

Online appendix: The political geography of cities

Capital cities, economic fundamentals, and urban growth

Richard Bluhm ^{*} Christian Lessmann [†] Paul Schaudt [‡]

June 2025

Contents

A Data Appendix	ii
A-1 Remotely-sensed data	ii
A-2 Cross-country data	iii
A-3 DHS data	iv
A-4 Investment data	vi
A-5 Descriptive statistics	vii
B Tracking capital cities and subnational units	x
B-1 Administrative units over time	x
B-2 Capital cities over time	xii
C Capital locations	xiv
D Selection issues: City detection	xvi
E Additional results	xvii
E-1 Additional figures	xvii
E-2 Additional tables	xxviii
F Former capitals and rump capitals	xxxiv
F-1 Former capitals	xxxiv
F-2 “Rump” capitals	xxxvi

^{*}University of Stuttgart, Institute of Economics and Law, e-mail: richard.bluhm@ivr.uni-stuttgart.de

[†]Technische Universität Dresden, Ifo Institute for Economic Research & CESifo Munich, e-mail: christian.lessmann@tu-dresden.de

[‡]University of St.Gallen, Department of Economics, KU Leuven, e-mail: paul.schaudt@unisg.ch

A. Data Appendix

A-1. Remotely-sensed data

Light density: We calculate our light density measures by taking the average value per pixel within the year (averaging over multiple satellites) and then summing the average pixel values across our city shapes before dividing by the city area. The baseline luminosity data is based on the raw data of the Version 4 DMSP-OLS (stable light) product ([U.S. Air Force Weather Agency, 1992-2013](#)). The bottom correction is implemented following [Storeygard \(2016\)](#) and the top coding correction is based on the data provided by [Bluhm and Krause \(2022, 2025\)](#).

Population density: is calculated by first taking the sum of population based on the Global Human Settlement Layer population raster ([Schiavina, Freire and MacManus, 2019](#)) and then dividing by the area of our cities.

Ruggedness: We calculate average ruggedness within 25km of our cities by taking the average pixel value of the terrain ruggedness index computed using elevation data from the SRTM Version 4.1 raster ([Jarvis et al., 2008](#)).

Malaria suitability: Malaria Ecology Index from [Kiszewski et al. \(2004a,b\)](#).

Market access: Own calculation based on the GHSL population raster ([Schiavina, Freire and MacManus, 2019](#)) and our city shapes. See main text for details.

River within 25km: We generate a dummy for all cities located within 25km of a river, based on our city coordinates and the river shapes from Natural Earth 10m rivers and lakes centerlines, version 4.1.0 ([Natural Earth, 2018](#)).

Lake within 25km: We generate a dummy for all cities located within 25km of a lake, based on our city coordinates and the lake centerlines from Natural Earth 10m rivers and lakes centerlines, version 4.1.0 ([Natural Earth, 2018](#)).

Port within 25km: We generate a dummy for all cities located within 25km of a port, based on our city coordinates and port locations obtained from the World Port Index 2010 ([National Geospatial-Intelligence Agency, 2010](#)).

Coast within 25km: We generate a dummy for all cities located within 25km of the coast, based on our city coordinates and the Natural Earth 10m coastline vectors, version 4.1.0 ([Natural Earth, 2018](#)).

Precipitation: Average precipitation is calculated within 25km buffers of our city coordinates; we average yearly values from Jan 1990 to Dec 2014 from the monthly totals. The precipitation data are obtained from the Center for Climatic Research, Department of Geography, University of Delaware and NOAA, version 4.01 ([University of Delaware, 2015](#)).

Elevation: Average elevation within a 25km buffer of the city is calculated based on the SRTM Version 4.1 raster ([Jarvis et al., 2008](#)).

Temperature: Average temperature is calculated for 25km buffers around our cities. We use the average temperature from Jan 1990 to Dec 2014 from the monthly totals as inputs, which are obtained from the Center for Climatic Research, Department of Geography, University of Delaware and NOAA, version 4.01 ([University of Delaware, 2015](#)).

Wheat suitability: Average wheat suitability is calculated for 25km buffers around our city coordinates. The wheat suitability values are obtained from the FAO GAEZ v3.0 Agro-climatically attainable yield for intermediate input level rain-fed wheat for baseline period 1961–1990 at a resolution of five arc minutes ([IIASA/FAO, 2012](#)).

Built-up: We calculate built-up and vegetation measures using the entire archive of Landsat images from 1987 until 2018, available at a resolution of 30m from Landsat 5 and 7 in *Google Earth Engine* ([U.S. Geological Survey, 1984–2012, 1999–2021](#)). The measures are based on spectral bands, denoted by ρ_x , and calculated as follows: $UI = (\rho_{SWIR2} - \rho_{NIR}) / ((\rho_{SWIR2} + \rho_{NIR}))$. Before calculating the UI, we create a cloud-free annual composite of the Landsat input.

ELF: Ethno-linguistic fractionalization of subnational regions. Calculated following [Eberle et al. \(2020\)](#) using data from Ethnologue, 17th edition ([Lewis, 2009](#)).

A-2. Cross-country data

Democracy: Indicator variable equal to one for countries with a polity V score ≥ 6 and zero otherwise ([Marshall and Gurr, 2020](#)).

Early (late) urbanizer: Indicator variable equal to one if a country is classified as an early (or late) urbanizer in 1950 by [Henderson et al. \(2017, 2018\)](#).

Federal: Indicator variable equal to one if a country is classified as federal by [Treisman \(2008\)](#).

Fiscal decentralization: Subnational revenue share as a percentage of GDP averaged between 1994 and 2000 or subnational government employment share in 1997 (both from Treisman, 2008).

A-3. DHS data

DHS wealth index: is taken directly from the DHS surveys (v190). The DHS (ICF, 1986-2019) describes their wealth index as: “... a composite measure of a household’s cumulative living standard. The wealth index is calculated using easy-to-collect data on a household’s ownership of selected assets, such as televisions and bicycles; materials used for housing construction; and types of water access and sanitation facilities” (see <https://www.dhsprogram.com/topics/wealth-index/wealth-index-construction.cfm>). Note that the specific assets considered are dependent on the country.

Electricity indicator: is an indicator variable for the availability of electricity in the household (V119).

Save water indicator: is an indicator variable set to unity if the respondent household has access to either protected wells or springs, boreholes, packaged water, and rainwater (v113) (see Henderson et al., 2020, for a similar classification).

Improved sanitation indicator: is an indicator variable equaling unity if the respondent household has access to either shared or non-shared facilities that flush/pour to piped sewer systems, septic tank, pit latrine; ventilated improved pit latrine, pit latrine with slab and composting toilets, as well as flushing to unknown locations (v116). We follow Henderson et al. (2020) and use the DHS-WHO joint monitoring program definitions.

At least eight years of schooling indicator: Is a dummy variable unity if the respondent has completed eight or more years of schooling (based on V107) and zero otherwise. It is only defined for respondents who are at least 16 years old.

Infant mortality: is defined as the probability of dying before the first birthday. The corresponding rate is normalized as a ratio per 1000 live births. The variable is constructed based on the “age at death” responses about the children of female respondents (variables b13-1 to b13-20). As is standard in the literature, we use the individual-child-level data to compute this measure and multiply the resulting dummy by 1000 to estimate a rate (“per thousand births”).

Log household size: is the log of the number of household members (v136).

Female head of household indicator: defined according to the reported gender of the household head (v151).

Log head of household age: is the log of age (in years) of the household head (v152).

Household head completed primary education indicator: is calculated based on the educational achievement variable (v149) if the respondent is the household head (v150). The indicator is unity if the household head has completed primary education or started but not finished secondary education (v149). The indicator is zero otherwise.

Household head completed secondary education indicator: is calculated based on the educational achievement variable (v149) if the respondent is the household head (v150). The indicator is unity if the household head has completed secondary education (v149). The indicator is zero otherwise.

Household head completed higher education indicator: is calculated based on the educational achievement variable (v149) if the respondent is the household head (v150). The indicator is unity if the household head has completed higher education (v149). The indicator is zero otherwise.

Age in years of the respondent (v012 and mv012) in the DHS. Also included as a squared term.

