Online Appendix

Bai and Jia, The Economic Consequences of Political Hierarchy

Contents

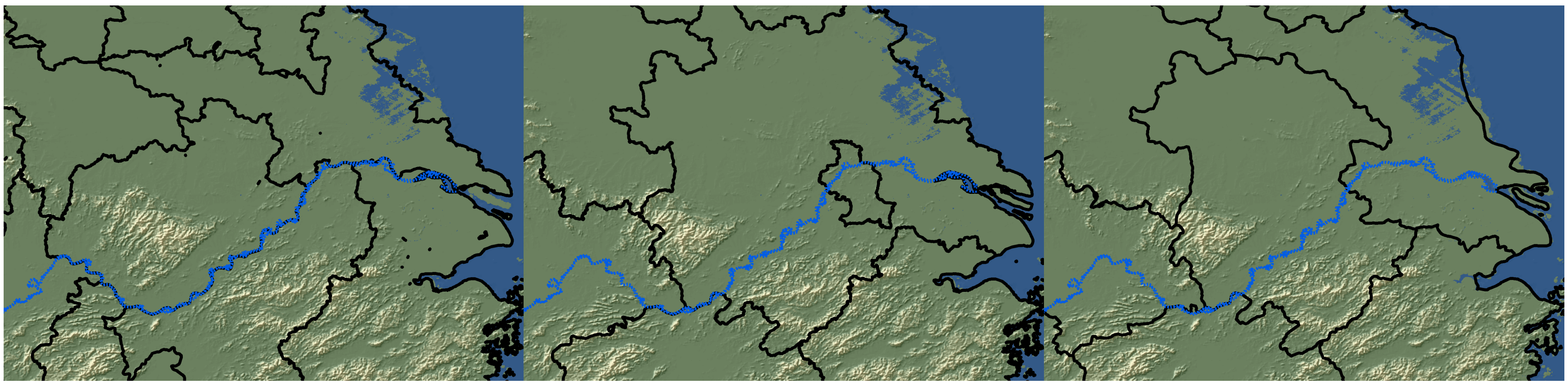
A Additional Information on the Background

A.1 Two Principles to Define Provincial Boundary: An Example

Figure A1 presents an example where changes in provincial boundaries were driven by the evolution from "*following the natural mountains and rivers*" to "*interlocking like dog's teeth*". As shown, the Yangtze River used as a provincial boundary by the (pre-Mongol) Song but included within provinces by the (post-Mongol) Ming and Qing.

Figure A1: Two Principles of Provincial Boundary Revision

(a) The Yangtze River: Song (b) The Yangtze River: Ming (c) The Yangtze River: Qing



A.2 Examples of Provincial Capital Relocation

Figure A2 presents two examples of provincial capital relocation. The cross indicates the provincial centroid, the hollow/solid star indicates the past/current national capital, and the hollow/solid square indicates the past/current provincial capital. Luzhou and Changsha were capitals in the Song. Both lost their capital status in the Ming. Changsha regained the capital status in the Qing but Luzhou didn't. These patterns are driven by the relocation of national capitals and redivision of provinces across regimes, which can be captured by our algorithm.

Figure A2: Examples of Provincial Capital Relocation

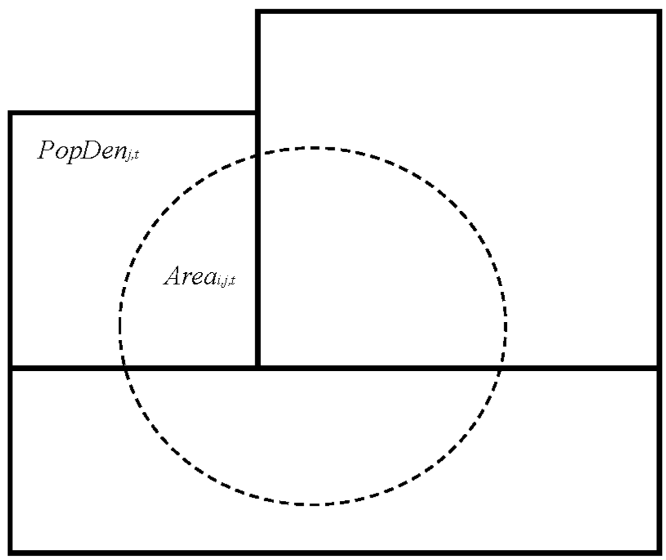
(d) Song (e) Ming (f) Qing

B Data Construction

B.1 Data on Population and Urbanization

We collect the information on prefecture population in 980, 1078, 1102, 1393, 1580, 1776, 1820, 1851, 1910, 1964, and 2000 according to the historical boundaries in each census. To construct a panel dataset, we map the original data to the prefecture boundary in 2000. The idea could be illustrated by the following figure.

