([Will, 2002; Hu, 2021](#)). Nonetheless, in Section 4, we test whether this was the case using the real assignment records. The randomization in assigning officers to counties provides a unique setting to study how exogenous variation in officers’ recruitment background affects government performance.

¹³ For sequences of civil exam qualifications, officers were ordered by exam year and their ranking in the civil exam.

¹⁴ For a detailed description of the assignment procedure, see [Liu \(1988, pp. 27–32\)](#) and [Zhang \(2010, pp. 237–247\)](#).

¹⁵ The remaining 25% of positions were under a more discretionary process delegated to provincial governors. These positions were located in counties with an importance rating higher than two or belonged to special position categories. We do not include these counties in our analysis.

3 Data

Position information. We collect information on county government positions from the *Roster of Government Personnel (Jinshenlu)* in 1766. The *Roster* was issued by the Department of Selection under the Ministry of Personnel, which lists all civil positions and officers in local and central government. For each county, the *Roster* provides information on the type of appointment, the classification of regional governance rating, and special position categories (if any). We also collect the number of subordinate officer positions for each county government.

While the appointment method remained fixed for most counties, some counties experienced adjustments during our study period. For robustness, we also collect information on county positions in 1820 from the ChinaW Dataset (Yue et al., 2007) and classify county positions under the random assignment accordingly. We link the data at the two time points to define an alternative sample of 968 counties without changes in the appointment method.

Personnel data. We collect 9,984 individual appointment records, containing the near universe of county administrators for the period 1736–1795 (Emperor Qianlong’s reign) across 622 counties of eight provinces, namely Anhui, Fujian, Guangdong, Guangxi, Hunan, Shandong, Shaanxi, and Sichuan. Among them, 493 counties adopted the random assignment of administrators. This constitutes the main sample in our analysis, corresponding to 7,546 appointment records. We collect this data from various provincial gazetteers. Each gazetteer contains a chapter on the history of official appointments in a region in a standard format. Appendix Figure C1 shows a sample page of appointment records. The appointment records provide rich information on officers’ names, ethnicity, hometown, and their recruitment backgrounds. We use this information to divide officers into two broad types: merit-based recruitment by civil service exams and other informal and opaque recruitment channels. Specifically, we consider an administrator to be merit-recruited if the officers had merit civil exam qualifications (i.e., 2nd-tier qualification (*Juren*) or 1st-tier qualification (*Jinshi*)). Appendix Figure A4 shows the composition of county administrators by qualifications. The share of merit-recruited administrators is around 70% with small fluctuation over time. The vast majority of administrators (94%) were ethnically Han Chinese.

We supplement our administrator personnel data with a dataset for senior officials from the Authoritative Biographical Database constructed by the Institute of History and Philology, Academia Sinica. This database provides rich information on higher-ranking officials, including their positions, personal backgrounds, and career tracks. The coverage

is more systematic for officials ranking higher than 5th. We mainly use this data in constructing elite family backgrounds, family connections, and exam-based classmate networks.

Public goods. First, we obtain data on disaster relief from [Chen et al. \(2012\)](#), who compile comprehensive records of government-conducted disaster relief action. We use the information on the timing of each relief record and pin down the location delivering relief at the county level. We complement this data with relief actions involving higher levels of government (the provincial or central government). We extract this information from text mining the official correspondences that provincial top leaders sent to the central government, and code the timing and location of relief actions. We draw these text data from the catalogue of the First Historical Archives of China. Second, we collect data on the construction of schools from [Ji \(1996\)](#).

Disaster and climate data. For the occurrence of the disaster, we rely on the record of natural disasters from [Zhang \(2004\)](#). This data provides a record of various types of natural disasters in historical China, including drought, floods, plagues, locust infestations, hurricanes, earthquakes, and snowstorms. Drought and flood make up the majority of disasters. We aggregate all records at the prefecture-year level and construct a binary variable for disaster occurrence.

To construct a measure of disaster severity, we link disaster occurrence to extreme weather from the meteorological data from [Chinese Academy of Meteorological Sciences \(1981\)](#). This data provides annual precipitation records for 120 “observation stations”, each covering one to three prefectures. The raw data include a five-level index measuring the extent to which the rainfall deviates from long-average normal levels in the harvest season (i.e., from May to September): extreme rain (1), moderate rain (2), normal (3), moderate drought (4), and extreme drought (5).¹⁶ Combining the data on disaster occurrence and rainfall extremes, we code the severity of the disaster on a three-point scale. A level of 1 is assigned if the disaster occurs during normal rainfall, 2 if the disaster occurrence coincides with moderate rain or drought, and 3 if the disaster occurrence coincides with extreme rain or drought.

Tax collection records. To measure performance in tax collection, we use newly collected administrative data on public finance. We derive the universe of *Ministry of Finance Archive (Huketiben)* from the online catalogue of the First Historical Archives of China, with a total of 225,519 pieces. We extract records on tax deficit based on the text in the records’ title, with a total amount of 22,047 reports in the entire Qing dynasty. The raw

¹⁶ Levels 1 and 5: above or below the normal rainfall by 1.17 standard deviations; levels 2 and 4: above or below the normal rainfall by 0.33 s.d.; level 3: within 0.33 s.d. from the normal rainfall.

data is at a highly granular level corresponding to reports on tax deficit (i.e., shortfall from the required revenue target) in a given year and a given county, including the names of corresponding local officers who were accountable.¹⁷ This dataset has several advantages. First, it contains the name of the specific administrator who was responsible for the deficit in tax revenues. This information allows us to precisely match the tax deficit records with personnel data at the individual level. Second, since administrators faced predetermined, explicit, and fixed revenue targets, the presence of deficits in revenues is a direct measure of performance in tax collection. We use these records to construct an indicator variable for tax deficit as a measure of fiscal performance.

Misconduct and discipline records. We collect official information on officer misconduct and discipline records from the Administrative Archives of the Ministry of Personnel (*Liketiben*). We also derive data from the First Historical Archives of China by a similar approach. This data contains the records of all kinds of disciplinary actions on officers processed by the Ministry of Personnel, including records of misconduct, impeachment charges, fines, suspensions, and dismissals, among other punishments. Matched with administrator personnel data by positions and officer names, we obtain a total of 1,244 misconduct and discipline records. This provides an individual-level measure capturing the overall performance of officers evaluated by the organization. Moreover, we also collect 1,826 misconduct records for the subordinate officers of county administrators from the same source.

Appendix Table A1 provides the descriptive statistics for the main variables in our sample of analysis.

4 Identification strategy

Our empirical strategy leverages the unique random assignment rule for county administrators. Periodical turnovers of administrators combined with the random assignment of successors generated exogenous time and cross-sectional variation in their recruitment backgrounds. We identify the causal effects of merit-recruited administrators by comparing counties under administrators recruited from civil service exams (treatment group) with those under administrators from other informal entry paths, which relied more on personal connections and patronage (control group). Our sample is a county-year panel composed of 493 counties under the random assignment rule between 1736–1796. We estimate the following regression model:

¹⁷ See Appendix C.1 for examples of tax deficit records.

$$Y_{ct} = \beta \times \text{Merit}_{i(ct)} + \theta_t + \lambda_c + \epsilon_{ct} \quad (1)$$

where Y_{ct} is the outcome of county c in year t . $\text{Merit}_{i(ct)}$ is a dummy equal to one if the county's administrator i was merit-recruited by civil service exams (as opposed to other informal channels), measured by whether the officer held merit civil exam qualification. θ_t denotes year fixed effects controlling for common temporal shocks. By including county fixed effects λ_c , we are comparing administrators working in the same position. The standard errors ϵ_{ct} are clustered at the county level. β identifies the effect of merit-recruited administrators.

Identification checks. Under the assumption that the assignment of administrators with different backgrounds to counties was exogenous to pre-existing county characteristics, this strategy allows us to estimate the causal impact of merit-recruited administrators. We conduct a series of checks to validate the identifying assumption. First, using the appointment records, we conduct balance checks comparing the characteristics of treated counties to those of control counties. Table 1 presents the results testing whether treated counties differ from control counties in various characteristics, including area, longitude, latitude, terrain ruggedness, distance to the provincial capital and Beijing, population, land tax quota, agricultural suitability, storage of local granary, the height of city wall, number of city gates, and number of subordinate officers. Treated counties appear no different from control counties along a wide range of characteristics with insignificant p-values (column 4).

Second, we examine whether the distributions of county characteristics are balanced across two groups. Figure 2 plots the distribution of land tax quota, population, distance to the provincial capital, granary storage, and rice and wheat suitability. As the figures show, counties in the merit group and control group exhibit extremely similar distribution patterns, consistent with the random nature of the assignment process.

Third, we further corroborate the random assignment pattern with a more demanding balance test. We first divide all assignments into 60 cohorts by the year of appointment. Then we conduct t-tests separately for each cohort to examine whether the characteristics (land tax quota, population, distance to the provincial capital, granary storage, and rice and wheat suitability) of counties in merit group differ from those in the control group. Figure 3 plots the distribution of the corresponding t-statistics from these t-tests. We find that the majority of t-statistics fall in the range threshold corresponding to 95% confidence interval, strongly supporting the random nature of administrator assignments.

Fourth, we can also test the validity of this randomization in the time dimension by examining the correlation between the assignment of merit-recruited administrators and

outcome variables prior to the turnover. Appendix Table A2 shows the results. For each appointment, we regress the indicator for merit-recruited administrators on the outcomes of interest measured in one, two, and three years before the turnover in columns 1, 2, and 3, respectively. In columns 4 and 5, we regress against the mean over the predecessor's tenure. The results show the absence of pre-trends. Overall, the patterns in the data strongly indicate that the randomization of the administrator assignment was successful.

5 Merit-based recruitment and government performance

5.1 Main results

Relief provision. We start by studying the impacts of meritocratic recruitment on local public goods provision. We first focus on disaster relief as a performance measure, which was the most important local public good in our context. Since climate extremes and disasters represent a serious threat to livelihoods in the agricultural economy, a sufficient and timely response to natural shocks is also a crucial indicator of government responsiveness (Besley and Burgess, 2002).

Columns 1–3 of Table 2 present the results on relief provision from estimating equation 1. The dependent variable is a dummy equal to one if the county received disaster relief. Column 1 shows the baseline result using the full sample, controlling for an indicator for disaster occurrence to capture the need for government intervention. We find that counties under merit-recruited administrators increased the probability of relief by 0.65 percentage points. The estimated effect is substantial in magnitude, amounting to a 22% increase relative to the mean. In column 2, we further focus on the sample when a disaster hit the region to study whether merit-recruited administrators were associated with greater responsiveness to disaster. We find that counties under merit-recruited administrators were 1.54 p.p. more likely to provide relief. As a benchmark, compared to the average relief probability of the control group (5.48%), this amounts to a 28% increase in responsiveness. In column 3, we find similar results after further including administrators' tenure year fixed effects to control for the potential effect of experience.

One natural concern is that the increase in relief provision might reflect administrators' lobbying for aid from higher levels of government instead of their own performance in relief policy. We provide evidence suggesting that this is unlikely to be the case. In Appendix Table A3, we find that merit administrators had no impact on the likelihood of relief actions involving the provincial or central government. The effect is small in magnitude and statistically insignificant. In addition, our findings are not driven by differences in the demand side. In Appendix Table A4, we use a range of measures of disaster occurrence

and extreme weather as outcomes. We find no evidence that merit-recruited administrators were associated with higher or lower incidences of disaster, drought, flood, precipitation level, the likelihood of extreme rainfall, or disaster occurrence by severity. All estimates are insignificant and very small in magnitude.

Disaster severity. We can also examine the effect on relief provision by the severity of the disaster. A more catastrophic disaster means that the situation is more complex and the relief tasks are more demanding, for which an administrator’s ability and effort is likely to be more important. To examine the heterogeneity in disaster severity, we interact the disaster severity with dummies for merit-recruited administrators and non-merit administrators separately and estimate the following specification:

$$Y_{ct} = \sum_{s=1}^3 \beta_s \cdot Disaster_{cts} \times Merit_{i(ct)} + \sum_{s=1}^3 \alpha_s \cdot Disaster_{cts} \times Non-merit_{i(ct)} + \theta_t + \lambda_c + \epsilon_{ct} \quad (2)$$

where β_s and α_s estimate the relief responsiveness when a disaster of severity s occurred for merit and non-merit administrators, compared to the case when disaster is absent.

Figure 4 presents the results of estimating equation 2. There are two takeaways from this figure. First, the likelihood of relief increases with disaster intensity. Second, compared to the control group, the impacts of merit-recruited administrators are stronger when the disaster is more severe. While merit-recruited and non-merit administrators show similar effects under modest disaster (severity 1), merit-recruited administrators increase the relief responsiveness to extreme disaster (severity 3) by 3.2 p.p. (from 10.5 p.p. to 13.7 p.p.), amounting to a 30.5% increase.

School construction. Second, we study administrators’ role in promoting education, which was also a declared responsibility of county governments but was given small weight in the evaluation of administrators. One potential concern is that the performance improvement in one task could come at the expense of other less-incentivized tasks. Administrators might reallocate effort from non-core tasks to those incentivized core tasks—as theory and recent empirical evidence on multitasking problems show that agents will exert effort for incentivized tasks and substitute effort away from non-contractable tasks (Holmstrom and Milgrom, 1991; Giné et al., 2022).

We examine the impacts on the construction of schools in columns 4 and 5 of Table 2. We find that merit-recruited administrators were also more likely to construct schools. The estimate is significant with a sizable magnitude, amounting to a 32% increase from the average probability. This result suggests that the performance gains in core tasks did not undermine other less-incentivized tasks.

Tax collection. We then look at fiscal performance. As the tax agents of the crown,

tax revenue collection was a crucial responsibility of county administrators, as well as a key dimension in their performance evaluation. The central government designated a fixed aggregate land tax quota for each county, which was the explicit revenue target faced by administrators every year. Using the tax collection records from the *Ministry of Finance Archive* matched with the administrator data, we measure fiscal performance using an indicator for tax revenue deficit, meaning that the administrator failed to meet the revenue target. Table 3 reports the results. We do not observe effects on tax deficit in the full sample (columns 1), potentially because revenue generation was under strict evaluation and most administrators could meet the revenue target despite their capability. We expect that administrators would have a greater impact in areas where tax collection was more challenging. To this end, we divide our sample counties into groups with higher and lower difficulty in taxation tasks, based on the central government's classification of whether a county was characterized by demanding taxation tasks. Consistent with this view, we find significant effects concentrated in these counties where fulfilling tax tasks was more difficult (columns 4 and 5). Merit-recruited administrators were 1.21 percentage points (or 44%) less likely to have deficits in tax revenues, indicating better performance in tax collection.

5.2 Additional outcomes

Honorary record. We complement the above performance outcomes with an additional measure of subjective performance. We make use of the chapter on officials' achievement in the local gazetteers, which is a list of officers who made remarkable contributions to the local welfare, along with a brief description of their deeds and achievements when serving in the locality. Appendix Figure C2 shows a sample page of the honorary records. Typical description includes their activities in promoting local public welfare, including providing relief to coping disasters, building schools and irrigation facilities, fighting crime and bandit, crackdown on subordinate corruption, as well as recognition of their personal integrity. Since the local gazetteers were *non-official* publications mainly compiled by local people, we interpret being listed in the honorary record as a subjective evaluation of individual performance in the view of the local public. In columns 1 and 2 of Table 4, we examine the effect on subjective performance at the appointment level, using a dummy equal to one if the administrator was listed in the honorary record. We find that merit-recruited administrators were 1.7 p.p. more likely to receive the honorary record (a 24% increase)—leaving their names in the history book, consistent with results on objective performance outcomes we documented above.

Management practices. We also investigate the impact on management practices. Re-

cent literature on the economics of management shows that better managers improve the outcomes and productivity of subordinates in private firms (Lazear et al., 2015; Hoffman and Tadelis, 2021). Leaders of public organizations can also affect organizational performance by management practices (Bloom et al., 2020; Fenizia et al., 2022; Otero and Muñoz, 2022). As managers of county government agencies, administrators lead and supervise subordinate officers in daily operations. We examine the management implication by looking at the impacts of merit-recruited administrators on the performance indicators of lower-level subordinate officers. In a similar vein as Bertrand et al. (2020) and Riaño (2022), we use misconduct and suspension records of subordinate officers as a proxy of (non)performance.

Columns 3 and 4 of Table 4 present the results. We find that subordinate officers under merit-recruited administrators were less likely to have any misconduct records (column 1). In column 2, we examine suspension, the severest disciplinary punishment. Subordinates under merit-recruited administrators were 0.22 percentage points less likely to be suspended. These findings are intriguing, considering the fact that administrators had limited power to make personnel decisions on their subordinates since their appointment, rotation, and promotion were controlled by the central government. It is likely that merit-recruited administrators could monitor and motivate their subordinates more effectively or better match subordinates with tasks.

