

Isolated Capital Cities, Accountability, and Corruption: Evidence from US States (Online Appendix)

By FILIPE R. CAMPANTE AND QUOC-ANH DO

This appendix contains two parts: first, the Data Appendix describes the variables and sources used in the main analysis in “Isolated Capital Cities, Accountability, and Corruption: Evidence from US States,” from Campante and Do. The Online Appendix Tables then contain the additional results that are referred to in the text of the paper

I. Data Appendix

AvgLogDistance

This measure is based on what Campante and Do (2010) call the “Gravity-based” CISC (G-CISC). They show that averaging (over the distribution of population) iso-elastic transformations of distance to the capital uniquely guarantees the following properties: First, that if a subset of a population becomes more concentrated around the capital, so does the overall distribution (subgroup consistency); second, that moving people closer to the capital implies a more concentrated distribution (monotonicity); and third, that the measure is independent of the units in which distance is measured (rank invariance). In addition, it also satisfies a property of convexity that seems reasonable in our context, namely that a given movement of a person towards the capital has a greater impact on concentration if it happens close to the capital. The G-CISC, in particular, can be shown to be uniquely unbiased to that type of measurement error, in that (uniformly) “random” deviations do not change the rankings of distributions as measured by it: technically, it is invariant under mean-preserving spreads of the distribution around points other than the capital.

More formally, let $\Omega \subset \mathbb{R}^2$ be a convex set that can contain (the representation of) all states, p_x denote the distribution of individuals in a given state, defined over Ω , and $z_{\phi x}$ be the distance between a point $x \in \Omega$ and the state capital, which we fix at a point $\phi \in \Omega$. We then have: $GCISC = \int_{\Omega} (1 - \alpha \log z_{\phi x}) dp_x$, where $\alpha > 0$ is a normalization parameter to be adjusted so that the measure is contained between 0 and 1. In practice, following Campante and Do (2010), we have the formula $GCISC_1 = \sum_i s_i (\alpha_1 \log(d_i) + \beta_1)$, where s_i is the share of the state’s population living in county i and d_i is the distance between county i ’s centroid and the point of interest (capital or centroid or largest city). (The capital is defined by the coordinates of the State House or Assembly.)

The parameters (α_1, β_1) are $\left(-\frac{1}{\log(\bar{d}_1)}, 1\right)$, where \bar{d}_1 is the maximum distance, across all states, between a state's capital (or another point of interest) and another point in that state. By the same token, $GCISC_2 = \sum_i s_i (\alpha_2 \log(d_i) + \beta_2)$. The parameters (α_2, β_2) are $\left(-\frac{1}{\log(\bar{d}_2)}, 1\right)$, where \bar{d}_2 is the maximum distance, for each state, between the state's capital (or other point of interest) and another point in that state. In this way, $GCISC_2$ controls for the state's size, while $GCISC_1$ does not. $LCISC$ replaces $\log(d)$ with d in the above formulae. We then define $AvgLogDistance_{not} = 1 - GCISC_1$ and $AvgLogDistance_{adj} = 1 - GCISC_2$. The population data are from the US Census, from 1920 to 2000. (As we go all the way back to 1920, we exclude Alaska and Hawaii, which were not states at the beginning of the period.) Centroid CISCs are calculated in similar fashion, with distances measured with respect to the state centroid.

AvgLogDistance_{not} 1920-1970: n= 48 , Mean: 0.689, S.D.: 0.095, Min: 0.423, Max: 0.842.

AvgLogDistance_{adj} 1920-1970: n= 48 , Mean: 0.797, S.D.: 0.073, Min: 0.589, Max: 0.920.

CentroidAvgLogDistance_{not} 1920-1970: n=48 , Mean: 0.758, S.D.: 0.084, Min: 0.474, Max: 0.910.

CORRUPTION

Number of federal convictions for public corruption 1976-2002 relative to average population in the state 1976-2002, from the 1989, 1999 and 2002 issues of the Report to Congress on the activities and operations of the Public Integrity Section, issued by the Department of Justice. The data were generously shared by Ed Glaeser and Raven Saks (GS), and is further described in Glaeser and Saks (2006).

n= 50 , Mean: 0.275, S.D.: 0.132, Min: 0.073, Max: 0.625.