Figure B1(a): Illustration of Population Mapping



Here, the circle (dash line) represents prefecture i in 2000, which is overlapped with J historical prefectures in year t (represented by the rectangular). Prefecture i in 2000 is then divided into J polygons by the historical boundary in year t. The area of each polygon is denoted as $Area_{i,j,t}$. We first calculate the population density in historical prefecture j in year t (each rectangular), denoted by $PopDen_{j,t}$ $(j = 1, 2, ..., J)$. Second, we assume the population is evenly distributed in each rectangular and hence calculate the population in each small polygon as $Area_{i,j,t} \times PopDen_{j,t}$. The population density of prefecture i in year t could be thus calculated as:

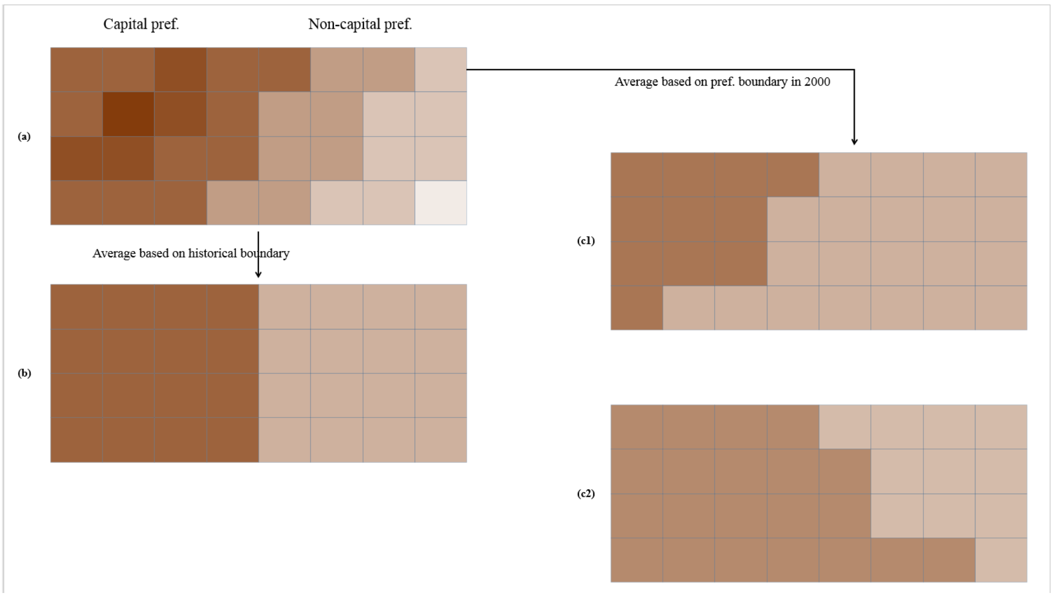
$$PopDen_{i,t} = \frac{\sum_{j=1}^{J} Area_{i,j,t} \times PopDen_{j,t}}{Area_i}$$

To check whether our findings are sensitive with mapping historical original data into the boundary in 2000, we use the same method convert the data into 1-degree X 1-degree grid or 2-degree X 2-degree grid level.

We collect information on urban population in 1580, 1820, 1964 and 2000, and map the original data into the prefecture in 2000 using the same method applied to population data.

If such construction causes measurement error in population density, the measure error may lead to a downward bias of the estimated effect of capital status. We illustrate the reason using the following graph.

Figure B1(b): Possible Measurement Error



Graph (a) includes two historical prefectures with equal size. The left sixteen grids represent a capital prefecture, while the right sixteen grids represent a non-capital prefecture. The darkness represents population density. Due to the lack of information on population density in each grid, we can only calculate the population density of each historical prefecture (sixteen grids), as graph (b) shows. To construct a panel, we follow the method introduced above to map original data (graph (b)) into the prefecture in 2000 (graph (c1) or (c2)). There are two possible cases. The first is that the capital prefecture was much larger in the history as graph (c1) shows. The non-capital prefecture will include some parts of capital prefecture, and the average population density of the non-capital prefecture should be over-estimated (much darker in graph (c1) than (b)). The second is that the capital prefecture was much smaller in the history as graph (c2) shows. The capital prefecture then include some grids of non-capital prefecture, and the average population density of the capital prefecture should be under-estimated (much lighter in graph (c2) than (b)). In both cases, the difference in population densities between the capital-prefecture and the non-capital prefectures (in graph (c1) or (c2)) is much smaller than the true difference (in graph (b)).

Therefore, no matter whether the prefecture becomes smaller or bigger, simple OLS regressions will underestimate the effect of capital status on population density.

B.2 Other Variables

Provincial capital We collect information on the geolocation of provincial capitals in each regime. If a prefecture (or a grid) contains a provincial capital, we term it a capital prefecture (grid).

Hierarchical distance We use the geographical coordinates of prefecture centroid to calculate hierarchical distance, which comprises of two components. The first is the great circle distance between prefecture i to a peer prefecture j in the same province in regime r (denoted as $D_{i,j,r}$). The second is the great circle distance between prefecture i to the national capital in regime r (denoted as $D_{i,NationalCap,r}$). The hierarchical distance is then calculated by $\sum_{j=1}^{N} A_j D_{i,j,r} + \lambda_r \sum_{j=1}^{N} A_j D_{i,NationalCap,r}$, in which A_j represents the area of the peer prefecture j. We employ λ_r in two ways: (1) allowing λ_r to vary and searching for the optimal λ_r which has the highest prediction power on capital status; and (2) fixing it to be a specific value.