5.3 Robustness checks

We conduct a range of robustness checks to corroborate our findings on multiple government performance outcomes. First, we drop all counties that experienced changes in appointment method during 1776–1820 and use an alternative sample of 440 counties constantly under the random assignment rule. Appendix Table B1 shows that the effects of meritocracy on disaster relief, tax collection, non-incentivized tasks, and subjective performance evaluation are all similar using the alternative county sample. Second, Appendix Table B2 shows that our results are robust to dropping counties with few turnovers. Third, our findings hold when comparing the performance among administrators taking office in the same year by including assignment year fixed effects (Appendix Table B3).

Our results are also robust to other administrators' individual characteristics. In Appendix Table B4, we show that the results are robust to dropping officers with previous prestigious experiences in the Hanlin Academy, who were likely to have better career prospects and strong connections in the central government. Our findings also remain unchanged when controlling for officers' ethnicity (Appendix Table B5).

6 Merit-based recruitment and judicial performance

In this section, we explore the dual role of county administrators as judges and study the effects of meritocracy on judicial performance. As judges of the local court, administrators were also in charge of adjudicating civil disputes and making conviction and sentencing decisions for criminal cases. They had a critical bearing on both the efficiency and quality of the judiciary system.

6.1 Criminal case data

To study the implications for performance in the judiciary, we use unique criminal case data from the Case Summary Books (*Huiti Huangce*) from [Chen et al. \(2020c\)](#). Starting from 1745, the Ministry of Justice conducted yearly summaries and statistics of all serious criminal cases involving the death sentence. The resulting publications are the Case Summary Books, which contain the universe of serious criminal cases with a short text summary for each case in a standard format. This data provides case-level information on crime types, sentence results, the number of criminals, the use of violence, and other case characteristics. During our study period, the data is available for 17 separate years, from 1744 to 1794. We link our administrator personnel data to the criminal case data based on the county and date of cases. Our final data covers 4,766 cases, of which 3,139 cases took place in counties with random assignment of administrators. Homicide cases account for the largest proportion (77%) among the total cases. Appendix Figure [A5](#) shows the number of cases by year.

To measure the efficiency of the judicial process, we follow the literature on the judiciary ([Ponticelli and Alencar, 2016](#); [Kondylis and Stein, 2021](#)) to use case duration as a measure of performance. For each case, we record the date of the crime and the date of the decision by local courts. We calculate the duration of each case (in months) by the gap between these two dates. It is worth mentioning that the speed of the judicial process was also subject to the performance evaluation of county administrators. For serious criminal cases, they were required to finish the trial and make a judgment within a stipulated time limit. Failure to meet deadlines would lead to disciplinary punishment ([Ch’u, 1962](#)). In our case data, the average case duration is 12.6 months. Appendix Figure [A6](#) plots the distribution, showing a large variation in duration time.

6.2 Effects on judicial efficiency

Before formally testing the effects of meritocracy on judicial efficiency, we first study the extent to which variation in case duration can be explained by administrators’ individual

effects. Using an approach similar to [Bertrand and Schoar \(2003\)](#) and [Fenizia et al. \(2022\)](#), we compare the adjusted R-square from regressions of case duration when we gradually include case characteristics, county, time, and administrator fixed effects. Appendix Table [A6](#) reports the results. In column 3, we include a set of case characteristics, county fixed effects, and year fixed effects. After adding administrator individual fixed effects, the adjusted R-square increases from 0.195 to 0.303 (column 4), implying that administrator effects account for a sizable variation in judicial efficiency. Formally, an F-test strongly rejects the null hypothesis that all the administrator effects are zero (p-value=0.00).

We proceed to examine the effects of meritocratic recruitment on judicial performance by estimating the following case-level regression:

$$Duration_j = \beta \times Merit_{i(j)} + \theta_t + \lambda_c + \delta X_j + \epsilon_j \quad (3)$$

where $Duration_j$ is the duration of case j , measured in months. $Merit_{i(j)}$ is a dummy equal to one if the judge (i.e., the administrator) of case j had merit civil exam qualification. θ_t denotes year fixed effects and λ_c denotes county fixed effects. X_j is a vector of case characteristics.

Table [5](#) reports the case-level results. Cases judged by merit-recruited administrators had a 0.6 month shorter duration, a 5% reduction from the mean duration. The results are robust to controlling for administrators' experience (column 2). In column 3, we further control for a wide range of case characteristics, including the number of criminals, the indicator for a probationary sentence, sentence level, charge types, the use of violence, and whether family relatives were involved. The coefficient of interest remains nearly unchanged, suggesting that the decrease in case duration is unlikely to be driven by differences in case composition.

6.3 Validity checks

Case composition and workload. One potential concern is that the higher judicial performance at the case level might be driven by differential case workload or case composition. Merit-recruited administrators might face a lighter case workload, or they selectively judged simpler cases, resulting in faster judicial processes at the case level. First, Appendix Table [A7](#) compares the characteristics of cases judged by merit-recruited administrators and those judged by others. Simple t-tests show that the cases of the two groups were quite similar. Second, we more formally test whether the number and types of cases vary with administrators' recruitment backgrounds in Table [A8](#). In column 1, we aggregate the number of cases at the county-year level. We find no evidence that merit-recruited ad-

ministrators had a lighter workload. The estimate is small in magnitude and insignificant. In columns 2-5, we show that recruitment background does not relate to the number of criminals, sentencing level, or indicator for case types.

Decision quality. Did the improvement in judicial efficiency sacrifice decision quality? We also investigate another potential concern of a trade-off between speed and quality. A natural measure of decision quality is the likelihood that the judgment got appealed and reversed in higher courts (Coviello et al., 2015; Kondylis and Stein, 2021). Unfortunately, our criminal case data from the Case Summary Books does not provide such information. To test the effects on the quality of judicial decisions, we make use of another case data. We digitize the archive of the Ministry of Justice (*Xingketiben*) accessed from the First Historical Archives of China. This data contains comprehensive records of criminal cases reviewed by the Ministry of Justice (MOJ) in a more scattered way, allowing us to track the evolution of the case from the submission of the initial decision by the local court, the review by the MOJ, and the subsequent reversal and retrial.¹⁸

We measure the decision quality by the probability that the decision gets reversed. If administrators with merit qualifications hastened the judicial process and sacrificed quality by making rushed and sloppy judgments, we would expect they had a higher probability of case reversal. Appendix Table A9 presents the results. If anything, cases judged by merit-recruited administrators are associated with a lower likelihood of reversal, though the point estimate is statistically insignificant (p -value=0.29). The results suggest that increasing the speed of the judgment process did not lead to a decline in decision quality.

7 Mechanisms

Our results above indicate that meritocratic recruitment of administrators by civil service exams led to better government performance. In the last part of this paper, we investigate the underlying mechanisms driving the results. First, the performance improvement may be due to their advantaged family backgrounds or connections to incumbent top officials. Second, administrators may benefit from their quasi-classmate networks created from civil service exams. Third, merit-recruited administrators might face stronger promotion incentives. Finally, the competitive exams could select higher-ability bureaucrats.

¹⁸ As discussed in Section 2.2, the local court submitted its initial decision on criminal cases involving the death sentence to the MOJ for review. The MOJ could overrule the decision due to wrongful convictions, incorrect legal code citations, or insufficient evidence. Reversed cases would be sent back to the original local court for retrial. The local court would then resubmit the decision of retrial to the MOJ, which means that the same case would appear multiple times in the MOJ archives.

7.1 Family backgrounds and connections

Individuals who stood out from the competitive civil service exams could be more likely to come from existing elite families with wide connections in the bureaucracy. Thus, merit-recruited administrators might also benefit from their elite family background by the advantage of knowledge and skills in officialdom or direct support from family connections to other influential officials, resulting in better performance.

We examine whether administrators from meritocratic civil service exams were more or less likely to come from elite families. There is concern that meritocracy could create barriers for low-SES individuals, perpetuate the advantage of existing elites, and function as a tool for elite reproduction (Bourdieu and Passeron, 1990; Markovits, 2019; Sandel, 2020). Therefore, this test per se is interesting as it speaks to the question of whether meritocratic recruitment effectively curbs nepotism and breaks the dominance of incumbent elites in public organizations.

To explore this hypothesis, we construct a measure of the elite family background using the information of high-ranking officials with expanding time coverage (1644–1796). First, following recent works on family network (e.g., Cruz et al., 2017; Geys, 2017; Ferraz et al., 2020; Gagliarducci and Manacorda, 2020), we identify (presumed) family tie using individuals' surname and hometown county. Specifically, we assume officials belonged to the same lineage if they shared the same surname and were born in the same county. We define an elite family indicator equal to one if a “family” (i.e., a county-surname unit) ever had an official ranking higher than 5th.¹⁹ This practice yields 1,698 elite families from 2,258 top officials, spreading across 629 counties. Appendix Figure A7 shows the geographic distribution of elite families, with each circle drawn in proportion to the number of elite families by county. In the second step, we match the elite family status with individual administrators using their surname and hometown county.²⁰ To gauge the extent to which elite families are over-represented in the officialdom, we estimate the population share using the full sample of 536,915 offender records from the archive of the Ministry of Justice. Elite families made up 4.78% of the population, but overall accounted for 22.0% of administrators.²¹

We investigate the family background of administrators at the individual level in Table

¹⁹ The ranking of county administrator is 7th. A smaller number means a higher ranking.

²⁰ Our approach shares a similar fashion to Geys (2017) and Ferraz et al. (2020) using the surname of politicians to infer political dynasty family and Gagliarducci and Manacorda (2020) who identify family ties between officials and workers based on surname and birthplace.

²¹ When the sample is restricted to individuals with less common surnames which were ranked higher than 10 in terms of population share, the population share of elite families decreased to 2.54%, whereas they accounted for 19.4% of administrator.

6. Now, we refine the elite family definition by adding a temporal restriction. We consider an officer to have an elite family background if he belonged to a “family” that ever had a high-ranking official during the past 60 years. In contrast to the elite reproduction hypothesis, we find that merit-recruited administrators were 5.4 percentage points less likely to come from elite families, a 38% decrease relative to the mean (column 1 of panel A). This is consistent with the more patronage and nepotism nature of informal recruitment backgrounds. The result is similar when we compare administrators in the same entry cohort by including entry year fixed effects in column 2. Panel A in Figure 5 provides visual evidence complementing the regression results. Merit-recruited administrators had a smaller proportion of elite family backgrounds in each decade. One natural concern is that individuals with very common surnames are more likely to be falsely linked to a family. In column 3, we add surname fixed effects to compare administrators within the same surname. The estimate remains similar. To further address the potential bias due to frequent surnames, in panel B, we exclude individuals with top 10 most common surnames which have population shares higher than 1.5%. We observe a 6.7 p.p. reduction in the likelihood of elite family background, amounting to a large decrease by 55%. The magnitude of estimates is slightly larger after excluding common surnames, suggesting that our proxy for elite family does not substantially suffer from the potential bias due to frequent surnames.²²

In addition to the elite family background, we can also examine the time-varying family connection to incumbent top officials. Using a similar approach, we consider an administrator as family-connected if he had any family tie to an incumbent top official (ranking higher than 5th). Panel B in Figure 5 shows the share of connected administrators over the entry year by recruitment backgrounds. The figure presents a clear pattern that administrators from meritocratic recruitment were less likely to have family connections. Appendix Table A10 reports the regression results examining the current family tie. Panel A uses the full sample and panel B excludes the top 10 most common surnames. Column 1 in panel B shows that merit-recruited administrators were 2.5 percentage points (or 53%) less likely to have family connections to incumbent top officials at the time of the first appointment. We find similar results when extending the sample to their career of administrators at the individual-year level (columns 3–4).

Taken together, our results suggest that advantaged family backgrounds or stronger family connections are unlikely to drive the effects on government performance. Moreover, the results imply that impersonal and competitive recruitment by civil service exam did

²² This is consistent with discussion by [Gagliarducci and Manacorda \(2020\)](#) that the measurement error of family tie using surname would lead to attenuation bias in the estimates.

help to curb nepotism and patronage in the selection into the bureaucracy.

7.2 Exam-based connections

Second, we consider another type of connection created by the civil service exam system. In our context, successful examinees of the same civil service exam would forge a “quasi-classmate” relationship with each other (Miyazaki, 1981). This personal network remained coherent and could play an important role during their political career (Bai et al., 2023). Merit-recruited administrators may have more support from their exam-based networks within the bureaucracy, especially when their connected classmates worked as their superiors in the same province.

To assess the role of exam-based networks in explaining government performance, we examine whether merit-recruited administrators with more exam connections to superior officials in the province had larger effects. Now we focus on the subsample of merit-recruited administrators with 1st-tier exam qualification, for which the information on national-level exams is available.²³ We measure the strength of exam networks by an indicator for the presence of classmates in the superior officials working in the same province. We interact the merit dummy with exam connections to examine the heterogeneous effect of administrators who *happened to* work with exam-connected superior officials in the same province. Appendix Table A11 shows that the exam connection has small and insignificant effects on government performance outcomes. The sign of interaction coefficients shows no clear pattern. Similarly, we find no evidence that the number of exam-connected officials in central governments affected the performance of county governments (Appendix Table A12). These findings suggest that the exam-based network plays a limited role in explaining our results.

7.3 Promotion incentives

Another possible explanation is differential promotion incentives. Administrators from meritocratic recruitment might enjoy favorable career prospects. They could face implicit greater promotion incentives if merit civil exam qualifications are complementary to performance for career advance, inducing them to exert more effort.

We first note that it is unlikely that merit-recruited administrators faced greater promotion incentives or enjoyed favored career prospects due to two observations. First, career

²³ We do not have similar information for administrators with 2nd-tier exam qualification to measure the provincial-level exam networks. However, provincial-level exam networks were much smaller in size and less important than national-level exam networks. For instance, Bai et al. (2023) show that the impact of national-level exam networks is larger than other types in war mobilization.

progression of administrators to higher positions was primarily determined by performance and experience (seniority). In the formal personnel policy, recruitment backgrounds were not taken into account in promotion decisions (Ch’u, 1962, pp. 32–33; Wu, 2013, p. 245). Second, officers with merit-recruited backgrounds made up a smaller share of higher-ranking positions. Merit-recruited officers accounted for 70% of county administrators, but their share of prefecture governors fell to 42% (Appendix Figure A8).²⁴ The smaller proportion of merit-recruited officers in higher positions is in contrast with a more favorable prospect for climbing the hierarchy.

Nonetheless, we can test whether the link between promotion and performance measures varies with officers’ recruitment backgrounds. Greater incentives would predict a stronger link between performance and promotion. To capture promotion, we link the county administrator data with the personnel records of prefecture governors. We define promotion as a dummy variable equal to one if the administrator was promoted to a prefecture position in the next move. Promotion is a rare event: the average promotion probability at the county-year level is 0.4%.²⁵

We examine the relationship between promotion and performance indicators as well as their interaction with merit qualification in Table 7. We use the misconduct records as overall (non)performance indicators. Specifically, we look at indicators for whether the administrator had any misconduct records over the tenure. Column 1 shows that performance indeed significantly predicts promotion. Having misconduct records is associated with a 0.34 percentage points decrease in promotion probability, or an 82% decrease relative to the mean rate. However, we find no evidence that the link between promotion and performance measures changes markedly with administrators’ recruitment backgrounds. Column 2 shows that the estimate of interaction with merit qualification is small in magnitude and statistically insignificant. We also examine officers’ experience (measured by the cumulative years of service as administrator) as another relevant metric for promotion in columns 3 and 4. Similarly, we find that experience is positively correlated with promotion, but the link does not change with officers’ recruitment backgrounds. These results suggest that differential promotion incentives are unlikely to be a major driver of our results.

²⁴ We refer to prefecture as all prefecture-level jurisdictions including independent departments (*Zhilizhou*) and prefectures (*Fu*).

²⁵ At the individual level, 5.28% of administrators (365 out of 6,907 officers) in our full sample were eventually promoted to prefecture level positions.

7.4 Selection for quality

Finally, merit-based recruitment by civil service exams could select higher-quality officers for better on-the-job performance. This is plausible, given the extreme competitiveness of the exams that tested general cognitive ability. Recent empirical work in private sector leaders shows that cognitive ability measured by exam scores positively correlates with entrepreneurial success in different contexts ([Levine and Rubinstein, 2017](#); [Bai et al., 2021](#)). Although screening solely by standardized exams testing literacy and cognitive ability might not be the best way to recruit “merit”, such rule-based and impersonal recruitment can substantially limit the scope for favoritism and corruption, ultimately resulting in better selection.