CORRUPTION (STATE OFFICIALS)

Number of federal convictions of state officials for public corruption 1986-2011, obtained separately for each year from the Transactional Records Access Clearinghouse (TRACfed) at Syracuse University, per 100 state government employees (as of 1980). Specifically, using the "Analyzer - Data" tool at <http://tracfed.syr.edu/index/index.php?layer=cricri>, we select for each state all the corresponding districts in the "Select geographical area" field, each individual year in the "Select fiscal year" field, and "Corruption(Govt Off)-State" in the "Select program area" field, and then count the number of convictions in the output file. This lets us obtain the information from the database's lead program on "Official Corruption".

n= 50; Mean: 0.0304; S.D.: 0.0223; Min: 0; Max: 0.0874.

CORRUPTION (EXALEAD)

Following Saiz and Simonsohn (2008), we use the *Exalead* (www.exalead.com) search tool to search for the word “corruption” near the name of the state (*corruption NEAR name of state*), and divide the number of hits by the number of hits for the name of the state alone. (In the case of Washington, we search for “Washington State”.) The search was performed in September 2009.
n= 50, Mean: 0.00023, S.D.: 0.00034, Min: 0.00004, Max: 0.00177.

INCOME

Median household income, from US Census. (GS)

% COLLEGE

Share of the population aged 25 and up with four or more years of college, from US Census. (GS)

POPULATION

From Bureau of Economic Analysis (BEA). (GS)

AREA

From US Census.
n= 50; Mean: 70748; S.D.: 85987; Min: 1045; Max: 571951.

MAXIMUM DISTANCE

Maximum distance between capital and county centroids, from US Census.
n= 48; Mean: 388.1; S.D.: 180.1; Min: 42.8; Max: 824.7.

SHARE OF GOVERNMENT EMPLOYMENT

Number of government employees relative to total employment, from BEA. (GS)

% URBAN

Urban population relative to total population in the state, from US Census. (GS)

REGIONAL DUMMIES

US Census regions: South, Midwest, West, Northeast.

RACIAL DISSIMILARITY

Racial heterogeneity: Dissimilarity index = $1 - \sum s_i^2$, where s_i is the population share of group i . (GS)

REGULATION INDEX

Component of economic freedom index published in 1999 by Clemson University, <http://freedom.clemson.edu>. (GS)

SHARE OF VALUE ADDED IN MINING

In 2007, from BEA.

CAPITAL SHARE

Population of capital city relative to total population in the state (1980), from US Census.

n= 50, Mean: 0.063, S.D.: 0.064, Min: 0.004, Max: 0.280.

CAPITAL LARGEST

Dummy equal to 1 is capital is largest city in the state.

n= 50, Mean: 0.34, S.D.: 0.478, Min: 0, Max: 1.

DRUG CASES

Criminal defendants charged with “drug offenses” and commenced at US District Courts during the 12-month period ending in September 30, 2011, per 100,000 inhabitants (as of 2000). The number of cases is taken from Table D-3 in the *2011 Annual Report of the Director: Judicial Business of the United States Courts*, available at:

<http://www.uscourts.gov/Statistics/JudicialBusiness.aspx#appTables>

n= 50, Mean: 6.351, S.D.: 6.379, Min: 1.040, Max: 37.845.

SuitCentroidAvgLogDistance

We compute *AvgLogDistance_{not}* as described before, but with land suitability instead of populations. We use the gridded land suitability data from the *Atlas of the Biosphere* published by the Center for Sustainability and the Global Environment (SAGE) at the University of Wisconsin-Madison (available at http://www.sage.wisc.edu/iamdata/grid_data_sel.php). (The original source of the data is Ramankutty et al (2002).) Land suitability is measured as the probability that a given area will be cultivated, calculated as a function of climate and soil properties. Climate properties included are annual sum of days with mean

temperature above 5 degrees Celsius and the ratio of actual to potential evapotranspiration; soil properties are soil carbon density and soil acidity (pH). We match the grid with coordinates of US counties, so that we have the average land suitability for each county, and we use the distance from each county's centroid to the state centroid as our measure of distance.
n=48, Mean: 0.752, S.D.: 0.079, Min: 0.466, Max: 0.880.

