Political Power and Collectivism: The Persistent Impact of Exposure to Historical Centralized Political Regimes*

Yuting Chen† Liqiang Liu‡ Jiarui Wang§

September 10, 2025

Abstract

Collectivism profoundly influences economic behavior, institutional development, and social interactions. Yet the role of politics in its emergence has received little attention. This paper identifies a novel determinant of collectivism: *exposure to centralized political power*. Our identification strategy exploits variation in county distances to the historical capitals of centralized Chinese dynasties, instrumented by proximity to the dynasties' initial territorial centroids. Using data from the 2005 China census, we find that individuals residing in counties closer to historical capitals of centralized regimes exhibit a stronger collectivist orientation than their cohort peers in more distant counties within the same province. This finding is robust to alternative measures of collectivism and exposure to centralized regimes, more conservative inference approaches, and the inclusion of additional covariates. We further show that the effect varies across regions and between the periods before and after the PRC's establishment in 1949, but not by gender.

Keywords: Collectivism, Political Regimes, Centralized Chinese Dynasties

JEL Classification: P48, Z13, N45

*We thank Ray Fisman, Martin Fiszbein, Nathan Nunn, and Daniele Paserman for their valuable comments. All errors are our own.

†University of British Columbia. Email: yuting37@student.ubc.ca.

‡Independent Researcher. Email: liqiang.liu6@outlook.com.

§Boston University. Email: jiaruiw@bu.edu.

"The virtue of the ruler is like the wind; the virtue of the common people is like the grass. When the wind blows, the grass bends."

—— Confucius (*The Analects*)

1 Introduction

Collectivism, as a fundamental cultural dimension, influences economics, politics, psychology, and social interactions. It shapes individual behavior and decision-making (Hofstede, 1984; Alesina and Giuliano, 2015), informs the functioning of economies and institutions (Greif, 1994; Kyriacou, 2016), drives innovation and long-run growth (Gorodnichenko and Roland, 2017, 2011), and affects cross-cultural interaction, cooperation, and conflict resolution (Hall, 1976; Pankaj, 2001; Samuel, 1993; Jiao and Zhao, 2023).

Despite the growing body of literature, the determinants of collectivism remain underexplored. China has historically been, and continues to be, characterized by a strong collectivist orientation with notable variation across regions (Hofstede, 2001; Hofstede et al., 2010; Zeng and Greenfield, 2015; Van de Vliert et al., 2013), making it an ideal context for studying the causes of collectivism. Motivated by China's long history of centralized rule (e.g., Twitchett et al., 1978; Li, 2023) and by theoretical work linking political systems to cultural beliefs and values (e.g., Alesina and Giuliano, 2015; Acemoglu and Robinson, 2025), this paper offers a novel perspective by examining whether political power shapes individuals' collectivist orientations. Specifically, we ask whether greater exposure to historical centralized political regimes leads individuals to be more collectivist.

Our empirical strategy is to leverage within-province, within-birth-year variation across counties. Intuitively, our analysis compares individuals from the same cohort within a given province, but from counties with differing degrees of exposure to historical centralized political regimes, to assess differences in collectivism. We achieve this by controlling for province-by-birth-year fixed effects and, in our more demanding specifications, by including county-level control variables (rice theory, lineage theory, and local economic activities) identified in the literature as determinants of collectivism.

While this empirical strategy offers a compelling comparison in political exposure, measured by the weighted geographical distance between counties and historical centralized national capitals, it does not eliminate all endogeneity concerns. To address the confounding issue stemming from the endogenous choice of historical capital locations, we employ an instrumental variables approach that exploits plausibly exogenous variation based on the centroids of the initial territories of ten centralized Chinese dynasties. We further control for migration rates in the most demanding specification to mitigate confounding concerns arising from the endogenous choice of parental residence. Additionally, two placebo

exercises—using "fake" capitals and "randomized" distances—mitigate the concern that the estimated effects are driven by factors other than exposure to historical centralized political regimes. Using data from the 2005 China Census, the main results are based on two-stage least squares (2SLS) from our most demanding specification.

Our estimates are consistent across all specifications for both measures of collectivism: name frequency and name-character frequency. They remain robust when collectivism is constructed at different geographical levels (region, province, or nation), when employing more conservative inference approaches (such as spatial heteroskedasticity-autocorrelation standard errors and two-way clustered standard errors), and when including additional control variables.

The estimated effects indicate that, within the same province, individuals from the same cohort in counties with greater exposure to historical centralized national capitals exhibit stronger collectivism than their counterparts from counties with less exposure. These effects are not driven by previously identified determinants of collectivism in the literature. Furthermore, the estimates are not only statistically significant and stable but also quantitatively meaningful. For instance, when collectivism is measured by name frequency, our most demanding specification estimate suggests that a one–standard deviation increase in exposure to historical centralized political regimes corresponds to a 0.50–standard deviation increase in collectivism.

To further strengthen the credibility of our estimates, we employ an alternative measure of political exposure: the duration of centralized regime coverage, defined as the total number of years a county was governed by the ten centralized Chinese dynasties. The results from this alternative estimation are consistent with our findings based on the primary measure of political exposure.

In addition to estimating the overall treatment effect, we examine heterogeneity along three dimensions—gender, region, and the pre- versus post-1949 period—where differences in effects are expected. The results reveal significant heterogeneity across regions and between the periods before and after the establishment of the People's Republic of China (PRC), while no systematic differences are observed between males and females. Specifically, we find that the effect is most pronounced in the southwest and northwest regions, where average levels of collectivism are relatively lower. Furthermore, individuals born after 1949 exhibit greater responsiveness to political exposure, a period marked by a stable and centralized system that emphasized collectivist and socialist values.

In summary, we find that political exposure—specifically, exposure to historical centralized political regimes—affects individuals' collectivism orientation. However, our analysis does not uncover the mechanisms through which political exposure shapes collectivism,

leaving this as an important avenue for future research.

Related Literature. Our paper relates to several strands of literature. First, there is a growing body of work that seeks to explain regional variation in collectivism in China, including the climato-economic theory[1] proposed by Van de Vliert et al. (2013), the rice theory[2] advanced by Talhelm et al. (2014), and the lineage theory[3] discussed by Hsu (1963) and Gong et al. (2021). We contribute to this literature by introducing a novel perspective—exposure to political regimes—to explain the origins of collectivism in China. Our approach differs from existing studies by exploiting cross-county variation within provinces.

Our paper also relates to a large literature that examines the interaction between political institutions and culture (e.g., Tabellini, 2010; Alesina and Giuliano, 2015; Bisin and Verdier, 2024). While much of the literature provides rich theoretical insights into these interactions, the empirical evidence remains relatively limited. We contribute to this literature by providing empirical evidence that exposure to historical political regimes shapes individuals' cultural orientations and beliefs. There is a body of work examining how political rule shapes individuals' cultural preferences (e.g., Acemoglu et al., 2001; Alesina and Fuchs-Schündeln, 2007), but this literature has primarily focused on the Western world. Our study instead examines the Eastern context, highlighting how historical political systems continue to shape cultural orientations and preferences in modern times.

Finally, our paper contributes to the literature on measuring collectivism and individualism. In addition to using name frequency as in Bazzi et al. (2020), we introduce an alternative measure based on the frequency of name characters, which is particularly suitable in the Chinese context where parents choose both the characters and their combinations when naming their children.

The remainder of the paper is organized as follows. Section 2 introduces the relevant historical background. Section 3 describes our main data sources, the measurement of collectivism, and the construction of political exposure. Section 4 outlines our empirical strategy and provides supporting evidence for identification. Section 5 presents our results on the effects of exposure to political power on collectivism, along with robustness checks, placebo exercises, and heterogeneity analyses. Section 6 reports results using an alternative measure of political exposure. Section 7 concludes and discusses limitations.

[1]The climato-economic theory states that 22 °C represents the optimal temperature for human survival; greater deviations from this level increase the challenges of survival, thereby compelling individuals to cooperate in the production of food and other essential goods (i.e., collectivism).

[2]The rice theory argues that a more collectivist culture emerged over the long history of rice cultivation, as rice farming requires intensive irrigation and thus necessitates greater cooperation among cultivators.

[3]The lineage theory posits that family and kinship networks serve as the primary social organizations through which individuals address problems. These structures embody mutual dependence in interpersonal ties, reproduce the social foundations of collectivist culture, and foster a worldview centered on group spirit.

2 Historical Background

2.1 Centralized Political Regimes in Imperial China

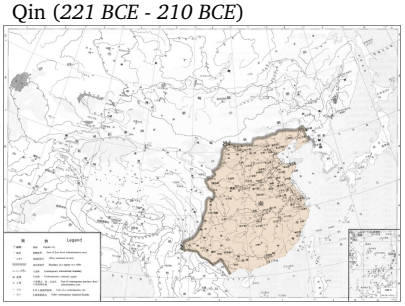
Centralized political regimes, in contrast to decentralized arrangements, refer to a system in which all state power is concentrated in the central government while local authorities function only as executors of central directives. Such regimes define the structural relationship between central and local governments as well as the mechanisms through which the ruling class governs society (Bai, 2011). The defining feature of centralized political regimes is the absence of local autonomy in political, economic, military, and cultural affairs. Local administrations are strictly subordinate to central authority. Local officials are generally appointed by the central government, held accountable to it, and closely supervised (Bai, 2011).

Throughout Chinese history there were fifty-eight dynasties (see Appendix Table 1; Twitchett et al., 1978). Of these, ten are generally regarded as having established unified centralized political regimes, as listed in Table 1. Their territorial boundaries are depicted in Figure 1.[4] The first such regime was founded by Qin Shihuang in 221 BCE. He introduced the commandery-county system, placing local administration under direct central control. A series of unification policies, including the standardization of script, weights and measures, currency, axle widths, and laws, ensured the dominance of the central government (Sima, 1995). The centralized political regime created by the Qin profoundly shaped Chinese political institutions for the next two millennia. Although the dynasty lasted only two emperors from 221 to 207 BCE, its institutional legacy endured. The Han dynasty inherited and consolidated the centralized regime created by the Qin. During the Western Han (202 BCE–9 CE), the imperial court curtailed the power of regional lords through the "Decree of Enfeoffment," thereby strengthening central authority. The dynasty persisted across both Western and Eastern periods for nearly four centuries, each ruled by around fourteen emperors. Subsequent dynasties established unified centralized regimes and further developed these institutions. The Sui (581–618 CE) introduced the "Three Departments and Six Ministries" system, refining the structure of central government. The Tang (618–907 CE), one of the most powerful and enduring dynasties, expanded administrative control through a three-tier system of circuits, prefectures, and counties, while appointing military governors to reinforce central oversight. The Northern Song (960–1127 CE)

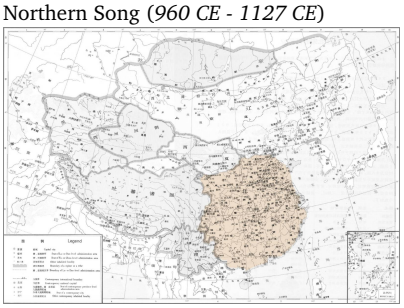
[4]Three Kingdoms, Jin, Northern and Southern Dynasties, the Five Dynasties and Ten Kingdoms, and other dynasties saw multiple competing regimes. While centralized arrangements existed within individual states, no unified centralized regime governed the entirety of China.

Figure 1: Territorial Boundaries of Centralized Political Regimes, 221 BCE – 1911 CE

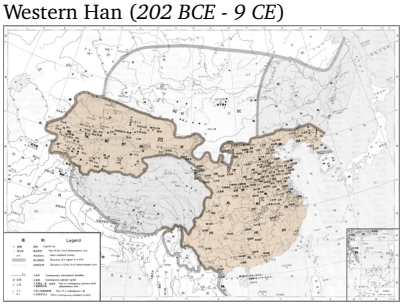
Qin (*221 BCE - 210 BCE*)



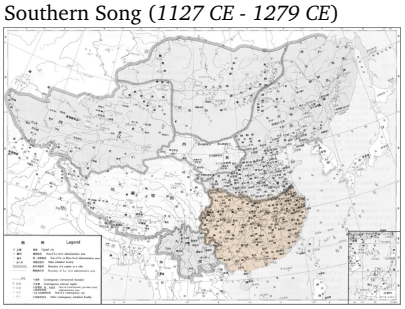
Northern Song (*960 CE - 1127 CE*)



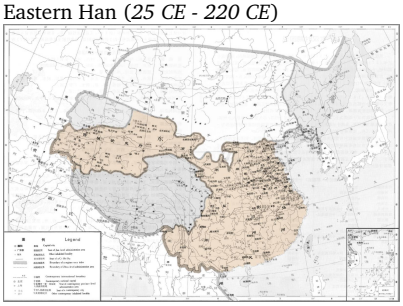
Western Han (*202 BCE - 9 CE*)



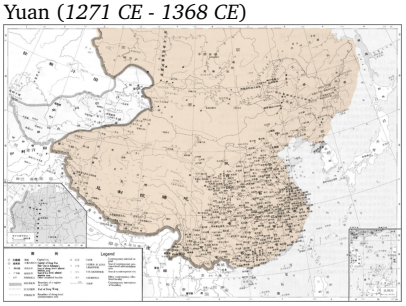
Southern Song (*1127 CE - 1279 CE*)



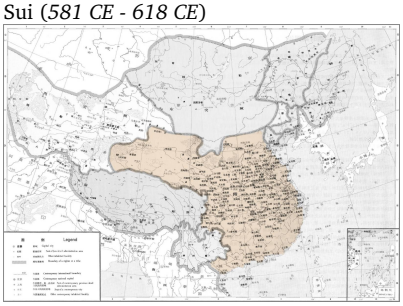
Eastern Han (*25 CE - 220 CE*)



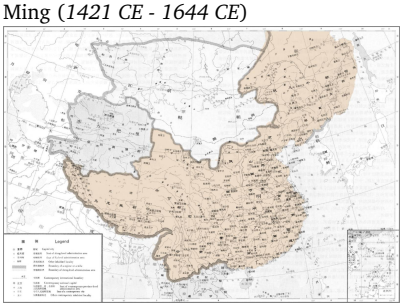
Yuan (*1271 CE - 1368 CE*)



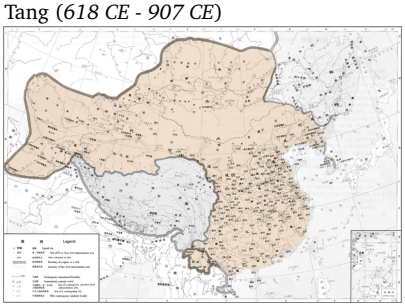
Sui (*581 CE - 618 CE*)



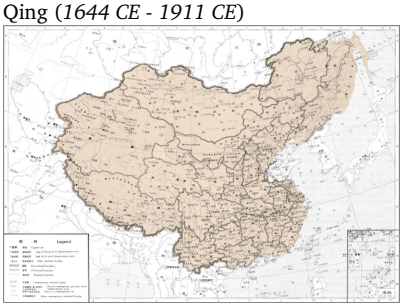
Ming (*1421 CE - 1644 CE*)



Tang (*618 CE - 907 CE*)



Qing (*1644 CE - 1911 CE*)



Notes: Maps drawn by authors using data from Tan (1987), and the territories are shaded in brown.

enhanced central authority by instituting fiscal commissioners to ensure that local revenues were remitted directly to the capital, a system largely preserved under the Southern Song (1127–1279 CE). The Yuan (1271–1368 CE) established the provincial system, under which centrally appointed officials supervised military, administrative, and fiscal affairs, creating the framework of the modern provincial structure. The Ming (1368–1644 CE) and Qing (1636–1912 CE) inherited these institutions and further strengthened centralized rule. The durability of these regimes rested not only on institutional design but also on the suppression of intellectual and cultural dissent. Qin Shihuang's burning of books and burying of scholars, along with the literary inquisitions under the Song, Ming, and Qing, enforced ideological conformity and reinforced centralized authority (Twitchett et al., 1978).