We shed light on this mechanism by exploring the intensive margin of competitive selection in civil service exams. If the exams were able to select quality in a way consistent with on-the-job performance, we would expect the effects on performance to be greater for those who ranked higher in competitive exams. In our setting, merit civil exam backgrounds included 1st-tier and 2nd-tier qualifications, where the former required passing another higher layer of exam (with an overall passing rate of 17.7%) and was hence more selective. Officers with 1st-tier qualifications accounted for 32% of merit-recruited county administrators. We test this by interacting the merit dummy with an indicator for 1st-tier exam qualification. The interaction captures the differential effects between administrators with 1st-tier and 2nd-tier civil exam qualifications. [Table 8](#) reports the results. While we find no significant difference for relief provision, we observe starker heterogeneous effects across other three outcomes. Overall, we find that administrators with higher-ranking qualification appear to have greater effects, suggesting that selection on individual competence was an important mechanism driving the difference in performance. These results are in line with [Dahis et al. \(2020\)](#) who find that judges’ rank in entry exams is positively associated with their performance.

8 Conclusion

How to recruit managers and workers is a central question for any organization. Recruitment can be particularly relevant for public sector managers who oversee policy implementation and supervise the functioning of state agencies. Open, competitive, and rule-based recruitment is considered a fundamental feature of a well-functioning bureaucracy ([Weber, 1922](#); [Dahlström and Lapuente, 2017](#); [Besley et al., 2022](#)). However, nepotism, patronage, and corruption are pervasive in many nowadays countries and throughout

history.²⁶

We investigate the causal impacts of meritocratic recruitment by civil service exams on government performance in the context of imperial China, where, like in many developing countries today, informal recruitment coexisted with merit-based civil service exams in recruiting paths. Using a unique random assignment of administrators to counties, we provide a systemic account that meritocratic recruitment improved government performance in multiple tasks. Combining personnel data with various novel administrative archive data, we show that merit-recruited administrators from civil service exams (as opposed to other informal and discretionary recruitment) were more active and responsive in providing disaster relief. They also performed better in tax collection, increased school construction, and were more likely to receive honorary recognition for their achievements in promoting local welfare. Moreover, exploring the dual role of administrators as judges, we use criminal case data to show that meritocratic recruitment also improved judicial efficiency. As the world's earliest prototype of civil service exams, the Chinese imperial exam institution had a profound and multifaceted influence (Yu, 2005; Zhao, 2015; Bai and Jia, 2016; Chen et al., 2020b; Xu, 2022; Jia et al., 2021; Hao et al., 2022). Yet, we know little about its effectiveness in selecting high-performing bureaucrats—its original function. We provide the first rigorous quantitative examination of its impacts on government performance.

Moreover, we show that merit-recruited administrators were less likely to come from advantaged elite families and also less likely to have family connections to incumbent top officials, implying that civil service exams did help to curtail nepotism and patronage in the selection into the bureaucracy.

Our findings have some general implications. All organizations face both information and agency problems in the allocation of jobs, as they decide how much discretion and personal judgment to allow, as opposed to the extent to which they rely on impersonal and rule-based selection criteria like competitive exams. Such trade-off tends to vary across different contexts, depending on the level of selection (e.g., entry recruitment, internal posting, or promotion) and the mission of the organization. In the context of recruitment in bureaucratic organizations where rent-seeking and favoritism are likely to pervade, particularly in developing countries, open and competitive selection by civil service exams could be conducive to government performance. Our results suggest that the gains from exams outweigh the drawbacks, even if the exams mainly test general cognitive

²⁶ For example, before the twentieth century, the U.S. bureaucracy was ruled by the “spoils system,” under which hiring and firing were based on politicians’ discretion and used for bribing and rewards for parties. Appointments based on personal connections and systemic sale of public offices were also common practice in premodern European countries like France, Spain, and Britain (Fukuyama, 2011; Grindle, 2012).

ability rather than screening for specialized skills. Taken together, our results echo [Weber \(1922\)](#) seminal work and highlight the importance of impersonal, competitive, and open recruitment as a key building block of effective public organizations.

References

- Aneja, Abhay, and Guo Xu.** 2023. "Strengthening State Capacity: Civil Service Reform and Public Sector Performance during the Gilded Age." Working Paper.
- Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S. Lee.** 2020. "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services." *American Economic Review*, 110(5): 1355–94.
- Autor, David H, and David Scarborough.** 2008. "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments." *Quarterly Journal of Economics*, 123(1): 219–277.
- Bai, Chong-En, Ruixue Jia, Hongbin Li, and Xin Wang.** 2021. "Entrepreneurial reluctance: Talent and firm creation in China." National Bureau of Economic Research Working Paper 28865.
- 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.
- Bergeron, Augustin, Pedro Bessone, John Kabeya Kabeya, Gabriel Z Tourek, and Jonathan L Weigel.** 2022. "Optimal assignment of bureaucrats: Evidence from randomly assigned tax collectors in the DRC." National Bureau of Economic Research Working Paper 30413.
- Bertrand, Marianne, and Antoinette Schoar.** 2003. "Managing with style: The effect of managers on firm policies." *Quarterly journal of economics*, 118(4): 1169–1208.
- Bertrand, Marianne, Robin Burgess, Arunish Chawla, and Guo Xu.** 2020. "The glittering prizes: Career incentives and bureaucrat performance." *Review of Economic Studies*, 87(2): 626–655.
- Besley, Timothy, and Robin Burgess.** 2002. "The Political Economy of Government Responsiveness: Theory and Evidence from India." *Quarterly Journal of Economics*, 117(4): 1415–1451.
- Besley, Timothy, Robin Burgess, Adnan Khan, and Guo Xu.** 2022. "Bureaucracy and Development." *Annual Review of Economics*, 14: 397–424.
- Best, Michael Carlos, Jonas Hjort, and David Szakonyi.** 2023. "Individuals and Organizations as Sources of State Effectiveness." *American Economic Review*, 113(8): 2121–67.
- Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen.** 2015. "Does management matter in schools?" *Economic Journal*, 125(584): 647–674.
- Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen.** 2020. "Healthy business? managerial education and management in health care." *Review of Economics and Statistics*, 102(3): 506–517.

- Bodde, Derk, and Clarence Morris.** 1967. *Law in Imperial China: exemplified by 190 Ch'ing Dynasty cases (translated from the Hsing-an hui-lan), with historical, social, and juridical commentaries.* Cambridge, MA and London, England: Harvard University Press.
- Boix, Carles, Michael Miller, and Sebastian Rosato.** 2013. "A complete data set of political regimes, 1800–2007." *Comparative political studies*, 46(12): 1523–1554.
- Bostashvili, David, and Gergely Ujhelyi.** 2019. "Political budget cycles and the civil service: Evidence from highway spending in US states." *Journal of Public Economics*, 175: 17–28.
- Bourdieu, Pierre, and Jean-Claude Passeron.** 1990. *Reproduction in education, society and culture.* SAGE Publications.
- Brassiolo, Pablo, Ricardo Estrada Martinez, Gustavo Fajardo, and Julian Martinez-Correa.** 2021. "Family Rules: Nepotism in the Mexican Judiciary." Working paper.
- Burks, Stephen, Bo Cowgill, Mitchell Hoffman, and Michael Housman.** 2015. "The Value of Hiring through Employee Referrals." *Quarterly Journal of Economics*, 130(2): 805–839.
- Chen, Bijia, Cameron Campbell, Yuxue Ren, and James Lee.** 2020a. "Big data for the study of Qing officialdom: The China government employee database-Qing (CGED-Q)." *Journal of Chinese History*, 4(2): 431–460.
- Chen, Ting, James Kai-sing Kung, and Chicheng Ma.** 2020b. "Long live Keju! The persistent effects of China's civil examination system." *Economic Journal*, 130(631): 2030–2064.
- Chen, Zhenhan, Guoliang Xiao, and Zhengwen Xiong.** 2012. *Qingshilu jingjishi ziliao [Economic history materials from the Veritable Records]*. Beijing: Peking University Press.
- Chen, Zhiwu, Zhan Lin, and Kaixiang Peng.** 2020c. "The Statistical Characteristics of Serious Criminal Cases in Qing Dynasty." *New History [Xinshixue]*, Vol. 12, 101–149. Shekewenxian Press.
- Chinese Academy of Meteorological Sciences.** 1981. *Zhongguo jinwubainian hanlao fenbu tuji [Yearly Charts of Dryness and Wetness in China for the Last 500-Year Period]*. Beijing: Ditu Press.
- Ch'u, T'ung-tsu.** 1962. *Local Government in China under the Ch'ing.* Cambridge and London: Harvard University East Asia Center.
- Colonnelli, Emanuele, Mounu Prem, and Edoardo Teso.** 2020. "Patronage and selection in public sector organizations." *American Economic Review*, 110(10): 3071–99.
- Cornell, Agnes, Carl Henrik Knutsen, and Jan Teorell.** 2020. "Bureaucracy and Growth." *Comparative Political Studies*, 53(14): 2246–2282.
- Coviello, Decio, Andrea Ichino, and Nicola Persico.** 2015. "The inefficiency of worker time use." *Journal of the European Economic Association*, 13(5): 906–947.

- Cruz, Cesi, Julien Labonne, and Pablo Querubin.** 2017. "Politician family networks and electoral outcomes: Evidence from the Philippines." *American Economic Review*, 107(10): 3006–3037.
- Dahis, Ricardo, Laura Schiavon, and Thiago Scot.** 2020. "Selecting Top Bureaucrats: Admission Exams and Performance in Brazil." Working Paper.
- Dahlström, Carl, and Victor Lapuente.** 2017. *Organizing Leviathan: Politicians, Bureaucrats, and the Making of Good Government*. Cambridge: Cambridge University Press.
- Dal Bó, Ernesto, 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.
- De Janvry, Alain, Guojun He, Elisabeth Sadoulet, Shaoda Wang, and Qiong Zhang.** 2023. "Subjective performance evaluation, influence activities, and bureaucratic work behavior: Evidence from China." *American Economic Review*, 113(3): 766–799.
- Deserranno, Erika.** 2019. "Financial incentives as signals: experimental evidence from the recruitment of village promoters in Uganda." *American Economic Journal: Applied Economics*, 11(1): 277–317.
- Deserranno, Erika, Philipp Kastrau, and Gianmarco León-Ciliotta.** 2021. "Promotions and Productivity: The Role of Meritocracy and Pay Progression in the Public Sector." Working Paper.
- Dustmann, Christian, Albrecht Glitz, Uta Schönberg, and Herbert Brücker.** 2016. "Referral-based job search networks." *Review of Economic Studies*, 83(2): 514–546.
- Elman, Benjamin A.** 2000. *A cultural history of civil examinations in late imperial China*. CA: University of California Press.
- Elman, Benjamin A.** 2013. *Civil Examinations and Meritocracy in Late Imperial China*. Cambridge, MA: Harvard University Press.
- Evans, Peter, and James E Rauch.** 1999. "Bureaucracy and growth: A cross-national analysis of the effects of "Weberian" state structures on economic growth." *American Sociological Review*, 64(5): 748.
- Fenizia, Alessandra, et al.** 2022. "Managers and Productivity in the Public Sector." *Econometrica*, 90(3): 1063–1084.
- Ferraz, Claudio, Frederico Finan, and Monica Martinez-Bravo.** 2020. "Political Power, Elite Control and Long-Run Development: Evidence from Brazil." National Bureau of Economic Research Working Paper 27456.
- Finan, Frederico, Benjamin A Olken, and Rohini Pande.** 2017. "The personnel economics of the developing state." In *Handbook of Economic Field Experiments, Volume 2.*, ed. Abhijit Banerjee and Esther Duflo, 467–514. Elsevier.

- Finer, Samuel E.** 1997. *The History of Government from the Earliest Times: Ancient monarchies and empires*. Vol. 1, Oxford, UK: Oxford University Press.
- Fisman, Raymond, Jing Shi, Yongxiang Wang, and Rong Xu.** 2018. "Social ties and favoritism in Chinese science." *Journal of Political Economy*, 126(3): 1134–1171.
- Folke, Olle, Shigeo Hirano, and James M Snyder.** 2011. "Patronage and Elections in US States." *American Political Science Review*, 105(3): 567–585.
- Fukuyama, Francis.** 2011. *The Origins of Political Order: From Prehuman Times to the French Revolution*. New York: Farrar, Straus and Giroux.
- Gagliarducci, Stefano, and Marco Manacorda.** 2020. "Politics in the family: Nepotism and the hiring decisions of Italian firms." *American Economic Journal: Applied Economics*, 12(2): 67–95.
- Geys, Benny.** 2017. "Political dynasties, electoral institutions and politicians' human capital." *Economic Journal*, 127(605): F474–F494.
- Giné, Xavier, Ghazala Mansuri, and Slesh A Shrestha.** 2022. "Mission and the bottom line: Performance incentives in a multigoal organization." *Review of Economics and Statistics*, 104(4): 748–763.
- Grindle, Merilee S.** 2012. *Jobs for the Boys: Patronage and the State in Comparative Perspective*. Cambridge, MA: Harvard University Press.
- Hao, Yu, Zhengcheng Liu, Xi Weng, and Li-an Zhou.** 2022. "The Making of Bad Gentry: The Abolition of Keju, Local Governance, and Anti-Elite Protests, 1902–1911." *The Journal of Economic History*, 82(3): 625–661.
- Heath, Rachel.** 2018. "Why do firms hire using referrals? Evidence from Bangladeshi garment factories." *Journal of Political Economy*, 126(4): 1691–1746.
- He, Huaihong.** 1998. *Selection Society: a Study on Social Structure from Qin-Han to Late Qing Dynasty [Xuanju shehui: qinhanzhiwanqing shehuixintai yanjiu]*. Beijing: Sanlian Shudian.
- Hoffman, Mitchell, and Steven Tadelis.** 2021. "People management skills, employee attrition, and manager rewards: An empirical analysis." *Journal of Political Economy*, 129(1): 243–285.
- Hoffman, Mitchell, Lisa B Kahn, and Danielle Li.** 2018. "Discretion in Hiring." *Quarterly Journal of Economics*, 133(2): 765–800.
- Holmstrom, Bengt, and Paul Milgrom.** 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization*, 7: 24–52.
- Ho, Ping-Ti.** 1962. *The Ladder of Success in Imperial China: Aspects of Social Mobility, 1368–1911*. New York: Columbia University Press.

- Horton, John J.** 2017. "The effects of algorithmic labor market recommendations: Evidence from a field experiment." *Journal of Labor Economics*, 35(2): 345–385.
- Hu, Cunlu.** 2021. "Research on the System of Drawing Lots for Monthly Appointment in Qing Dynasty [Qingdai Yuexuan Zhiqian Zhidu Kaolu]." *Qing History Journal*, , (1): 42–51.
- Janke, Katharina, Carol Propper, and Raffaella Sadun.** 2019. "The impact of ceos in the public sector: Evidence from the english nhs." National Bureau of Economic Research Working Paper 25853.
- Jiang, Qin, and James Kai-sing Kung.** 2021. "Social mobility in late imperial China: reconsidering the "ladder of success" hypothesis." *Modern China*, 47(5): 628–661.
- Jia, Ruixue, Gérard Roland, and Yang Xie.** 2021. "A theory of power structure and institutional compatibility: China vs. Europe revisited." National Bureau of Economic Research Working Paper 28403.
- Ji, Xiaofeng.** 1996. *Zhongguo Shuyuan Cidian [A Compendium on the Chinese Academies]*. Hangzhou: Zhejiang Jiaoyu Chubanshe.
- Khan, Adnan Q, Asim Ijaz Khwaja, and Benjamin A Olken.** 2019. "Making moves matter: Experimental evidence on incentivizing bureaucrats through performance-based postings." *American Economic Review*, 109(1): 237–70.
- Khan, Adnan Q, Asim I Khwaja, and Benjamin A Olken.** 2016. "Tax farming redux: Experimental evidence on performance pay for tax collectors." *Quarterly Journal of Economics*, 131(1): 219–271.
- Kondylis, Florence, and Mattea Stein.** 2021. "The speed of justice." *Review of Economics and Statistics*, 1–46.
- Kung, James Kai-Sing.** 2021. "The World's First Meritocracy Through the Lens of Institutions and Cultural Persistence." In *The Palgrave Handbook of Comparative Economics*. 159–184. Springer.
- Lazear, Edward P, Kathryn L Shaw, and Christopher T Stanton.** 2015. "The value of bosses." *Journal of Labor Economics*, 33(4): 823–861.
- Leaver, Clare, Owen Ozier, Pieter Serneels, and Andrew Zeitlin.** 2021. "Recruitment, Effort, and Retention Effects of Performance Contracts for Civil Servants: Experimental Evidence from Rwandan Primary Schools." *American Economic Review*, 111(7): 2213–46.
- Levine, Ross, and Yona Rubinstein.** 2017. "Smart and illicit: who becomes an entrepreneur and do they earn more?" *The Quarterly Journal of Economics*, 132(2): 963–1018.
- Li, Danielle, Lindsey R Raymond, and Peter Bergman.** 2020. "Hiring as exploration." National Bureau of Economic Research Working Paper 27736.
- Limodio, Nicola.** 2021. "Bureaucrat Allocation in the Public Sector: Evidence from the World Bank." *The Economic Journal*.