TOTAL BORDER

Number of total miles of borders with another state (from Holmes 1998), with another country (from US Census, as reported in the *US Statistical Abstract*, 2012, Table 363), and of general coastline, including Great Lakes (from US Census, as reported in the *US Statistical Abstract*, 2012, Tables 361 and 364).
n= 49, Mean: 1039, S.D.: 480, Min: 103, Max: 2780.

Reference: Holmes, Thomas J. (1998), "The Effects of State Policies on the Location of Industry: Evidence from State Borders." *Journal of Political Economy* 106(4): 667-705.

LATITUDE AND LONGITUDE

Latitude of northernmost internal point, longitude of westernmost internal point. (GS)
n= 50 and 50, Mean: 39.5 and 93.6, S.D.: 6.1 and 19.1, Min: 20.8 and 69.4 , Max: 61.6 and 157.0.

ELEVATION SPAN

Difference between state's highest and lowest point, in feet, from US Geological Survey.
n= 50, Mean: 5494, S.D.: 4736, Min: 345, Max: 20320.

PERCENTAGE OF WATER AREA

Percentage of state area that is inland water (i.e. surrounded by US land), from US Census.
n= 50, Mean: 2.97, S.D.: 3.06, Min: 0.20, Max: 17.

NAVIGABLE WATERWAYS

Mileage of inland waterways, determined by including the length of channels 1) with a controlling draft of nine feet or greater, 2) with commercial cargo traffic reported for 1998 and 1999, but 3) were not offshore (i.e., channels in coastal areas included only the miles from the entrance channel inward). Channels within major bays are included (e.g., Chesapeake Bay, San Francisco Bay, Puget Sound,

Long Island Sound, and major sounds and straits in southeastern Alaska). Channels in the Great Lakes are not included, but waterways connecting lakes and the St. Lawrence Seaway inside the United States are included. From US Army Corps of Engineers, available at http://www.statemaster.com/graph/trn_inl_wat_mil-transportation-inland-waterway-mileage.
n= 50, Mean: 592.4, S.D.: 919.3, Min: 0, Max: 5497.

SHARE OF ARABLE LAND, 1950

Agricultural land as a share of total area, in 1950, from US Department of Agriculture.
n= 48, Mean: 0.616, S.D.: 0.218, Min: 0.107, Max: 0.984.

NEWSPAPER COVERAGE

We start by searching for each of the terms under consideration – “state elections,” “state budget,” “state government,” and the name of the governor – in the online archives of all of a state’s newspapers that are available on NewsLibrary.com, for articles published between 01/01/2008 and 12/31/2009. (In the case of the governor’s name, we search for the name of each governor during the sub-period, within that period, in which he or she was in office. We look for different combinations, including first name (or popular nickname) and last name, with or without middle initial. The full list is available upon request.) We also search for the neutral term “Monday”. In order to compute the CISC of circulation, we use county-level circulation data from the *U.S. County Penetration Report* (Spring 2010), from the Audit Bureau of Circulations. Our final sample includes 436 newspapers, from 49 states (all except Montana).
“State Elections”: n= 431, Mean: 997.8, S.D.: 900.2, Min: 0; Max: 7019.
Governor’s name: n= 436, Mean: 719.8, S.D.: 736.5, Min: 0, Max: 5096.

KNOWLEDGE

From the American National Election Studies (ANES), 1998 Pilot Study. We code a dummy equal to 1 if the respondent answers “Yes” to question 98P130 (“Do you happen to remember the names of the candidates for Governor who are running in the November election?”) *and* provides a correct name in response to question 98P131 (which follows up asking for the name of such a candidate).

INTEREST

From the American National Election Studies (ANES), 1998 Pilot Study. We code a dummy equal to 1 if the respondent reports to care “a great deal,” “quite a bit,” or “some” to question 98P105 (“How much attention did you pay to newspaper articles about the campaign for Governor?”). We limit the sample to

individuals who report to have read newspapers at least once in the past week, in response to question 98P103 (“How many days in the past week did you read a daily newspaper?”).

GENERAL INTEREST

From the American National Election Studies (ANES), 1998 Pilot Study. We code a dummy equal to 1 if respondents answer “most of the time” or “some of the time” to question 98P376 (“Some people seem to follow what’s going on in government and public affairs most of the time, whether there’s an election going on or not. Others aren’t that interested. Would you say you follow what’s going on in government and public affairs most of the time, some of the time, only now and then, or hardly at all?”).