Table 1: Historical Centralized Political Regimes and Capital Cities

Political Regime	Start Year	End Year	Duration (Years)	Capital City
Qin	-221	-210	11	Xianyang
Western Han	-202	9	211	Chang'an
Eastern Han	25	220	195	Lo-yang
Sui	581	618	37	Chang'an
Tang	618	907	289	Chang'an
Northern Song	960	1127	167	K'ai-feng
Southern Song	1127	1279	152	Lin'an
Yuan	1271	1368	97	Da-du
Ming	1421	1644	223	Beijing
Qing	1644	1911	267	Beijing

Notes: The table reports the ten major dynasties generally recognized as unified centralized political regimes in imperial China. For a comprehensive list of all Chinese dynasties between 221 BCE and 1911 CE, including fragmented regimes and competing states, see Appendix Table 1. Capital city names follow conventional English transliterations used in the historical literature.

2.2 Capital Cities of Centralized Political Regimes

The capitals of the ten centralized dynasties are illustrated in Figure 2 and detailed in Table 1. Broadly speaking, the capitals of China's centralized regimes were located in the northern and central regions of today's Chinese territory, rather than on the periphery. The Qin, Western Han, Sui, and Tang all established their capitals near Chang'an (present-day Xi'an). The Eastern Han shifted the capital eastward to Luoyang, while the Northern Song located its capital in Kaifeng. The Southern Song established its seat in Lin'an, corre-

sponding to present-day Hangzhou. In the later dynasties, the Yuan, Ming, and Qing chose Beijing as their capital, which remained the political center through the late imperial era.

Figure 2: Historical Capital Cities of Centralized Political Regimes

Notes: The figure presents the historical capital cities of China's centralized dynasties from 221 BCE to 1911 CE. The size of each circle is determined by two factors: (1) the number of times a location served as a capital and (2) the duration of the dynasty during which it was the capital.
Data Sources: The authors' construction using historical records from The Cambridge History of China (Twitchett et al., 1986).

In the pre-modern period, under the institutional framework of centralized rule, the choice of capital was critical to national security, the management of political conflict, and the effectiveness of governance (Campante et al., 2019). Several factors consistently influenced the location of capitals. First, military security: the capital needed to be defensible against external threats, often relying on natural barriers such as mountains and rivers for protection (Hou, 1986; Campante and Do, 2014). Second, economic and resource considerations: fertile plains surrounding the capital, such as the Guanzhong and North China Plains, were essential to ensuring reliable supplies of grain and other resources. Third, centrality and connectivity: locating the capital near the geographic center of the polity facilitated the dispatch of orders, the movement of troops, and the integration of trade routes; transport accessibility further reinforced its role as an administrative hub (Hou, 1986; Zhao, 2021). Fourth, climate and environmental shocks: floods, droughts, and famines often precipitated dynastic decline, making climatic conditions a crucial fac-

tor in the rise and fall of capitals in traditional societies (Diamond, 2011). Finally, the Emperor's First Seat of Power Hypothesis emphasizes the ruler's political base. Founding emperors frequently chose to locate the capital in the region where they first consolidated their authority, taking advantage of existing local networks, familiarity with the terrain, and reduced informational asymmetry. For instance, Liu Bang, the founding emperor of the Han, established his base in the Guanzhong region and ultimately selected Chang'an as the capital (Zhao, 2021).

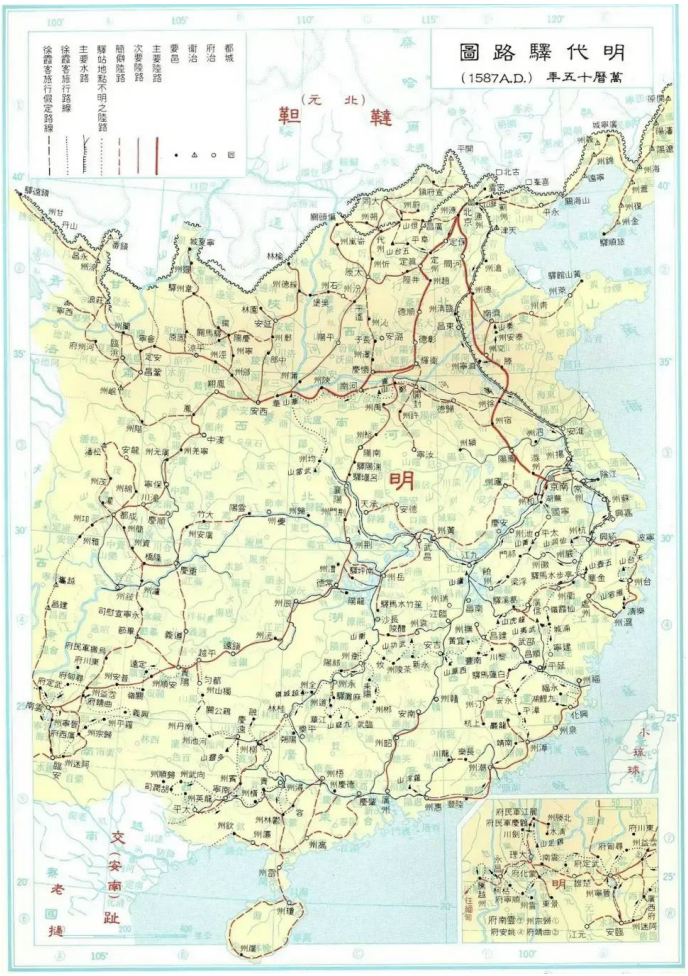
2.3 Spatial Reach of Centralized Political Power

In imperial China, the spatial extension of centralized authority relied primarily on the postal relay system, which transmitted imperial edicts from the capital city to local administrations, radiating outward from the political center to the more distant peripheries.

The Qin dynasty under the First Emperor Qinshihuang established the foundations of this system by constructing the *chidao* (imperial highways), a nationwide transport network centered on the capital of Xianyang, and instituting a tightly organized courier system. This network consisted of "posts" (*you*), which specialized in transmitting urgent state documents—required by law to cover up to 400 *li* (about 125 miles) per day through horse relays—and "pavilions" (*ting*), smaller stations embedded in local communities that combined security functions with routine information transmission. For the first time, a standardized communication infrastructure spanned an empire of unprecedented scale, laying the foundation for what later dynasties described as "the Han inheriting the Qin system." (Twitchett et al., 1986)

The relay network reached its apogee under the Sui and Tang dynasties, when it was institutionalized and systematized to an unprecedented degree. By the mid-Tang, more than 1,600 relay stations spanned the empire, including over 1,200 land-based stations and more than 250 water stations, employing more than 20,000 personnel. Regulations mandated strict transmission speeds: the most urgent military dispatches, for instance, were required to travel 500 *li* (about 155 miles) per day. Stations were placed roughly every 30 *li* (about 9 miles), with their scale and resources varying according to geographic and strategic importance, and were equipped with horses, boats, or staff as appropriate. The Song dynasty developed a more differentiated system, structured around three levels of urgency: the standard relay (*dipu*), the emergency relay (*jijiaodi*), and the most urgent "gold-token" relay (*jinpai jijiaodi*). The *dipu*, staffed by infantry couriers known as "relay soldiers," formed the backbone of the network, with relay points spaced at intervals of 10–20 *li* (3–6 miles) to ensure rapid stepwise transmission. The Yuan dynasty expanded

Figure 3: Relay Routes of the Ming Dynasty



Note: This figure is sourced from A Study of Post Stations in the Ming Dynasty (Yang, 1994).

the system to an intercontinental scale. Its *zhanchi* network extended across Eurasia, integrating horse relays, water stations, wagon depots, and sedan-chair stations into a vast communication infrastructure. Access to the system required an imperial pass (*paifu*) issued directly by the court, underscoring its role as both a logistical tool and an instrument of political control. The Ming and Qing dynasties inherited and institutionalized this system into a stable and mature "dual-track" framework. The *yizhan* (relay stations) handled urgent military dispatches, major political edicts, and the reception and forwarding of high officials, tribute, and state goods, while the *dipu* functioned as a complementary system for transmitting routine documents among different levels of government. Together, these mechanisms ensured a highly structured, hierarchical, and durable channel through which the central state projected authority across a vast territorial expanse. Figure 3 illustrates the relay routes of the Ming dynasty.

3 Data and Measurement

In this section, we first describe the data sources used in the analysis. We then introduce the measurement of the treatment, the outcome, and the key control variables. We also provide supporting evidence showing that our collectivism measures validate each other. Lastly, we present descriptive evidence for the correlation between collectivism and exposure to centralized political regime.

3.1 Data

Historical Capital Cities of Imperial China. The data is constructed from historical records documented in The Cambridge History of China (Twitchett et al., 1986). It covers the period from the Qin Dynasty (221 BCE), when China first established a centralized political system, through the fall of the Qing Dynasty in 1911, which marked the collapse of the last imperial regime. In total, the dataset encompasses 59 historical regimes and 65 capital cities, spanning both major dynasties and short-lived states, including those of the Five Dynasties and Ten Kingdoms period. Since some dynasties shifted their capitals during their rule, a single regime may correspond to more than one capital city. For the purposes of analysis, we focus on ten dynasties that maintained a centralized political system and their recorded capital cities, as summarized in Table 1. Because several dynasties shared the same capital location, the final dataset contains seven distinct cities.

The geographic coordinates of these cities are mapped to their corresponding modern locations, with refinements planned to incorporate greater historical precision where warranted. In addition, we use the recorded start and end years of each dynasty to calculate its duration. On average, these centralized dynasties persisted for 154 years, with the Tang Dynasty lasting the longest at 289 years and the Qin Dynasty the shortest at just 11 years.

Population Census of China. The data is sourced from the 2005 census of China. We include all individuals with Han-Mandarin given names with a length of at most three characters, except for those from Anhui province. The given name data from Anhui is subject to measurement errors – more than half of the names are recorded as numbers in the dataset. Since we focus on the frequency of given names as our individualism-collectivism measure, this measurement error will mechanically bias our name-based measure towards collectivism in Anhui. Therefore, we exclude Anhui province in the analysis.

3.2 Measuring Exposure to Centralized Political Power

Weighted Geographical Distance. We define the political exposure of a county as the factor-weighted geographical distance between the county's centroid and the location of a historical centralized national capital, with Euclidean distance employed as the measurement. In this context, we consider two factors that may affect the magnitude of exposure: the duration of the dynasty and the time gap between the midpoint of the dynasty and the year (i.e., 1912) following the fall of the Qing dynasty, the last imperial dynasty of China. Although China has had multiple centralized national capitals throughout its history, these capitals never coexisted during the same time period. Therefore, the average of the summed individual exposures for all historical centralized national capitals provides a reasonable representation of the aggregate exposure. Specifically, we calculate the weighted average exposure (WAE) of county c using the following formula:

$$WAE_c = \frac{1}{n}\sum_{i=1}^{n}\left(weight_{c,i} \times \frac{1}{\varphi\left(Euclidean\ distance\ to\ capital_{c,i}\right)}\right) \tag{1}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{f\left(duration_i\right) \times g\left(time\ gap_i\right) \times \sigma_{c,i}}{\varphi\left(Euclidean\ distance\ to\ capital_{c,i}\right)}\right) \tag{2}$$

where i denotes an individual historical national capital, and n represents the total number of historical centralized national capitals.

The factor weights are determined by three components: $f\left(duration_i\right)$, $g\left(time\ gap_i\right)$, and $\sigma_{c,i}$. In our baseline results, we set $f\left(duration_i\right) = duration_i$, capturing the positive relationship between the duration of the dynasty and the extent of exposure. Furthermore, we set $g\left(time\ gap_i\right) = \frac{1}{time\ gap_i}$, reflecting the negative relationship between the time gap and the extent of exposure. Additionally, we set $\sigma_{c,i} = 1$, assuming that there are no additional crucial factors to determine the magnitude of exposure.[5] Lastly, we set $\varphi\left(Euclidean\ distance\ to\ capital_{c,i}\right) = Euclidean\ distance\ to\ capital_{c,i}$, capturing the idea that the closer a county is to the national capital, the greater its influence from the national capital. In the Appendix, we explore alternative functional forms for robustness. The baseline formula for calculating the weighted average exposure is, therefore, as follows:

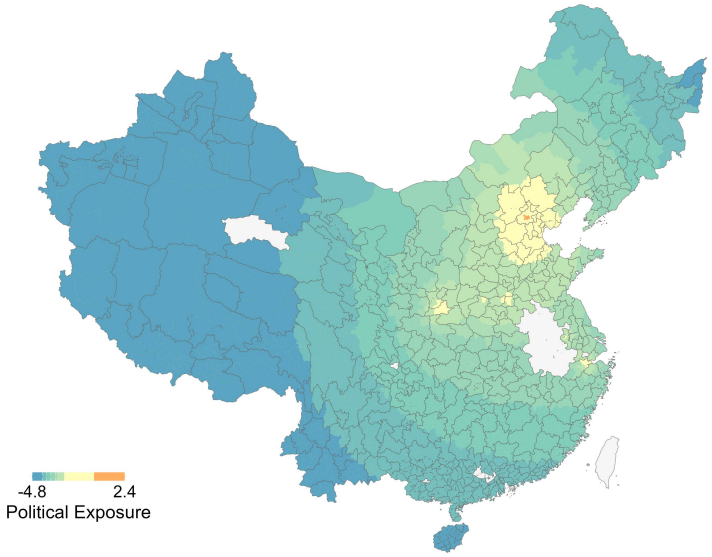
$$WAE_c = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{duration_i}{time\ gap_i} \times \frac{1}{Euclidean\ distance\ to\ capital_{c,i}}\right) \tag{3}$$

[5]This assumption may not be entirely accurate, as other factors, such as the distribution of the population discussed in Campante and Do (2010), could also influence the extent of exposure. However, the results are expected to be similar.

We interpret a higher value of the measure, WAE_c, as indicating that county c is more exposed to the set of historical national capitals. For instance, individuals in a county are more exposed to political power if their county is located closer to a national capital with a longer duration and a smaller time gap from the year following the fall of the Qing dynasty.

Figure 4 displays geographical variations in this distance measure. A brighter shading on the map corresponds to a closer proximity to the historical national capitals as shown in Figure 2, indicating a higher degree of exposure. For example, counties closer to Beijing and Shaanxi experience greater exposure than those in Xinjiang.

Figure 4: Variation in Political Exposure: Weighted Geographical Distance



Notes: The figure displays the map of political exposure measured by weighted distance in China, where counties with brighter shading have more political exposure. The boundaries on the map are at the prefectural level. The light gray areas represent regions where data are missing.

Centralized Regime Coverage. In addition to the distance measure, we introduce a second metric for exposure to centralized political power: centralized regime coverage. We calculate this measure by multiplying two components: (1) an indicator variable indicating whether county c was part of dynasty d, and (2) the duration of dynasty d. The average centralized regime coverage measure ($ACRC$) is specifically calculated using the following formula:

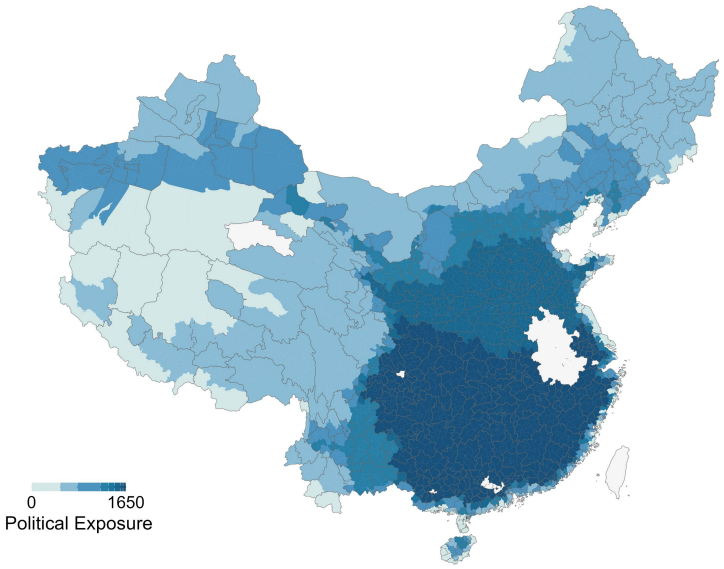
$$ACRC_c = \frac{1}{n} \sum_{d=1}^{n} \left(\mathbf{1}(covered_{c,d}) \times duration_d \right) \tag{4}$$

where $\mathbf{1}(covered_{c,d})$ is an indicator variable equal to 1 if county c was part of dynasty d, $duration_d$ is the duration of dynasty d, and n represents the total number of historical

centralized national capitals.