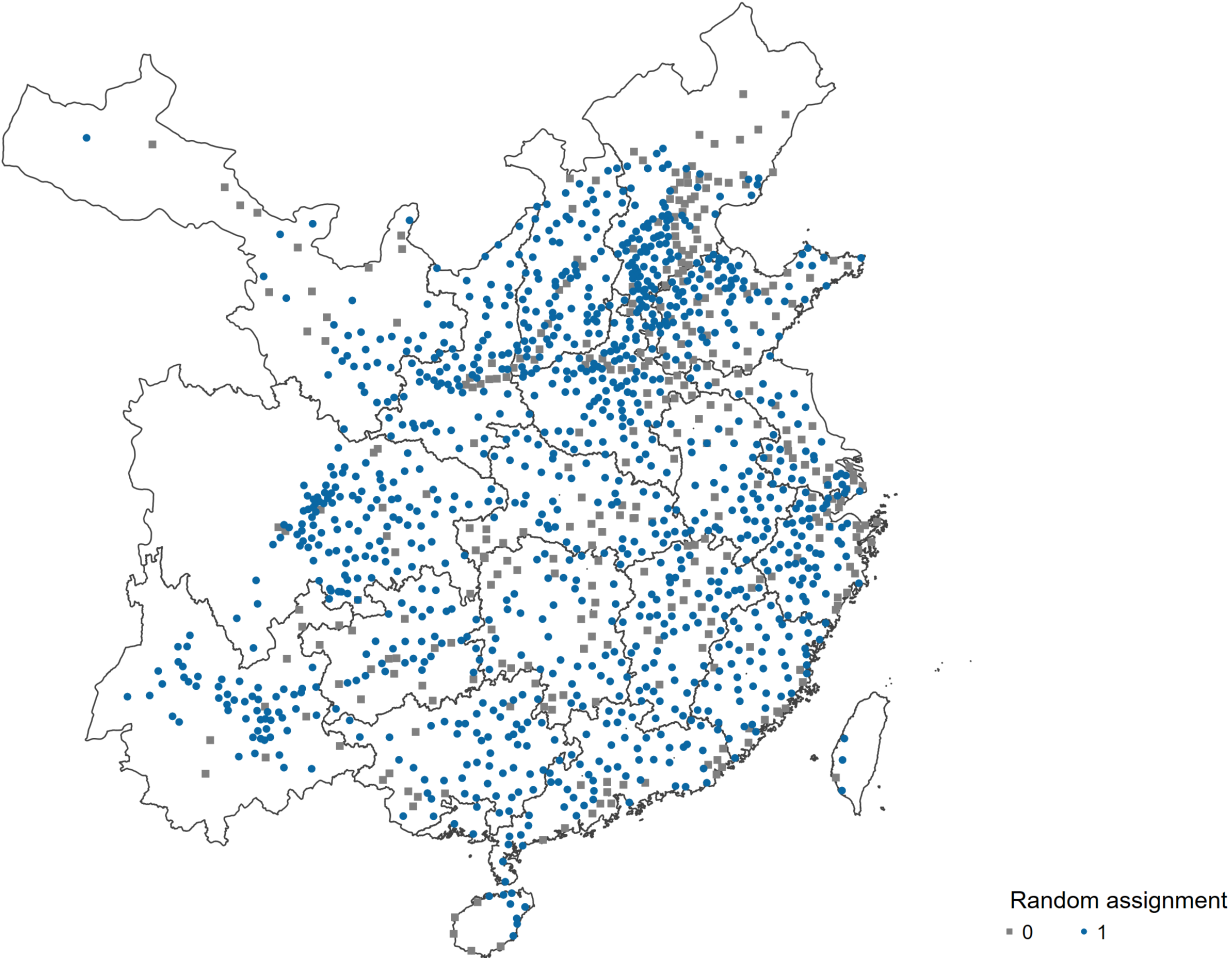
- Liu, Kevin Zhengcheng, and Xiaoming Zhang.** 2023. "Discretion, Talent Allocation, and Governance Performance: Evidence from China's Imperial Bureaucracy." Working Paper.
- Liu, Ziyang.** 1988. *A Study on Local Bureaucratic Institution in Qing Dynasty [Qingdai Difang Guanzhikao]*. Beijing: Zijingcheng Press.
- Li, Xuan.** 2020. "The Costs of Workplace Favoritism: Evidence from Promotions in Chinese High Schools." Working Paper.
- Markovits, Daniel.** 2019. *The Meritocracy Trap: How America's Foundational Myth Feeds Inequality, Dismantles the Middle Class, and Devours the Elite*. Penguin Press.
- Meyer-Sahling, Jan-Hinrik, Christian Schuster, and Sass Kim Mikkelsen.** 2018. *Civil Service Management in Developing Countries: What Works? Evidence From a Survey with 23,000 Civil Servants in Africa, Asia, Eastern Europe And Latin America*. UK Department for International Development (DFID).
- Miyazaki, Ichisada.** 1981. *China's Examination Hell: The Civil Service Examinations of Imperial China*. New Haven: Yale University Press.
- Moreira, Diana, and Santiago Pérez.** 2022. "Who Benefits from Meritocracy?" National Bureau of Economic Research Working Paper 30113.
- Moreira, Diana, and Santiago Pérez.** 2021. "Civil Service Reform and Organizational Practices: Evidence from the Pendleton Act." National Bureau of Economic Research Working Paper 28665.
- Muñoz, Pablo, and Mounu Prem.** 2022. "Managers' productivity and recruitment in the public sector: The case of school principals." Working Paper.
- Ornaghi, Arianna.** 2019. "Civil service reforms: Evidence from US police departments." Working Paper.
- Otero, Cristóbal, and Pablo Muñoz.** 2022. "Managers and public hospital performance." Working Paper.
- Pallais, Amanda, and Emily Glassberg Sands.** 2016. "Why the referential treatment? Evidence from field experiments on referrals." *Journal of Political Economy*, 124(6): 1793–1828.
- Ponticelli, Jacopo, and Leonardo S Alencar.** 2016. "Court enforcement, bank loans, and firm investment: evidence from a bankruptcy reform in Brazil." *Quarterly Journal of Economics*, 131(3): 1365–1413.
- Rasul, Imran, and Daniel Rogger.** 2018. "Management of bureaucrats and public service delivery: Evidence from the nigerian civil service." *Economic Journal*, 128(608): 413–446.
- Rasul, Imran, Daniel Rogger, and Martin J Williams.** 2021. "Management, Organizational Performance, and Task Clarity: Evidence from Ghana's Civil Service." *Journal of Public Administration Research and Theory*, 31(2): 259–277.

- Rauch, James E, and Peter B Evans.** 2000. "Bureaucratic structure and bureaucratic performance in less developed countries." *Journal of Public Economics*, 75(1): 49–71.
- Riaño, Juan Felipe.** 2022. "Bureaucratic nepotism." Working Paper.
- Rowe, William T.** 2009. *China's Last Empire: the Great Qing*. Vol. 6, Cambridge, MA: Harvard University Press.
- Sandel, Michael J.** 2020. *The Tyranny of Merit: What's Become of the Common Good?* New York: Farrar, Straus and Giroux.
- Stasavage, David.** 2020. *The Decline and Rise of Democracy: A Global History from Antiquity to Today*. NJ: Princeton University Press.
- Ujhelyi, Gergely.** 2014. "Civil Service Rules and Policy Choices: Evidence from US State Governments." *American Economic Journal. Economic Policy*, 6(2): 338.
- Voth, Joachim, and Guo Xu.** 2022. "Discretion and Destruction: Promotions, Performance, and Patronage in the Royal Navy." Working Paper.
- Wang, Yeh-Chien.** 1973. *Land Taxation in Imperial China, 1750-1911*. Cambridge, MA: Harvard University Press.
- Weber, Max.** 1922. *Economy and Society*. Berkeley: University of California Press.
- Will, Pierre-Etienne.** 1990. *Bureaucracy and Famine in Eighteenth-century China*. Stanford: Stanford University Press.
- Will, Pierre-Étienne.** 2002. "Creation, Conflict, and Routinization: The Appointment of Officials by Drawing Lots, 1594–1700." *Ming Qing Yanjiu*, 11(1): 73–121.
- Will, Pierre-Etienne, and Roy Bin Wong.** 1991. *Nourish the People: The State Civilian Granary System in China, 1650-1850*. Ann Arbor: University of Michigan Center for Chinese.
- World Bank.** 2020. *Enhancing Government Effectiveness and Transparency: The Fight Against Corruption*. World Bank Group.
- Wu, Yue.** 2013. *Zhongguo de juannazhidu yu shehui [China's system of office sale and society]*. Nanjing: Jiangsu Renmin press.
- Xu, Chenggang.** 2022. "The Origin of China's Communist Institutions." *The Cambridge Economic History of China*, , ed. Devin Ma and Richard von Glahn, 531–564. Cambridge University Press.
- Xu, Guo.** 2018. "The Costs of Patronage: Evidence from the British Empire." *American Economic Review*, 108(11): 3170–98.
- Xu, Guo, Marianne Bertrand, and Robin Burgess.** 2021. "Organization of the State: Home Assignment and Bureaucrat Performance." *Journal of Law, Economics, and Organization*, ewab022.

- Xu, Huayu, and Achyuta Adhvaryu.** 2023. "The Human Capital Effects of Access to Elite Jobs." *American Economic Journal: Applied Economics*.
- Yue, Zumou, G. William Skinner, and Mark Henderson.** 2007. "ChinaW Dataset: Cities, County Seats and Yamen (1820-1893)." Davis: University of California, Regional Systems Analysis Project.
- Yu, Ying-shih.** 2005. "The Functions and Significance of the Keju System in Chinese History." *Twenty-First Century*, 4–18.
- Zhang, Deer.** 2004. *Zhongguo Sanqiannian Qixiang Jilu Zongji [Comprehensive Compilation of Weather Records in Last Three Millenniums of China]*. Nanjing: Jiangsu Educational Press.
- Zhang, Zhenguo.** 2010. *Qingdai Wenguan Xuanren Zhidu Yanjiu. [The Study on the Civilian Designating System of Qing Dynasty]*. PhD dissertation, Nankai University.
- Zhao, Dingxin.** 2015. *The Confucian-Legalist State: A New Theory of Chinese History*. Oxford: Oxford University Press.
- Zheng, Xiaoyou.** 2022. *Renming Guantian: Qingdai Xinbu de Zhengwu yu Guanyuan [Life Matters: The Administration and Officials of the Ministry of Justice in the Qing Dynasty]*. Shanghai: Shanghai Renming Press.

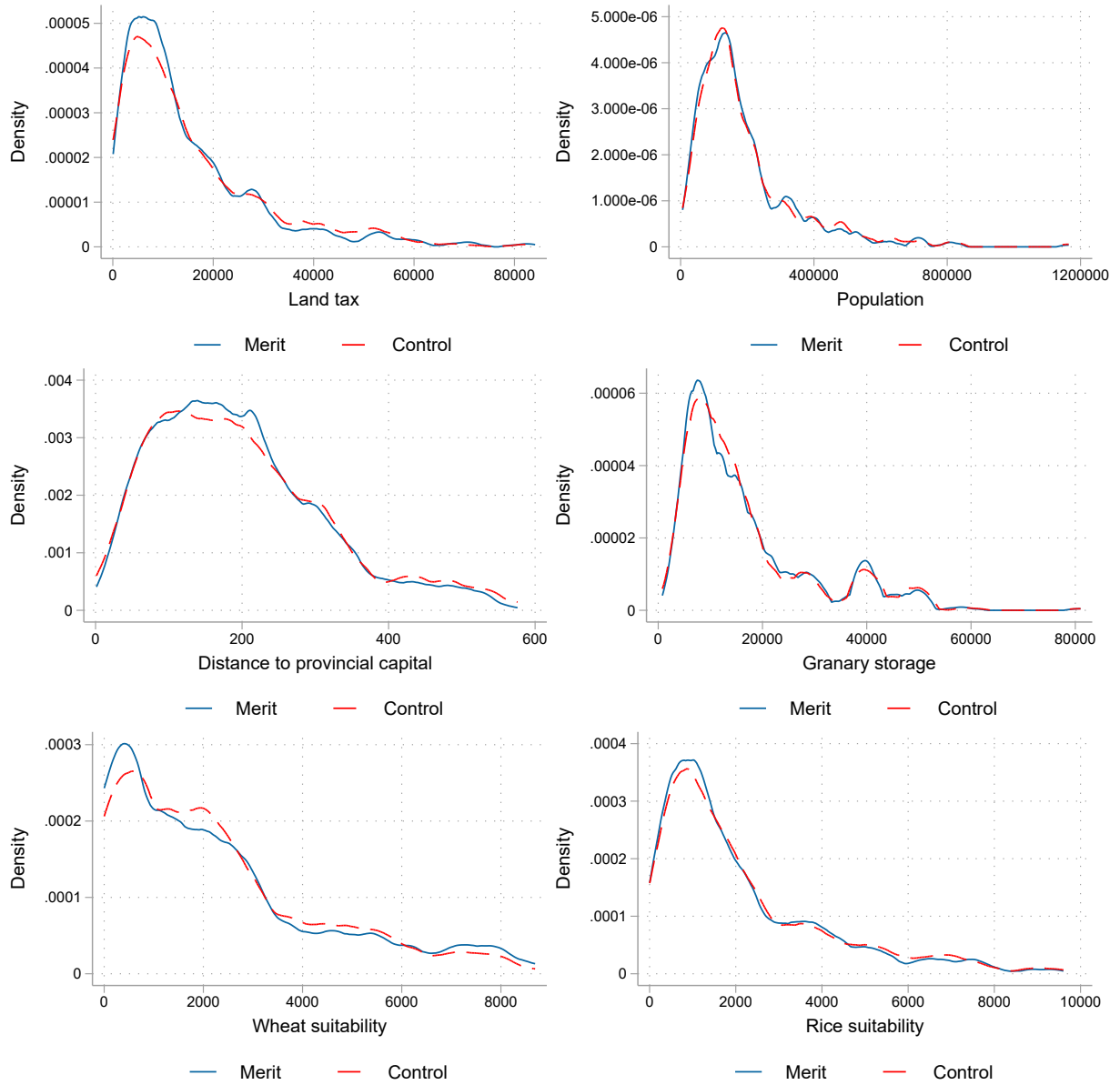
Figures and tables

Figure 1. Spatial distribution of counties with random assignment



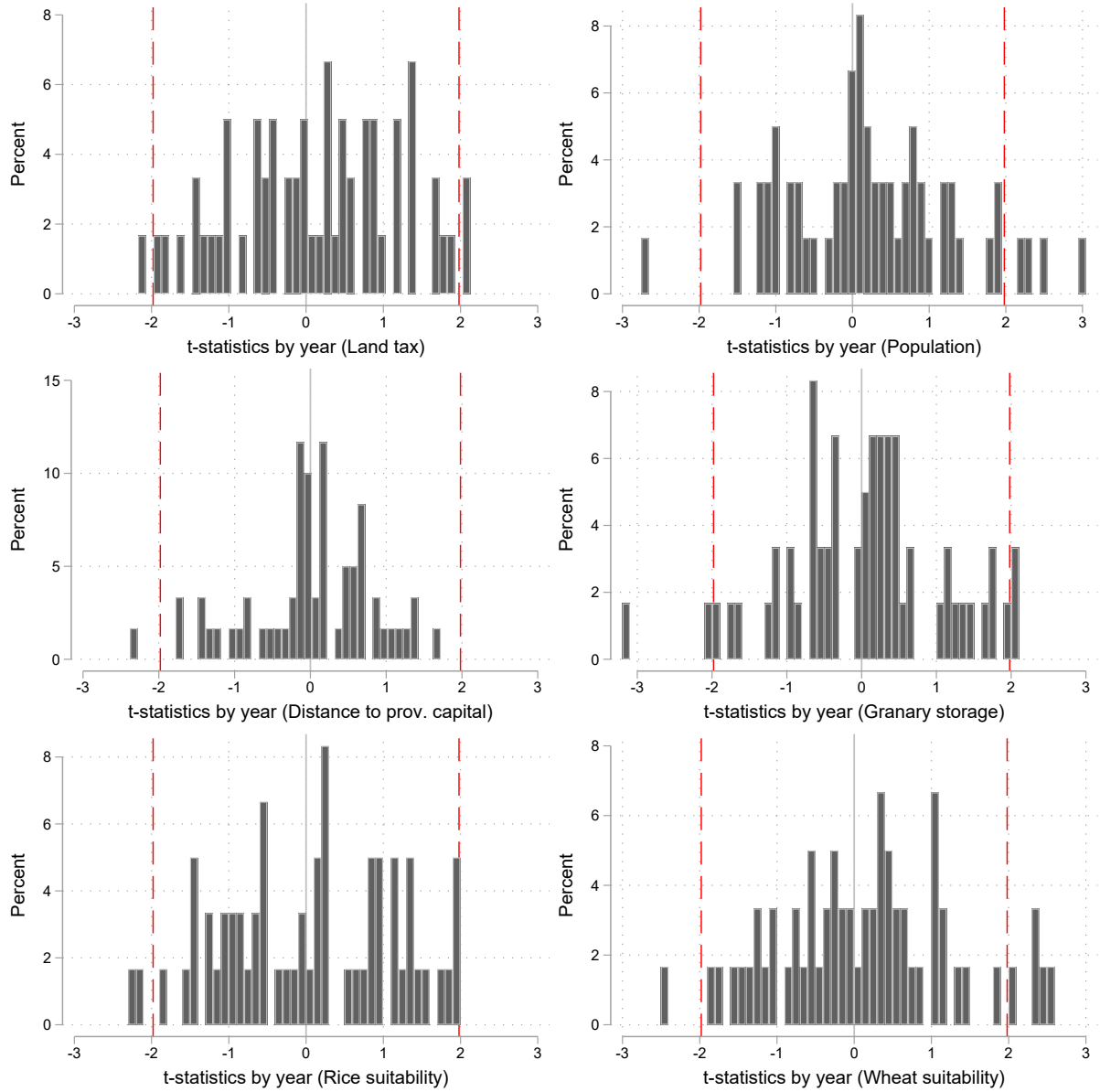
Notes: This figure shows the location of county governments. The blue circles denote counties under random assignment of administrators. We discuss this in Section 2.3.