GENERAL LEVEL OF INFORMATION

From the American National Election Studies (ANES), 1998 Pilot Study. Question 98P440, asked of interviewers (“Respondent’s general level of information about politics and public affairs seemed: 1 Very high, 2 Fairly high, 3 Average, 4 Fairly low, 5 Very low”). We drop individuals whom the interviewers stated they could not rate (6).

TURNOUT IN STATE ELECTIONS

Total popular vote cast for candidates for governor at the county level in 1990-2012 (from David Leip’s Atlas of US Presidential Elections, <http://uselectionatlas.org>), divided by total voting-age population in the county (from US Census). This covers at least six elections for the vast majority of states; the exceptions are NJ and VA (the only states that vote for governor in the year immediately after a presidential election). A few other states have more than six elections: NH and VT have two-year governor terms, and CA, RI, UT, WV, and WI had one extra election on top of the regularly scheduled.
n= 18837, Mean: 0.462, S.D.: 0.127, Min: 0.041, Max: 1.0.

STATE CAMPAIGN CONTRIBUTIONS

Sum of state-wide campaign contributions to state office between 2001 and 2010, from the National Institute on Money in State Politics (available at www.followthemoney.org).
n= 50, Mean: 102.4 million, S.D.: 135.81 million, Min: 9.9 million, Max: 824.9 million.

REAL ESTATE CAMPAIGN CONTRIBUTIONS

Contributions to state office by individuals and companies in the real estate industry between 2001 and 2010, from the National Institute on Money in State Pol-

itics (www.followthemoney.org.) Summed by county: n= 3104, Mean: \$165,206 , S.D.: \$1,246,983, Min: 0, Max: \$42,248,116.

PRESIDENTIAL CAMPAIGN CONTRIBUTIONS

Total donations to presidential campaigns in 2008 cycle, from Federal Elections Commission (FEC) data compiled by the Center for Responsive Politics, available at http://www.opensecrets.org/pres08/pres_stateAll.php?list=all n= 50, Mean: 17.5 million, S.D.: 27.96 million, Min: 0.41 million, Max: 153.2 million.

PUBLIC GOOD EXPENDITURES

Share of state general expenditures assigned to the categories “Education,” “Public Welfare,” “Health,” and “Hospitals,” from *US Statistical Abstract* 2012, Table 454.
n= 50, Mean: 0.654, S.D.: 0.074, Min: 0.366, Max: 0.766.

OTHER EXPENDITURES

Share of state general expenditures assigned to the categories “Governmental Administration,” “Interest on General Debt,” and “Other,” from *US Statistical Abstract* 2012, Table 454.
n= 50, Mean: 0.179, S.D.: 0.056, Min: 0.109, Max: 0.400.

PUBLIC GOOD PROVISION

First principal component of: “Smartest State” index, from Morgan Quitno Corporation (2005); % Insured: 100 - Persons Without Health Insurance Coverage in 2008-2009 (Percent Total) from *US Statistical Abstract* 2012, Table 156; and log of Hospital Beds in 2009, from *US Statistical Abstract* 2012, Table 194 (divided by population).
n= 50, Mean: 0.237, S.D.: 1.232, Min: -1.944, Max: 3.494.

Table OA1. Correlations between Corruption Measures

	Federal Convictions	State Officials	Exalead	Local Officials	Perceptions
Federal Convictions	1				
State Officials	0.5953 [0.000]	1			
Exalead	0.2486 [0.082]	0.2327 [0.104]	1		
Local Officials	0.4113 [0.003]	0.3327 [0.018]	0.1469 [0.309]	1	
Perceptions (Boylan-Long)	0.1804 [0.230]	0.4942 [0.001]	0.1207 [0.424]	0.2109 [0.159]	1

p-values in parentheses. See Data Appendix for definitions of variables.