Female: is an indicator that is unity for all respondents in the IR dataset of the DHS and zero for all respondents in the MR dataset of the DHS.

Sex: indicates if the respondent child is female (b4-01 to b4-20).

Multiple births: indicates if a respondent's child was born as a twin or multiple (b0-01 to b0-20).

Period of birth indicator: Indicator for the period of the birth of the reported children (by decade, i.e., 1990s).

A-4. Investment data

Development aid (World Bank): Geocoded development aid provided by the World Bank is obtained from [AidData \(2017\)](#). This geocoded dataset includes all projects approved from 1995-2014 in the World Bank IBRD/IDA lending lines. It tracks more than \$630 billion in commitments for 5,684 projects across 61,243 locations. We construct several aid variables in the following sectors: Education; Health; Water Supply & Sanitation; Government & Civil Society; Other Social Infrastructure & Services; Economic Infrastructure & Services; Agriculture, Forestry & Fishing; Industry, Mining & Construction; and Environmental Protection. They correspond to the broadest classification of the project types provided by the World Bank. Note that any project can have multiple (up to 5) classifications. In such cases, the same project appears under multiple headings.

Development aid (China): Geocoded development aid-like financial flows for China are obtained from [Bluhm et al. \(2024, 2025\)](#). This dataset geolocates Chinese Government-financed projects implemented between 2000-2014. It captures 3,485 projects worth \$273.6 billion in total official financing. The dataset includes both Chinese aid and non-concessional official financing. We construct several aid variables in the following sectors: Education; Health; Water Supply & Sanitation; Government & Civil Society; Other Social Infrastructure & Services; Economic Infrastructure & Services; Agriculture, Forestry & Fishing; Industry, Mining & Construction; and Environmental Protection. They correspond to the broadest classification of the project types provided by the World Bank. Note that any project can have multiple (up to 5) classifications. In such cases, the same project appears under multiple headings.

FDI: The raw data for our FDI outcomes (dummy, log investment value, and log estimated jobs) comes from the fDi Markets database ([Financial Times Ltd., 2020](#)). The database contains detailed information on FDI projects worldwide from 2003 until 2018, including information about the investing company, the origin country in which the company is based, and much more. Essential for us is that the database has the estimated jobs created, the value spent, the host city name, and if the project is a greenfield investment. We geocoded the projects using the same OSM algorithm we employed for the location of the capital cities using the host city information. In the next step, we match the FDI to our cities if the project's host city (which does not need to meet any population threshold) falls within a 10km buffer of our detected cities. Finally, we summarize the invested dollar value and the estimated jobs by the host city location and take their logs. Note that we only gathered data for our reformed areas since the terms of use allow us to use 10% of their sample. The data is then aggregated to the NAICS 2-digit level. The 2-digit NAICS classifications we use are: Agriculture, Forestry, Fishing and Hunting;

Mining, Quarrying, and Oil and Gas Extraction; Utilities; Construction; Manufacturing; Wholesale Trade; Retail Trade; Transportation and Warehousing; Information; Finance and Insurance; Real Estate and Rental and Leasing; Professional, Scientific, and Technical Services; Administrative and Support and Waste Management and Remediation Services; Educational Services; Health Care and Social Assistance; Arts, Entertainment, and Recreation; Accommodation and Food Services; and Public Administration.

A-5. Descriptive statistics

LIST A-1 Countries in sample

Afghanistan (L,A,U); Albania (L,A,U); Algeria (L,A,U); Angola (L,A,U); Argentina (E,D,F); Australia (E,D,F); Austria (E,D,F); Bangladesh (L,A,U); Belarus (L,,U); Belgium (E,D,F); Benin (L,A,U); Bolivia (L,D,U); Brazil (L,D,F); Bulgaria (L,D,U); Burkina Faso (L,A,U); Burundi (L,A,U); Cambodia (L,A,U); Cameroon (L,A,U); Canada (E,D,F); Central African Republic (L,A,U); Chad (L,A,U); Chile (E,D,U); China (L,A,U); Colombia (L,D,U); Congo (L,A,.); Costa Rica (L,D,U); Cuba (E,A,U); Czech Republic (E,D,U); Côte d'Ivoire (L,A,U); Democratic Republic of the Congo (L,A,U); Denmark (E,D,U); Dominican Republic (L,A,.); Ecuador (L,D,U); Egypt (L,A,U); Eritrea (L,A,U); Estonia (E,D,U); Ethiopia (L,A,F); Finland (E,D,U); France (E,D,U); Georgia (E,A,U); Germany (E,D,F); Ghana (L,A,U); Greece (E,D,U); Guatemala (L,A,U); Guinea (L,A,U); Haiti (L,A,U); Honduras (L,D,U); Hungary (E,D,U); India (L,D,F); Indonesia (L,A,U); Iran (L,A,U); Iraq (L,A,U); Ireland (E,D,U); Italy (E,D,U); Japan (E,D,U); Kazakhstan (E,A,U); Kenya (L,A,U); Korea, North (L,A,U); Korea, South (L,D,U); Kyrgyzstan (L,A,U); Lao People's Democratic Republic (L,A,U); Latvia (E,D,U); Lesotho (L,A,U); Liberia (L,A,U); Madagascar (L,A,U); Malawi (L,A,U); Malaysia (L,A,F); Mali (L,A,U); Mauritania (L,A,U); Mexico (E,A,F); Moldova (L,A,U); Mongolia (L,A,U); Morocco (L,A,U); Mozambique (L,A,U); Myanmar (L,A,U); Nepal (L,A,U); Netherlands (E,D,U); New Zealand (E,D,U); Nicaragua (L,D,U); Niger (L,A,U); Nigeria (L,A,F); Norway (E,D,U); Oman (L,A,U); Pakistan (L,D,F); Panama (L,D,U); Papua New Guinea (L,A,U); Paraguay (L,A,U); Peru (E,D,U); Philippines (L,D,U); Poland (E,A,U); Portugal (L,D,U); Romania (L,A,U); Russian Federation (E,,F); Rwanda (L,A,U); Saudi Arabia (L,A,.); Senegal (L,A,U); Sierra Leone (L,A,U); Slovakia (L,D,U); Somalia (L,A,U); South Africa (E,A,U); Spain (E,D,F); Sri Lanka (L,A,U); Sudan (L,A,F); Sweden (E,D,U); Switzerland (E,D,F); Syrian Arab Republic (L,A,U); Taiwan (L,A,U); Tajikistan (L,A,U); Thailand (L,A,U); Togo (L,A,U); Tunisia (L,A,U); Turkey (L,D,U); Turkmenistan (E,A,U); U.K. of Great Britain and Northern Ireland (E,D,U); Uganda (L,A,U); Ukraine (L,,U); United Arab Emirates (E,A,F); United Republic of Tanzania (L,A,U); United States of America (E,D,F); Uruguay (E,D,U); Uzbekistan (L,A,U); Venezuela (E,D,F); Viet Nam (L,A,U); Yemen (L,A,U); Zambia (L,A,U); Zimbabwe (L,A,U);

Notes: The list depicts the countries covered in our study. The letter in parenthesis indicate to which cross-country classification with respect to early-late urbanizer (E/L), political system (autocracy (A)/democracy (D)) and federal (F) vs. unitary (U) country they are assigned.