Figure 2. Balance of the distribution of county characteristics



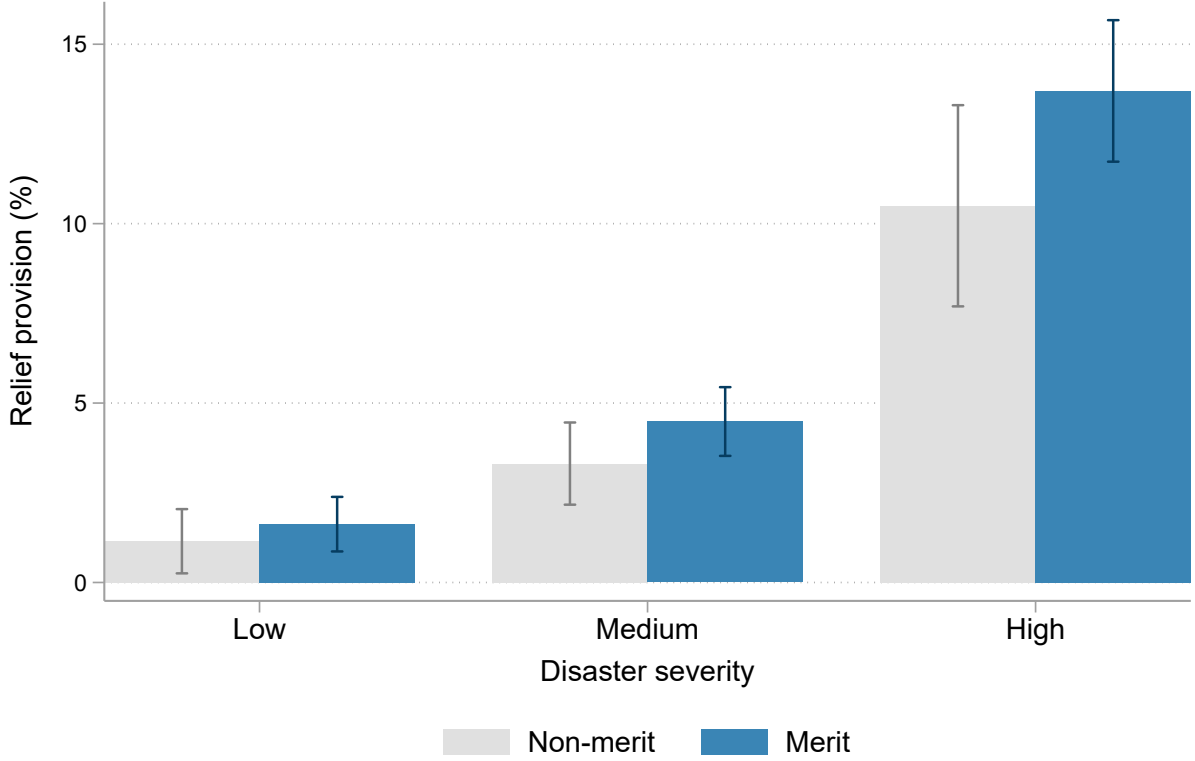
Notes: This figure shows the distribution of land tax quota, population, distance to the provincial capital, granary storage, and rice and wheat suitability of treated counties that were assigned merit-recruited administrators (blue solid line) and control counties that were assigned administrators from other informal recruitment paths (red dash line). We discuss these results in Section 4.

Figure 3. Balance tests by year cohort



Notes: This figure shows the distribution of t-statistics from balance tests by year cohort. In this exercise, we divide all appointments into 60 cohorts by the year of appointment. Then we conduct t-tests separately for each year cohort to examine whether counties in the merit group differ from those in the control group in land tax quota, population, distance to the provincial capital, granary storage, and rice and wheat suitability. We plot the distribution of the corresponding t-statistics from these t-tests. We discuss these results in Section 4.

Figure 4. Effects on relief provision by disaster severity

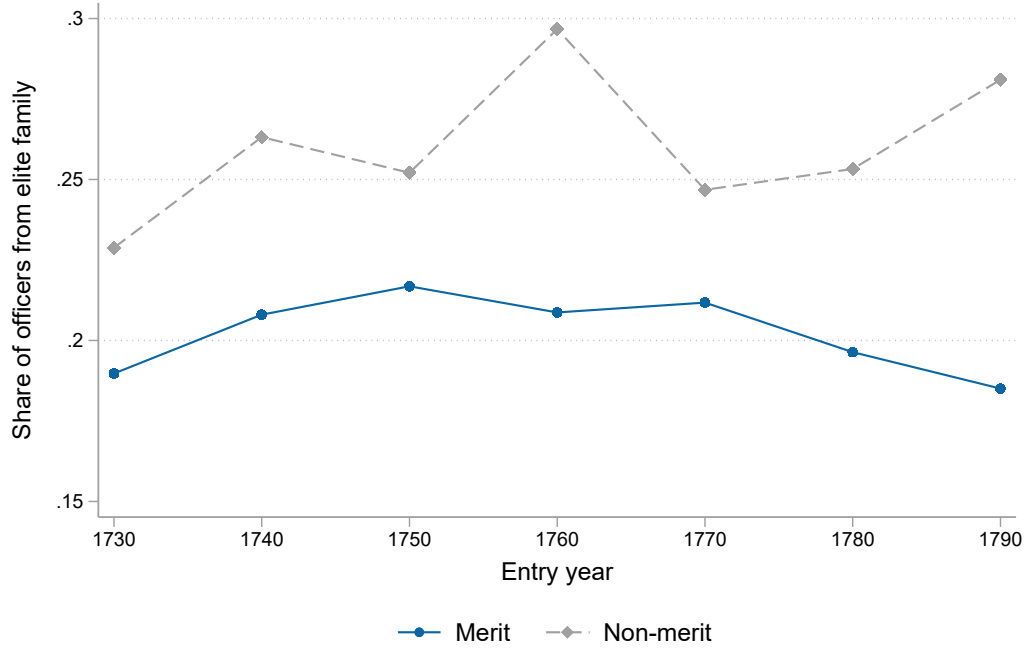


Notes: This figure presents the impacts of merit-recruited administrators on relief responsiveness by disaster severity from equation 2:

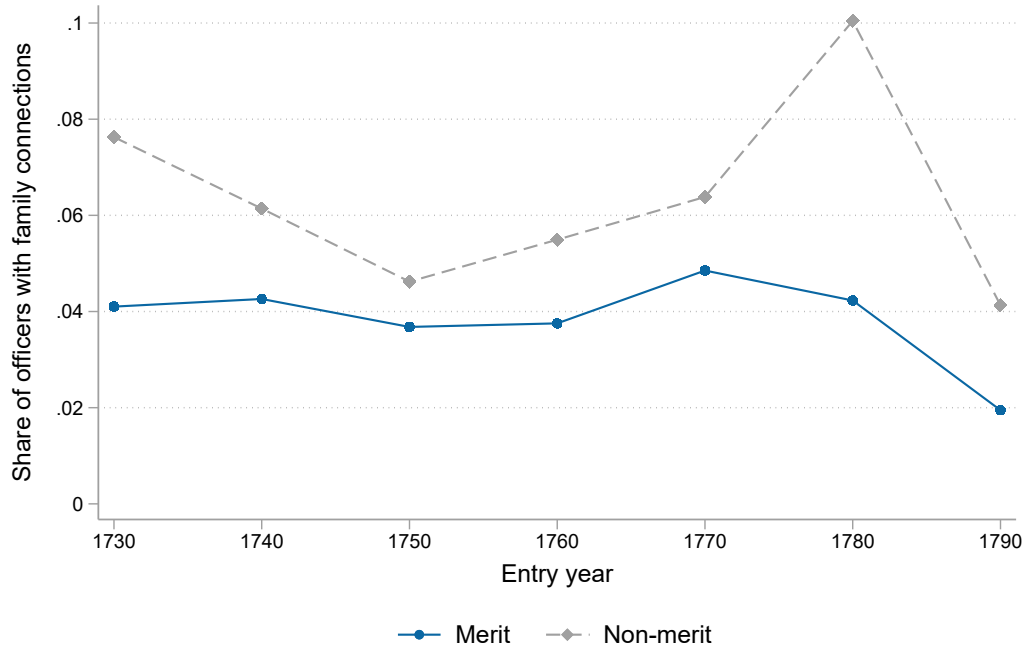
$$Y_{ct} = \sum_{s=1}^3 \beta_s \cdot Disaster_{cts} \times Merit_{i(ct)} + \sum_{s=1}^3 \alpha_s \cdot Disaster_{cts} \times Non-merit_{i(ct)} + \theta_t + \lambda_c + \epsilon_{ct}$$

Bars represent coefficients of $Disaster_{cts} \times Merit_{i(ct)}$ and $Disaster_{cts} \times Non-merit_{i(ct)}$ by different disaster severity s . The dependent variable is the likelihood of relief provision, rescaled by 100 for legibility. The estimating equation includes county, year, and turner fixed effects. Vertical lines show the 95% confidence interval. Robust standard errors in parentheses are clustered at the county level. See Appendix Table A5 for the corresponding regression table. We discuss these results in 5.1.

Figure 5. Meritocracy and family ties of administrators
(a) Share of administrators from elite families



(b) Share of administrators with family connections



Notes: Panel A shows the share of administrators with elite family backgrounds over time (binned by the decade of entry time). Elite family is a dummy equal to one if the administrator belonged to a family (measured by surname-county group) that ever had an official ranking higher than 5th. Panel B shows the share of individuals who had family connections to an incumbent official ranking higher than 5th when they were appointed as county administrators for the first time. We discuss these results in Section 7.1.

Table 1. Balance on county characteristics

	(1) Control	(2) Merit	(3) Diff.	(4) Diff.
	Mean	Mean		p-value
Log(Area)	7.441	7.406	-0.036 (0.023)	0.122
Longitude	112.362	112.503	0.141 (0.181)	0.436
Latitude	29.206	29.156	-0.051 (0.185)	0.785
Ruggedness	389.462	396.707	7.245 (7.663)	0.345
Log(Distance to prov. capital)	5.06	5.063	0.002 (0.027)	0.928
Log(Distance to Beijing)	7.038	7.06	0.021 (0.022)	0.33
log(Population in 1776)	11.887	11.855	-0.032 (0.028)	0.25
log(Land taxes)	9.168	9.151	-0.016 (0.042)	0.692
log(Maize suitability)	4.049	3.882	-0.166 (0.131)	0.203
log(Rice suitability)	6.95	6.937	-0.013 (0.058)	0.82
log(Sweet-potato suitability)	1.849	1.794	-0.055 (0.102)	0.586
log(Wheat suitability)	4.165	4.057	-0.108 (0.136)	0.426
log(Granary storage)	9.452	9.449	-0.003 (0.022)	0.902
Height of city wall	6.604	6.579	-0.024 (0.072)	0.735
Number of city gates	4.211	4.196	-0.015 (0.040)	0.709
No. of subordinate officers	2.100	1.942	-0.158 (0.126)	0.211
No. sheriffs	0.811	0.717	-0.094 (0.119)	0.428
Number of appointments	2,167	5,379		

Notes: Columns 1 and 2 show the mean for the control and treated groups. Column 3 shows the difference between the two groups. Robust standard errors clustered at the county level are reported in parentheses. Column 4 reports the p-value from a pairwise t-test of the equality of means across the two groups. We discuss these results in Section 4.

Table 2. Effects on public goods: disaster relief and school construction

	Relief provision			School construction	
	(1)	(2)	(3)	(4)	(5)
Merit	0.645*** (0.227)	1.536*** (0.493)	1.560*** (0.492)	0.423*** (0.157)	0.424*** (0.157)
Disaster	4.750*** (0.391)				
County FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Tenure FE			Y		Y
Sample	Full	Disaster=1	Disaster=1	Full	Full
Mean of D.V.	2.936	6.171	6.171	1.446	1.446
Observations	28,611	12,002	11,999	28,611	28,611

Notes: Notes: This table presents estimates of the effects of merit-recruited administrators on public goods provision. The unit of observation is the county-year. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Relief provision is a dummy that is 1 if the county received disaster relief, rescaled by 100 for legibility. School construction is a dummy that is 1 if any schools were newly constructed or expanded, rescaled by 100. Tenure is the number of years in the current position. Columns 2 and 3 use the sample when a disaster occurred. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 5.1

Table 3. Effects on fiscal (non)performance: tax revenue deficit

	Tax deficit				
	(1)	(2)	(3)	(4)	(5)
Merit	0.039 (0.179)	0.239 (0.188)	0.241 (0.188)	-1.210** (0.527)	-1.209** (0.528)
County FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Tenure FE			Y		Y
Sample	Difficulty in taxation				
	Full	Low	Low	High	High
Mean of D.V.	1.187	0.929	0.929	2.734	2.734
Observations	26,879	23,039	23,039	3,840	3,838

Notes: This table presents estimates of the effects of merit-recruited administrators on performance in tax collection. The dependent variable is an indicator equal to 1 if the county experienced deficits in tax revenue collection (i.e., the shortfall from the revenue target), rescaled by 100 for legibility. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Column 1 reports the result in the full sample of counties. Columns 2–3 and 4–5 restrict the sample to counties where difficulty in tax collection was low and high (based on the central government's binary classification), respectively. Tenure is the number of years in the current position. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 5.1

Table 4. Effects on additional outcome: subjective performance and management practices

	Honorary record		Subordinate misconduct	
	(1)	(2)	Any (3)	Suspension (4)
Merit	1.725*** (0.630)	1.403** (0.631)	-0.797*** (0.292)	-0.217*** (0.077)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Tenure FE		Y	Y	Y
Mean of D.V.	7.037	7.037	3.46	0.227
Observations	7,546	7,544	28,611	28,611

Notes: This table presents estimates of the effects of merit-recruited administrators on subjective performance and management practices. In columns 1 and 2, the unit of observation is individual-appointment. The dependent variable is a dummy equal to 1 if the administrator was listed in the honorary record in local gazetteers, rescaled by 100. In columns 3 and 4, the unit of observation is the county-year. The dependent variables are dummies that take value 1 if the subordinate officers under the administrators had any misconduct records (column 3) and were suspended (column 4) in a given year, rescaled by 100. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Tenure is the number of years in the current position. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 5.2

Table 5. Effect on judicial efficiency: case disposal duration

	Case duration (months)		
	(1)	(2)	(3)
Merit	-0.601** (0.268)	-0.598** (0.263)	-0.561** (0.265)
County FE	Y	Y	Y
Year FE	Y	Y	Y
Tenure FE		Y	Y
Case controls			Y
Mean of D.V.	12.58	12.58	12.58
Observations	3,089	3,088	3,032

Notes: This table presents estimates of the effects of merit-recruited administrators on judicial performance from estimating equation 3. The unit of observation is the case. The dependent variable is the duration of case measured in months. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Tenure is the number of years in the current position. Case controls include the number of criminals, an indicator for a probationary sentence, sentence levels, charge types, the use of violence, and an indicator for involving family relatives. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 6.2