Table OA2. Correlations with Predetermined Variables including Historical Variables

Variable	(1)	(2)	(3)	(4)	(5)	(6)
	Centroid AvgLogDist _{not} (population)	Centroid AvgLogDist _{not} (suitability)	Centroid AvgLogDist _{not} (population)	Centroid AvgLogDist _{not} (suitability)	AvgLogDist _{not}	AvgLogDist _{not}
	Individual	Individual	Joint	Joint	Individual	Joint
Log Total Border	0.0115 [0.565]	0.0080 [0.418]	0.0134 [0.558]	0.0161 [0.262]	-0.0363 [0.242]	-0.0463 [0.147]
Latitude	0.0004 [0.590]	-0.0001 [0.810]	0.0009 [0.360]	-0.0004 [0.568]	-0.0020 [0.231]	0.0003 [0.883]
Longitude	-0.0003 [0.539]	0.0003 [0.305]	-0.0001 [0.921]	0.0004 [0.224]	-0.0009 [0.254]	0.0003 [0.737]
Log Distance to DC	-0.0027 [0.742]	0.0009 [0.834]	0.0070 [0.567]	-0.0011 [0.804]	-0.0140 [0.193]	-0.0099 [0.441]
Date of Statehood	-0.0002 [0.177]	0.0000 [0.612]	-0.0004* [0.053]	-0.0002 [0.139]	-0.0004* [0.090]	-0.0001 [0.656]
Log Elevation Span	-0.0049 [0.307]	0.0004 [0.881]	-0.0104 [0.101]	-0.0053 [0.134]	-0.0257*** [0.009]	-0.0204* [0.057]
Percentage of Water Area	-0.0017 [0.181]	-0.0004 [0.682]	-0.0034 [0.121]	-0.0008 [0.527]	0.0001 [0.979]	-0.0019 [0.542]
Log Navigable Waterways	0.0016 [0.287]	-0.0006 [0.316]	0.0008 [0.762]	-0.0005 [0.748]	0.0071* [0.087]	0.0036 [0.452]
Share of Arable Land (1950)	-0.0036 [0.869]	-0.0116 [0.354]	-0.0429* [0.072]	-0.0155 [0.302]	0.0478 [0.343]	0.0151 [0.756]
Log Population (1870)	0.0040 [0.225]	-0.0021 [0.309]	0.0019 [0.768]	-0.0013 [0.659]	0.0091 [0.123]	0.0035 [0.844]
Log Railroads (1870)	0.0022 [0.388]	-0.0019 [0.295]	-0.0030 [0.345]	-0.0030 [0.193]	0.0024 [0.375]	-0.0148 [0.134]
F statistic (Geographic)			1.02	1.25		3.10***
P-value			0.448	0.304		0.009
F statistic (Historical)			0.46	0.98		1.79
P-value			0.634	0.388		0.184

p-values in brackets. Columns (1), (2), (5): Coefficients from individual regressions of AvgLogDistance on Log Area, Log Maximum Distance, and reported variable. Columns (3), (4), (6): Coefficients from multiple regression of AvgLogDistance on Log Area, Log Maximum Distance, and all reported variables. F statistic and p-value are for the joint hypothesis of significance of reported coefficients. *** p<0.01, ** p<0.05, * p<0.1.

Table OA3. Additional Robustness Checks I

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.: Corruption	No controls	No size controls	Including size controls	Time-specific isolation	Excluding West	Excluding South	Excluding Northeast	Excluding Midwest
AvgLogDistance _{not} (1920-1970)	0.4439*** [0.140]	0.6134** [0.252]			1.3938*** [0.422]	1.0019** [0.393]	0.9424** [0.373]	0.7876** [0.346]
AvgLogDistance _{adj} (1920-1970)			0.8818*** [0.272]					
AvgLogDistance _{not} (1970)				0.9057*** [0.286]				
Observations	48	48	48	48	37	32	39	36
R-squared	0.114	0.458	0.529	0.533	0.614	0.567	0.560	0.495

Robust standard errors in brackets. Dependent variable: Corruption = Federal convictions for corruption-related crime relative to population, avg. 1976-2002. Independent variables as of 1970. Size Control Variables (not included in Columns (1) and (2)): Log Area and Log Maximum Distance. Other Control Variables (not included in Column (1)): Log Income, Log Population, % College, Share of government employment, % Urban, Census Region dummies. *** p<0.01, ** p<0.05, * p<0.1.