LIST A-2
DHS survey sample

AGO (2006,2007,2011); ALB (2008); ARG (2008); BDI (2010,2011,2012,2013); BEN (1996,2001,2011,2012); BFA (1992,1993,1998,1999,2003,2010,2014); BOL (2008); BRA (2008); CAF (1994,1995); CIV (1994,1998,1999,2011,2012); CMR (1991,2004,2011); COD (2007,2013,2014); COL (2010); DOM (2007,2013); EGY (1992,1995,2000,2003,2005,2008,2014); GHA (1993,1994,1998,1999,2003,2008,2013,2014); GIN (1999,2005,2012); HND (2011); HTI (2000,2006,2007,2012); IDN (2003); KEN (2003,2008,2009,2014); KGZ (2012); LBR (2006,2007,2008,2009,2011,2013); LSO (2004,2005,2009,2010,2014); MAR (2003); MDA (2005); MDG (1997,2008,2009,2011,2013); MLI (1995,1996,2001,2006,2012,2013); MOZ (2009,2011); MWI (2000,2004,2005,2010,2012,2014); NER (1992,1998,2012); NGA (1990,2003,2008,2010,2013); PAK (2006); PER (2000,2004,2009); PHL (2003,2008); RWA (2005,2008,2010,2011,2014); SEN (1992,1993,1997,2005,2008,2009,2010,2011,2012); SLE (2008,2013); TCD (2014); TGO (1998,2013,2014); TJK (2012); TZA (1999,2003,2004,2007,2008,2009,2010,2011,2012); UGA (2000,2001,2006,2008,2009,2011,2014); ZMB (2007,2013,2014); ZWE (1999,2005,2006,2010,2011)

Notes: The list depicts the countries and survey years for which we match DHS clusters to our cities.

TABLE A-1
Summary statistics: Fundamentals

	Mean	SD	Min	Max	N
<i>Panel A. Cities (all)</i>					
Log light density	2.96	1.29	1.26	7.65	524,867
Log population 1990	10.84	0.88	9.25	17.06	524,867
Ruggedness	14.45	15.39	0.46	120.22	524,867
Malaria suitability	0.01	0.02	0.00	0.17	524,867
Market access (pop 1990 based)	10.29	1.32	3.46	13.55	524,867
River within 25km	0.35	0.48	0.00	1.00	524,867
Lake within 25km	0.02	0.14	0.00	1.00	524,867
Port within 25km	0.10	0.30	0.00	1.00	524,867
Coast within 25km	0.16	0.37	0.00	1.00	524,867
Distance to coast	379.98	377.04	2.57	2,504.02	524,867
Average precipitation	9.20	5.39	0.05	81.39	524,867
Average elevation	460.49	576.87	-27.67	5,023.05	524,867
Average temperature	19.80	6.96	-7.59	32.09	524,867
Wheat suitability	2,309.79	2,076.13	0.00	7,252.34	524,867
<i>Panel B. Cities (within reformed areas)</i>					
Log light density	2.36	1.13	1.26	7.51	186,019
Log population 1990	10.79	0.82	9.90	16.80	186,019
Ruggedness	13.40	15.52	0.53	110.43	186,019
Malaria suitability	0.01	0.03	0.00	0.16	186,019
Market access (pop 1990 based)	10.55	1.36	3.48	13.55	186,019
River within 25km	0.38	0.49	0.00	1.00	186,019
Lake within 25km	0.02	0.13	0.00	1.00	186,019
Port within 25km	0.06	0.23	0.00	1.00	186,019
Coast within 25km	0.10	0.30	0.00	1.00	186,019
Distance to coast	492.10	395.75	2.57	2,442.78	186,019
Average precipitation	9.62	4.56	0.05	75.78	186,019
Average elevation	491.80	603.09	-25.44	5,023.05	186,019
Average temperature	22.00	6.00	-5.49	30.60	186,019
Wheat suitability	1,991.65	1,768.36	0.00	6,886.30	186,019

Notes: Panel A of the table reports the summary statistics for our sample of all cities. Panel B reports summary statistics for the sample of cities located within reformed regions.

B. Tracking capital cities and subnational units

We separately track changes in the geography of subnational units and capitals over time and cross-reference both results at the end to minimize the scope for error. We start cataloging subnational capitals using two of the most comprehensive databases available at the time (i.e., the Statoids, [Law, 2010, 2020](#) and the City Population database, [Brinkhoff, 2020](#)). We use the Global Administrative Unit Layers (GAUL) vector data ([FAO, 2015](#)) as a baseline to track subnational units over time. It only records the spatial extent of administrative units but contains no information on their capitals. The three databases have varying temporal coverage. The Statoids data often track capitals and subnational units back to the founding of a country and are usually accurate (up until 2013/2014) but lack any spatial information. The City Population and GAUL data cover short periods, from 1998 until 2020 and 1990 until 2014, respectively.

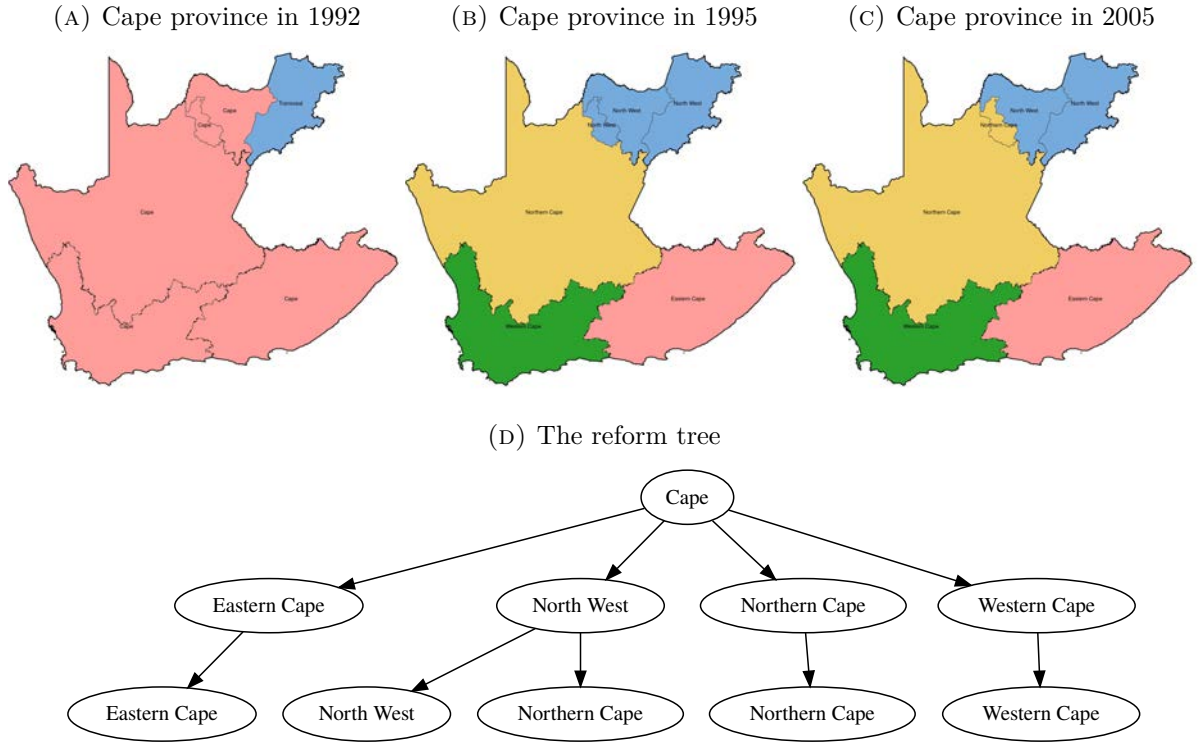
B-1. Administrative units over time

We begin by backing out a reform tree from the GAUL data using a simple spatial algorithm. We create the spatial intersection of the two vector data sets for any pair of two years. This creates new areas or affiliations whenever a border is moved, deleted, or created. We then cycle forward by intersecting the result of the previous intersection with the next year of official data and so on. During each iteration, we also record the current region identifier and add it to an identification string that contains 25 (i.e., 2014-1990) identifiers in the last year.

We obtain two data sets in this manner. The first is a spatial data set of micro-regions, which in the final year contains the smallest spatial unit whose borders were *not reformed* in any preceding years. We call this unit a splinter. The second is a kind of evolutionary tree for each contemporary splinter, summarizing its entire history of regional affiliations and its respective administrative center back to 1990. Note that splinters only result from border reforms that cut across borders from the previous year. Abolishing a border does not create new splinters but changes the region’s identity. The combination of the spatial splinter data set and the reform tree identifies all administrative reforms in a general and spatially consistent manner. Moreover, the reform tree lets us quickly compare the results to other non-spatial data sources, such as City Population or Statoids.