We interpret $ACRC_c$ as the total number of years a county was governed by the ten centralized Chinese dynasties. A higher value of $ACRC_c$ indicates a greater exposure to these centralized political regimes. Figure 5 illustrates the geographical variation of the coverage measure, with darker shading representing a longer period a county was governed by the centralized Chinese dynasties. As shown on the map, counties in Central and East China had a greater duration of centralized dynastic rule than those in Northeast and Northwest China.

Figure 5: Variation in Political Exposure: Centralized Regime Coverage



Notes: The figure displays the map of political exposure measured by centralized regime coverage in China, where counties with darker shading have more political exposure. The boundaries on the map are at the prefectural level. The light gray areas represent regions where data are missing.

3.3 Measuring Collectivism

We use multiple measurement approaches for collectivism to strengthen the robustness of our results.

Name Frequency. We measure collectivism based on the frequency of given names, following the approach as in Bazzi et al. (2020). Specifically, for each unique name in our sample, we first calculate its frequency among the group of people born in the same year, with the same sex, and in the same region.[6] The underlying assumption is that adopting common names reflects collectivist tendencies, while choosing unique names reflects the

[6]We check robustness by measuring collectivism at different geographic levels—province and nation, respectively.

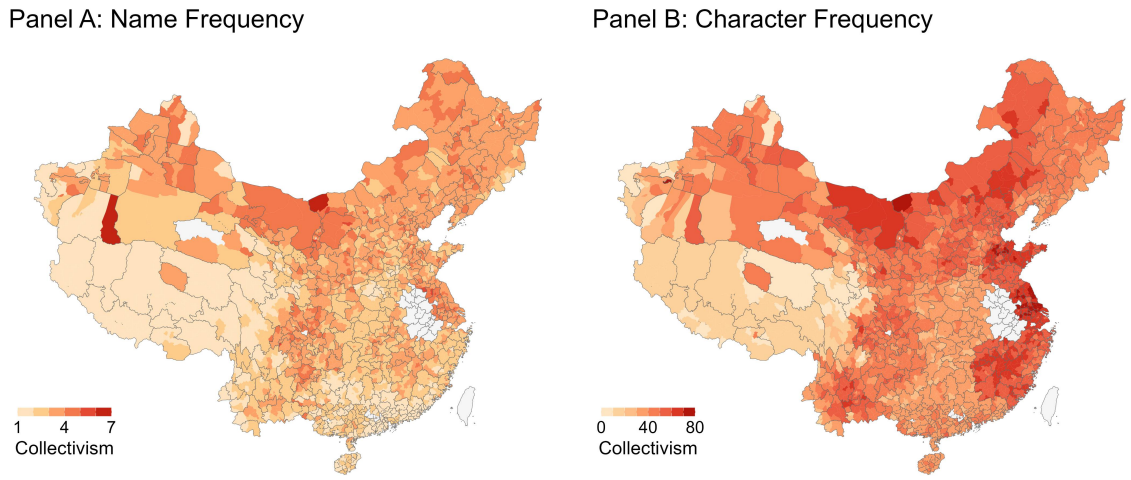
inclination to stand out. We then calculate the average name frequency for each county-by-birth-year cell, to obtain an aggregated individualism-collectivism measure for each cell.

Character Frequency. Unlike English names, which are typically drawn from an established set of words, Chinese given names usually consist of one or two characters (Cai et al., 2018). This makes it possible to use the uniqueness of characters in given names as an alternative measure of parents' collectivism–individualism values (Cai et al., 2018). While Cai et al. (2018) rely on character frequency from the Modern Chinese Character Frequency of Use Dictionary (State Language Commission and State Bureau of Standards, 1986), our approach uses frequency data specific to names. This adjustment is important because some characters may occur often in everyday language, but rarely appear in names. We therefore construct a name-based measure of individualism–collectivism that parallels the given-name-level frequency measure. For each unique character in the sample, we calculate its frequency within the same birth year, sex, and region group. For two-character names, we take the mean frequency of the two characters. The resulting average character frequency, aggregated at the county-by-birth-year level, serves as an alternative metric of the local individualism–collectivism propensity.

Correlation Between Collectivism Measures. While our two collectivism measures come from different perspectives, we expect them to be correlated because they measure the same underlying outcome. To confirm the correlation, we present binned scatterplots and estimate regressions with and without province-by-birth-year fixed effects, presenting the results in Appendix Figure 1. The results indicate a statistically significant correlation between collectivism measured by name frequency and collectivism measured by name character frequency in both cases.

Geographical Distribution of Collectivism Measures. Figure 6 shows the county-level variation in the two collectivism measures: the name-frequency measure in Panel A and the character-frequency measure in Panel B. As the figure indicates, the northern and central regions of China are generally more collectivistic. Moreover, within provinces, counties located closer to historical centralized national capitals as shown in Figure 2 tend to be more collectivistic.

Figure 6: County-Level Variation in Collectivism Measures



Notes: The figure illustrates the county-level variation in two measures of collectivism in China: Panel A uses name frequency, while Panel B uses name character frequency. The darker the county, the more collectivistic it is. The boundaries on the map are at the prefectural level. The light gray areas represent areas where data are missing.
Source: Authors' calculations using a representative random sample of the 2005 census data.

3.4 Control Variables

We next introduce data sources for the key control variables identified in the literature as determinants of collectivism in China.

Rice Suitability. We calculate the average rice suitability for each county using the rice cultivation suitability index from the Global Agro-Ecological Zones (GAEZ) Version 4,[7] provided by the Food and Agriculture Organization of the United Nations (FAO) and the International Institute for Applied Systems Analysis (IIASA). The original suitability index was derived for the period 1971–2000, based on climatic, soil, and topographic factors, at a spatial resolution of 5 arcminutes and a scale ranging from 0 to 10,000.

Clan Strength. We use the county-level clan strength measure, obtained from Chen et al. (2024). It is constructed as the first principal component of four proxies: (i) an indicator for provinces in southern China based on Tang and Zhao (2023); (ii) the number of provincially or nationally recognized ancestral temples in a county; (iii) the number of genealogy books for clans in the county per 10,000 residents; and (iv) the climatic and soil suitability for growing rice rather than wheat.[8]

[7] Global Agro-Ecological Zones: Version 4.

[8] We impute missing county-level clan strength measure using the prefectural mean; when the prefectural mean is unavailable, we use the provincial mean. We prefer imputation to dropping missing values, as the missingness is likely non-random.

Night-Time Light. We use night-time light (NTL) intensity as a proxy for local economic activity. We compute average night-time light intensity at the county level. The data come from the Defense Meteorological Satellite Program – Operational Linescan System (DMSP-OLS) satellite program and cover the years 1992 to 2005. Specifically, we use the average visible light product provided by the National Oceanic and Atmospheric Administration's (NOAA) National Centers for Environmental Information (NCEI), which reports annual average light intensity at approximately 1 km resolution.[9] The data are recorded as digital number values ranging from 0 to 63, filtered to include only stable light sources such as cities and infrastructure.

3.5 Descriptive Evidence

Does exposure to historical centralized political power cause individuals to become more collectivist? We provide descriptive evidence suggesting that this can indeed be the case. Figure 7 illustrates the relationship between collectivism and political exposure, with both variables being residualized with respect to province-by-birth-year fixed effects. All four associations are positive and statistically significant. The findings indicate that individuals residing in counties with higher exposure to historical centralized political power tend to be more collectivist than individuals from the same province but residing in different counties with less exposure.

4 Empirical Strategy

In this section, we discuss our main empirical strategy by first introducing the baseline specification and then addressing possible identification challenges.

4.1 Baseline Specification

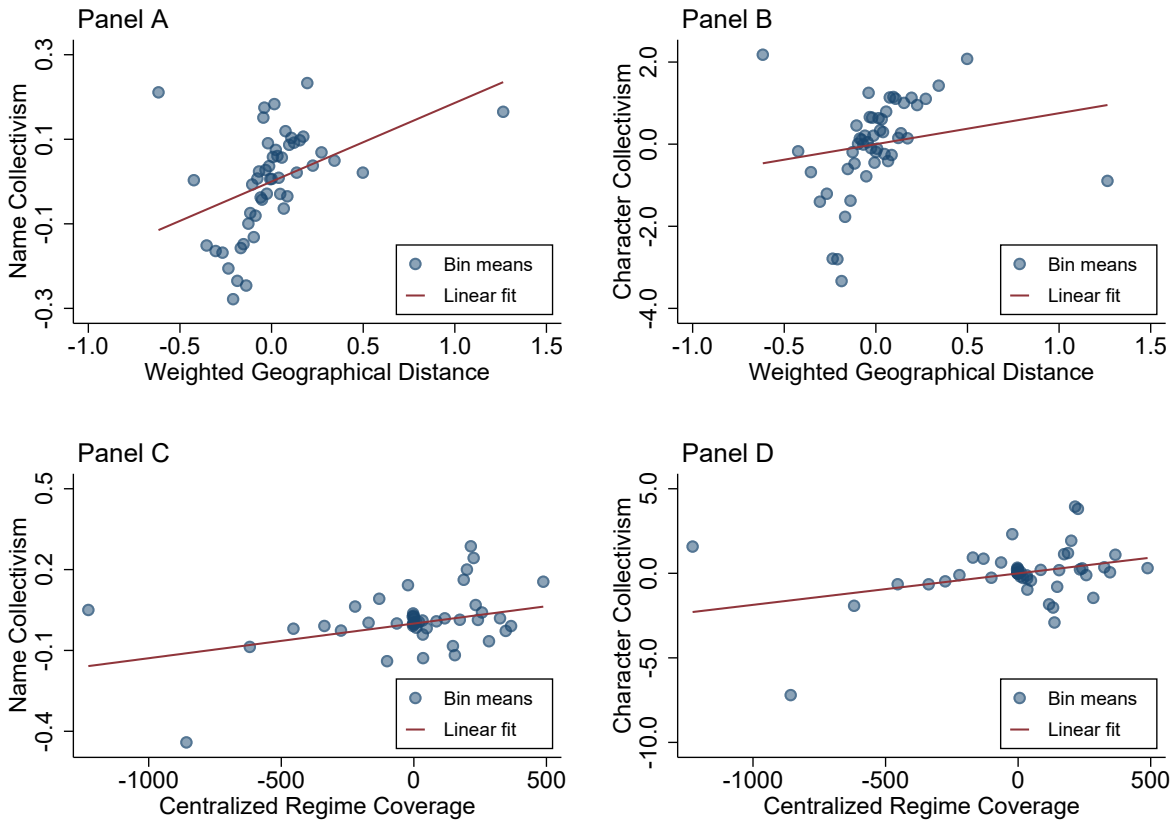
Our observations are the county-by-birth-year cells observed in calendar year 2005. Specifically, we estimate the following regression model:

$$Y_{c,b} = \alpha_0 + \alpha_1 \log(Exposure_c) + \phi_{pro \times b} + \varepsilon_{c,b} \tag{5}$$

where the outcome variables, $Y_{c,b}$, represent collectivism measures, and $Exposure_c$ is the variable of interest, capturing average exposure to historical centralized national capitals.

[9] Nighttime Lights.

Figure 7: Correlation Between Collectivism and Political Exposure



Notes: This figure shows binned scatterplots (50 bins) of the relationships between collectivism measures (name-based and character-based) and political exposure measures (distance-based and coverage-based). Blue dots represent bin means, and the red line shows the linear fit. All measures are residualized with respect to province-by-birth-year fixed effects, and all correlations are positive and statistically significant at the 1% level ($p < 0.01$).

In the regression, we include province-by-birth-year fixed effects, $\phi_{pro\times b}$, to rule out comparisons across provinces or birth cohorts, as individuals (cells) from different provinces or birth cohorts may differ in factors other than political exposure that could affect their collectivism. Additionally, we incorporate county-level controls, $\mathbf{X}_c$, into the regression model specified in Equation (6).

$$Y_{c,b} = \beta_0 + \beta_1 \log(Exposure_c) + \phi_{pro\times b} + \beta_2^{'}\mathbf{X}_c + \varepsilon_{c,b} \tag{6}$$

where the controls used in the model include rice suitability, clan strength, and night-time light, which aim to address confounding issues related to societal patterns of farming rice versus wheat, cultural factors (e.g., lineage), and economic activity, as identified determinants of collectivism in the literature (e.g., Talhelm et al., 2014; Gong et al., 2021; Van de

Vliert et al., 2013).

β_1 is the coefficient of interest, representing the relationship between collectivism and exposure to historical centralized national capitals.

4.2 Instrumental Variables Strategy

The parameter β_1 is identified through a within-province comparison, where we compare outcomes for county-by-birth-year cells from the same birth cohort but in different counties. For a causal interpretation of β_1, the key identifying assumption is that $\log(Exposure_c)$ is not correlated with $\varepsilon_{c,b}$, conditional on the covariates included in Equation (6). In other words, the differential exposure to historical centralized national capitals between the counties used for comparison is not due to endogenous selection by the ruling dynasty's choice of capital sites or by the location choices of individuals' parents.

Reverse Causality. Theoretical work implies the possibility of a two-way causality between culture and political institutions (e.g., Alesina and Giuliano, 2015; Bisin and Verdier, 2024). However, reverse causality is less of a concern here, as the treatment—exposure to historical centralized national capitals—is predetermined. We estimate its impact on the collectivism of individuals born from 1912 to 2005. While confounding factors may jointly determine a dynasty's capital site and the future collectivism of individuals in given counties, it is highly improbable that the choice of capital was influenced by future collectivism.

Identification Challenge. There are two possibilities in which the identification assumption can be violated, given that the difference in political exposure between two within-province counties is determined by their distance to the historical national capitals.[10] The first possibility—and our primary concern—is that confounding factors which determined the location of historical national capitals could be correlated with collectivism, thus biasing the OLS estimates. For example, historical national capitals may have been strategically located by ruling dynasties in areas characterized by higher levels of collectivism, as this may have contributed to greater regime stability.

The second possibility is that confounding unobserved factors influencing residential choices could be also correlated with collectivism, thereby biasing the OLS estimates. For example, more collectivist individuals' parents may be more likely to reside in counties within a province located closer to historical national capitals than those who are less collectivist.

[10]The distance is calculated between the location of the county centroid and the location of the historical national capital.

IV Strategy. To address the primary concern, we employ an instrumental variables (IV) strategy inspired by the work of Campante and Do (2014). Specifically, we use the plausibly exogenous centroid of the dynasty's initial national territory as a source to construct an instrument for the potentially endogenous location of the national capital. The centroid, representing the average coordinate of the dynasty's territory, is determined solely by the dynasty's geographical shape.[11] As this is an essentially arbitrary location for an established territory, it is unlikely to directly affect any relevant outcomes. Therefore, the centroid can be considered plausibly exogenous and uncorrelated with unobserved factors, $\varepsilon_{c,b}$.

Additionally, we use the centroid of the dynasty's initial national territory, as shown in Figure 1, to establish a connection with the national capital's location. This is because the capital was chosen during the period when the dynasty's initial territory was established, with its initial territory serving as a consideration for the capital's location.

Using this source of exogenous variation, we construct our instrumental variable using the following formula:

$$WAE_c^z = \frac{1}{n}\sum_{d=1}^{n}\left(\frac{duration_d}{time\ gap_d} \times \frac{1}{Euclidean\ distance\ to\ capital_{c,d}^z}\right) \tag{7}$$

where $Euclidean\ distance\ to\ capital_{c,i}^z$ captures the Euclidean distance between the centroid of county c and the centroid of dynasty d.

To ensure the validity of our instrumental variable, two key assumptions must be satisfied: the relevance assumption and the exclusion restriction assumption.

(a) *Relevance Assumption*. Does the variation in the instrumental variable generate sufficient variation in the treatment variable? We estimate the first-stage regression model using a specification similar to Equation (6), controlling for fixed effects and other relevant controls. The results are presented in Table 2. The estimates suggest a statistically significant positive correlation between exposure to historical centralized national capitals and our instrumental variable. Moreover, the first-stage F-statistics are consistently above the conventional threshold of 10, which indicates that our instrumental variable is not weak.[12]

(b) *Exclusion Restriction Assumption*. Does our instrumental variable affect collectivism

[11] The dynasty's geographical shape may have been influenced by military and economic factors. However, such effects would be expected to operate at the national level and thus remain constant when comparing counties within the same province.