Table OA4. Additional Robustness Checks II

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. Var.:	Corruption	Corruption	State Officials (Unweighted)	Corruption Perceptions	Local Officials	Corruption	Corruption
AvgLogDistance _{not} (1920-1970)	0.9949*** [0.338]	0.9385*** [0.344]	0.0918* [0.049]	4.6039** [1.963]	0.5722*** [0.165]		
Political Competitiveness	0.1371 [0.308]						
Taxes/Expenditures		-0.1353 [0.202]					
Capital Share						-0.1542 [0.297]	
Capital Largest							-0.0215 [0.035]
Observations	48	48	48	45	48	50	50
R-squared	0.535	0.537	0.615	0.644	0.509	0.412	0.415

Robust standard errors in brackets. Dependent variable: Corruption = Federal convictions for corruption-related crime relative to population, avg. 1976-2002. State Officials = Federal convictions of state public officials for corruption-related crime per 100 state government employees, avg. 1986-2011. Local Officials = Federal convictions of local public officials for corruption-related crime per 100 state government employees, avg. 1986-2011. Political Competitiveness and Taxes/Expenditures: See Data Appendix. Control Variables (as of 1970 except for Columns (6) and (7), 2000): Log Area and Log Maximum Distance, Log Income, Log Population, % College, Share of government employment, % Urban, Census Region dummies. *** p<0.01, ** p<0.05, * p<0.1.

Table OA5. Instrumenting using Centroid AvgLogDistance_{adj} (population)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var.: Corruption	1st Stage	1st Stage	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
AvgLogDistance _{not}	-0.3666		1.8817*	1.3854	1.1876			
	[0.217]		[0.899]	[0.835]	[0.783]			
AvgLogDistance _{adj}		-0.4850*				1.5151*	1.0518	0.8976
		[0.252]				[0.739]	[0.628]	[0.582]
Basic Control	X	X	X	X	X	X	X	X
Control Var. I	X	X		X	X		X	X
Control Var. II	X	X			X			X
Observations	48	48	48	48	48	48	48	48
R-squared	0.823	0.608	0.374	0.515	0.591	0.391	0.524	0.598
F-statistic	2.86	3.69	-	-	-	-	-	-
AR p-value	-	-	0.069	0.129	0.136	0.07	0.129	0.136

Robust standard errors in brackets. Dependent variable: Corruption = Federal convictions for corruption-related crime relative to population, avg. 1976-2002. Independent variables as of 1970 (GCISC: avg. 1920-1970). Basic Control variables: Log Income, Log Population, % College, Log Area, Log Maximum Distance. Control Variables I: Share of government employment, % Urban, Census Region dummies. Control Variables II: Racial dissimilarity, Regulation index, Share of value added in mining. IV: Centroid AvgLogDistance_{adj} of population. AR p-value: p-value from Anderson-Rubin (minimum distance) test. *** p<0.01, ** p<0.05, * p<0.1.

Table OA6. Distance to the Capital and Turnout in State Elections: Presidential vs. Off Years across Election Cycles

Dep. Var.:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Turnout in State Elections	1990 - 1992		1993 - 1996		1997 - 2000		2001 - 2004		2005 - 2008		2009 - 2012	
	Pres. Years	Off Years	Pres. Years	Off Years	Pres. Years	Off Years	Pres. Years	Off Years	Pres. Years	Off Years	Pres. Years	Off Years
Log Distance to Capital	-0.0032	-0.0501***	-0.0041	-0.0465**	-0.0048	-0.0439***	-0.0034	-0.0337***	-0.0048	-0.0400***	-0.0061	-0.0275***
	[0.004]	[0.005]	[0.003]	[0.012]	[0.003]	[0.008]	[0.004]	[0.008]	[0.005]	[0.007]	[0.004]	[0.006]
Test of equal coefficients: p-value	0.000		0.000		0.000		0.000		0.000		0.000	
Observations	571	265	566	385	566	385	566	428	566	385	638	440
R-squared	0.809	0.842	0.826	0.695	0.707	0.858	0.78	0.652	0.736	0.622	0.798	0.79

Robust standard errors in brackets, clustered by state. OLS regressions. Dependent variable: Turnout in state election, county-level (1990-2012). Independent variables: Log Distance to Capital, Log Distance to Centroid. Control variables: Log Density of Population over 18, % High School and above, % College and above, Log Median Household Income, Poverty Rate, Shares of population under 5, 5-17, 18-24, 25-44, 45-64, 65-74, 75-84, 85 and above, Shares of census-defined races, all from the preceding census, and Gini coefficient, Racial fractionalization, Religious fractionalization from 1990. The p-values are from tests of equal coefficients of Log Distance to Capital between samples of presidential election years and off years. All columns include state fixed effects. *** p<0.01, ** p<0.05, * p<0.1.