[Figure B-1](#) illustrates the two data sets created by this process. It shows the history of reforms in Cape Province in South Africa from 1992 onward (the green area in panel A). Cape Province was split into four new regions in 1994 (panel B). Three successor provinces are congruent with the former province. The fourth region (North-West) includes some areas of the former Transvaal (the neighboring province to the northeast, marked in yellow in panel A). Furthermore, a part of the North-West was assigned to the Northern Cape in 2005 (see the yellow area in panel B, which turns purple in panel C). As a result, all

FIGURE B-1
Reform history of Cape Province, South Africa



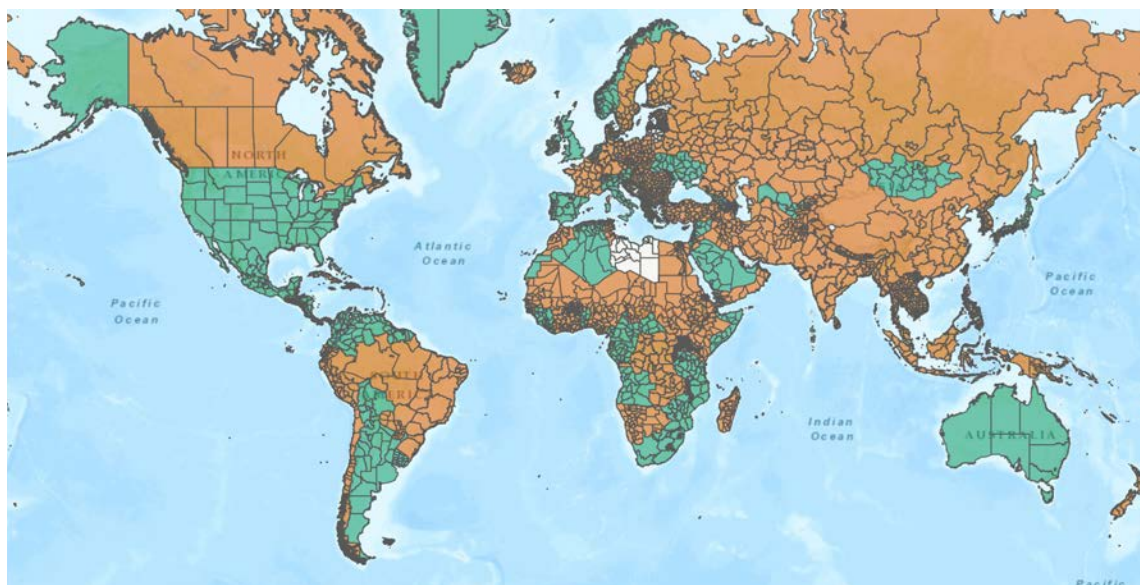
Notes: Panels A to C illustrate initial and successor regions of the Cape Province in South Africa. Panel D illustrates the evolutionary tree for the splinters that comprised Cape Province, South Africa. The last level represents the situation after the 2005 reform.

splinters of Cape Province are affiliated with at least two different administrative regions over this period (panel D).

Next, we compare the resulting reform tree with Statoids and City Population to document discrepancies (of which there are many). First, the different sources do not always agree on what unit constitutes the first-order administrative level. GAUL sometimes contains macro-regions, which have no political function and are easily identified using other data sources. Whenever we detect a case where GAUL disagrees with other sources or misses a reform entirely, we collect additional spatial data for these regions. From 2000 onward, AidData’s GeoBoundaries database ([Runfola et al., 2020a,b](#)) and GADM ([GADM, 2018–2022](#)) provide high-quality data, although neither is without error. Data from the early 1990s is more challenging and sometimes requires us to digitize offline maps. In rare cases, we recovered the correct shapes by merging regions. Uganda, for example, consecutively split its larger regions into smaller units so that the most recent vector data was sufficient to reconstruct an administrative map for each year. In summary, we found that around 40% of all countries in GAUL had missing or incomplete data from 1990 to 2014 (see [Figure B-2](#) for an illustration).

Finally, we extended the corrected sample to the period from 1987 to 2018. Extending

FIGURE B-2
Corrections made in GAUL data from 1990-2014



Notes: The figure plots the corrected GAUL countries. Countries correct in GAUL are green, those we fixed are orange, and those we could not fix because we lacked data for one or more years are transparent. The topographic base map is provided by Esri, using source material from Esri, TomTom, FAO, NOAA, and USGS.

the sample from 2014 onward is straightforward since many statistical offices upload official vector files, and we could use a newer version of AidData’s GeoBoundaries database and GADM to fill in the gaps. Extending backward from 1989 to 1987 was more cumbersome. We relied on older vector data sourced individually for each country and early editions of the Atlas Britannica.

B-2. Capital cities over time

This workflow starts with two lists of capital city years obtained from Statoids and City Population. The lists were provided to two trained coders, who independently cross-referenced and checked each entry for inconsistencies. The coders resolved any differences using additional data sources such as the CIA Factbook, Wikipedia, or secondary literature. A third coder compared these two sets of results and resolved differences, if there were any, in a final arbitration process.

Next, the two expert coders geocoded the locations of all administrative cities, i.e., the longitude and latitude of the city centroids, using OpenStreetMap’s (OSM) Nominatim API and Google Maps’ geocoding API. OSM and Google accurately identified the coordinates of most cities without any problems. Unfortunately, not all cities are coded automatically, and some cities are coded incorrectly. In those cases, we manually identified the city coordinates. In Uganda, for example, we had to geocode around 60 out of 136 administrative centers manually. The manual coding included another arbitration

layer in case of disagreements.

Finally, we merge the remotely-sensed universe of city cores and envelopes with the coordinates of administrative cities. We consider exact matches in all cases where the centroid of a capital city falls within 3 km of a city core or envelope. We proceed by matching names in the few instances where no town is within this distance of an administrative center. Any cluster within 50 km of a capital city with almost the same name, defined as a Levenshtein edit distance of less than 3, is considered a match.

C. Capital locations

We now examine the political geography determinants of capital locations within regions and provide some descriptive statistics of which cities will likely become capitals within a new administrative region.