Geography To construct the dummy indicating whether the prefecture contains a plain, we first use CHGIS V4 DEM (Digital Elevation Model) to calculate the average slope for all 0.25-degree X 0.25-degree grids. If the average slope is less than 1, the grid will be defined as a plain. We then match the grid to the prefecture boundary in 2000, and construct a dummy indicating whether the prefecture contains at least one plain. By matching CHGIS V4 DEM with prefecture boundary in 2000, we could calculate the average elevation and slope within each prefecture. The other geographic variables, such as dummy indicating whether a prefecture contains major river or is on the coast, longitude, latitude, and area are directly generated from CHGIS V4 (2007) using ArcGIS software.

Crop suitability Based on the suitability index from the Food and Agriculture Organization's 2012 Global Agro-Ecological Zones database, which ranges from 1 ("not suitable") to 8 ("very high") in each 0.5-degree x 0.5-degree grid cell, we measure prefecture-level crop suitability as the average for all cells located in each prefecture with a primary focus on the suitability for wheat, rice, fox millet, maize, and sweet potato.

Macroregions By analyzing the urban and the associated local and regional system hierarchies, Skinner (1977) divides traditional China (today's China excluding Inner Asian territories, or China proper) into nine physiographic macro-regions presented in Figure B2. Specifically, they include Northeast China, North China, Northwest China, Upper Yangzi, Middle Yangzi, Lower Yangzi, Southeast Coast, Lingnan, and Yungui (see the map below). We construct a set of dummy variables indicating in which region the prefecture is located.

Figure B2: Macroregions in Skinner (1977)

Public employment The data in 1776 comes from the Complete Directory of Qing Officials. A limitation of these records is that they only include individuals holding offices. The data in 2000 comes from the population census. We include the population employed in administrations and public institutions as public employment. The former includes those working in government agencies, Party agencies and social organizations; the latter refers to those employed in health care, sports and social welfare; education and culture; and scientific research and polytechnic services.

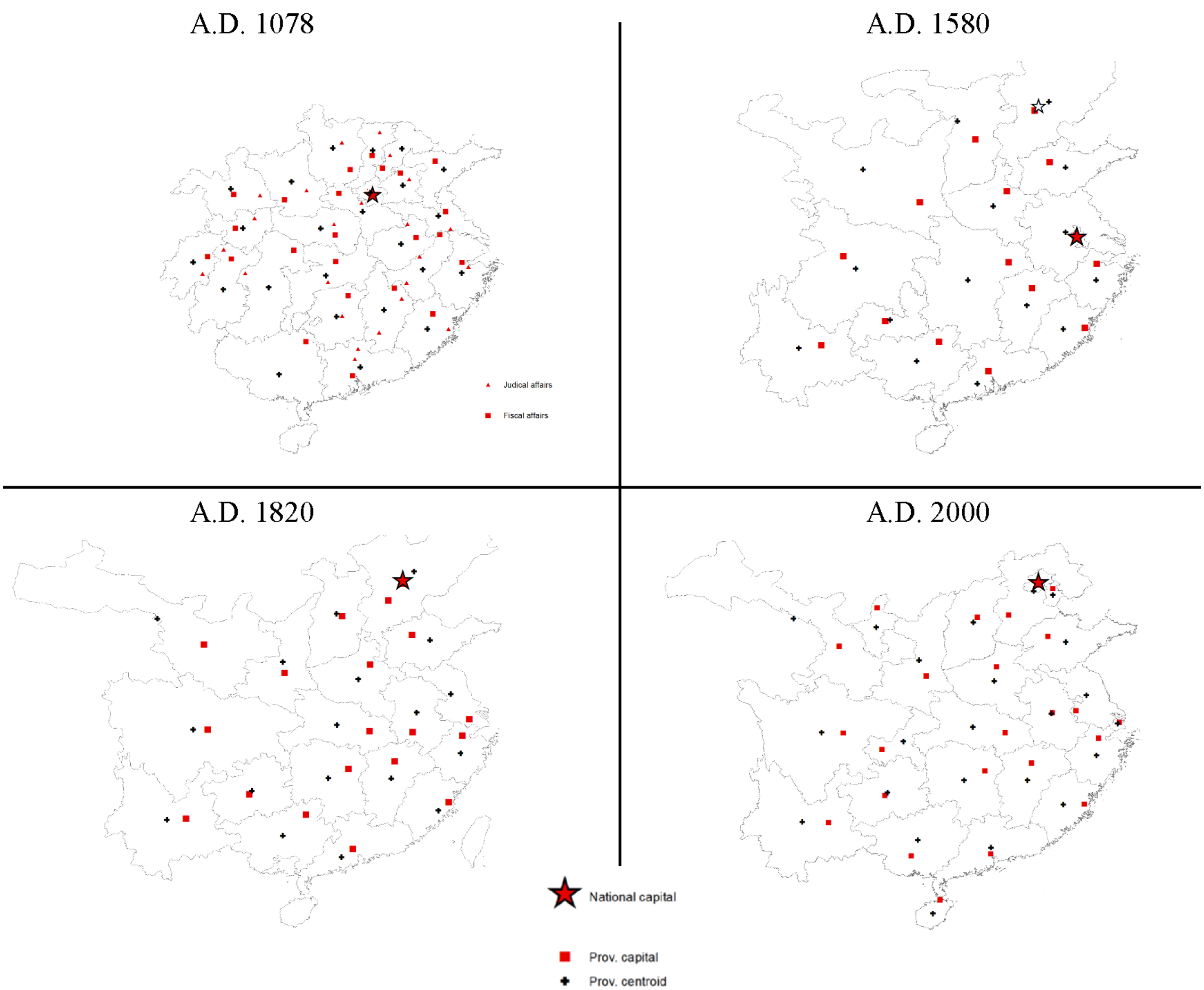
Transportation networks We digitize historical maps on the roads (both water and land way) in 1078, 1580 and 1820, as Figure 7 shows, and calculate the centrality measures as discussed in Section 6. For modern transportation networks, we focus on railways in 1990 because we are interested in those monopolized by the state.