[12] We also report the p-values of the Anderson–Rubin (AR) tests (Anderson and Rubin, 1949) in Table 3. The results indicate that our analysis is not affected by a weak instrument issue.

Table 2: First Stage Results for the Instrumental Variables Estimates in Table 3

	Dependent Variable: log(exposure)	
	(1)	(2)
log(Instrument)	0.178***	0.160***
	(0.041)	(0.040)
Rice suitability ($\times$ 1000)		–0.006
		(0.005)
Clan		–0.019**
		(0.007)
Night light ($\times$ 1000)		0.045***
		(0.008)
Province-by-Birth-Year FE	✓	✓
Observations	219,197	219,197
Adj. R^2	0.866	0.872
First stage F-statistic	19.207	16.170

Notes: Standard errors are reported in brackets and clustered at the county level, and the first-stage F statistics are cluster-robust. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

exclusively through its effect on the treatment variable? First, the exogenous origin of the centroid of the dynasty's initial national territory rules out the possibility that the instrumental variable and collectivism are jointly determined by confounding factors, thereby aligning with the assumption. The remaining question is whether there are other channels through which the instrumental variable may affect collectivism. By construction, the location of the centroid should affect only the distance to each given county; therefore, we argue that the treatment variable, measured by the factor-weighted distance, is the sole channel through which the instrumental variable affects collectivism.

However, a second identification concern remains: the residing counties of an individual's parents may still be endogenous, even if our instrumental variable is valid for a given county. We address this concern by conducting the following analyses: (1) we document that internal migration was a relatively small phenomenon in our sample, with only about 9% of individuals migrating to another county and about 5% moving to a different province. This suggests that the residing county concern is unlikely to significantly impact our results. (2) We analyze the correlations between county-level migration rates[13] and our outcome, treatment, and instrument variables. Our analysis in Appendix B indicates that county-level migration rates are statistically significantly correlated with collectivism. However, they are only statistically significantly correlated with political exposure, as mea-

[13]The migration rate is defined as the percentage of individuals within a given county who relocated to a different county.

sured by weighted distance, for some specifications. Critically, county-level migration rates are not statistically significantly correlated with our instrumental variable.

While our analysis suggests that choice of individuals' parents' residing counties is less likely to be a concern, we include the county-level migration rates in our more demanding specification. The main results presented in the next section are estimated using two-stage least squares (2SLS) with the following specification:

$$\log(Exposure_c) = \beta_0 + \beta_1 \log(Exposure_c^z) + \phi_{pro \times b} + \beta_2^{'} \mathbf{X}_c + \nu_{c,b} \tag{8}$$

$$Y_{c,b} = \gamma_0 + \gamma_1 \log(\widehat{Exposure}_c) + \phi_{pro \times b} + \gamma_2^{'} \mathbf{X}_c + \varepsilon_{c,b} \tag{9}$$

where $\log(Exposure_c^z)$ is the logarithm of the instrumental variable constructed using Formula 7, and $\log(\widehat{Exposure}_c)$ denotes the predicted value of the average exposure to historical centralized national capitals from the first-stage regression in Equation (8).

Throughout, we allow the error term in all specifications to be correlated among individuals from the same county and compute standard errors clustered at the county level.

5 Exposure to Centralized Political Power and Collectivism

In this section, we begin by presenting the baseline estimates, moving from less demanding to more demanding specifications with additional controls, followed by the IV estimates. Next, we provide additional evidence to support the robustness of our results. Lastly, we explore heterogeneity in the treatment effects.

5.1 Main Results

Baseline Estimates. Columns 1-3 of Table 3 present the baseline estimates using OLS regressions, with control variables added sequentially across columns. Panels (a) and (b) report the results for our two measures of collectivism: the frequency of given names and the frequency of given names' characters.

The results in Table 3 are consistent across all specifications: exposure to historical centralized national capitals is statistically significantly associated with our measures of collectivism. As an illustration, the coefficient for our name frequency measure in Column 1 is 0.197 (SE $=$ 0.033). This coefficient decreases to 0.165 (SE $=$ 0.035) after introducing key control variables established in the literature. The coefficient remains statistically significant with the inclusion of the migration control.

Table 3: The Impact of Exposure to Centralized Political Regime on Collectivism

	(a) Collectivism Measure: Name Frequency					
	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Log(exposure)	0.197***	0.162***	0.165***	1.619***	1.741***	1.763***
	(0.033)	(0.035)	(0.035)	(0.470)	(0.548)	(0.546)
Rice suitability ($\times$ 1000)		0.100***	0.096***		0.112***	0.108***
		(0.010)	(0.010)		(0.013)	(0.013)
Clan		–0.073***	–0.073***		–0.028	–0.028
		(0.015)	(0.015)		(0.024)	(0.025)
Night light ($\times$ 1000)		0.030***	0.023***		–0.044	–0.053
		(0.008)	(0.007)		(0.030)	(0.031)
Migration rate			0.333***			0.382***
			(0.059)			(0.070)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓	✓
Observations	219,202	219,202	219,149	219,202	219,202	219,149
Outcome mean	2.725	2.725	2.725	2.725	2.725	2.725
Adj. R^2	0.304	0.306	0.306	0.281	0.279	0.279
First stage F-statistic				19.207	16.170	16.206
Anderson-Rubin, p-val				0.000	0.000	0.000
	(b) Collectivism Measure: Name Character Frequency					
	OLS			IV		
	(1)	(2)	(3)	(4)	(5)	(6)
Log(exposure)	0.961**	1.313***	1.310***	10.136***	11.464**	11.511***
	(0.393)	(0.382)	(0.382)	(3.796)	(4.417)	(4.420)
Rice suitability ($\times$ 1000)		0.658***	0.659***		0.732***	0.730***
		(0.110)	(0.110)		(0.119)	(0.119)
Clan		–0.738***	–0.737***		–0.448	–0.446
		(0.174)	(0.174)		(0.232)	(0.232)
Night light ($\times$ 1000)		–0.342***	–0.339***		–0.816***	–0.822***
		(0.088)	(0.093)		(0.256)	(0.259)
Migration rate			–0.842			0.231
			(0.775)			(0.830)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓	✓
Observations	219,202	219,202	219,149	219,202	219,202	219,149
Outcome mean	38.114	38.114	38.114	38.114	38.114	38.114
Adj. R^2	0.491	0.492	0.492	0.484	0.483	0.483
First stage F-statistic				19.207	16.170	16.206
Anderson-Rubin, p-val				0.004	0.003	0.003

Notes: This table reports the estimated effects of exposure to historical centralized political regime on collectivism. Standard errors are reported in brackets and clustered at the county level, and the first-stage F statistics are cluster-robust. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

IV Estimates. Although our most demanding baseline specification in column (3) helps to ensure the identifying variation is largely comparable, the main concern remains that the location of dynasties' capital sites, and thus exposure to historical centralized national capitals, may be correlated with other confounding factors that also cause changes in collectivism.

Columns 4-6 of Table 3 present the 2SLS estimates of Equation (9) with political exposure instrumented according to Equation (8), and also cluster-robust first-stage F-statistics and p-values from the Anderson–Rubin (AR) weak IV robust test. The 2SLS estimates confirm the findings from the OLS results: exposure to historical centralized national capitals does have a statistically significant impact on collectivism. The first-stage F-statistics of the instrumental variable across specifications are all above the conventional threshold, suggesting that the instrument is not weak. Moreover, the Anderson–Rubin test's p-value confirms the validity of our finding, even with a potentially weak instrument.

Across specifications, the 2SLS coefficient estimates remain stable for the two measures of collectivism reported in the two panels of Table 3. For example, the coefficient in column 4 for name frequency measure, is 1.619 (SE = 0.470). It increases slightly to 1.763 (SE = 0.546) when all control variables are included. Importantly, the minimal change in point estimates between specifications with and without key controls indicates that the estimated effect is not primarily driven by the rice theory (Talhelm et al., 2014), the lineage theory (Gong et al., 2021), or the economic factors (Van de Vliert et al., 2013) identified in the literature, though some of them do have an impact in our setting.

Furthermore, the 2SLS point estimates are generally considerably larger in absolute value than the corresponding OLS results. This difference may suggest that our political exposure measure, based on weighted geographical distance, is endogenous, thus leading to Omitted Variable Bias (OVB) in the OLS estimates.

Additionally, our estimates are not only statistically significant, but also quantitatively meaningful. For example, for collectivism measured by name frequency, our most demanding specification (Panel (a), column 6) shows a coefficient of 1.763. This suggests that a one-standard deviation increase in exposure to historical centralized political regimes would yield a 0.50-standard deviation increase in collectivism. Similarly, when collectivism is measured by character frequency (Panel (b)), the coefficient of 11.511 suggests that a one-standard deviation increase in political exposure would yield a 0.29-standard deviation increase in collectivism.[14]

[14]Appendix Table 2 reports the means and standard deviations (SD) of the outcome and treatment variables. The standard deviation of name-frequency collectivism is 2.595, that of character-frequency collectivism is 29.448, and that of $\log(\text{exposure})$ is 0.739.

In summary, our findings indicate that individuals (cells) from counties with greater exposure to historical centralized national capitals exhibit stronger collectivism than those in counties with less exposure in the same province. These estimates are both statistically and economically significant.

5.2 Robustness

Next, we examine the robustness of our results using multiple exercises, including collectivism measures constructed at different geographic levels, different inference methods, additional control variables, and placebo exercises. The results consistently support the robustness of our findings.

Collectivism Measures. Following Bazzi et al. (2020), our primary collectivism measures are calculated at the regional level. However, in the absence of clear theoretical or institutional guidance, it is difficult to determine the appropriate level of geographic aggregation. To assess robustness in this dimension, we replicate our estimates using collectivism constructed at the provincial level and at the national level.

Table 4 presents the corresponding 2SLS estimates of the effect of political exposure on collectivism using the IV specification in Equation (9). Across all measures and specifications, the results indicate that our findings are consistent and robust. Thus, the estimated effects are not sensitive to the choice of geographic level at which collectivism is measured.

Alternative Inference. In our main results, we cluster standard errors at the county level, which corresponds to the level at which the treatment is assigned (Abadie et al., 2023). To examine the robustness of statistical inference, we conduct two additional exercises: one that addresses potential spatial correlation and another that applies two-way clustering.

Table 5 reports the 2SLS estimates of the effect of political exposure on the two collectivism measures, using the IV specification in Equation (9). Following Conley (1999), columns 2–4 report spatial heteroskedasticity-autocorrelation (HAC) standard errors, which allow for arbitrary correlation in unobservables across counties within 50, 90, and 130 km, respectively.[15] Additionally, column 5 reports two-way clustered standard errors, by county and by birth year. This accounts for the possibility of correlated unobservables across counties that are subject to shocks affecting individuals born in the same year. As Table 5 shows, our findings are generally robust to these two alternative, more conservative inference approaches.

[15]These cutoffs are chosen based on the average county size in China. Note that county sizes vary substantially; this inference method clusters relatively small counties together but may split larger counties.

Table 4: The Impact of Exposure to Centralized Political Regime on Collectivism (Constructed at Different Geographic Levels)

	(a) Collectivism Measure: Name Frequency					
	Within Province			Within Nation		
	(1)	(2)	(3)	(4)	(5)	(6)
Log(exposure)	0.290***	0.300***	0.303***	10.556***	11.515***	11.659***
	(0.099)	(0.112)	(0.112)	(2.818)	(3.325)	(3.298)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓	✓
Key controls		✓	✓		✓	✓
Migration rate			✓			✓
Observations	219,202	219,202	219,149	219,202	219,202	219,149
Outcome mean	1.535	1.535	1.535	10.150	10.150	10.150
First stage F-statistic	19.207	16.170	16.206	19.207	16.170	16.206
Anderson-Rubin, p-val	0.001	0.003	0.002	0.000	0.000	0.000
	(b) Collectivism Measure: Name Character Frequency					
	Within Province			Within Nation		
	(1)	(2)	(3)	(4)	(5)	(6)
Log(exposure)	2.363*	2.547*	2.553*	79.030***	88.678***	89.175***
	(1.265)	(1.425)	(1.413)	(25.203)	(29.526)	(29.491)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓	✓
Key controls		✓	✓		✓	✓
Migration rate			✓			✓
Observations	219,202	219,202	219,149	219,202	219,202	219,149
Outcome mean	11.629	11.629	11.629	219.219	219.219	219.219
First stage F-statistic	19.207	16.170	16.206	19.207	16.170	16.206
Anderson-Rubin, p-val	0.048	0.049	0.047	0.000	0.000	0.000

Notes: This table reports the IV estimates using Equation (9). Key controls include rice suitability, clan strength and night light. Standard errors are reported in brackets and clustered at the county level, and the first-stage F statistics are cluster-robust. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

Additional Controls. Although we incorporate a set of key control variables that substantially reduce the risk of omitted variable bias, it is possible that additional confounders remain unaccounted for in our most demanding specification. Here, we assess the robustness of our main results by incorporating three additional control variables that could, in theory, bias our estimates. The first potential confounder is population. For instance, counties with larger populations may be located closer to historical centralized national capitals, and population size may also be correlated with stronger collectivist orientations. We address this concern by controlling for the logarithm of population, and column 2 of

Table 5: Alternative Inference for Robustness

	(a) Collectivism Measure: Name Frequency				
	County	Spatial HAC (Conley)			Two-Way
		50 km	90 km	130 km	County + Birth Year
	(1)	(2)	(3)	(4)	(5)
Log(exposure)	1.763***	1.763**	1.763**	1.763*	1.763***
	(0.546)	(0.708)	(0.839)	(0.976)	(0.638)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓
Key controls	✓	✓	✓	✓	✓
Migration rate	✓	✓	✓	✓	✓
Observations	219,149	219,149	219,149	219,149	219,149
	(b) Collectivism Measure: Character Frequency				
	County	Spatial HAC (Conley)			Two-Way
		50 km	90 km	130 km	County + Birth Year
	(1)	(2)	(3)	(4)	(5)
Log(exposure)	11.511***	11.511**	11.511*	11.511	11.511**
	(4.420)	(5.652)	(6.855)	(7.668)	(4.641)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓
Key controls	✓	✓	✓	✓	✓
Migration rate	✓	✓	✓	✓	✓
Observations	219,149	219,149	219,149	219,149	219,149

Notes: This table reports the IV estimates using Equation (9). Key controls include rice suitability, clan strength and night light. Standard errors (SE) are reported in brackets. The estimates in column 1, serving as the baseline, are taken from Column 6 of Table 3. SEs in Columns 2–4 are estimated using the Conley (1999) GMM-based HAC procedure for spatial dependence. SEs in Column 5 are two-way clustered by county and birth year. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

Table 6 indicates that the estimated effect remains similar to our main results.

The second potential confounder is county size, as it may mechanically correlate with our measure of political exposure. This could bias our estimates if county size is also associated with collectivist orientations. Column 3 of Table 6 reports the results including county size, which suggest that it poses little concern in our setting. The third potential confounder is distance to the provincial capital. For example, more collectivist individuals may prefer to reside in counties closer to provincial capitals, and the locations of provincial capitals may themselves be correlated with those of historical centralized national capitals. The results reported in column 4 of Table 6 confirm this possibility. A comparison between these estimates and our main results suggests that our main findings rest on more conservative estimates.

Lastly, we include all three additional controls into our most demanding specification, with the results reported in column 5 of Table 6. Overall, this exercise shows that our main findings remain robust to these potential confounders and represent conservative estimates.