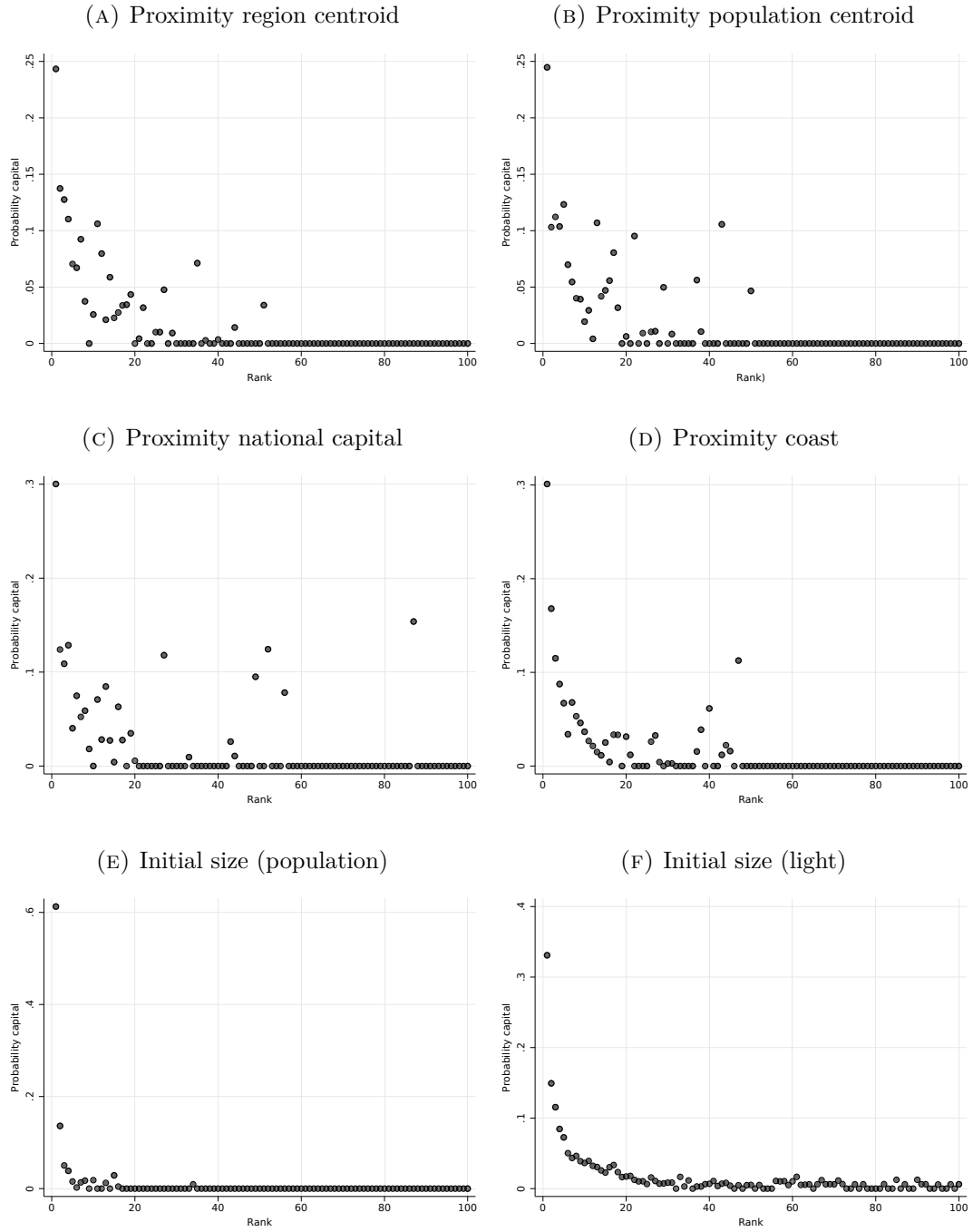
We take our inspiration from [Bai and Jia \(2023\)](#), who propose that central government planners in historical China faced a trade-off when determining the location of regional capital cities. Being close to citizens implies that the administrative location can efficiently exercise control (levy taxes and provide services at a low cost). Proximity to the national capital, in turn, makes the local administration more accountable to the national government and minimizes the cost of delivering local taxes to the central government (for similar arguments see [Bardhan and Mookherjee, 2000](#); [Campante and Do, 2014](#)). The optimal solution to this problem minimizes a location’s ‘hierarchical distance:’ the distance between all citizens within a province and the national capital (with some weight on either objective). Of course, other factors are likely to play a role in these location decisions today, so we consider a range of additional variables, from proximity to the coast to the size distribution of cities in the initial region.

Panels A to C of [Figure C-1](#) provide some evidence that hierarchical distance also matters in our global sample of contemporary capital city reforms. We rank cities within regions according to their distance to the region centroid in panel A, their distance to the population-weighted centroid in panel B, or their distance to the national capital in panel C. In all three cases, cities that occupy lower ranks (are closer) are considerably more likely to become a capital when a region is split. Panel D adds the proximity to the coast as a proxy for the external trade orientation and documents a similar pattern. We find a few outliers where high ranks have a high probability of becoming a capital (due to a few regions in South Asia with relatively “remote” capitals).

Finally, we examine the initial size, either based on population or light density, as a predictor of gaining the status of a regional capital.¹ Panel E shows a strong relationship between the initial size of a city and the probability of becoming the region’s capital. The largest city in a region is also the region’s capital in almost 61.2% of cases, the second-largest city in around 13.6% of cases, while the chances of being a capital for the third and fourth-largest cities are in the single digits. Cities that rank five or higher have an average probability below 1%. The relationship weakens if we rank by initial light, where the decline in the probability is smoother, and the largest city becomes the capital in only 33.9% of cases (panel F).

¹Note that the largest city does not minimize the distance to all citizens by definition, although there is a high correlation of 0.63.

FIGURE C-1
Determinants of capital locations within regions: City ranks



Notes: This figure shows scatter plots of the average probability that a city becomes a capital over the distribution of city characteristics along various dimensions. Panel A ranks cities in terms of proximity to the regional centroid. Panel B ranks cities according to the proximity to the population-weighted centroid of a region. Panel C uses the proximity to the national capital, and panel D uses the proximity to the coast. Panels E and F rank cities by population or light density based on their initial size.

D. Selection issues: City detection

The main text focuses on the cities detected in 1990. We then analyze changes in the core and the larger agglomeration, including new developments in these cities from 1990 until 2015. Determining the sample of cities avoids a sample selection issue that we illustrate in more detail in this appendix.

The selection effect arises since the status of a city as a subnational capital also influences the likelihood of detection in 2015. Our main result is that cities grow faster once they gain capital city status. Recall that we only observe urban boundaries in two periods (1990 and 2015). If a small city becomes a subnational capital in the interim and grows faster, it is more likely to cross our detection thresholds and be classified as a city in 2015. Suppose we track light density (or other outcomes) in these cities over the entire period, even though they are only detected later. In that case, we include this dynamic selection bias in our estimation and, with it, the possibility of pre-trends.

We design a simple test to illustrate this selection effect. We regress the change in detection status from 1990 to 2015 on the share of years a city is a subnational capital during the same period. The change in status is the first difference of a binary variable indicating whether we detected a city in 1990 or 2015. [Table D-1](#) reports the results from several specifications, where we incrementally add country and initial-region fixed effects for our two samples. Columns 1 to 3 show that a city that becomes a capital halfway through the period from 1990 to 2015 has a 7.4 to 11.8 percentage points higher probability of being detected in 2015. The estimated effect sizes are smaller for the sample of cities in reformed regions, but the overall pattern remains the same. Obtaining the status as a first-order capital during the sample significantly increases the likelihood of detection in 2015.

TABLE D-1
City detection probability

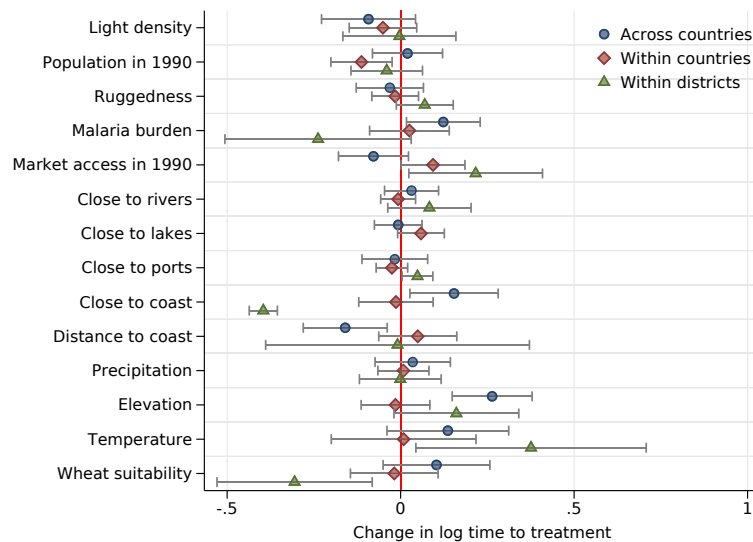
	<i>Dependent Variable: $\Delta \text{DETECTED}_{ci}$</i>					
	All Cities			Reformed Regions		
	(1)	(2)	(3)	(4)	(5)	(6)
CAPITAL	0.1473 (0.0384)	0.2177 (0.0386)	0.2349 (0.0411)	0.1283 (0.0522)	0.1431 (0.0480)	0.1680 (0.0471)
Fundamentals	✓	✓	✓	✓	✓	✓
Country FE	–	✓	✓	–	✓	✓
Initial-Region FE	–	–	✓	–	–	✓
City-unions	28009	28009	28009	10904	10904	10904

Notes: The table reports results from a regression of the change in detection status of a city between 1990 and 2015 on the fraction of years in which a city is a capital. Standard errors clustered on initial regions are provided in parentheses.