C Additional Results on Hierarchical Distance

C.1 Maps Across Regimes

As long as $\lambda > 0$, the definition of hierarchical distance implies that provincial capitals should be located away from the provincial centroid and toward the national capital. This prediction is confirmed by the pattern in each regime. Figure C1 plots the locations of provincial capitals (indicated by red squares), national capitals (indicated by red stars), and the provincial centroids (indicated by black crosses) regime by regime. See Table 3A for the estimation results by regime.

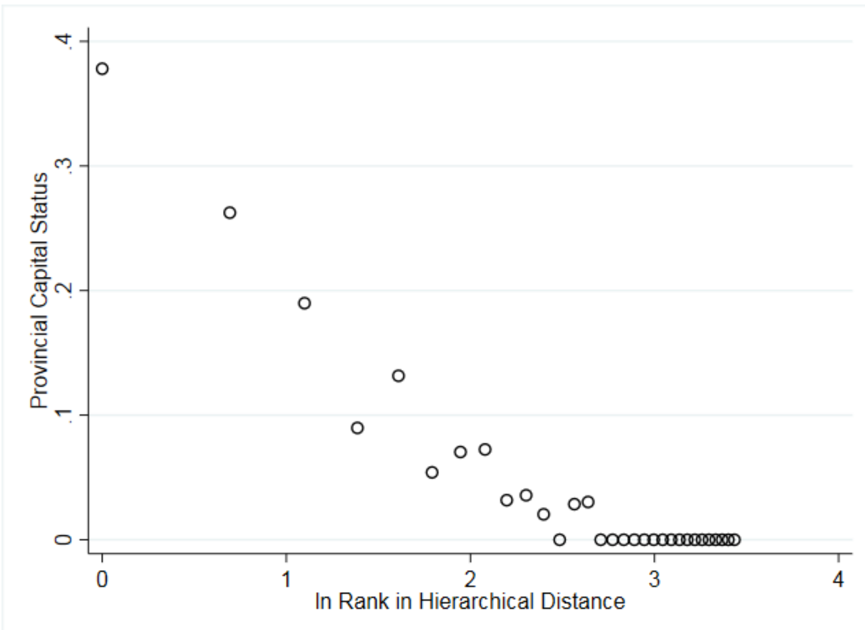
Figure C1: Provincial Capital Locations by Regime



C.2 Fixing the Value of λ

In our main analysis, we use the λ_r that gives the largest prediction power for each regime r. This assumption is not essential for our analysis. The main patterns look similar if we vary λ from 0.1 to 0.9. Figure C2 plots the case when λ=0.2 as an example, which is similar to Figure 3 in our main text.

Figure C2: ln Rank in Hierarchical Distance and the Probability of Being a Provincial Capital



D Additional Results for the DID Analysis

D.1 Accessing the Importance of Omitted Variable Bias

We evaluate our baseline's robustness to omitted variable bias using the method proposed by Oster (2017), which assumes that if the relation between treatment and unobservables can be fully captured from the relation between treatment and observables, then the omitted variable bias will be proportional to the coefficient movements scaled by the change in R-squared when controls are included. Specifically, Oster (2017) proposes the following equation to calculate an approximation of the bias-adjusted treatment effect:

$$\beta^* \approx \tilde{\beta} - \delta \frac{R_{\max} - \tilde{R}}{\tilde{R} - \dot{R}} (\tilde{\beta} - \dot{\beta})$$

in which β^* represents the bias-adjusted effect of capital status on population density, while $\dot{\beta}$ and $\dot{R}$ ($\tilde{\beta}$ and $\tilde{R}$) denote the coefficients of capital status and R-squared with the fewest (most) controls.

In Table 4, we obtain $\dot{\beta} = 0.622$ and $\dot{R} = 0.757$ when we regress logged population density on capital status with only controlling prefecture and year fixed effects. As we include more controls, the coefficient of provincial capital decreases while R-squared increases. Specifically, $\tilde{\beta} = 0.470$ and $\tilde{R} = 0.882$ when we controlled the most controls as column (4) of Table 4 shows. In the equation above, δ measures the relative importance of unobservables to observables, and R_{max} is the R-squared if the regression controls for both these latter. Using the above equation with only the upper bounds of δ and R_{max} enables derivation of a set of lower bounds for the bias-adjusted effect of capital status (β^*). Following Oster (2017), we take 1 as the upper bound of δ and assume that $R_{max} = 1$. Then, based on the two assumptions, our results show that the lower bound of bias-adjusted effect of capital on logged population density is 0.327.[1] The robustness of this approximation can be evaluated using two standards (Oster 2017): First, because the capital effect moves toward zero when more variables are controlled for, we check whether the lower bound of β is smaller than zero. The answer is clearly negative. Second, we consider whether the lower bound of β is within the +/- 2.8 standard error of the capital coefficient with the most controls. The answer is positive, suggesting that the size of the estimate from the regression with the most controls is similar to the bias-adjusted estimate.[2] These results imply a lower likelihood that the estimated capital effect is fully driven by unobservables.