Table 6: Robustness Checks with Additional Controls

	(a) Collectivism Measure: Name Frequency				
	(1)	(2)	(3)	(4)	(5)
Log(exposure)	1.763***	1.708***	1.778***	2.234***	2.139***
	(0.546)	(0.557)	(0.525)	(0.853)	(0.804)
Log(population)		✓			✓
County size			✓		✓
Log(distance to provincial capital)				✓	✓
Province-by-Birth-Year FE	✓	✓	✓	✓	✓
Key controls	✓	✓	✓	✓	✓
Migration rate	✓	✓	✓	✓	✓
Observations	219,149	219,149	218,322	219,149	218,322
	(b) Collectivism Measure: Character Frequency				
	(1)	(2)	(3)	(4)	(5)
Log(exposure)	11.511***	9.070**	11.512***	16.689**	13.666**
	(4.420)	(4.058)	(4.371)	(6.780)	(5.915)
Log(population)		✓			✓
County size			✓		✓
Log(distance to provincial capital)				✓	✓
Province-by-Birth-Year FE	✓	✓	✓	✓	✓
Key controls	✓	✓	✓	✓	✓
Migration rate	✓	✓	✓	✓	✓
Observations	219,149	219,149	218,322	219,149	218,322

Notes: This table reports the IV estimates using Equation (9). Key controls include rice suitability, clan strength and night light. The estimates in column 1, serving as the baseline, are taken from Column 6 of Table 3. Standard errors are reported in brackets and clustered at the county level. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

Placebo Exercise. As further robustness checks, we conduct two placebo exercises to rule out the possibility that the estimated effects are driven by factors other than political exposure. The first placebo exercise tests whether exposure to large non-capital cities affects collectivism. To implement this, we use cities that are economically important but never served as centralized national capitals.[16] The results, reported in Panel (a) of Table

[16]These cities include Shanghai, Shenzhen, Guangzhou, Chongqing, Suzhou, Chengdu, Wuhan,

7, indicate that exposure to these "fake" capitals has no statistically significant effect on collectivism.[17]

The second placebo exercise tests whether exposure to historical centralized national capitals, when the distance between counties and capitals is randomly reassigned, has any effect on collectivism. To implement this, we randomly shuffle county coordinates to generate artificial weighted distances (i.e., "fake" political exposure). Similarly, the results reported in Panel (b) of Table 7 show that counties with greater political exposure, as measured by artificial weighted distances, are not statistically more collectivistic, compared to counties with less political exposure from the same province.

Table 7: Results for Placebo Exercise

	(a) "Fake" Capitals			
	Name Frequency		Character Frequency	
	(1)	(2)	(1)	(2)
Log(exposure)	34.397	13.253	215.325	86.516
	(83.914)	(12.581)	(533.114)	(85.358)
Province-by-Birth-Year FE	✓	✓	✓	✓
Key controls		✓		✓
Migration rate		✓		✓
Observations	219,202	219,149	219,202	219,149
	(b) "Randomized" Distance			
	Name Frequency		Character Frequency	
	(1)	(2)	(1)	(2)
Log(exposure)	0.003	0.010	–0.265	–0.205
	(0.038)	(0.036)	(0.432)	(0.427)
Province-by-Birth-Year FE	✓	✓	✓	✓
Key controls		✓		✓
Migration rate		✓		✓
Observations	219,202	219,149	219,202	219,149

Notes: This table reports the IV estimates of the placebo exercises using Equation (9). Key controls include rice suitability, clan strength and night light. Standard errors are reported in brackets and clustered at the county level. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

Hangzhou, Wuxi, and Qingdao, which are among the most economically advanced cities in China and could plausibly have been national capitals.

[17]We also conduct the same exercise by randomly assigning "fake" capitals within stratified regions (i.e., Southwest China and Northeast China) that never hosted a centralized national capital in China's history. The results consistently show no statistically significant effects.

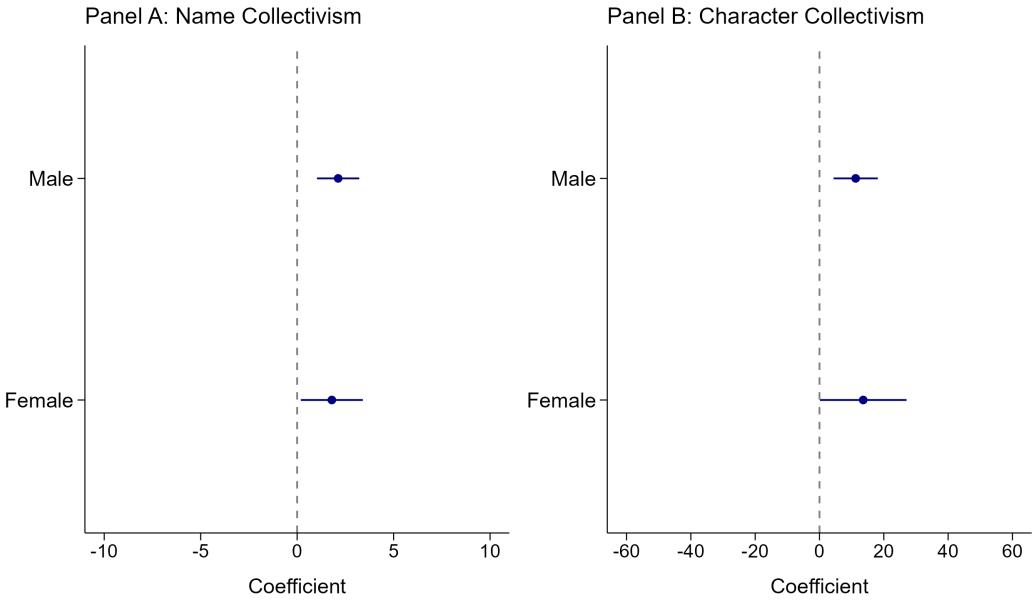
5.3 Heterogeneity

Beyond the overall treatment effect, we also examine treatment effect heterogeneity across three dimensions: gender, region, and pre- or post-1949 period.

Gender. There are at least two possible hypotheses for gender differences in the treatment effect of political exposure. First, names in Chinese culture are not only personal identifiers but also reflect sociocultural norms, including gender norms (Gao, 2011). Parents may have expressed collectivist values differently in sons' versus daughters' names, leading to heterogeneous treatment effects. Second, female names tend to be more formulaic, often drawing on a smaller pool of conventional characters associated with nature and beauty, whereas male names are typically more diverse and aspirational. (Zhigang and Micklin, 1996). This difference in naming conventions could mechanically contribute to treatment heterogeneity.

Figure 8 presents coefficient plots from the IV regression specified in Equation (9). For both collectivism measures, the results reject our hypothesis and consistently suggest no systematic gender differences in the treatment effect, indicating a general effect of exposure to historical centralized regimes along this dimension.

Figure 8: Impact of Political Exposure on Collectivism by Gender

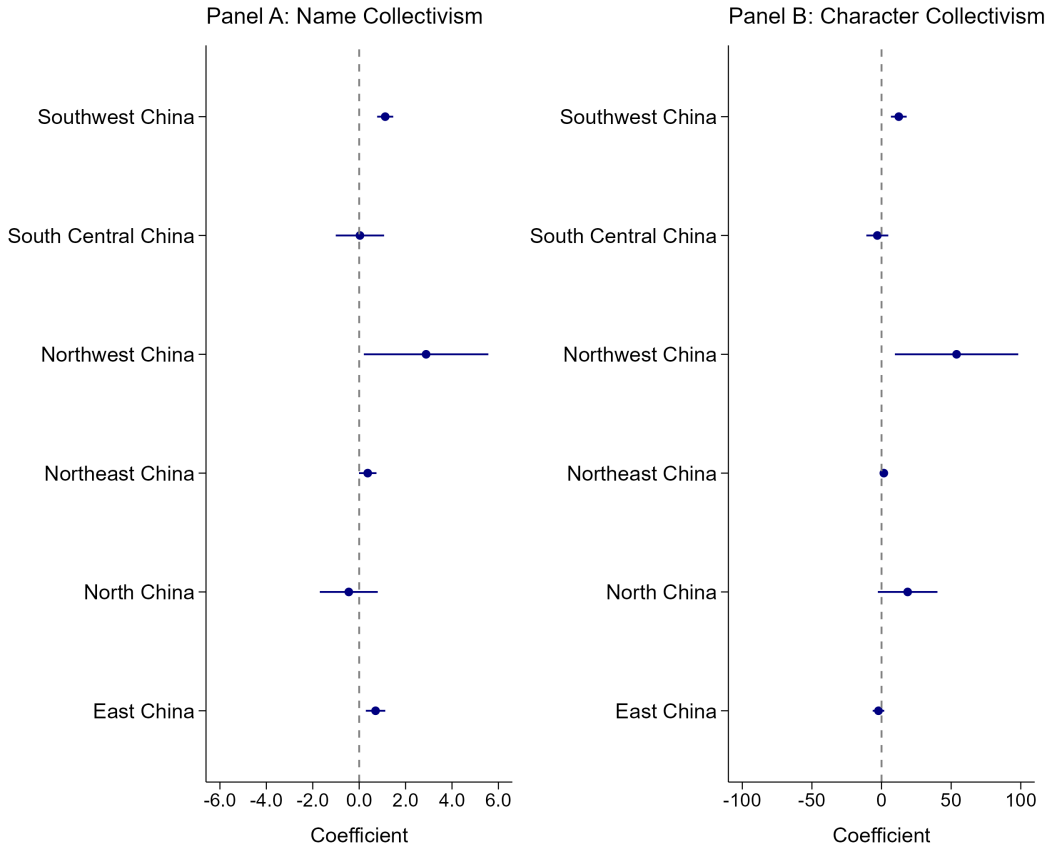


Notes: This figure reports the estimates of the effect of political exposure (measured by weighted geographical distance) on collectivism by gender. Panel A measures collectivism by given-name frequency, while Panel B by given-name character frequency. The coefficients and their 95% confidence intervals are estimated using IV regression from Equation (9).

Region. Regional differences in cultural norms are prominent throughout China (e.g., ZHU et al. (2022)). We therefore anticipate that this regional heterogeneity may influence the treatment effect of political exposure, with regions characterized by less collectivistic and more flexible cultural norms demonstrating a greater responsiveness to political exposure (e.g., Acemoglu and Robinson (2025)).

Figure 9 presents coefficient plots from the IV regression specified in Equation (9). Across two measures of collectivism, our results suggest the treatment effect is heterogeneous across regions, with the effect is most pronounced in the southwest and northwest, in contrast to other regions where the effect is either marginal or statistically insignificant. As shown in Appendix Figure 3, regions in the southwest and northwest are characterized by lower average collectivism values, which may explain their greater responsiveness to the treatment effect of exposure to the centralized national regime.

Figure 9: Impact of Political Exposure on Collectivism by Region

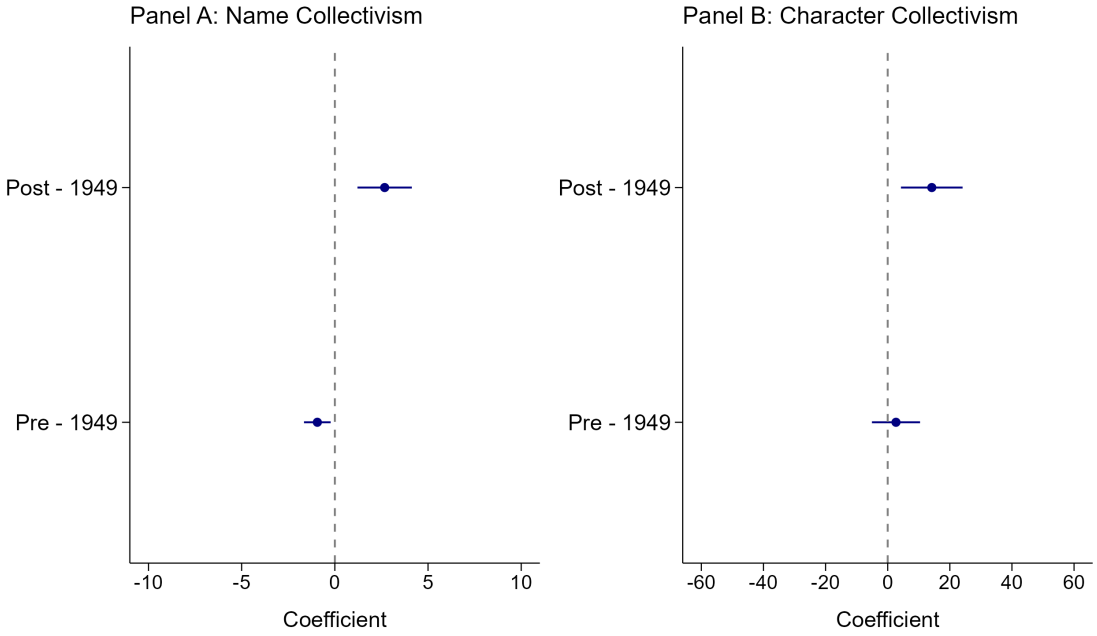


Notes: This figure reports estimates of the effect of political exposure (measured by weighted geographical distance) on collectivism by region. Panel A measures collectivism by given-name frequency, while Panel B by given-name character frequency. The coefficients and their 95% confidence intervals are estimated using IV regression from Equation (9). The regional division follows the classification of the National Bureau of Statistics (NBS) of China.

Pre-1949 vs. Post-1949. Another interesting dimension is the difference in the treatment effect between individuals born before and after the establishment of the People's Republic of China, as these two groups were socialized under distinct political systems and cultural norms.[18]

Figure 10 presents coefficient plots from the IV regression specified in Equation (9). The results suggest that individuals born after 1949 are more responsive to the effect of political exposure than those born before 1949. A potential explanation is that the treatment effect may be more pronounced within a stable, centralized system that emphasizes collectivism and socialist values, while the effect may be diminished or reversed under a fragmented governance model that promotes Western liberal ideas.

Figure 10: Impact of Political Exposure on Collectivism, Pre-1949 vs. Post-1949



Notes: This figure illustrates the estimated effects of political exposure (measured by weighted geographical distance) on collectivism for individuals born before and after the establishment of the People's Republic of China. Panel A measures collectivism by given-name frequency, while Panel B by given-name character frequency. The coefficients and their 95% confidence intervals are estimated using IV regression from Equation (9).

[18]Prior to 1949, China was governed by the Republic of China (1912–1949), a period characterized by fragmented governance. This era included the Warlord period, the Nationalist (Kuomintang) government, and a protracted civil war, all of which contributed to significant regional autonomy and political decentralization. Furthermore, foreign concessions and cultural pluralism exposed the country to various influences, including Western liberal, republican, and Confucian revivalist ideas. Following the founding of the People's Republic of China (PRC) in 1949, the country underwent profound institutional, political, and cultural transformations. The Communist Party established a centralized, one-party socialist system, which emphasized collectivism, egalitarianism, and socialist values (e.g., Myers (1991), Sun and Ryder (2016)).

6 Alternative Estimate: Centralized Regime Coverage and Collectivism

In this section, we estimate the effects of political exposure on collectivism by employing an alternative definition of political exposure: the total number of years a county was governed by centralized Chinese dynasties. We first introduce the empirical specification and discuss identification. We next present the results.

6.1 Empirical Specification

We estimate the following regression model:

$$Y_{c,b} = \gamma_0 + \gamma_1 Coverage_c + \phi_{pro \times b} + \gamma_2^{'} \mathbf{X}_c + \varepsilon_{c,b} \tag{10}$$

where $Coverage_c$ represents political exposure, measured by the duration of centralized regime coverage, as constructed using Formula 4. All other variables are the same as those defined in Equation (6).

Given that the centralized regime coverage measure is pre-determined by the territories of the dynasties, we expect it to be uncorrelated with other relevant outcomes, provided we control for the determinants of these territories. While some of the territory determinants could be correlated with collectivism, these factors exist at the national level. We argue that our within-province comparison mitigates such confounding issues, as the counties being compared should experience similar effects from the omitted territory determinants.

The remaining concern is that individuals' parents may be endogenously choosing their residing counties. For example, parents with a stronger collectivistic orientation may have been more likely to settle in counties that were under the governance of centralized dynasties for a longer period. Our analysis in Appendix B confirms this possibility. To address this identification concern, we include county-level migration rates in Equation 10.

6.2 Results

The results from this alternative estimate are consistent with our main findings in Section 5 regarding the impact of centralized political regimes on collectivism.

Baseline Results. Table 8 presents the estimated impact of exposure to centralized political regimes on collectivism, with the first three columns showing results for the name frequency measure and the last three for the character frequency measure. We sequentially

add control variables for both collectivism measures. Columns 3 and 6 present the results for our most demanding specification, as defined in Equation (10).