Table 6. Merit-based recruitment, elite family backgrounds, and family connections

	Elite family			Family connection		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A: All officers</i>						
Merit	-0.054*** (0.010)	-0.055*** (0.010)	-0.042*** (0.011)	-0.025*** (0.006)	-0.024*** (0.006)	-0.018*** (0.007)
Entry year FE		Y	Y		Y	Y
Surname FE			Y			Y
Mean of D.V.	0.141	0.141	0.141	0.047	0.047	0.047
Observations	5,825	5,825	5,702	5,825	5,825	5,702
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel B: Excluding common surnames</i>						
Merit	-0.067*** (0.012)	-0.072*** (0.012)	-0.055*** (0.013)	-0.033*** (0.007)	-0.034*** (0.007)	-0.026*** (0.008)
Entry year FE		Y	Y		Y	Y
Surname FE			Y			Y
Mean of D.V.	0.121	0.121	0.121	0.036	0.036	0.036
Observations	3,535	3,535	3,412	3,535	3,535	3,412

Notes: This table shows that merit-recruited administrators were less likely to come from elite family backgrounds and less likely to have family connections to incumbent senior officials. The unit of observation is individual. Panel A use all administrators. Panel B excludes administrators with the top 10 most common surnames which have population shares higher than 1.5%. *Elite family* is a dummy equal to 1 if the administrator belonged to a family (measured by surname-county group) that ever had an official ranking higher than 5th during the past 60 years. *Family connection* is a dummy equal to 1 if the administrator had family connections to an incumbent official ranking higher than 5th. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Standard errors in parentheses. Standard errors in parentheses. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 7.1

Table 7. Promotion incentives do not vary with merit qualifications

	Promotion			
	(1)	(2)	(3)	(4)
Misconduct record	-0.337*** (0.091)	-0.362** (0.162)		
Merit × Misconduct record		0.035 (0.189)		
Experience			0.092*** (0.016)	0.086*** (0.032)
Merit × Experience				0.009 (0.038)
Merit		0.030 (0.117)		-0.021 (0.138)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Mean of D.V.	0.409	0.409	0.409	0.409
Observations	28,611	28,611	28,611	28,611

Notes: This table examines the link between promotion and performance or seniority and shows that the link does not change with administrators' recruitment backgrounds. Promotion is a dummy that is 1 if the officer was promoted to prefecture position in the next year, rescaled by 100 for legibility. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Misconduct record is a dummy variable that takes value 1 if the officer had any misconduct records over the tenure. Experience is the cumulative years of service as a county administrator. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 7.3

Table 8. Heterogeneity in performance by civil service exam qualification tiers

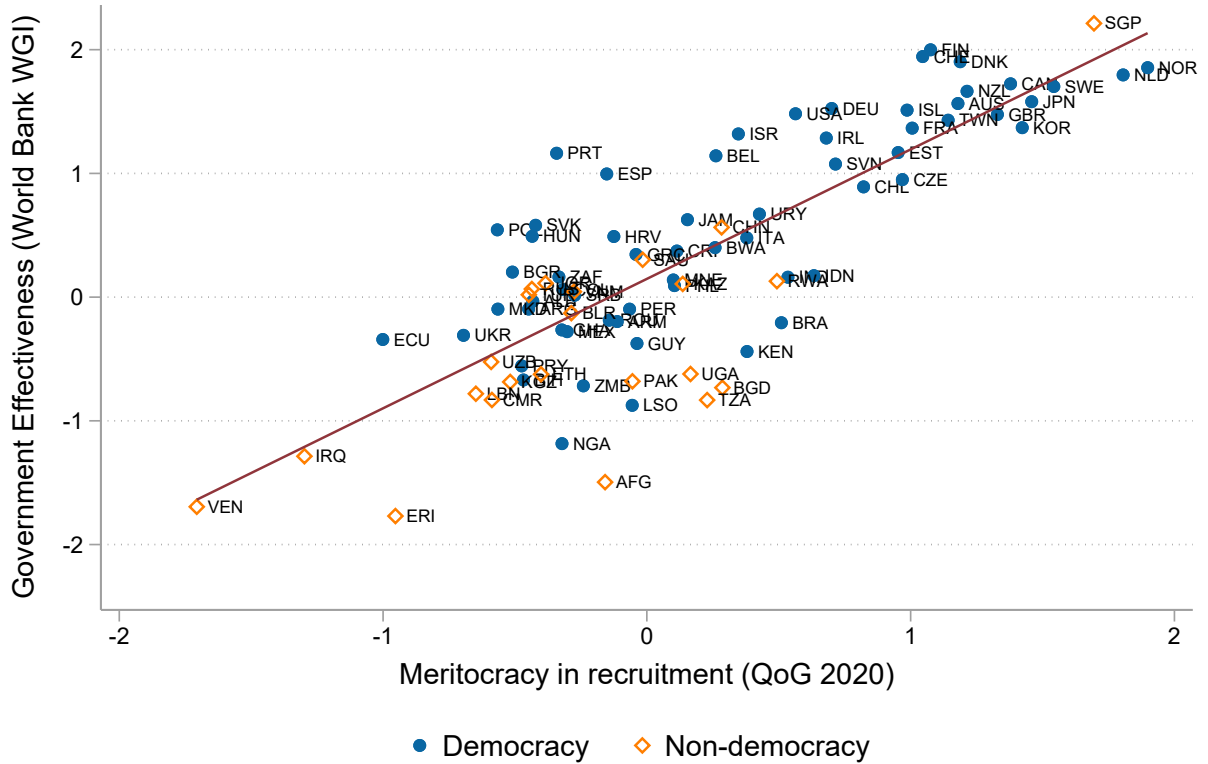
	Relief provision (1)	Tax deficit (2)	School construction (3)	Honorary Record (4)
Merit	1.569*** (0.532)	-0.656 (0.614)	0.244 (0.170)	0.279 (0.642)
Merit $\times$ 1st tier	-0.028 (0.589)	-1.696** (0.813)	0.516*** (0.198)	3.475*** (0.914)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Tenure FE	Y	Y	Y	Y
Sample	Disaster	Tax-difficult		
Mean of D.V.	6.171	2.734	1.446	7.037
Observations	11,999	3,838	28,611	7,544

Notes: This table examines the heterogeneous effects of merit-recruited administrators by the civil exam qualification tier. The sample is restricted to counties constantly under the random assignment rule. The unit of observation is the county-year in columns 1–3 and individual-appointment in column 4. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. 1st-tier is an indicator for administrators with 1st-tier civil exam qualification (*Jinshi*). Column 1 uses the sample when a disaster occurred. Column 2 restricts the sample to counties characterized by high difficulty in tax collection according to the central government's classification. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 7.4

Appendix

A Appendix Figures and Tables

Figure A1. Meritocracy and government performance: cross-country correlation



Notes: This figure plots the cross-country correlation between meritocracy and government effectiveness by regime type. Data on government effectiveness come from the World Bank's Worldwide Governance Indicators in 2019. Meritocracy is measured by The Quality of Government Expert Survey 2020 as an index for the extent to which the appointments to bureaucratic positions are based on individuals' merits. Higher values stand for more merit-based appointment. We use the dichotomous measure of democracy in 2019 from (Boix et al., 2013).

Figure A2. Levels of civil service exam and passing rate

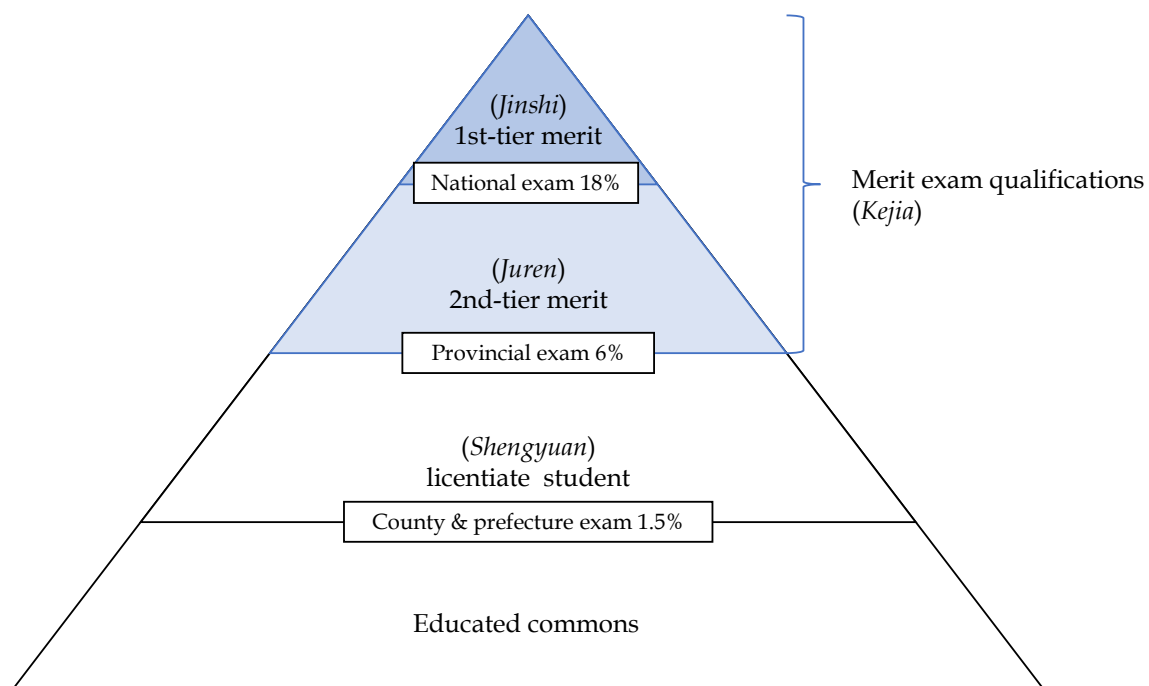
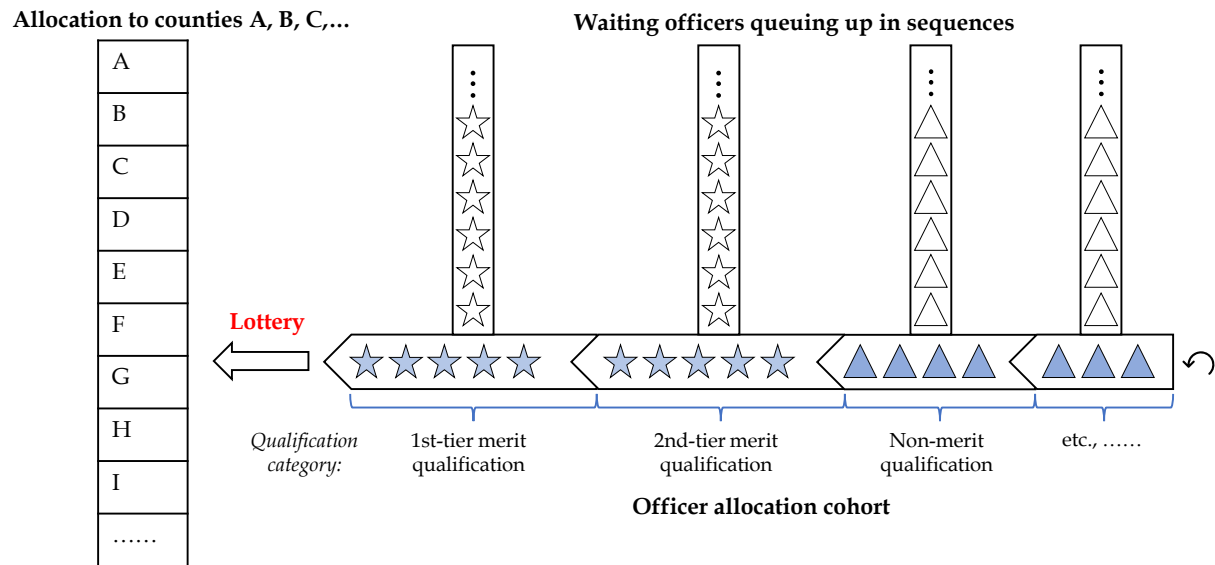
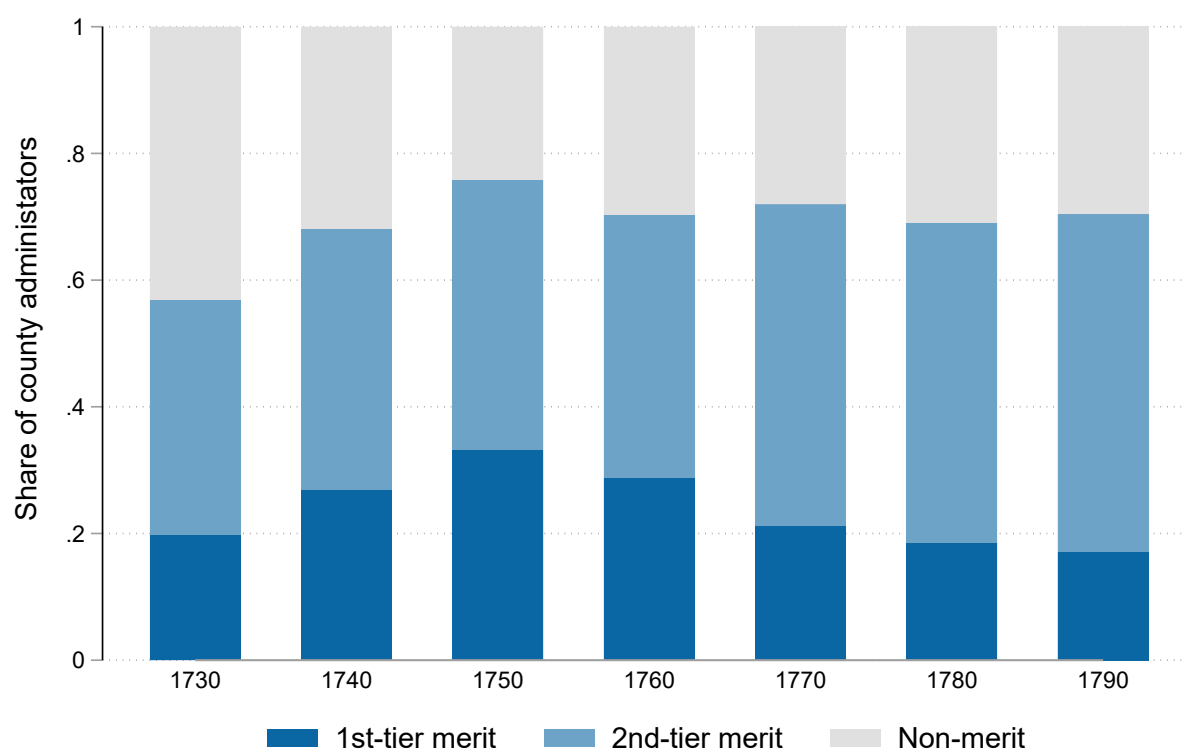


Figure A3. Illustration of the random assignment process



Notes: This figure illustrates the process of random assignment. The assignment took place on a monthly basis: (1) The Ministry of Personnel (MOP) determined the size of the allocation cohort for this month based on the number of vacancies; (2) waiting officers filled in the allocation cohort following a stipulated sequence and quota of qualification categories; (3) officers of this allocation cohort were assigned to vacant counties by random lottery. See Section 2.3 for details.

Figure A4. Composition of county administrators by recruitment qualifications



Notes: This figure shows the composition of county administrators by recruitment backgrounds over time. See details in Section 3.