D.2 Gaining vs. Losing Capital Status

We separately estimate the impacts of gaining and losing capital status on population density by using a change-on-change specification. Columns (1)-(2) in Table D2 show that change in capital status is associated with around 40% change in population density. Columns (3)-(4) includes lagged ln population density in $t-1$ and $t-2$. The negative coefficients on the lagged variables suggest a pattern of convergence in population density. Nevertheless, change in capital status still has a sizable impacts after controlling for these lagged dependent variables.

Columns (5)-(8) show that both gaining and losing provincial capital status matter. The finding on losing capital status suggests that the omitted variable concern may not be essential; otherwise, we would observe that losing capital status matters little.

[1] The bias-adjusted effect is 0.470-(0.622-0.470)*(1-0.882)/(0.882-0.757)=0.327

[2] The +/- 2.8 standard error of the coefficient of capital status based on column (4) of Table 3 is [0.178, 0.762].

Table D2: Gaining vs. Losing Status

	The impact of change in capital status				The impact of gaining vs. losing			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
δ prov. cap	0.431***	0.415***	0.334***	0.305***				
	(0.106)	(0.110)	(0.086)	(0.080)				
Gaining capital status					0.513**	0.465**	0.425**	0.385***
					(0.221)	(0.230)	(0.165)	(0.148)
Losing capital status					-0.390***	-0.389***	-0.287***	-0.263**
					(0.117)	(0.124)	(0.105)	(0.102)
Lag. ln pop. density			-0.428***	-0.425***			-0.428***	-0.425***
			(0.015)	(0.017)			(0.015)	(0.017)
Lag 2 ln pop. density				-0.096***				-0.096***
				(0.020)				(0.020)
Pref. FE		Y	Y	Y		Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE * Controls	Y	Y	Y	Y	Y	Y	Y	Y
Observations	2,610	2,610	2,610	2,610	2,610	2,610	2,610	2,610
R-squared	0.715	0.715	0.789	0.800	0.715	0.715	0.789	0.800
# prefectures	261	261	261	261	261	261	261	261

Notes: Controls include (i) geography variables: log area, whether a prefecture contains a plain, a major river, whether it is on the coast, as well as its slope, elevation, longitude, and latitude; (ii) agriculture variables: the suitability of rice, wheat, millet, sweet potatoes and maize; and (iii) the 9-physiographic macroregions defined by Skinner (1977).
Standard errors presented in the paraphrases are clustered at the prefecture level. ***: significant at 1%, **: significant at 5%, *: significant at 10%.

D.3 Ever-Capital Prefectures

We also consider the role of ever-capital prefectures in two ways. First, we find that provincial capital status is associated with a higher population density within the subgroup of prefectures that have ever been a provincial capital (columns (1)-(2) of Table D3).

Second, our baseline finding also holds if we further control for all the ever-capital prefectures to exhibit different trends by controlling for ever-capital status status and its interaction with year dummies (columns (3)-(8)).

Table D3: The Impact of Capital Status on Population Density, Ever-Capital Prefectures

	Ever-Capital Pref. ln Pop Density		All Prefectures ln Pop Density		All Prefectures Δln Pop Density			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Prov. Capital	0.655***	0.311**	0.655***	0.514***				
	(0.198)	(0.124)	(0.195)	(0.142)				
Δ Prov. Capital					0.448**	0.453***		
					(0.178)	(0.141)		
Gaining Capital Status							0.218	0.450**
							(0.275)	(0.220)
Losing Capital Status							-0.649***	-0.455***
							(0.211)	(0.174)
Pref. FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE * Controls		Y		Y		Y		Y
Year FE * Ever-capital			Y	Y	Y	Y	Y	Y
Observations	693	693	2,871	2,871	2,610	2,610	2,610	2,610
R-squared	0.738	0.926	0.758	0.882	0.411	0.716	0.412	0.716
# prefectures	63	63	261	261	261	261	261	261

Notes: Controls include (i) geography variables: log area, whether a prefecture contains a plain, a major river, whether it is on the coast, as well as its slope, elevation, longitude, and latitude; (ii) agriculture variables: the suitability of rice, wheat, millet, sweet potatoes and maize; and (iii) the 9-physiographic macroregions defined by Skinner (1977).
Standard errors presented in the paraphrases are clustered at the prefecture level. ***: significant at 1%, **: significant at 5%, *: significant at 10%.

D.4 Period-by-Period Results

Using the period before the capital status changes as the reference group, the following table shows that there are no systematic differences in population density before the status change. The eight prefectures with multiples changes are excluded from this analysis.

Due to the frequency of data availability, we have more periods after capital status change than those before change. Thus, we divide the pre-periods before period0 into 1-75 years before 0, and 75+ years before 0; the post-periods into 1-75 years after 0, and 76-150 years after 0, 151-300 years after 0 and 300+ years after 0. Roughly speaking, a period of 75 years is about one fourth of a regime's length. As shown in Table D4, both the increase and the decrease in population density occur only after gaining and losing capital status.