E. Additional results

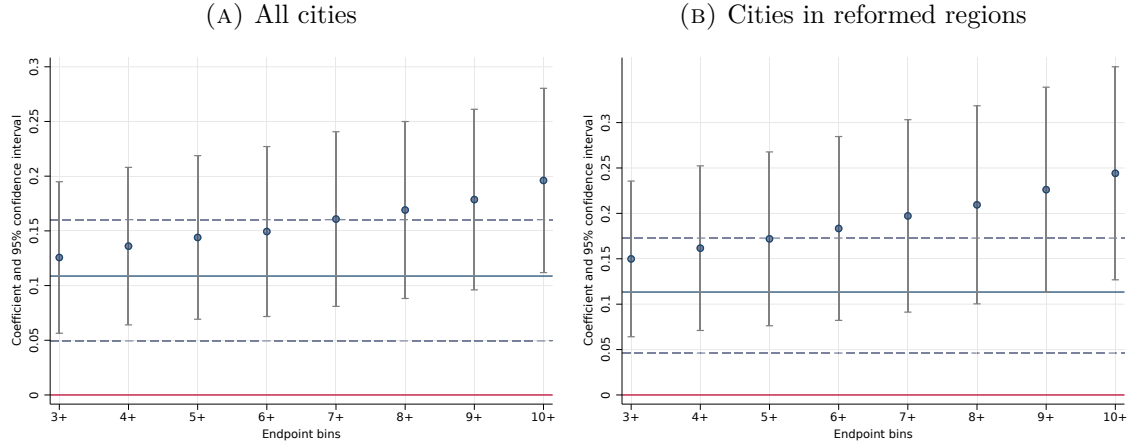
E-1. Additional figures

FIGURE E-1
Time to treatment



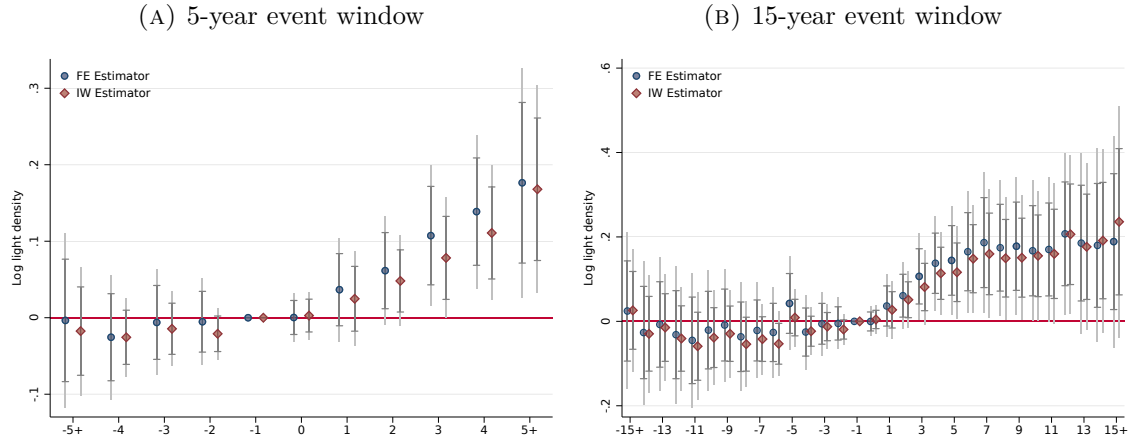
Notes: The figure illustrates results from cross-sectional regressions of the time to treatment (in logs plus one) on initial city characteristics. The regressions were run three times, once without fixed effects, once with country fixed effects, and once with initial region fixed effects. The coefficients are standardized beta coefficients. Some coefficients are omitted in the specification with initial region fixed effects due to a lack of within-region variation. 95% confidence intervals clustered on initial regions are indicated by the error bars.

FIGURE E-2
Endpoint binning and medium-run effect size: Event-study estimates



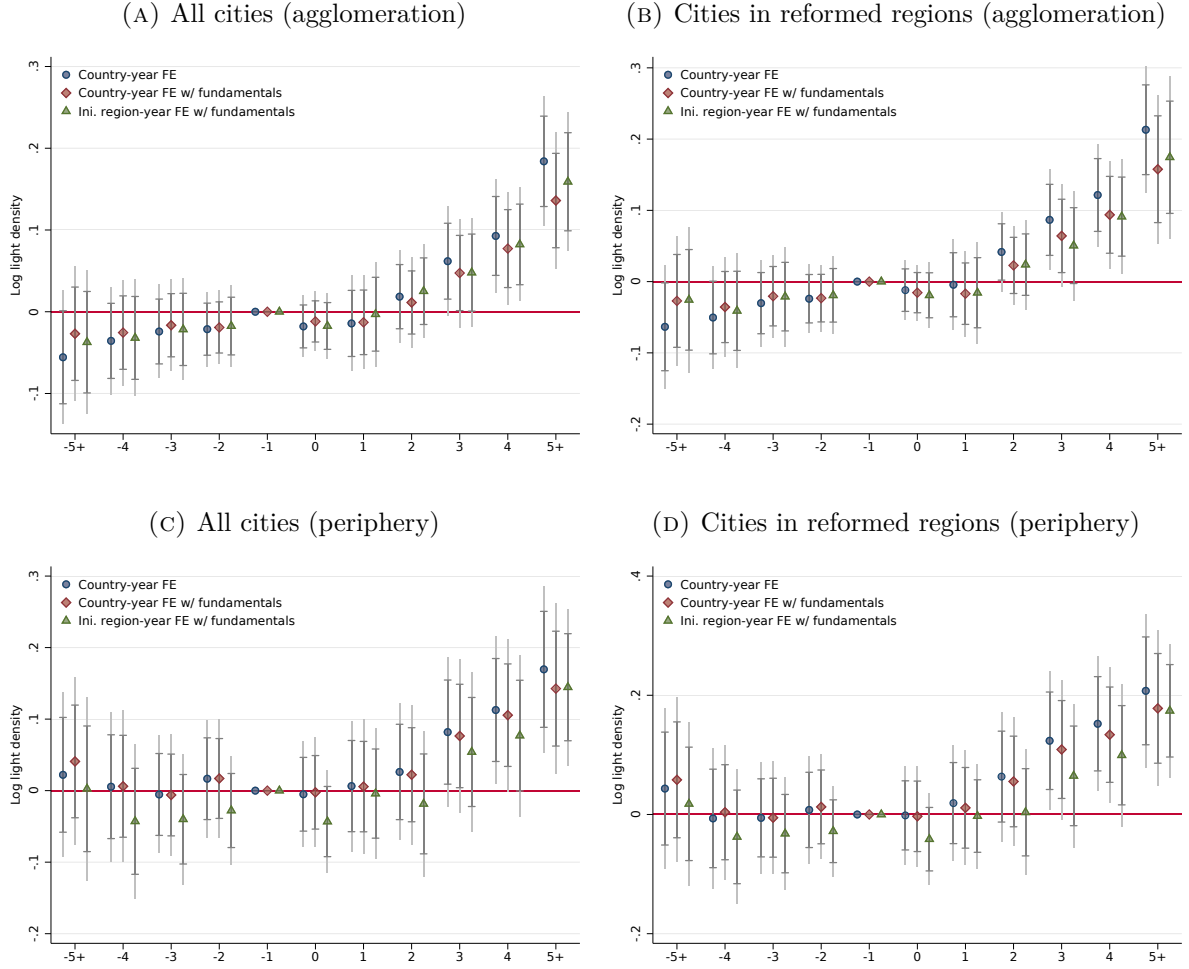
Notes: The figure shows point coefficients and 95% confidence intervals of the endpoint bins estimated in several event studies with varying window sizes. The underlying event studies use five pre-treatment periods and extend the event window from 3 (or more) to 10 (or more) periods. The effect in the last pre-period is normalized to zero. Panel A is based on column 3, and panel B is based on column 6 of [Table E-2](#). The blue line indicates the difference-in-differences estimate corresponding to each panel, and the dashed blue lines provide the 95% confidence intervals of these estimates.