Figure A5. The number of criminal cases in data

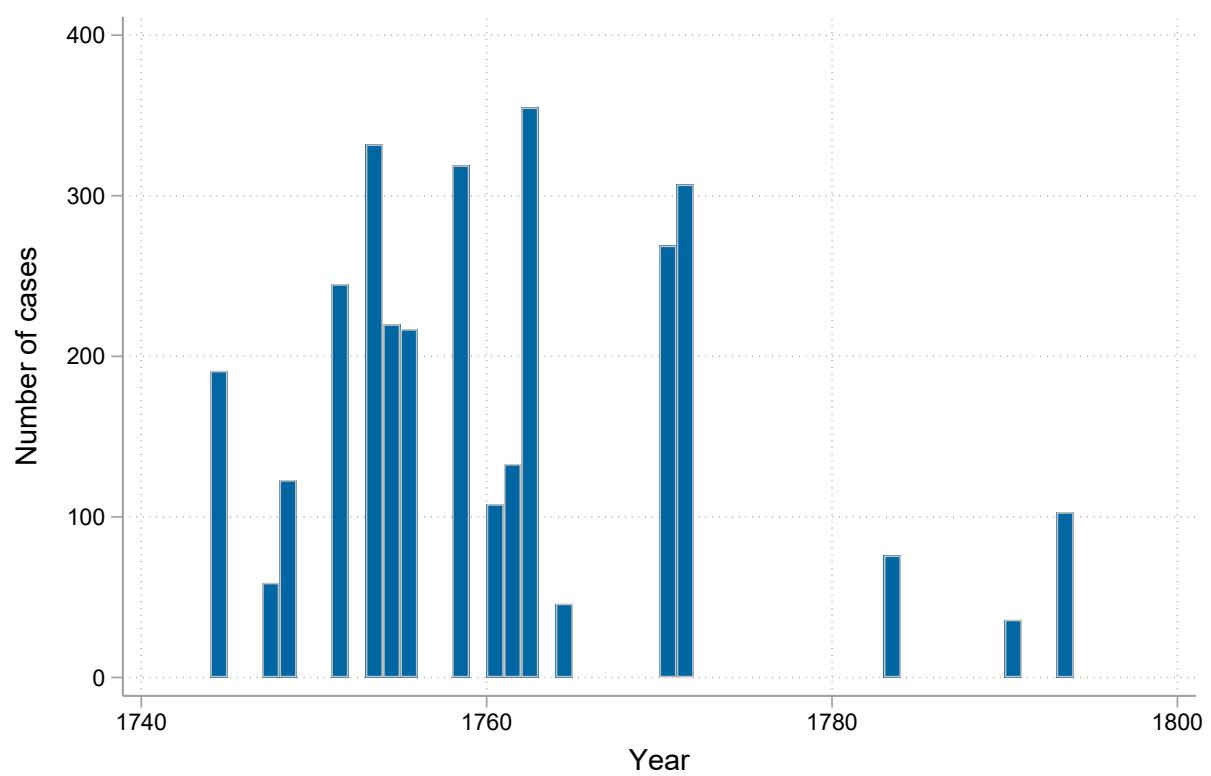
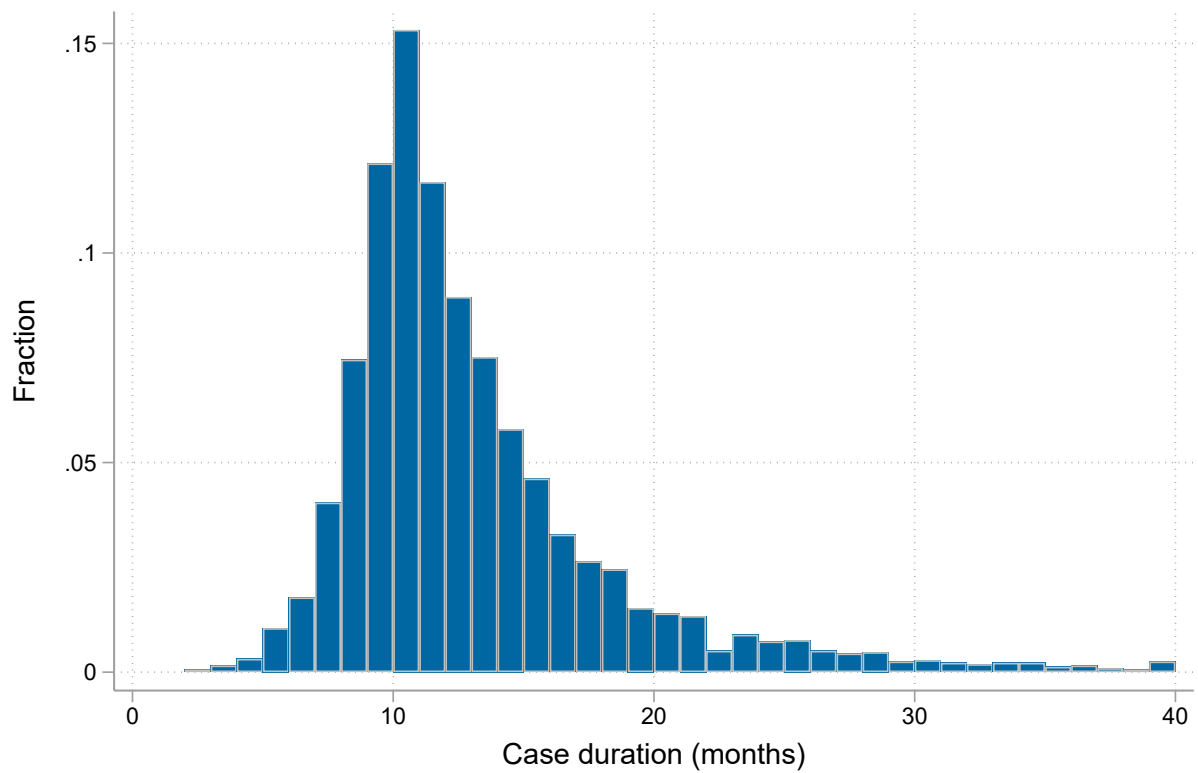
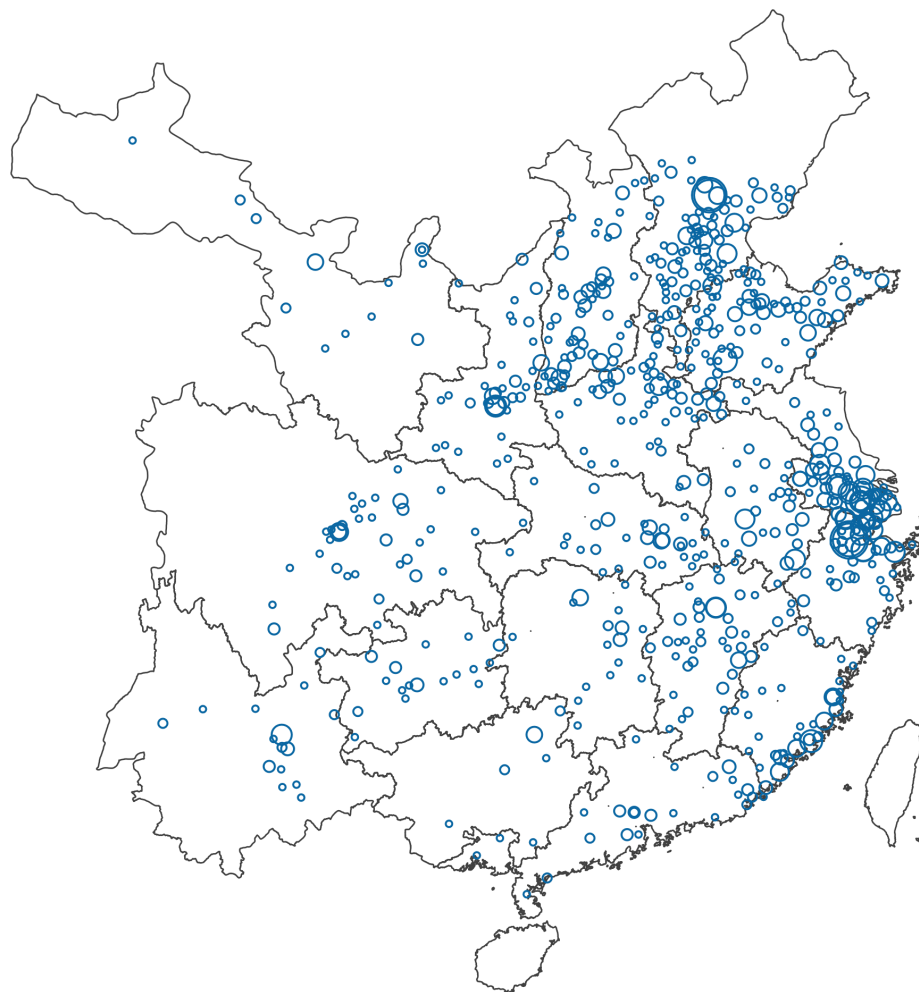


Figure A6. Distribution of case duration



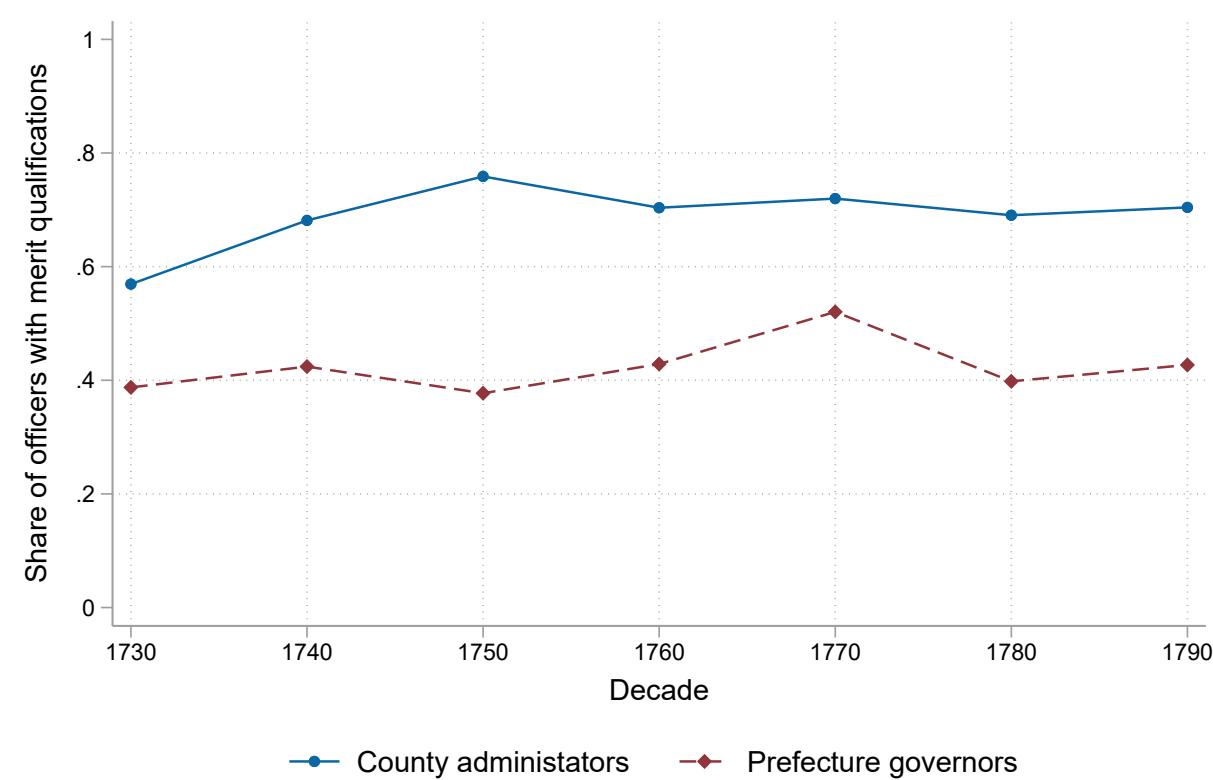
Notes: This figure plots the distribution of case duration measured in months. The sample covers 3,139 criminal cases. See Section 6.1 for details.

Figure A7. Location and number of elite families



Notes: This map shows the geographic distribution of elite families. The circles are drawn in proportion to the number of elite families in each county. Elite family is defined as a county-surname unit that ever had an official ranking higher than 5th. We discuss this in Section 7.1.

Figure A8. Share of officers with merit exam qualifications: county vs. prefecture level



Notes: This figure shows the share of officers with merit-recruited backgrounds by county/prefecture positions over time. We discuss these results in Section 7.3.

Table A1. Descriptive statistics

	Mean	Std. dev.	Min	Max	Obs.
	(1)	(2)	(3)	(3)	(3)
Panel A: County-year level					
Relief provision	2.936	16.881	0	100	28,611
Disaster	0.420	0.494	0	1	28,611
Disaster (level 1)	0.150	0.357	0	1	28,611
Disaster (level 2)	0.182	0.386	0	1	28,611
Disaster (level 3)	0.087	0.282	0	1	28,611
Tax revenue deficit	1.187	10.829	0	100	26,879
School construction	1.447	11.942	0	100	28,611
Subordinate misconduct record	3.460	18.277	0	100	28,611
Subordinate suspension	0.227	4.761	0	100	28,611
Promotion	0.538	7.317	0	100	28,611
Exam connections within province	0.148	0.355	0	1	7,154
Merit	0.727	0.446	0	1	28,611
Merit 1st-tier	0.228	0.420	0	1	28,611
Panel B: Officer level					
Merit	0.717	0.450	0	1	6,194
Merit 1st-tier	0.250	0.433	0	1	6,194
Ethnicity: Han	0.940	0.237	0	1	6,194
Ethnicity: Manchu or Bannerman	0.060	0.237	0	1	6,194
Elite family	0.141	0.348	0	1	5,825
Family connection in the entry year	0.047	0.211	0	1	5,825

Notes: This table presents the descriptive statistics for the variables in county-year panel data (Panel A) and administrator individual level data (Panel B). The sample covers 493 counties between 1736 and 1795.

Table A2. Check on pre-trends in outcomes

	<i>Merit_t</i>				
	(1)	(2)	(3)	(4)	(5)
Relief	0.022 (0.034)	-0.028 (0.038)	0.015 (0.041)	0.000 (0.001)	0.000 (0.000)
School	0.050 (0.046)	0.030 (0.053)	0.058 (0.049)	0.088 (0.066)	0.068 (0.061)
Tax deficit	0.039 (0.047)	-0.092 (0.070)	-0.083 (0.068)	0.006 (0.072)	-0.013 (0.072)
Honorary record	0.020 (0.023)	0.014 (0.024)	0.020 (0.026)	0.020 (0.022)	-0.001 (0.022)
Misconduct of subordinates	-0.015 (0.028)	-0.052 (0.033)	-0.036 (0.043)	-0.041 (0.032)	-0.046 (0.031)
Explanatory variables measured in	t-1	t-2	t-3	Mean of the predecessor	
County FE				Y	
Year FE				Y	
Observations	7,041	5,511	4,149	7,041	7,041

Notes: This table shows the absence of correlation between the assignment of merit administrator and pre-turnover outcome variables. The unit of observation is individual-appointment. For each appointment, we regress the indicator for whether the administrator had merit civil exam qualification on the outcomes of interest measured in one, two, and three years before the turnover in columns 1, 2, and 3, respectively. In columns 4 and 5, explanatory variables are measured by the mean over the predecessor's tenure. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 4

Table A3. No effect on relief involving higher levels of government

	Relief (higher levels of government)		
	(1)	(2)	(3)
Merit	0.017 (0.088)	-0.028 (0.173)	-0.024 (0.174)
Disaster	0.342*** (0.092)		
County FE	Y	Y	Y
Year FE	Y	Y	Y
Tenure FE			Y
Sample	Full	Disaster=1	Disaster=1
Mean of D.V.	0.322	0.508	0.508
Observations	28,611	12,002	11,999

Notes: This table presents estimates of the effects of merit-recruited administrators on disaster relief involving higher levels of government. The unit of observation is the county-year. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Relief provision is a dummy that is 1 if the county received disaster relief involving provincial or central government, rescaled by 100 for legibility. Tenure is the number of years in the current position. Columns 2 and 3 use the sample when a disaster occurred. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 5.1

Table A4. Meritocratic recruitment and disaster occurrence

	Disaster	Flood	Drought	Precipitation index
	(1)	(2)	(3)	(4)
Merit	-0.005 (0.007)	0.001 (0.006)	-0.005 (0.005)	-0.018 (0.014)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Mean of D.V.	0.42	0.244	0.128	2.899
Observations	28,611	28,611	28,611	28,611
	Weather extreme	Disaster (level 1)	Disaster (level 2)	Disaster (level 3)
	(5)	(6)	(7)	(8)
Merit	0.007 (0.005)	-0.005 (0.005)	-0.002 (0.005)	0.002 (0.004)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Mean of D.V.	0.119	0.15	0.182	0.087
Observations	28,611	28,611	28,611	28,611

Notes: This table shows the null effects of merit-recruited administrators on disaster occurrence. The unit of observation is the county-year. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 5.1

Table A5. Effects on relief provision by disaster severity

	Relief provision		
	(1)	(2)	(3)
α_1 : Non-merit $\times$ Disaster low	0.742 (0.458)	1.151** (0.455)	1.150** (0.455)
β_1 : Merit $\times$ Disaster low	1.351*** (0.388)	1.623*** (0.387)	1.625*** (0.387)
α_2 : Non-merit $\times$ Disaster middle	3.221*** (0.562)	3.324*** (0.583)	3.314*** (0.583)
β_2 : Merit $\times$ Disaster middle	4.465*** (0.465)	4.499*** (0.488)	4.486*** (0.488)
α_3 : Non-merit $\times$ Disaster high	10.792*** (1.471)	10.500*** (1.425)	10.498*** (1.427)
β_3 : Merit $\times$ Disaster high	14.484*** (1.064)	13.702*** (1.003)	13.698*** (1.004)
County FE	Y	Y	Y
Year FE		Y	Y
Tenure FE			Y
p-value: $\alpha_1 = \beta_1$	0.187	0.304	0.300
p-value: $\alpha_2 = \beta_2$	0.053	0.064	0.065
p-value: $\alpha_3 = \beta_3$	0.020	0.034	0.034
Mean of D.V.	1.446	1.446	1.446
Observations	28,611	28,611	28,611