Table 8: The Impact of Exposure to Centralized Political Regime on Collectivism

	Dependent Variable: Collectivism Measure					
	Name Frequency			Name Character Frequency		
	(1)	(2)	(3)	(4)	(5)	(6)
Coverage ($\times$ 1000)	0.138***	0.099**	0.120***	1.926***	2.713***	2.746***
	(0.034)	(0.035)	(0.035)	(0.465)	(0.491)	(0.505)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓	✓
Key controls		✓	✓		✓	✓
Migration rate			✓			✓
Observations	218,375	218,375	218,322	218,375	218,375	218,322
Outcome mean	2.725	2.725	2.725	38.114	38.114	38.114
Adj. R^2	0.304	0.306	0.306	0.491	0.492	0.492

Notes: This table reports the estimated effects of political exposure (measured by centralized regime coverage) on collectivism. Key controls include rice suitability, clan strength and night light. Standard errors are reported in brackets and clustered at the county level, and the first-stage F statistics are cluster-robust. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

As the table shows, the estimates for both collectivism measures are stable and statistically significant across all specifications. This indicates that individuals from counties with longer exposure to Chinese dynastic rule are more collectivistic than those from counties with shorter exposure within the same province.

While statistically significant, these estimates are quantitatively smaller compared to our main findings. For example, the coefficient of 2.746 (column 6) indicates that an additional 1,000 years of exposure to a centralized political regime would yield a 0.09-standard deviation increase in collectivism.[19]

Robustness. Similar to our main analysis, we assess the robustness of this alternative estimate by examining collectivism at the provincial and national levels, using alternative inference approaches, and including additional control variables. The results provide consistent evidence supporting the robustness of our findings.

Appendix Table 4 presents the estimated effects on collectivism at different geographic levels. The results are consistent with our baseline estimates at the regional level, showing

[19]For comparison, a one-standard deviation increase in political exposure, as measured by weighted geographical distance, is associated with a 0.29-standard deviation increase in collectivism. In contrast, the same increase in exposure, measured by centralized regime coverage, is associated with a smaller 0.04-standard deviation increase in collectivism.

a statistically significant impact of exposure to a centralized political regime on collectivism across all specifications and collectivism measures.

Appendix Table 5 presents the results from alternative inference approaches for the estimated effects of political exposure, including spatial heteroskedasticity-autocorrelation (HAC) standard errors (Conley, 1999) and two-way clustering by county and birth year. The results suggest that our estimates are generally robust to these two more conservative inference approaches. However, the estimated effect on name frequency becomes insignificant when the spatial correlation threshold is set above 130 km.

Appendix Table 6 presents the estimated effects on collectivism incorporating additional control variables. Across the two measures of collectivism, the results indicate that controlling for population, county size, and distance to the provincial capital does not significantly alter our estimates. This alleviates concerns that these potential confounders, rather than political exposure, drive the observed effects on collectivism.

Heterogeneity. We replicate our heterogeneity analysis for this alternative estimate across dimensions including gender, region, and pre- or post-1949 periods.

Appendix Figures 5, 6, and 7 present coefficient plots showing the estimated effects of political exposure (measured by centralized regime coverage) on two collectivism measures, separated by gender, region, and the period before or after the establishment of the People's Republic of China. The results suggest the effect is similar for males and females but differs across regions and for individuals born before and after 1949, which aligns with the findings from our main analysis.

7 Conclusion

Collectivism is a crucial cultural dimension that influences multiple aspects of social life. This paper examines China as an ideal setting to investigate the origins of collectivism. Contributing to the existing literature, we introduce a novel angle: political exposure as a determinant of collectivism.

We leverage within-province, within-birth-year variation across counties to identify the causal effects. Beginning with province-by-birth-year fixed effects and further controlling for determinants of collectivism identified in the literature (rice theory, lineage theory, and local economic activities), we obtain a compelling comparison in political exposure. To address remaining endogeneity concerns, we employ an instrumental variable strategy that exploits plausibly exogenous variation based on the centroids of the initial territories of ten centralized Chinese dynasties, and we account for migration-related concerns by

explicitly controlling for migration.

Our 2SLS estimates suggest that exposure to historical centralized national capitals increases individuals' collectivist orientations. The heterogeneity analysis indicates that these effects are more pronounced in the northwest and southwest regions and in the period following the establishment of the PRC in 1949, while no systematic differences are observed by gender.

To validate our findings, we employ two measures of collectivism and two measures of political exposure. To further assess robustness, we conduct multiple exercises, including constructing collectivism measures at different geographic levels, applying alternative inference methods, incorporating additional control variables, and performing placebo tests. Across these exercises, the results consistently support the robustness of our findings.

While our paper provides compelling empirical evidence that political power shapes cultural preferences such as collectivist orientation, it does not identify the specific mechanisms through which political exposure influences collectivism, leaving this as an important avenue for future research. Moreover, the documented relationship suggests a promising line of research—particularly in the context of China—on the cultural channel through which ancient political regimes may affect outcomes such as economic growth (e.g., Alesina and Giuliano, 2015; Bisin and Verdier, 2017; Touré, 2021) or the functioning of modern institutions (e.g., Tabellini, 2008; Guiso et al., 2016) by shaping cultural orientations and, consequently, individual behavior.

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics 138*(1), 1–35.

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The colonial origins of comparative development: An empirical investigation. *American economic review 91*(5), 1369–1401.

Acemoglu, D. and J. A. Robinson (2025). Culture, institutions, and social equilibria: A framework. *Journal of Economic Literature 63*(2), 637–692.

Alesina, A. and N. Fuchs-Schündeln (2007). Good-bye lenin (or not?): The effect of communism on people's preferences. *American Economic Review 97*(4), 1507–1528.

Alesina, A. and P. Giuliano (2015). Culture and institutions. *Journal of economic literature 53*(4), 898–944.

Anderson, T. W. and H. Rubin (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of mathematical statistics 20*(1), 46–63.

Bai, G. (2011). *Zhongguo zhengzhi zhidu tongshi: Di yi juan, Zonglun (xiudingban) [General History of Chinese Political Institutions: Vol. 1, General Theory (Revised Edition)]*. Beijing: Social Sciences Academic Press.

Bazzi, S., M. Fiszbein, and M. Gebresilasse (2020). Frontier culture: The roots and persistence of "rugged individualism" in the united states. *Econometrica 88*(6), 2329–2368.

Bisin, A. and T. Verdier (2017). On the joint evolution of culture and institutions. Technical report, National Bureau of Economic Research.

Bisin, A. and T. Verdier (2024). On the joint evolution of culture and political institutions: Elites and civil society. *Journal of Political Economy 132*(5), 1485–1564.

Cai, H., X. Zou, Y. Feng, Y. Liu, and Y. Jing (2018). Increasing need for uniqueness in contemporary china: Empirical evidence. *Frontiers in psychology 9*, 554.

Campante, F. and Q.-A. Do (2010). A centered index of spatial concentration: Expected influence approach.

Campante, F. R. and Q.-A. Do (2014). Isolated capital cities, accountability, and corruption: Evidence from us states. *American Economic Review 104*(8), 2456–2481.

Campante, F. R., Q.-A. Do, and B. Guimaraes (2019). Capital cities, conflict, and misgovernance. *American Economic Journal: Applied Economics 11*(3), 298–337.

Chen, S., R. Fisman, X. Lan, Y. Wang, and Q. Ye (2024). The costs and benefits of clan culture: Elite control versus cooperation in china. Technical report, National Bureau of Economic Research.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics 92*(1), 1–45.

Diamond, J. (2011). *Collapse: how societies choose to fail or succeed: revised edition*. Penguin.

Gao, G. (2011). Shall i name her “wisdom” or “elegance”? naming in china. *Names 59*(3), 164–174.

Gong, W., M. Zhu, B. Gürel, and T. Xie (2021). The lineage theory of the regional variation of individualism/collectivism in china. *Frontiers in Psychology 11*, 596762.

Gorodnichenko, Y. and G. Roland (2011). Which dimensions of culture matter for long-run growth? *American Economic Review Papers and Proceedings 101*(3), 492–498.

Gorodnichenko, Y. and G. Roland (2017). Culture, institutions, and the wealth of nations. *Review of Economics and Statistics 99*(3), 402–416.

Greif, A. (1994). Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. *Journal of political economy 102*(5), 912–950.

Guiso, L., P. Sapienza, and L. Zingales (2016). Long-term persistence. *Journal of the European Economic Association 14*(6), 1401–1436.

Hall, E. T. (1976). *Beyond culture*. Anchor.

Hofstede, G. (1984). *Culture's consequences: International differences in work-related values*, Volume 5. sage.

Hofstede, G. (2001). Culture's consequences: Comparing values, behaviors, institutions and organizations across nations. *International Educational and Professional*.

Hofstede, G., G. Hofstede, and M. Minkov (2010). Cultures and organizations: Software of the mind, 3rd mcgraw hill. *New York*.

Hou, Y. (1986, June). The fundamental principles of the location choice of ancient china's capital cities. In *Study on Ancient Chinese Capital Cities*, Volume 4, Xi'an, Shaanxi. China Conference.

Hsu, F. L. K. (1963). *Clan, Caste, and Club*. Princeton, NJ: D. Van Nostrand Co.

Institute of History, C. A. o. S. S. (Ed.) (1982). *Historical Atlas of China*. Beijing: China Cartographic Publishing House. Published in multiple volumes from 1982 to 1988.

Jiao, J. and J. Zhao (2023). Individualism, collectivism, and allocation behavior: Evidence from the ultimatum game and dictator game. *Behavioral Sciences 13*(2), 169.

Kyriacou, A. P. (2016). Individualism–collectivism, governance and economic development. *European Journal of Political Economy 42*, 91–104.

Li, W. (2023). On the formation, continuation, and contemporary enlightenment of the centralized system in ancient china. *Frontiers of History in China 18*(4).

Myers, R. H. (1991). *Two societies in opposition: the Republic of China and the People's Republic of China after forty years*. Hoover Press.

Pankaj, G. (2001). Distance still matters: the hard reality of global expansion. *Harvard business review 79*(8), 137–147.

Samuel, H. (1993). The clash of civilizations. *Foreign affairs 72*(3), 22–49.

Sima, Q. (1995). *Records of the Grand Historian: Qin Dynasty* (3rd ed.). New York: Columbia University Press.

State Language Commission and State Bureau of Standards (1986). *Modern Chinese Character Frequency Dictionary*. Beijing: Beijing Language Institute Press. Based on an 11.08 million character corpus (1977–1982).

Sun, J. and A. G. Ryder (2016). The chinese experience of rapid modernization: Sociocultural changes, psychological consequences? *Frontiers in psychology 7*, 109319.

Tabellini, G. (2008). Institutions and culture. *Journal of the European Economic association 6*(2-3), 255–294.

Tabellini, G. (2010). Culture and institutions: economic development in the regions of europe. *Journal of the European Economic association 8*(4), 677–716.

Talhelm, T., X. Zhang, S. Oishi, C. Shimin, D. Duan, X. Lan, and S. Kitayama (2014). Large-scale psychological differences within china explained by rice versus wheat agriculture. *Science 344*(6184), 603–608.

Tan, Q. (1982–1987). *Historical Atlas of China*. Beijing: China Cartographic Publishing House.

Tang, C. and Z. Zhao (2023). Informal institution meets child development: Clan culture and child labor in china. *Journal of Comparative Economics 51*(1), 277–294.

Touré, N. (2021). Culture, institutions and the industrialization process. *Journal of Economic Behavior & Organization 186*, 481–503.

Twitchett, D. C., J. K. Fairbank, A. Feuerwerker, et al. (1978). *The Cambridge History of China*, Volume 1991. Cambridge University Press Cambridge.

Twitchett, D. C., J. K. Fairbank, A. Feuerwerker, et al. (1986). *The Cambridge History of China*, Volume 1. Cambridge University Press Cambridge.

Van de Vliert, E., H. Yang, Y. Wang, and X.-p. Ren (2013). Climato-economic imprints on chinese collectivism. *Journal of Cross-Cultural Psychology 44*(4), 589–605.

Yang, Z. (1994). *A Study of Post Stations in the Ming Dynasty* . Shanghai: Shanghai Ancient Books Publishing House.

Zeng, R. and P. M. Greenfield (2015). Cultural evolution over the last 40 years in china: Using the google ngram viewer to study implications of social and political change for cultural values. *International Journal of Psychology 50*(1), 47–55.

Zhao, H. (2021). What matters in the choice of a national capital location in chinese history: an empirical analysis based on quantitative economic history. *Spatial Economic Analysis 16*(3), 355–377.

Zhigang, W. and M. Micklin (1996). The transformation of naming practices in chinese families: Some linguistic clues to social change. *International Sociology 11*(2), 187–212.

ZHU, X., X. ZHENG, H. XU, and Y. XU (2022). Within-country regional cultural differences and their organizational implications. *Advances in Psychological Science 30*(7), 1651.

Appendix

A Additional Figures and Tables

This Appendix presents additional figures and tables that complement the main results. A brief description of each is provided below.

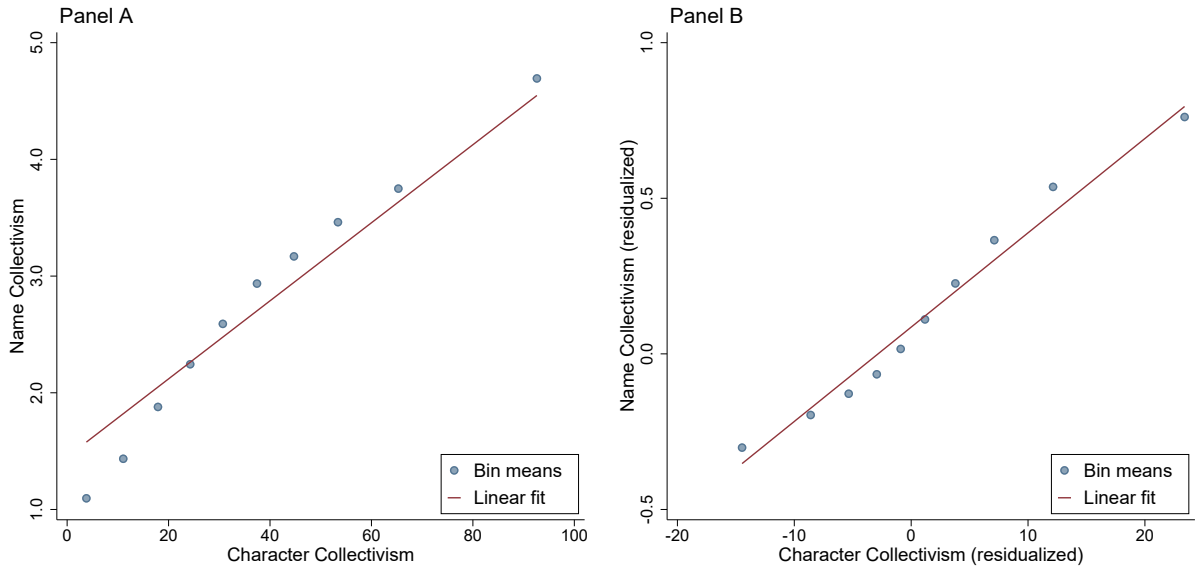
A.1 Figures

Appendix Figure 1 presents binscatter plots illustrating the correlation between our two collectivism measures: the frequency of given names and the frequency of characters in given names. Panel A shows the correlation using raw data, while Panel B uses residualized data after controlling for province-by-birth-year fixed effects.

Appendix Figures 2, 3, and 4 present the average levels of collectivism, measured by name frequency, across gender, region, and birth cohort, highlighting differences in collectivism levels along each dimension.