D.5 Using Subperiods of the Data

Our finding is robust to using different subperiods of the data. In columns (1)-(2) of Table D5, we show that including the data in 1880, 1953, 1982, 1990 does not change our main finding, even though doing so make the gaps across periods more uneven. In columns (3)-(4), we find that our results are robust to keeping subperiods of roughly equal length; columns (5)-(9) show that our results are also robust to dropping any specific regime.

Table D4: Period-by-Period Estimates (Dependent Var.: ln Pop Density)

	(1)	(2)	(3)
Pre-Gaining: -75+	-0.329		-0.239
	(0.205)		(0.202)
Pre-Gaining: -75 ∼-1	-0.064		-0.050
	(0.081)		(0.087)
Post-Gaining: 1 ∼75	0.589**		0.563**
	(0.237)		(0.237)
Post-Gaining: 76 ∼150	0.601**		0.569**
	(0.284)		(0.286)
Post-Gaining: 151 ∼300	1.010**		0.857*
	(0.470)		(0.484)
Post-Gaining: 300+	1.099*		0.921
	(0.563)		(0.591)
Pre-Losing: -75+		0.124	0.046
		(0.110)	(0.110)
Pre- Losing: -75 ∼-1		0.001	-0.009
		(0.053)	(0.057)
Post- Losing: 1 ∼75		-0.512***	-0.391***
		(0.144)	(0.128)
Post- Losing: 76 ∼150		-0.470***	-0.333***
		(0.142)	(0.128)
Post- Losing: 151 ∼300		-0.497*	-0.340
		(0.269)	(0.267)
Post- Losing: 300+		-0.611**	-0.317
		(0.273)	(0.268)
Pref. FE, Year FE	Y	Y	Y
Year FE * Controls	Y	Y	Y
Year FE * Ever-capital	Y	Y	Y
Observations	2,783	2,783	2,783
R-squared	0.882	0.881	0.883
# Prefectures	253	253	253

Notes: Controls include (i) geography variables: log area, whether a prefecture contains a plain, a major river, whether it is on the coast, as well as its slope, elevation, longitude, and latitude; (ii) agriculture variables: the suitability of rice, wheat, millet, sweet potatoes and maize; and (iii) the 9-physiographic macroregions defined by Skinner (1977).
Standard errors presented in the paraphrases are clustered at the prefecture level. ***: significant at 1%, **: significant at 5%, *: significant at 10%.

Table D5: Reducing Uneven Gaps and Using Different Subperiods (Dependent Var.: ln Pop Density)

		Baseline (1)	+smaller gaps (2)	Gap: ca400 (3)	Gap: ca120 (4)	Excluding: Song (5)	Ming (6)	Qing (7)	PR China (8)	Ming/Qing (9)
Prov. Capital		0.514***	0.501***	0.512***	0.575***	0.369***	0.485***	0.626***	0.473***	0.665***
		(0.142)	(0.131)	(0.154)	(0.143)	(0.123)	(0.138)	(0.154)	(0.172)	(0.173)
Song	980	Y	Y	Y	Y		Y	Y	Y	Y
	1078	Y	Y				Y	Y	Y	Y
	1102	Y	Y		Y		Y	Y	Y	Y
Ming	1393	Y	Y	Y	Y	Y		Y	Y	
	1580	Y	Y		Y	Y		Y	Y	
Qing	1776	Y	Y			Y	Y		Y	
	1820	Y	Y	Y	Y	Y	Y		Y	
	1851	Y	Y			Y	Y		Y	
	1880		Y			Y	Y		Y	
	1910	Y	Y			Y	Y		Y	
P R China	1953		Y			Y	Y	Y		Y
	1964	Y	Y			Y	Y	Y		Y
	1982		Y			Y	Y	Y		Y
	1990		Y			Y	Y	Y		Y
	2000	Y	Y		Y	Y	Y	Y		Y
Pref. FE		Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE		Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE*Controls		Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE*Ever-capital		Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations		2,871	3,915	783	1,566	3,132	3,393	2,610	2,610	1,827
R-squared		0.882	0.887	0.887	0.891	0.893	0.885	0.903	0.864	0.915
# prefectures		261	261	261	261	261	261	261	261	261

Notes: Controls include (i) geography variables: log area, whether a prefecture contains a plain, a major river, whether it is on the coast, as well as its slope, elevation, longitude, and latitude; (ii) agriculture variables: the suitability of rice, wheat, millet, sweet potatoes and maize; and (iii) the 9-physiographic macroregions defined by Skinner (1977). Standard errors presented in the paraphrases are clustered at the prefecture level. ***: significant at 1%, **: significant at 5%, *: significant at 10%.

D.6 Grid-level Analysis

We conduct a 1-degree × 1-degree grid analysis and a 2-degree × 2-degree grid analysis. The size of a prefecture in our main analysis is between the size of the two grids: instead of 261 prefectures, we have 361 1-degree × 1-degree grids and 94 2-degree × 2-degree grids. These two sets of grid-level analysis again generate patterns comparable with our prefecture-level analysis (presented in Table D6) indicating that our findings are robust to alternative boundary definitions.