Notes: This table the results of estimating equation 2, corresponding to Figure 4. The unit of observation is the county-year. Relief provision is a dummy that is 1 if the county received disaster relief, rescaled by 100 for legibility. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Tenure is the number of years in the current position. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 5.1

Table A6. Explanatory power of administrators for judicial efficiency

	Case disposal duration (months)			
	(1)	(2)	(3)	(4)
Adj. R-square	0.038	0.088	0.195	0.303
Observations	3,090	3,090	3,090	3,090
Case controls	Y	Y	Y	Y
County FE		Y	Y	Y
Year FE			Y	Y
Administrator FE				Y
P-value of F test				0.000

Notes: This table investigates how much of the variation in case duration is explained by the office, manager, and time fixed effects. The unit of observation is the case. Case controls include the number of criminals, an indicator for a probationary sentence, sentence levels, charge types, the use of violence, and an indicator for involving family relatives. The p-value at the bottom of the table tests the null hypothesis that administrator effects are jointly zero. We discuss these results in Section 6.2

Table A7. Balance of case characteristics

	Control (1)	Merit (2)	Diff. (3)
Num. of criminal	1.12 (0.389)	1.137 (0.452)	0.017 (0.262)
Sentence probation	0.897 (0.304)	0.898 (0.303)	0.001 (0.012)
Sentence level	1.425 (0.977)	1.423 (0.977)	-0.003 (0.040)
Case type: robbery	0.062 (0.242)	0.058 (0.233)	-0.004 (0.009)
Case type: adultery	0.033 (0.180)	0.032 (0.176)	-0.001 (0.008)
Case type: murder	0.067 (0.250)	0.076 (0.265)	0.009 (0.011)
Case type: homicide	0.758 (0.428)	0.769 (0.421)	0.011 (0.018)
Case type: kinship code	0.079 (0.270)	0.064 (0.245)	-0.015 (0.010)
Direct violence	0.933 (0.250)	0.937 (0.244)	0.004 (0.010)
Relatives involved	0.242 (0.428)	0.231 (0.422)	-0.011 (0.020)
Observations	836	2,303	

Notes: Columns 1 and 2 report the sample mean and standard deviation (in parentheses) for case characteristics. Column 3 shows the difference between the two groups. Robust standard errors clustered at the county level are reported in parentheses. Standard errors are reported in parentheses. We discuss these results in Section 6.3

Table A8. Case workload and case composition

	Workload	case composition			
	Number cases (1)	Number criminals (2)	Sentence level (3)	Robbery (4)	Murder (5)
Merit	0.0564 (0.055)	0.0065 (0.009)	0.0125 (0.050)	-0.0069 (0.011)	0.0017 (0.012)
County FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Mean of D.V.	1.57	1.024	1.423	0.059	0.074
Observations	1,943	2,808	3,057	3,089	3,089

Notes: The unit of observation in column 1 is the county-year. The unit of observation in columns 2-5 is the case. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualifications. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 6.3

Table A9. Impact on judicial decision quality: case reversal

	Reversal		
	(1)	(2)	(3)
Merit	-0.085 (0.082)	-0.086 (0.081)	-0.083 (0.081)
County FE	Y	Y	Y
Year FE	Y	Y	Y
Tenure FE		Y	Y
Case controls			Y
Mean of D.V.	0.259	0.259	0.259
Observations	27,361	27,361	27,361

Notes: The unit of observation is the case. Reversal is a dummy that is 1 if the initial decision of the case was reversed by the “Supreme Court” (i.e., Ministry of Justice), rescaled by 100 for legibility. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Tenure is the number of years in the current position. Case controls include number of criminals, indicator for probationary sentence, sentence level, charge types, the use of violence, and indicator for involving family relatives. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 6.3

Table A10. Merit-based recruitment and time-varying family connections: individual-year level results

	Family connection	
	(1)	(2)
<i>Panel A: All officers</i>		
Merit	-0.013*** (0.003)	-0.009*** (0.003)
Year FE		Y
Surname FE		Y
Mean of D.V.	0.044	0.044
Observations	26,879	26,863
	(1)	(2)
<i>Panel B: Excluding common surnames</i>		
Merit	-0.024*** (0.003)	-0.019*** (0.003)
Year FE		Y
Surname FE		Y
Mean of D.V.	0.033	0.033
Observations	16,320	16,304
Sample	Officer-year	

Notes: The unit of observation in columns 3–4 is individual-year. Panel A use all administrators. Panel B excludes administrators with the top 10 most common surnames which have population shares higher than 1.5%. Family connection is a dummy equal to 1 if the administrator had family connections to an incumbent official ranking higher than 5th. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Standard errors in parentheses. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 7.1

Table A11. Merit-based recruitment and government performance: the role of exam-based connections

	Relief provision	Tax deficit	School construction	Honorary record
	(1)	(2)	(3)	(4)
Merit	1.598** (0.728)	-1.294* (0.674)	0.664*** (0.230)	3.898*** (1.066)
Merit × Exam connections (province)	-0.559 (1.202)	-0.320 (0.925)	-0.039 (0.445)	-0.338 (2.054)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Tenure FE	Y	Y	Y	Y
Sample	Disaster	Tax-difficult	Full	Full
Mean of D.V.	5.879	2.526	1.523	7.861
Observations	6,185	2,137	14,974	3,914

Notes: The unit of observation is the county-year in columns 1–3 and individual-appointment in column 4. The sample consists of non-merit administrators and 1st tier merit-recruited administrators for which the exam network information is available. Column 1 uses the sample when a disaster occurred. Column 2 restricts the sample to counties characterized by high difficulty in tax collection according to the central government’s classification. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. *Exam connections (province)* is a dummy equal to 1 if the administrator had any exam connections (quasi-classmate relationships) to his superior officials within the province. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 7.2

Table A12. Merit-based recruitment and government performance: the role of exam-based connections in the central government

	Relief provision	Tax deficit	School construction	Honorary record
	(1)	(2)	(3)	(4)
Merit	1.517** (0.669)	-1.515** (0.672)	0.659*** (0.220)	3.856*** (1.012)
Merit × Exam connections (central)	0.495 (0.684)	0.838 (0.543)	0.044 (0.239)	-1.143 (0.984)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Tenure FE	Y	Y	Y	Y
Sample	Disaster	Tax-difficult	Full	Full
Mean of D.V.	5.879	2.526	1.523	7.861
Observations	6,185	2,137	14,974	3,914

Notes: The unit of observation is the county-year in columns 1–3 and individual-appointment in column 4. The sample consists of non-merit administrators and 1st tier merit-recruited administrators for which the exam network information is available. Column 1 uses the sample when a disaster occurred. Column 2 restricts the sample to counties characterized by high difficulty in tax collection according to the central government’s classification. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Exam connections (central) is the number of exam-connected officials in the central government, standardized by the mean. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level. We discuss these results in Section 7.2

B Additional robustness checks

Table B1. Effects on government performance: robustness to alternative sample without adjustments in appointment method

	Relief provision	Tax deficit	School construction	Honorary record
	(1)	(2)	(3)	(4)
Merit	1.324** (0.517)	-1.151* (0.609)	0.381** (0.166)	1.502** (0.680)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Tenure FE	Y	Y	Y	Y
Sample	Counties with no adjustments			
Mean of D.V.	6.171	2.734	1.446	7.037
Observations	10,808	3,103	25,537	6,648

Notes: The sample is restricted to counties constantly under the random assignment rule. The unit of observation is the county-year in columns 1–3 and individual-appointment in column 4. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Column 1 uses the sample when the disaster occurred. Column 2 restricts the sample to counties characterized by demanding taxation tasks according to the central government's classification. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level.

Table B2. Effects on government performance: dropping countries with fewer turnovers

	Relief provision	Tax deficit	School construction	Honorary record
	(1)	(2)	(3)	(4)
Merit	1.731*** (0.482)	-1.247** (0.539)	0.417** (0.162)	1.312** (0.650)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Tenure FE	Y	Y	Y	Y
Sample	Total turnover>10			
Mean of D.V.	6.171	2.734	1.446	7.037
Observations	11,388	3,713	26,467	7,216

Notes: The sample is restricted to counties that experienced more than 10 turnovers during 1736–1795. The unit of observation is the county-year in columns 1–3 and individual-appointment in column 4. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Column 1 uses the sample when the disaster occurred. Column 2 restricts the sample to counties characterized by demanding taxation tasks according to the central government's classification. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level.

Table B3. Effects on government performance: comparing within the same assignment year

	Relief provision	Tax deficit	School construction	Honorary record
	(1)	(2)	(3)	(4)
Merit	1.870*** (0.509)	-1.085** (0.529)	0.382** (0.157)	1.282** (0.632)
County FE	Y	Y	Y	Y
Assignment year FE	Y	Y	Y	Y
Tenure FE	Y	Y	Y	Y
Mean of D.V.	6.171	2.734	1.446	7.037
Observations	11,999	3,838	28,611	7,544

Notes: The unit of observation is the county-year in columns 1–3 and individual-appointment in column 4. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Column 1 uses the sample when a disaster occurred. Column 2 restricts the sample to counties characterized by demanding taxation tasks according to the central government's classification. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level.

Table B4. Effects on government performance: dropping officers with central gov't experiences

	Relief provision	Tax deficit	School construction	Honorary record
	(1)	(2)	(3)	(4)
Merit	1.523*** (0.491)	-1.274** (0.521)	0.404** (0.157)	1.329** (0.629)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Tenure FE	Y	Y	Y	Y
Sample	Dropping officers with Hanlin experiences			
Mean of D.V.	6.171	2.734	1.446	7.037
Observations	11,952	3,817	28,527	7,524

Notes: The sample excludes officers with previous prestigious experiences in the Hanlin Academy. The unit of observation is the county-year in columns 1–3 and individual-appointment in column 4. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Column 1 uses the sample when a disaster occurred. Column 2 restricts the sample to counties characterized by demanding taxation tasks according to the central government's classification. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level.

Table B5. Effects on government performance: robustness to ethnicity fixed effects

	Relief provision	Tax deficit	School construction	Honorary record
	(1)	(2)	(3)	(4)
Merit	1.461*** (0.492)	-1.209** (0.528)	0.419*** (0.157)	1.282** (0.636)
County FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Tenure FE	Y	Y	Y	Y
Ethnicity FE	Y	Y	Y	Y
Mean of D.V.	6.171	2.734	1.446	7.037
Observations	11,999	3,838	28,611	7,544

Notes: The unit of observation is the county-year in columns 1–3 and individual-appointment in column 4. Merit is an indicator that takes value 1 if the administrator had merit civil exam qualification. Column 1 uses the sample when a disaster occurred. Column 2 restricts the sample to counties characterized by demanding taxation tasks according to the central government's classification. Robust standard errors in parentheses are clustered at the county level. ***, **, * denote significance at 1%, 5%, 10% level.

C Data appendix

C.1 Examples of tax deficit records

Below are three examples of tax deficit records, listing the names of administrators who were accountable for the revenue deficit.

[Archive id: 02-01-04-15901-011] Report on **Zhou Fengqi**, the **administrator** of **Yuanhe County, Jiangsu Province**, failing to complete the grain tribute tax **for 1763** and the corresponding disciplinary measures

[Archive id: 02-01-04-13957-004] Report on **Liu Yujie**, the **administrator** of **She County, Anhui Province**, failing to complete the central land tax **for 1741 and 1742** and the corresponding disciplinary measures according to the rule

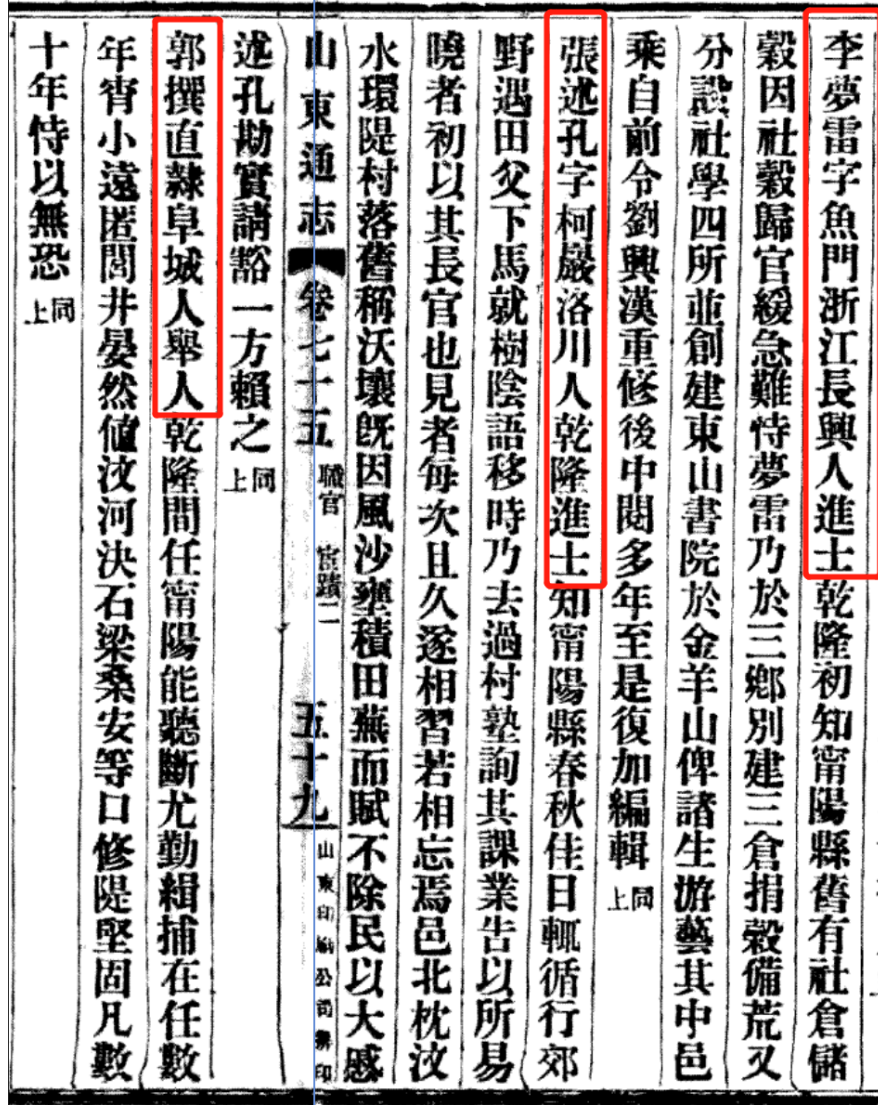
[Archive id: 02-01-04-14284-011] Report on **Zhu Shigu**, the **administrator** of **Gaotai County, Gansu Province**, failing to complete the central land tax **for 1746** and the corresponding disciplinary measures according to the rule

Figure C1. Sample page of administrator appointment records

咸寧縣知縣														
陳齊賢	喬履信	顧鴻	侯天章	柴華	臧應桐	紀虛中	張求垣	柳大任	伍錦	凌樹屏	劉度昭	顧世瑛	吳炳	沈元輝
廣西	河南	四川	山東	貴州	漢軍	乾隆	江蘇	山東	順天	浙江	漢軍	浙江	江蘇	河南
一	一	四	一	一	一	三	一	一	一	一	一	一	一	一
年	年	年	年	年	年	年	年	年	年	年	年	年	年	年
至	至	至	至	至	至	至	至	至	至	至	至	至	至	至
樂	樂	樂	樂	樂	樂	樂	樂	樂	樂	樂	樂	樂	樂	樂
縣	縣	縣	縣	縣	縣	縣	縣	縣	縣	縣	縣	縣	縣	縣
人	人	人	人	人	人	人	人	人	人	人	人	人	人	人
四	五	七	九	十	十	十	十	十	十	十	十	十	十	十
舉	進	舉	舉	舉	舉	舉	舉	舉	舉	舉	舉	舉	舉	舉
年	年	年	年	年	年	年	年	年	年	年	年	年	年	年
任	任	任	任	任	任	任	任	任	任	任	任	任	任	任
正														

Notes: This figure shows a sample page of appointment records for administrators of Xianning county. See details in Section 3.

Figure C2. Sample of honorary records in local gazetteers



Notes: This figure shows a sample page of honorary records for administrators of Ningyang county in *Gazetteer of Shandong*, listing their achievements and activities during their tenure in office. See details in Section 5.2.