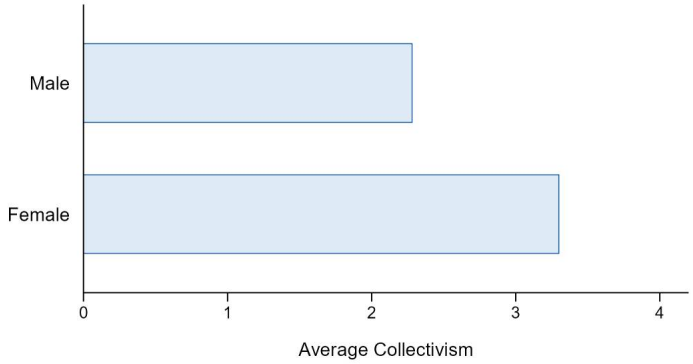
Appendix Figures 5, 6, and 7 present the results of heterogeneity analyses on the estimated effects of political exposure, measured by centralized regime coverage, on two collectivism measures: name frequency in Panel A and character frequency in Panel B. Specifically, Appendix Figure 5 shows gender differences in the treatment effect, Appendix Figure 6 shows regional differences, and Appendix Figure 7 shows differences in the treatment effect before and after the establishment of the People's Republic of China.

Appendix Figure 1: Correlation Between Residualized Collectivism Measures



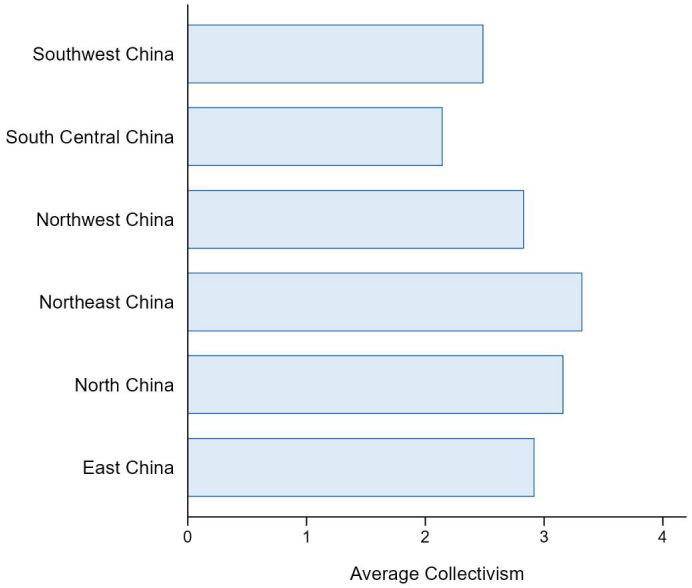
Notes: This figure shows binscatter plots (10 bins) of the relationship between character-based collectivism and name-based collectivism. Panel A presents the raw values of the two measures, while Panel B shows the residualized values after controlling for province-by-birth-year fixed effects. The fitted lines show a statistically significant positive relationship at the 1% level ($p < 0.01$). A few outliers in Panel B are excluded from the binned scatterplot to enhance readability.

Appendix Figure 2: Collectivism (Mean Values) by Gender



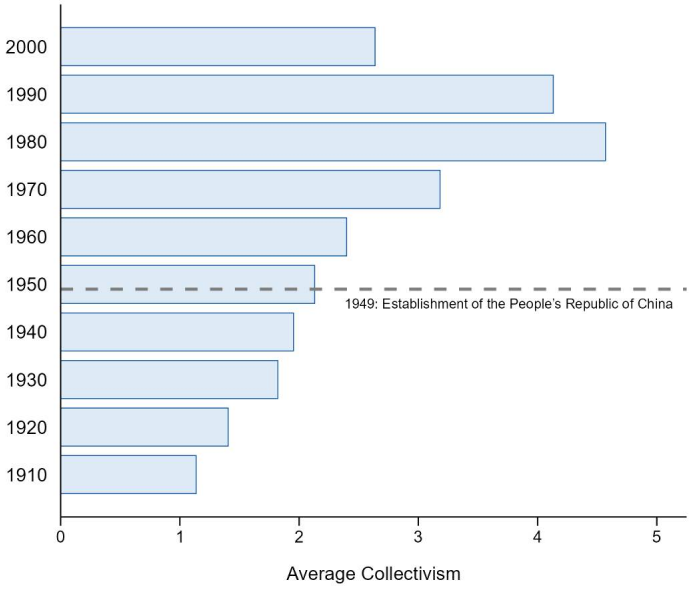
Note: This figure shows the average levels of collectivism, measured by name frequency, by gender.

Appendix Figure 3: Collectivism (Mean Values) by Region



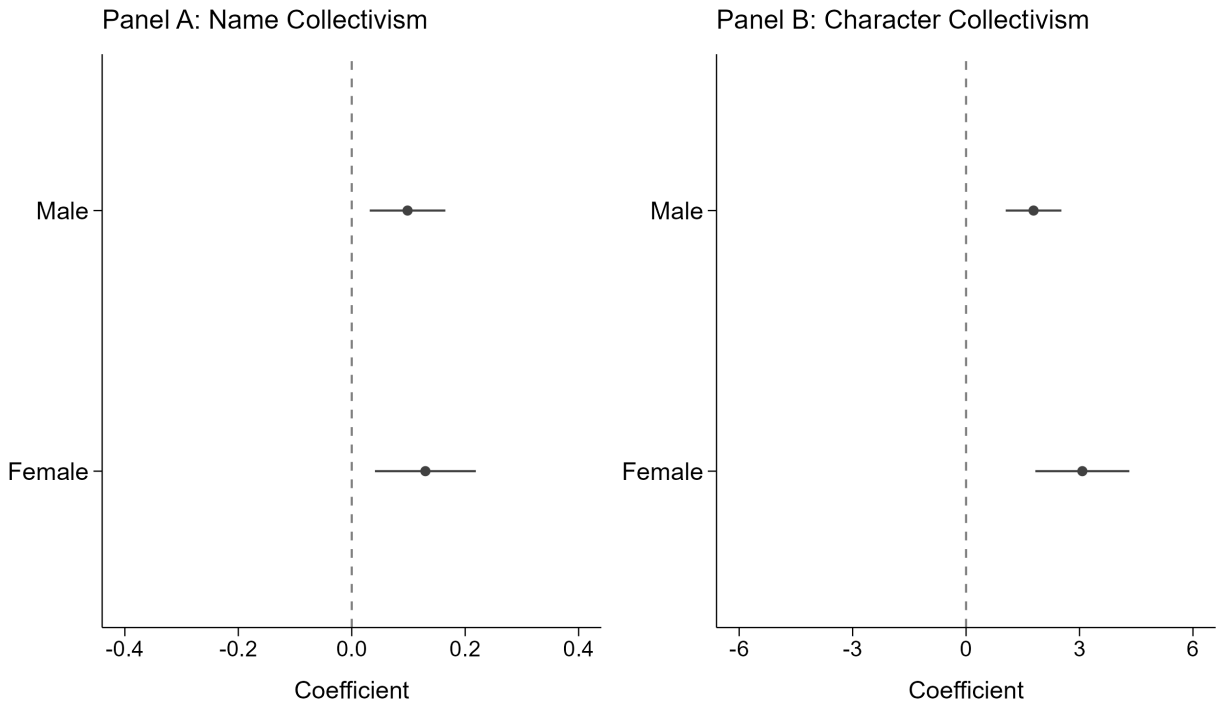
Notes: This figure shows the average levels of collectivism, measured by name frequency, by region. The regional division follows the classification of the National Bureau of Statistics (NBS) of China.

Appendix Figure 4: Collectivism (Mean Values) by Birth Cohort



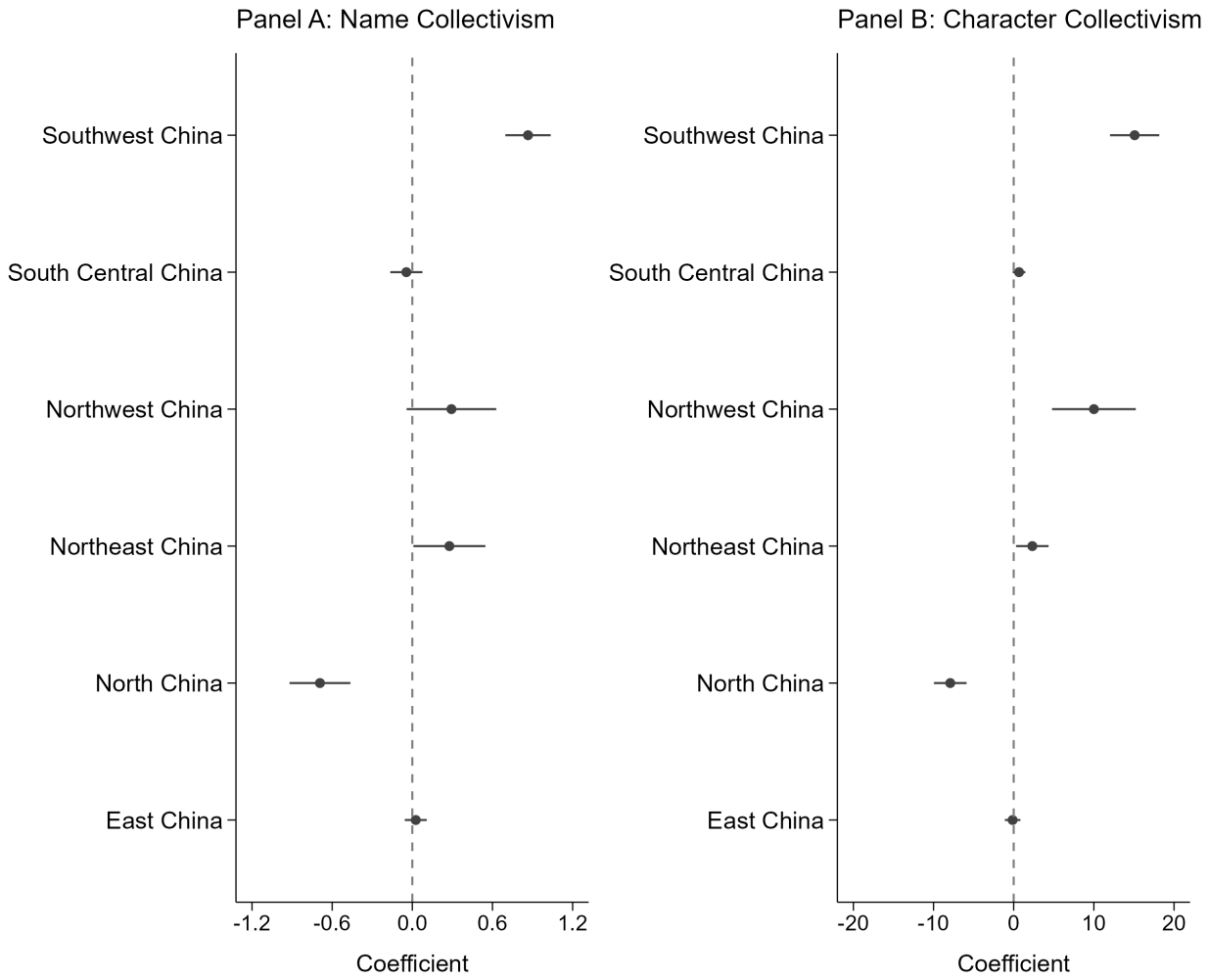
Notes: This figure shows the average levels of collectivism, measured by name frequency, by birth decade. Decades are used to group individuals by their birth year. For example, the 1910 decade includes individuals born between 1912 and 1919, while the 1920 decade includes those born from 1920 to 1929, and so on. Individuals born before 1912 are excluded from the analysis because the last dynasty ended in 1911.

Appendix Figure 5: Impact of Exposure to Centralized Political Regime on Collectivism by Gender



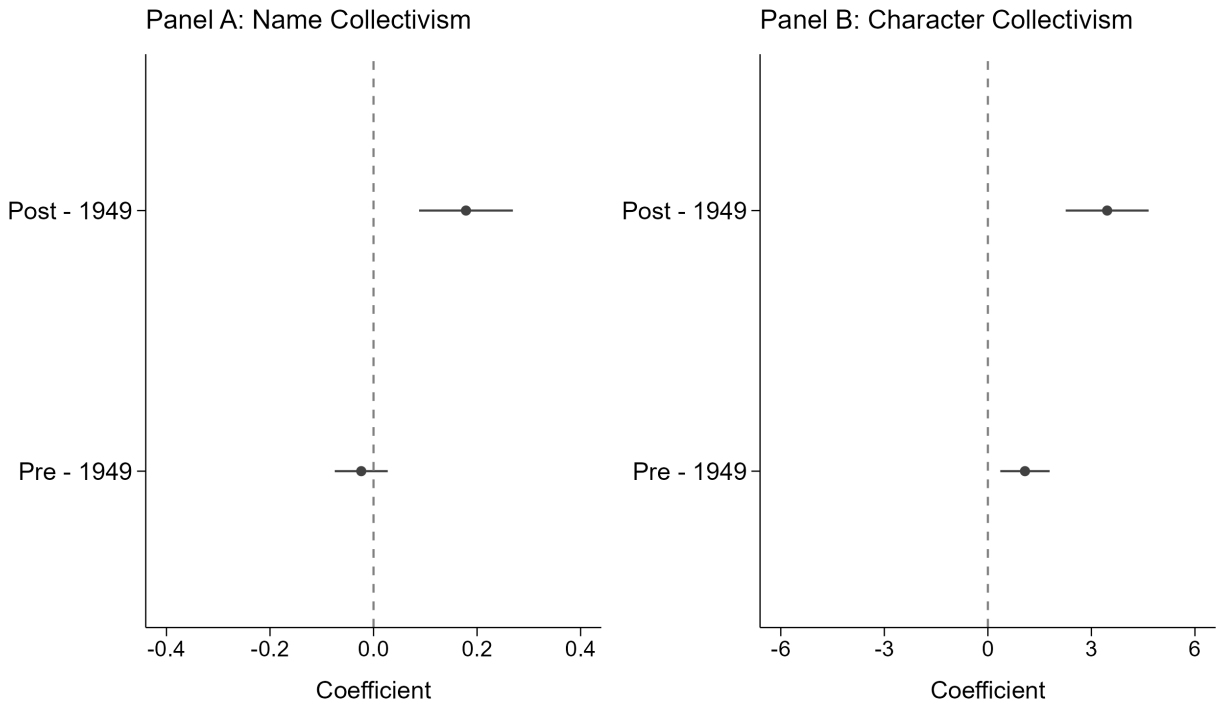
Notes: This figure reports the estimated effects of political exposure (measured by centralized regime coverage) on collectivism by gender. Collectivism in Panel A is measured by the frequency of given names, while collectivism in Panel B is measured by the frequency of given name Characters The coefficients and their 95% confidence intervals are estimated using regression from Equation (10). The coefficients and corresponding confidence intervals are multiplied by 1,000 to improve readability.

Appendix Figure 6: Impact of Exposure to Centralized Political Regime on Collectivism by Region



Notes: This figure reports the estimated effects of political exposure (measured by centralized regime coverage) on collectivism by region. Collectivism in Panel A is measured by the frequency of given names, while collectivism in Panel B is measured by the frequency of given name Characters The coefficients and their 95% confidence intervals are estimated using regression from Equation (10). The coefficients and corresponding confidence intervals are multiplied by 1,000 to improve readability. The regional division follows the classification of the National Bureau of Statistics (NBS) of China.

Appendix Figure 7: Impact of Exposure to Centralized Political Regime on Collectivism Pre-1949 vs. Post-1949



Notes: This figure reports the estimated effects of political exposure (measured by centralized regime coverage) on collectivism for individuals born before and after the establishment of the People's Republic of China. Collectivism in Panel A is measured by the frequency of given names, while collectivism in Panel B is measured by the frequency of given name Characters The coefficients and their 95% confidence intervals are estimated using regression from Equation (10). The coefficients and corresponding confidence intervals are multiplied by 1,000 to improve readability.

A.2 Tables

Appendix Table 1 presents the historical dynasties of China from 221 BCE to 1911 CE and their corresponding capital cities. For each dynasty, the table reports the start and end years (and thus the duration), the names of the capital cities, and an indicator variable *Centralized* that identifies dynasties with a centralized political system. We highlight these centralized dynasties in gray.

Appendix Table 2 presents the summary statistics for the outcome, treatment, and control variables, reporting the mean, standard deviation, median, and number of observations. The unit of observation is the county-by-birth-year cell.

Appendix Table 3 presents the OLS estimates of the correlation between collectivism (measured by name frequency and name character frequency) and exposure to historical centralized political regimes. Collectivism is measured by name frequency in Panel (a) and by name character frequency in Panel (b), with each measure constructed at multiple geographic levels, including the province, region, and nation.