Table D6: The Impact of Capital Status on Population Density: Grid-level Analysis

	1-degree X 1-degree				2-degree X 2-degree			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Prov. Capital	0.323***	0.288***	0.270***	0.338***	0.270**	0.197**	0.166*	0.323***
	(0.077)	(0.066)	(0.065)	(0.065)	(0.103)	(0.094)	(0.098)	(0.100)
Grid FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Year FE * Geography			Y	Y			Y	Y
Year FE * Agriculture			Y	Y			Y	Y
Year FE * Region FE				Y				Y
Observations	3,971	3,971	3,971	3,971	1,034	1,034	1,034	1,034
R-squared	0.755	0.792	0.806	0.845	0.821	0.877	0.889	0.920
# of grid	361	361	361	361	94	94	94	94

Notes: Standard errors presented in the paraphrases are clustered at the grid level. ***: significant at 1%, **: significant at 5%, *: significant at 10%.

E Additional Results for the IV Analysis

E.1 Validity Checks of the Instrument

For our IV analysis, we assume that the change of a prefecture's hierarchical distance is orthogonal to its own characteristics, because the change in a prefecture's hierarchical distance stems from regime change.

To check whether this is the case, we first show that the change in hierarchical distance is not significantly correlated with the levels and changes in population density in the past periods (presented in Table E1(a)).

In addition, we find that ln rank in hierarchical distance matters for both gaining capital status and losing capital status (presented in Table E1(b)). Since it seems unlikely that the central government would intentionally increase a prefecture's hierarchical distance to make it lose its capital status, this finding is reassuring for the validity of hierarchical distance.

Table E1(a): Validity Check I – Hierarchical Distance and Pre-change Characteristics

	Δln Hierarchical Distance					
	(1)	(2)	(3)	(4)	(5)	(6)
lag. ln Pop Density	-0.029		-0.030			
	(0.024)		(0.031)			
lag2. ln Pop Density		-0.021	-0.004			
		(0.022)	(0.018)			
lag. Δln Pop Density				-0.010		-0.014
				(0.018)		(0.021)
lag2. Δln Pop Density					-0.032	-0.034
					(0.038)	(0.039)
Prefecture FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Year FE*Controls	Y	Y	Y	Y	Y	Y
Observations	2,610	2,349	2,349	2,349	2,088	2,088
R-squared	0.278	0.278	0.279	0.278	0.278	0.279
#prefectures	261	261	261	261	261	261

Notes: Controls include (i) geography variables:log area, whether a prefecture contains a plain, a major river, whether it is on the coast, as well as its slope, elevation, longitude, and latitude; (ii) agriculture variables: the suitability of rice, wheat, millet, sweet potatoes and maize; and (iii) the 9-physiographic macroregions defined by Skinner (1977).
Standard errors presented in the paraphrases are clustered at the prefecture level. ***: significant at 1%, **: significant at 5%, *: significant at 10%.

Table E1(b): Validity Check II – Gaining vs. Losing Capital Status

		Seemingly Unrelated Regression	
	ΔCapital Status	Gaining Cap.	Losing Cap.
	(1)	(2)	(3)
Δln Hierarchical distance	-0.059***	-0.027***	0.032***
	(0.016)	(0.004)	(0.005)
Prefecture FE	Y	Y	Y
Year FE	Y	Y	Y
Year FE*Controls	Y	Y	Y
Observations	2,610	2,610	2,610
R-squared	0.187	0.195	0.300
#prefectures	261	261	261

Notes: Controls include (i) log area, geography variables: whether a prefecture contains a plain, a major river, whether it is on the coast, as well as its slope, elevation, longitude, and latitude; (ii) agriculture variables: the suitability of rice, wheat, millet, sweet potatoes and maize; and (iii) the 9-physiographic macroregions defined by Skinner (1977).
Standard errors presented in the paraphrases are clustered at the prefecture level. ***: significant at 1%, **: significant at 5%, *: significant at 10%.

E.2 ln Rank in $\sum_{j=1}^{N} A_j D_{i,j,t}$ × ln Rank in $\sum_{j=1}^{N} A_j D_{i,NatCap,t}$ as an instrument

Since our main concern is whether being closer to the national capital or the other prefectures affects development through channels beyond provincial capital status, we further employ ln Rank in $\sum_{j=1}^{N} A_j D_{i,j,t}$ × ln Rank in $\sum_{j=1}^{N} A_j D_{i,NatCap,t}$ as an instrument while controlling for the independent effects of the two components. These results are presented in Table E2.

Table E2: Exclusion Restriction? Alternative Specification of the Instrument

	Reduced form			IV	
	All	Ever-capital	Never-capital	First-stage	Second-stage
	(1)	(2)	(3)	(4)	(5)
Prov. Capital					0.850***
					(0.324)
ln Rank in dist to National Cap.*	0.057**	0.176***	0.006	0.067***	
ln Rank in dist to Prov Centroid	(0.023)	(0.047)	(0.029)	(0.008)	
ln Rank in dist to National Cap.	-0.197***	-0.489***	-0.049	-0.187***	-0.038
	(0.052)	(0.089)	(0.070)	(0.019)	(0.025)
ln Rank in dist to Prov Centroid	-0.161***	-0.391***	-0.039	-0.191***	0.001
	(0.047)	(0.090)	(0.061)	(0.167)	(0.032)
Pref. FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
Year FE*Controls	Y	Y	Y	Y	Y
Observations	2,871	693	2,178	2,871	2,871
R-squared	0.879	0.929	0.882	0.216	0.879
# prefectures	261	63	198	261	261
Weak instrument test				67.6	

Notes: Controls include (i) geography variables: log area, whether a prefecture contains a plain, a major river, whether it is on the coast, as well as its slope, elevation, longitude, and latitude; (ii) agriculture variables: the suitability of rice, wheat, millet, sweet potatoes and maize; and (iii) the 9-physiographic macroregions defined by Skinner (1977).
Standard errors presented in the paraphrases are clustered at the prefecture level. ***: significant at 1%, **: significant at 5%, *: significant at 10%.