FIGURE E-3
TWFE versus IW estimator of dynamic treatment effects



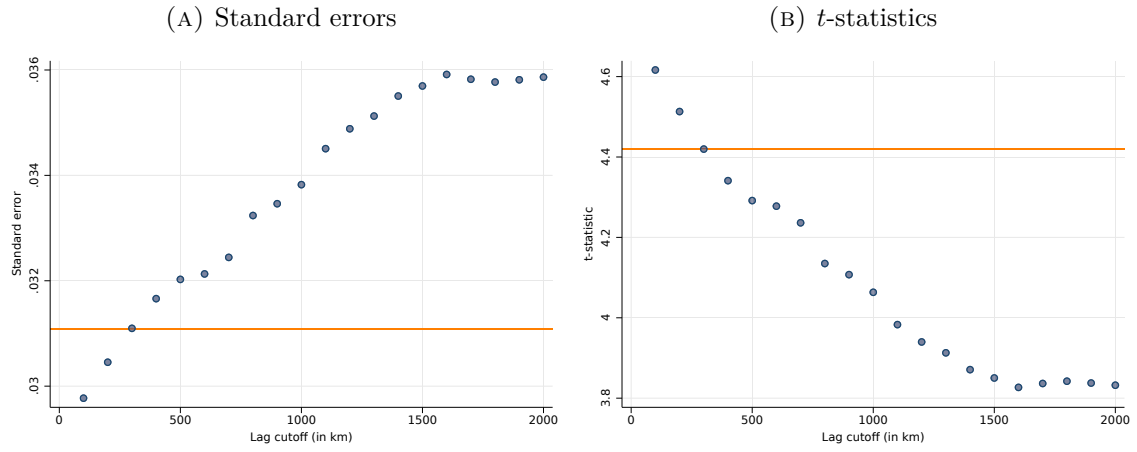
Notes: The figure illustrates event-study results from fixed effects regressions of the log of light intensity per square kilometer on a binned sequence of treatment change dummies, city fixed effects, initial-region-by-year fixed effects, time-varying locational fundamentals for a panel that is balanced in calendar time. Circles represent point estimates from two-way fixed effects estimation (TWFE). Diamonds represent point estimates from interaction-weighted (IW) estimation (see [Sun and Abraham, 2021](#)). Panel A shows estimates of a five-year event window. Panel A shows estimates of a 15-year event window. The gray error bars provide 95% confidence intervals based on standard errors clustered on initial regions. The whiskers indicate uniform 95% sup- t confidence bands computed using the plug-in method ([Montiel Olea and Plagborg-Møller, 2019](#)).

FIGURE E-4
Agglomerations: Event-study estimates



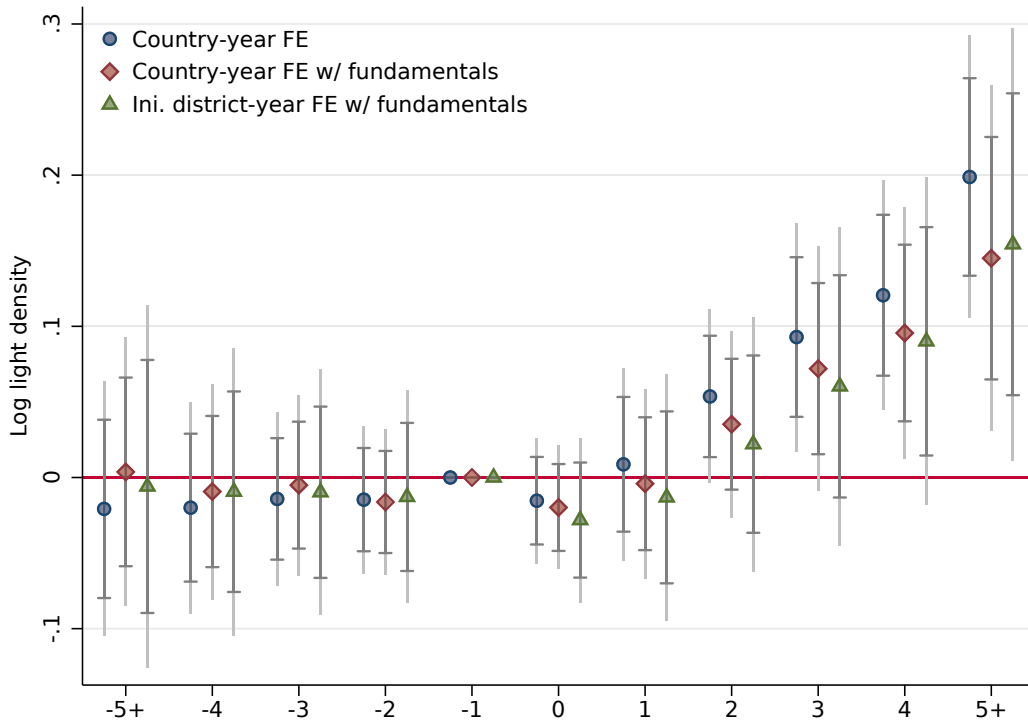
Notes: The figure reports event-study estimates corresponding to the difference-in-differences results presented in Table III. The upper panels report results for the larger agglomeration (envelope). The lower panels report results for the periphery (new parts added after 1990). Panels A and C show estimates for all cities. Panels B and D show estimates for cities in reformed regions. Circles represent point estimates from a regression with city and country-year fixed effects, diamonds represent specifications with additional controls for locational fundamentals, and triangles represent specifications with initial-region-by-year fixed effects. All regressions include city fixed effects. The gray error bars provide 95% confidence intervals based on standard errors clustered on initial regions. The whiskers indicate uniform 95% sup- t confidence bands computed using the plug-in method (Montiel Olea and Plagborg-Møller, 2019).

FIGURE E-5
Accounting for spatial autocorrelation



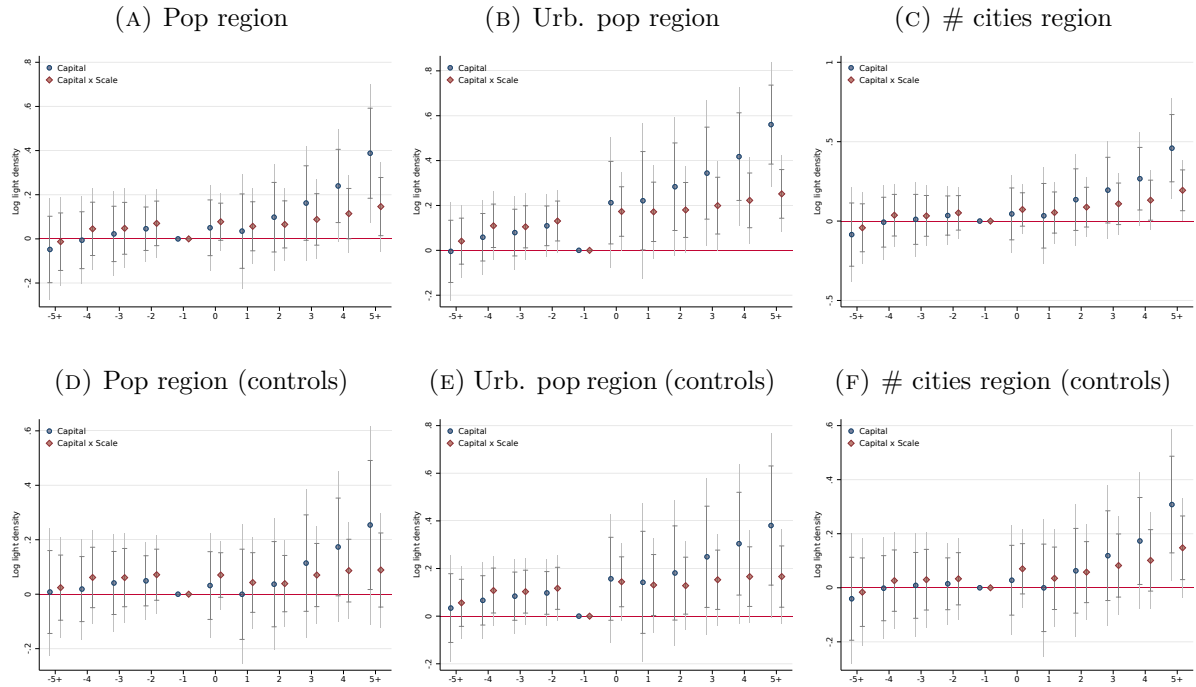
Notes: The figure illustrates results from varying the spatial lag cutoff when estimating standard errors, which allow for cross-sectional dependence. All results are based on a variant of column 6 in [Table E-2](#) where we restrict the sample to reformed areas and include city fixed effects and initial-region fixed effects. Here, we omit the time-varying effects of the fundamentals for computational reasons (to reduce the size of the regressor matrix). The estimated effect in this specification is 0.1427 with a standard error of 0.0316. All Conley errors are estimated with a uniform kernel and a time-series HAC with a cutoff of 1,000 years to allow for arbitrary dependence over time. Panel A shows estimates of the resulting standard errors, with the original error clustered on initial regions highlighted in orange. Panel B shows estimates of the resulting t -statistics, with the original t -statistic clustered on initial regions highlighted in orange.