Appendix Table 4 presents the estimated effects of political exposure, measured by centralized regime coverage, on two collectivism measures at the province and nation levels: name frequency in Panel (a) and character frequency in Panel (b).

Appendix Table 5 presents the results of alternative inference approaches for the estimated effects of political exposure, measured by centralized regime coverage, on collectivism. Collectivism is measured by name frequency in Panel (a) and by name character frequency in Panel (b).

Appendix Table 6 presents estimates incorporating additional control variables, where political exposure is measured by centralized regime coverage. Collectivism is measured by name frequency in Panel (a) and by name character frequency in Panel (b).

Appendix Table 1: Historical Dynasties and Capital Cities of China (221 BCE – 1911 CE)

Dynasty	Start Year	End Year	Duration (Years)	Capital City	Centralized	Dynasty	Start Year	End Year	Duration (Years)	Capital City	Centralized
Qin	-221	-210	11	Xianyang	Yes	Chen	557	589	32	Jiankang	No
Western Han	-202	9	211	Chang'an	Yes	Sui	581	618	37	Chang'an	Yes
Eastern Han	25	220	195	Luoyang	Yes	Tang	618	907	289	Chang'an	Yes
Wei	220	265	45	Luoyang	No	Wu	902	937	35	Yangzhou	No
Wu	220	280	60	Jianye	No	Later Liang	907	923	16	Kaifeng	No
Shu	221	263	42	Chengdu	No	Former Shu	907	925	18	Chengdu	No
Western Jin	265	317	52	Luoyang	No	Wu-Yueh	907	978	71	Hangzhou	No
Former Zhao	304	329	25	Pingyang	No	Min	909	945	36	Fuzhou	No
Cheng Han	304	347	43	Chengdu	No	Liao	916	1125	209	Shangjing	No
Former Liang	314	376	62	Guzang	No	Southern Han	917	971	54	Guangzhou	No
Eastern Jin	317	420	103	Jiankang	No	Later Tang	923	936	13	Luoyang	No
Later Zhao	319	351	32	Yecheng	No	Jingnan	924	963	39	Jiangling	No
Former Yan	349	370	21	Yedu	No	Chu	927	963	36	Changsha	No
Former Qin	351	384	33	Chang'an	No	Later Shu	934	965	31	Chengdu	No
Later Yan	384	409	25	Zhongshan	No	Later Jin	936	946	10	Kaifeng	No
Later Qin	384	417	33	Chang'an	No	Wu	937	937	0	Nanking	No
Western Qin	385	431	46	Jincheng	No	Southern Tang	937	975	38	Nanking	No
Later Liang	386	403	17	Guzang	No	Later Han	947	950	3	Kaifeng	No
Northern Wei	386	494	108	Pingcheng	No	Later Zhou	951	960	9	Kaifeng	No
Southern Liang	397	414	17	Ledu	No	Northern Han	951	979	28	Taiyuan	No
Southern Yan	400	410	10	Guanggu	No	Northern Song	960	1127	167	Kaifeng	Yes
Western Liang	400	422	22	Jiuquan	No	Western Xia	1038	1227	189	Xingqing	No
Northern Liang	401	439	38	Zhangyi	No	Jin	1115	1234	119	Huining	No
Xia	407	431	24	Tongwan	No	Southern Song	1127	1279	152	Lin'an	Yes
Northern Yan	409	436	27	Changli	No	Yuan	1271	1368	97	Da-du	Yes
Liu-Song	420	479	59	Jiankang	No	Ming	1368	1421	53	Nanjing	Yes
Southern Qi	479	502	23	Jiankang	No	Ming	1421	1644	223	Beijing	Yes
Northern Wei	494	534	40	Luoyang	No	Southern Ming	1644	1645	1	Nanjing	No
Liang	502	557	55	Jiankang	No	Southern Ming	1645	1646	1	Fuzhou	No
Eastern Wei	534	550	16	Yecheng	No	Southern Ming	1646	1647	1	Zhaoqing	No
Western Wei	535	557	22	Chang'an	No	Southern Ming	1647	1661	14	Guilin	No
Northern Qi	550	577	27	Yecheng	No	Qing Dynasty	1644	1911	267	Beijing	Yes
Northern Zhou	557	581	24	Chang'an	No						

Notes: Centralized dynasties are highlighted in gray.
Source: Historical Atlas of China (Institute of History, 1982), The Cambridge History of China (Twitchett et al., 1986).

Appendix Table 2: Summary Statistics

Variable	Mean	Std. Dev.	Median	Observations
Outcome Variables:				
Collectivism (name)	2.725	2.595	1.857	219,202
Collectivism (character)	38.114	29.448	34.000	219,202
Treatment Variables:				
log(exposure)	–3.143	0.739	–3.282	219,202
Coverage	1,292	422	1,497	218,375
Control Variables:				
Rice suitability	1,231	1,484	543	219,202
Clan	–0.212	1.448	–0.017	219,202
Night light	568	1,482	142	219,202
Migration rate	0.058	0.147	0.000	219,149

Notes: This table provides summary statistics for county-by-birth-year cells. Variable definitions: *Collectivism (name)* measures collectivism by the frequency of given names in a region; *Collectivism (character)* measures collectivism by the frequency of name characters in a region; *log(exposure)* captures counties' exposure to historical centralized national capitals; *Rice suitability* captures potential agricultural output (measured in calories); *Clan* is the first principal component of four clan proxies and reflects the strength of clans; and *Night light* is the average night-time light intensity, proxy for local economic activity. To address potential non-random missingness, we use prefecture or province average to impute missing values for these three control variables. Additionally, *Migration rate* is the county-level migration rates of moving to another county.

Appendix Table 3: Results for the OLS Estimates by Collectivism
(Constructed at Different Geographic Levels)

	(a) Name Frequency Collectivism Measure					
	Within Province		Within Region		Within Nation	
	(1)	(2)	(1)	(2)	(1)	(2)
log(exposure)	0.068***	0.060***	0.197***	0.165***	1.075***	0.854***
	(0.011)	(0.011)	(0.033)	(0.035)	(0.177)	(0.183)
Rice suitability ($\times$ 1000)		0.029***		0.096***		0.561***
		(0.004)		(0.010)		(0.057)
Clan		–0.024***		–0.073***		–0.316***
		(0.005)		(0.015)		(0.084)
Night light ($\times$ 1000)		0.006***		0.023***		0.157***
		(0.002)		(0.007)		(0.039)
Migration rate		0.037*		0.333***		2.884***
		(0.017)		(0.059)		(0.331)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓	✓
Adj. R^2	0.311	0.313	0.304	0.306	0.320	0.323
Observations	219,202	219,149	219,202	219,149	219,202	219,149
	(b) Name Character Frequency Collectivism Measure					
	Within Province		Within Region		Within Nation	
	(1)	(2)	(1)	(2)	(1)	(2)
log(exposure)	0.384**	0.496***	0.961**	1.310***	4.985**	6.790***
	(0.125)	(0.119)	(0.394)	(0.382)	(2.174)	(2.148)
Rice suitability ($\times$ 1000)		0.176***		0.659***		4.074***
		(0.034)		(0.110)		(0.646)
Clan		–0.302***		–0.737***		–4.786***
		(0.057)		(0.174)		(0.980)
Night light ($\times$ 1000)		–0.111***		–0.339***		–1.919***
		(0.034)		(0.093)		(0.469)
Migration rate		–0.098		–0.084		8.834***
		(0.600)		(0.775)		(3.014)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓	✓
Adj. R^2	0.207	0.208	0.491	0.492	0.566	0.567
Observations	219,202	219,149	219,202	219,149	219,202	219,149

Notes: This table reports the estimated associations between political exposure (measured by weighted geographical distance) and collectivism. Standard errors are reported in brackets and clustered at the county level. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table 4: The Impact of Exposure to Centralized Political Regime on Collectivism (Constructed at Different Geographic Levels)

	(a) Collectivism Measure: Name Frequency					
	Within Province			Within Nation		
	(1)	(2)	(3)	(4)	(5)	(6)
Coverage ($\times$ 1000)	0.060***	0.053***	0.056***	0.781***	0.456**	0.622***
	(0.012)	(0.012)	(0.012)	(0.170)	(0.175)	(0.177)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓	✓
Key controls		✓	✓		✓	✓
Migration rate			✓			✓
Observations	218,375	218,375	218,322	218,375	218,375	218,322
Outcome mean	1.535	1.535	1.535	10.150	10.150	10.150
Adj. R^2	0.312	0.313	0.313	0.320	0.322	0.323
	(b) Collectivism Measure: Name Character Frequency					
	Within Province			Within Nation		
	(1)	(2)	(3)	(4)	(5)	(6)
Coverage ($\times$ 1000)	0.874***	1.178***	1.186***	9.805***	13.929***	14.584***
	(0.164)	(0.179)	(0.197)	(2.583)	(2.744)	(2.790)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓	✓
Key controls		✓	✓		✓	✓
Migration rate			✓			✓
Observations	218,375	218,375	218,322	218,375	218,375	218,322
Outcome mean	11.629	11.629	11.629	219.219	219.219	219.219
Adj. R^2	0.207	0.207	0.207	0.567	0.568	0.568

Notes: This table reports the estimated effects of political exposure (measured by centralized regime coverage) using Equation (10). Key controls include rice suitability, clan strength and night light. Standard errors are reported in brackets and clustered at the county level, and the first-stage F statistics are cluster-robust. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table 5: Alternative Inference for Robustness

	(a) Collectivism Measure: Name Frequency				
	County	Spatial HAC (Conley)			Two-Way
		50 km	90 km	130 km	County + Birth Year
	(1)	(2)	(3)	(4)	(5)
Coverage ($\times$ 1000)	0.120***	0.120**	0.120*	0.120	0.120***
	(0.035)	(0.049)	(0.063)	(0.075)	(0.046)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓
Key controls	✓	✓	✓	✓	✓
Migration rate	✓	✓	✓	✓	✓
Observations	218,322	218,322	218,322	218,322	218,322
	(b) Collectivism Measure: Character Frequency				
	County	Spatial HAC (Conley)			Two-Way
		50 km	90 km	130 km	County + Birth Year
	(1)	(2)	(3)	(4)	(5)
Coverage ($\times$ 1000)	2.746***	2.746***	2.746***	2.746***	2.746***
	(0.505)	(0.699)	(0.893)	(1.043)	(0.527)
Province-by-Birth-Year FE	✓	✓	✓	✓	✓
Key controls	✓	✓	✓	✓	✓
Migration rate	✓	✓	✓	✓	✓
Observations	218,322	218,322	218,322	218,322	218,322

Notes: This table reports the estimated effects of political exposure (measured by centralized regime coverage) using Equation (10). Key controls include rice suitability, clan strength and night light. Standard errors (SE) are reported in brackets. The estimates in column 1, serving as the baseline, are taken from Columns 3 and 6 of Table 8. SEs in Columns 2–4 are estimated using the Conley (1999) GMM-based HAC procedure for spatial dependence. SEs in Column 5 are two-way clustered by county and birth year. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table 6: Robustness Checks with Additional Controls

	(a) Collectivism Measure: Name Frequency				
	(1)	(2)	(3)	(4)	(5)
Coverage (× 1000)	0.120***	0.111***	0.117***	0.104***	0.097***
	(0.035)	(0.034)	(0.035)	(0.035)	(0.034)
Log(population)		✓			✓
County size			✓		✓
Log(distance to provincial capital)				✓	✓
Province-by-Birth-Year FE	✓	✓	✓	✓	✓
Key controls	✓	✓	✓	✓	✓
Migration rate	✓	✓	✓	✓	✓
Observations	218,322	218,322	218,322	218,322	218,322
	(b) Collectivism Measure: Character Frequency				
	(1)	(2)	(3)	(4)	(5)
Coverage (× 1000)	2.746***	2.548**	2.718***	2.814***	2.641***
	(0.505)	(0.482)	(0.498)	(0.506)	(0.480)
Log(population)		✓			✓
County size			✓		✓
Log(distance to provincial capital)				✓	✓
Province-by-Birth-Year FE	✓	✓	✓	✓	✓
Key controls	✓	✓	✓	✓	✓
Migration rate	✓	✓	✓	✓	✓
Observations	218,322	218,322	218,322	218,322	218,322

Notes: This table reports the estimated effects of political exposure (measured by centralized regime coverage) using Equation (10). Key controls include rice suitability, clan strength and night light. The estimates in column 1, serving as the baseline, are taken from Columns 3 and 6 of Table 8. Standard errors are reported in brackets and clustered at the county level. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

B Supplementary Analysis

This Appendix presents additional results from the supplementary analysis to complement the main findings.

B.1 Is Migration a Potential Threat to Identification?

Parents' location choices may be endogenous, potentially biasing our estimates. To examine this possibility, we analyze the correlation between migration patterns and our outcome, treatment, and instrument variables.

Calculation of Migration Rate. We infer each individual's county of birth from their hukou registration location. For individuals born in 2004 and 2005 with pending hukou status (as registration is required within one year of birth and the survey was conducted in 2005), we treat their county of residence as their county of birth, assuming no migration within that period. We then compare recorded birth locations at the county, prefecture, and province levels with individuals' residence locations at the time of the survey to identify migration across county, prefecture, or province boundaries.

Percentage of Migration Rate. The results indicate that most individuals remained in their county of birth. On average nationwide, approximately 8.8% of the population had moved out of their county of origin, 6.8% had moved out of their birth prefecture, and 5.0% had moved out of their birth province.

Correlation of Migration Rates with Outcome, Treatment, and IV. To assess whether internal migration biases our results, we calculate the out-of-county migration rate at the county-by-birth-year level and examine the correlation between migration rates and name-based collectivism measures, exposure to centralized political regimes, and our instrumental variable. Panel (a) of Appendix Table 8 shows that migration rates are statistically significantly correlated with collectivism when measured by name frequency, but not when measured by character frequency. Panel (b) of Appendix Table 8 reveals similarly mixed and inconsistent correlations between migration rates and political exposure. Taken together with the consistent and robust association between political exposure and name-based collectivism, these results suggest that migration may not materially bias our findings. Moreover, as shown in the Appendix Table 8, our instrumental variable is not statistically significantly correlated with migration, further alleviating concerns that migration materially biases our estimates. In our most demanding specification, we show that our main findings remain robust when controlling for this rate (Section 5).

Appendix Table 7: Correlation of Migration with Outcome and Treatment

	(a) Outcome Variables: Collectivism			
	Name Frequency		Character Frequency	
	(1)	(2)	(1)	(2)
Migration rate	0.445***	0.328***	–0.404	–0.125
	(0.058)	(0.060)	(0.722)	(0.772)
Province-by-Birth-Year FE	✓	✓	✓	✓
Key controls		✓		✓
Adj. R^2	0.305	0.306	0.491	0.492
Observations	219,149	219,144	219,144	219,144
	(b) Treatment Variables: Political Exposure			
	Weighted Geographical Distance		Centralized Regime Coverage	
	(1)	(2)	(1)	(2)
Migration rate	0.064**	–0.031	–77.844**	–244.823***
	(0.026)	(0.024)	(32.773)	(30.710)
Province-by-Birth-Year FE	✓	✓	✓	✓
Key controls		✓		✓
Adj. R^2	0.861	0.868	0.505	0.565
Observations	219,144	219,144	218,322	218,322

Notes: This table reports the estimates of the relationships between migration rates and other variables, including both the outcome and treatment variables. Key controls include rice suitability, clan strength and night light. Standard errors are reported in brackets and clustered at the county level. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.

Appendix Table 8: Correlation of Migration with Instrument

	Dependent Variables: IV	
	(1)	(2)
Migration rate ($\times$ 1000)	–0.155	–10.869
	(14.119)	(14.662)
Province-by-Birth-Year FE	✓	✓
Key controls		✓
Adj. R^2	0.820	0.822
Observations	219,149	219,149

Notes: This table reports the estimates of the relationship between migration rates and instrument variable. Key controls include rice suitability, clan strength and night light. Standard errors are reported in brackets and clustered at the county level. Significance levels are denoted by * p<0.10, ** p<0.05, *** p<0.01.