Table OA7. Turnout and Isolation of the Capital City: Aggregate Results

Dep. Var.:	(1)	(2)	(3)	(4)
Turnout in State Elections	Excluding presidential elections		Full sample	
AvgLogDistance _{not}	-0.2627		-0.1492	
	[0.169]		[0.177]	
AvgLogDistance _{adj}			-0.2443	
			[0.143]	
Observations	35		48	
R-squared	0.563		0.408	

Robust standard errors in brackets. OLS regressions. Dependent variable: Turnout in state elections, averaged from 1990 to 2012. Independent variables as of 1970 (AvgLogDistance. 1920-2000). Control variables: Log Density and Log Maximum Distance, Log Income, % College, Share of government employment, Regional dummies. *** p<0.01, ** p<0.05, * p<0.1.

Table OA8. Governors' Salaries and Isolation of the Capital City

Dep. Var.: Log Governor Salary	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	OLS	OLS	2SLS Population	2SLS Population	2SLS Land Suit.	2SLS Land Suit.
AvgLogDistance _{not}	1.3051** [0.563]	1.2453* [0.709]			1.8045** [0.711]		1.5467* [0.842]	
AvgLogDistance _{adj}			0.7715* [0.429]	0.6453 [0.462]		1.4708*** [0.563]		1.2997* [0.701]
House Price Ratio	-0.0053 [0.064]	-0.0090 [0.060]	-0.0223 [0.067]	-0.0452 [0.062]	0.0178 [0.060]	0.0258 [0.060]	0.0055 [0.064]	0.0148 [0.067]
Observations	47	47	47	47	47	47	47	47
AR p-value	-	-	-	-	0.013	0.013	0.138	0.138
R-squared	0.629	0.639	0.524	0.556	0.627	0.639	0.635	0.643

Robust standard errors in brackets. Dependent variable: Log of Governor Salary (2008). Independent variables as of 2000 (GCISC: avg. 1920-2000). Control variables: Log Area and Log Maximum Distance (except AvgLogDistance_{adj} specifications), Log Income, Log Population, % College (all columns), Share of government employment, % Urban, Regional dummies (except for Columns (1) and (4)). IV: centroid AvgLogDistance_{not} of population. AR p-value: p-value from Anderson-Rubin (minimum distance) test. Sample excludes South Dakota, for which data on house prices is not available. *** p<0.01, ** p<0.05, * p<0.1.

Table OA9. Public Goods with adjusted Avg Log Distance

Dep. Var.:	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	2SLS	2SLS	2SLS
	PG Exp.	Oth. Exp.	PG Prov.	PG Exp.	Oth. Exp.	PG Prov.
AvgLogDistance _{adj}	-0.3078** [0.117]	0.1970** [0.085]	-2.9495* [1.566]	-0.4447** [0.176]	0.2658** [0.122]	-0.3261 [2.029]
Observations	48	48	48	48	48	48
AR p-value	-	-	-	0.021	0.041	0.874
R-squared	0.381	0.510	0.837	0.436	0.579	0.870

Robust standard errors in brackets. Dependent variables: PG Exp. (Public Good Expenditures) = Share of state expenditures on education, public welfare, health, and hospitals in 2008; Oth. Exp. (Other Expenditures) = Share of state expenditures on government administration, interest on debt, and "other" in 2008; PG Prov. (Public Good Provision) = First principal component of "Smart State" Index (2005), % of population with health insurance (2008-9), and log of hospital beds per capita (2009). Independent variables: AvgLogDistance_{adj} avg 1920-2000. Control variables: Log Area and Log Maximum Distance, Log Income, Log Population, % College, Share of government employment, Racial dissimilarity, % Urban, Regional dummies (all specifications). IV: centroid AvgLogDistance_{not} of population. AR p-value: p-value from Anderson-Rubin (minimum distance) test. *** p<0.01, ** p<0.05, * p<0.1.