As shown in column (4), since a higher value of rank in distance to national capital (or provincial centroid) indicates longer distance, each component decreases the probability of being a provincial capital. However, the marginal effect of each component declines with the other component. Intuitively, for a prefecture that is already far from the provincial centroid, the change in distance to the national capital matters less. This is why we obtain a positive interaction effect.

Two findings are worth emphasizing. First, both ln Rank in $\sum_{j=1}^{N} A_j D_{i,j,t}$ and ln Rank in $\sum_{j=1}^{N} A_j D_{i,NatCap,t}$ are indeed strongly correlated with population density (column (1)); however, when using their interaction as the instrument, we do not find that either part alone has an additional direct effect (column (5)). Once again, this finding suggests that the variation in the interaction mitigates the concern of exclusion restriction. Second, compared with the estimate using hierarchical distance, the IV estimate from using the interaction is slightly higher, but the first-stage is smaller (with a F-statistics of 67.6 in contrast with 131.9 in Table 4). This difference further suggests that our definition of hierarchical distance better captures the political logic of provincial capital choice by the central government.

E.3 Placebo Hierarchical Distances

The two-component structure of hierarchical distance also allows to create multiple placebo hierarchical distance ranks by exploiting the changes in national capital status. For instance, we calculate one such placebo to Kaifeng when Kaifeng was not a capital and similar ones for Nanjing and Beijing before they became national capitals.

Table E3: Considering Placebo Hierarchical Distance

	(1)	(2)	(3)
IV Estimates: ln Pop Density			
Prov. Capital	0.901***	0.981**	0.779***
	(0.206)	(0.418)	(0.229)
ln Rank in H dist. KF * Post-	0.017		
	(0.024)		
ln Rank in H dist. NJ * Pre-		0.030	
		(0.044)	
ln Rank in H dist. BJ * Pre-			-0.028
			(0.045)
Pref. FE	Y	Y	Y
Year FE	Y	Y	Y
Year FE * Controls	Y	Y	Y
Observations	2,871	2,871	2,871
R-squared	0.878	0.877	0.880
#Prefectures	261	261	261
F-Stat (Weak instrument test)	168.2	40.9	132.5

Notes: Controls include (i) geography variables: log area, whether a prefecture contains a plain, a major river, whether it is on the coast, as well as its slope, elevation, longitude, and latitude; (ii) agriculture variables: the suitability of rice, wheat, millet, sweet potatoes and maize; and (iii) the 9-physiographic macroregions defined by Skinner (1977).
Standard errors presented in the paraphrases are clustered at the prefecture level. ***: significant at 1%, **: significant at 5%, *: significant at 10%.

In Table E3, we find that including these placebo hierarchical distance ranks does not alter our IV estimate. Nor does it affect population density, implying that our findings are specific to these cities' political status.

E.4 Distance to Major Market Centers

To check whether our findings are confounded by distances to major market centers, we calculate a prefecture's (hierarchical) distance to three types of market centers: the north China plain during the Song Dynasty and the lower Yangtze after the Song (cf. Skinner 1977), Shanghai in the east, and Guangzhou in the south. To calculate the ranks of these distances, we replace $D_{i,NatCap,t}$ with $D_{i,Market,t}$. Like the placebo distances, these ranks are also correlated with our instrument and thus may be correlated with the probability of being a capital. However, once again, none explains the role of our hierarchical distance (see Table E4).

Table E4: Considering Distance to Major Market Centers

	(1)	(2)	(3)
IV Estimates: ln Pop Density			
Prov. Capital	0.716***	0.858***	0.814***
	(0.256)	(0.248)	(0.227)
ln Rank in H dist. to major econ region	-0.029		
	(0.026)		
ln Rank in H dist. to the East (Shanghai)		0.024	
		(0.027)	
ln Rank in H dist. to the South (Guangzhou)			0.017
			(0.023)
Pref. FE	Y	Y	Y
Year FE	Y	Y	Y
Year FE * Controls	Y	Y	Y
Observations	2,871	2,871	2,871
R-squared	0.881	0.879	0.880
#Prefectures	261	261	261
F-Stat (Weak instrument test)	105.2	113.1	134.9

Notes: Controls include (i) geography variables: log area, whether a prefecture contains a plain, a major river, whether it is on the coast, as well as its slope, elevation, longitude, and latitude; (ii) agriculture variables: the suitability of rice, wheat, millet, sweet potatoes and maize; and (iii) the 9-physiographic macroregions defined by Skinner (1977).
Standard errors presented in the paraphrases are clustered at the prefecture level. ***: significant at 1%, **: significant at 5%, *: significant at 10%.

Thus, once again, this finding shows that the hierarchical distance to the political center (instead of national economic centers) is the driver of our finding on provincial capital status.