FIGURE E-6
Event study within reformed child regions



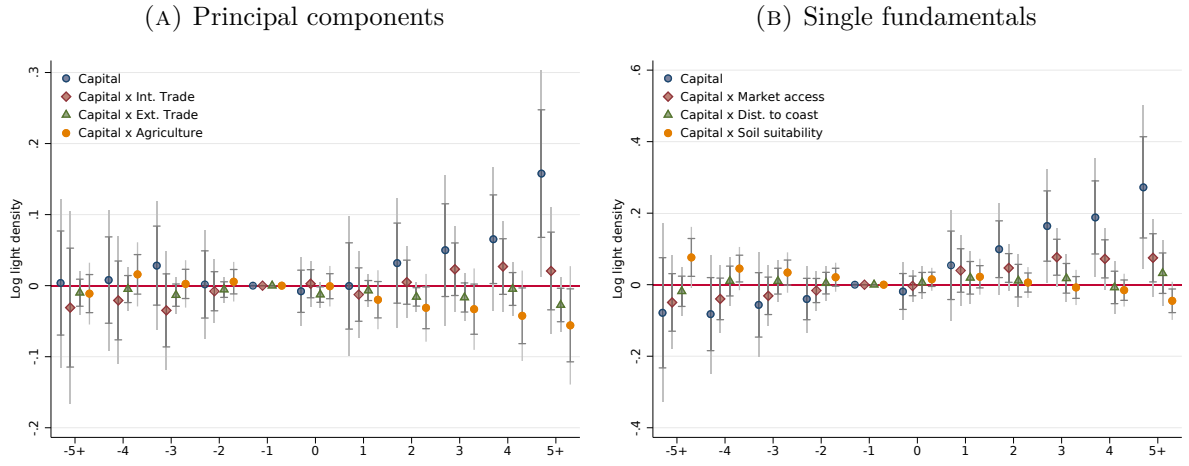
Notes: The figure illustrates event-study results from fixed effects regressions of the log of light intensity per square kilometer on a binned sequence of treatment change dummies, city fixed effects, child-region-by-year fixed effects, time-varying locational fundamentals for a panel that is balanced in calendar time. Circles represent point estimates from a regression with city and country-year fixed effects, diamonds represent specifications with additional controls for economic fundamentals, and triangles represent specifications with initial-region-by-year fixed effects. All regressions include city fixed effects. The gray error bars provide 95% confidence intervals based on standard errors clustered on initial regions. The whiskers indicate uniform 95% sup- t confidence bands computed using the plug-in method ([Montiel Olea and Plagborg-Møller, 2019](#)).

FIGURE E-7
Scale: Event-study estimates



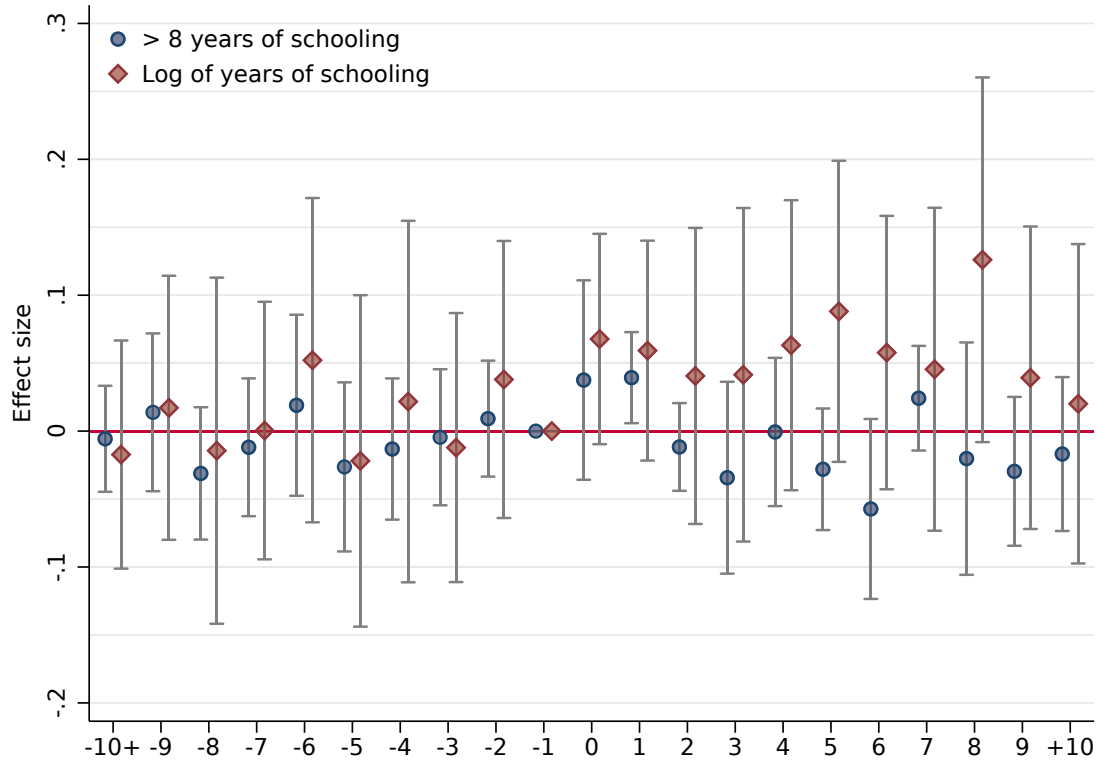
Notes: The figure reports event-study estimates corresponding to the difference-in-differences results presented in Table VI. Panels A to C report the event studies without controls (corresponding to columns 1, 3, and 5 in the table). Panels D to F report the event studies, including controls (corresponding to columns 2, 4, and 6). All regressions include city fixed effects and initial-region-by-year fixed effects. The gray error bars provide 95% confidence intervals based on standard errors clustered on initial regions. The whiskers indicate uniform 95% sup- t confidence bands computed using the plug-in method (Montiel Olea and Plagborg-Møller, 2019).

FIGURE E-8
Fundamentals: Event-study estimates



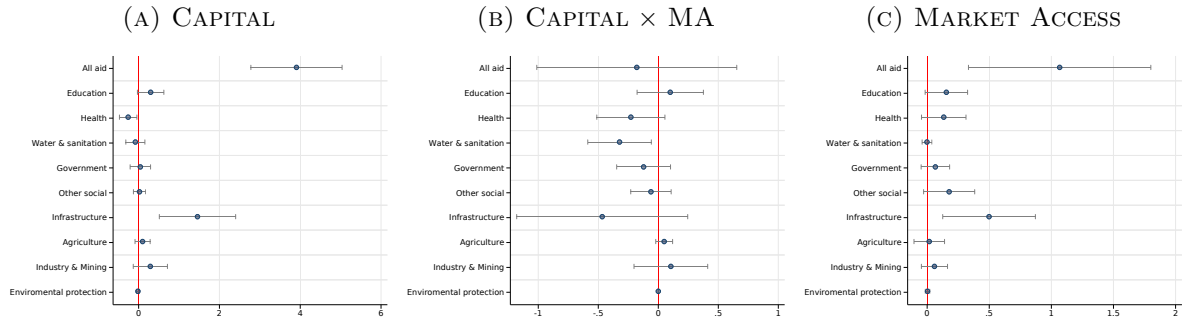
Notes: The figure reports event-study estimates corresponding to the difference-in-differences results presented in Table VII. Panel A reports estimates corresponding to column 5 of panel A, whereas panel B reports the estimates corresponding to column 5 of panel B of the table. All regressions include city fixed effects and initial-region-by-year fixed effects. The gray error bars provide 95% confidence intervals based on standard errors clustered on initial regions. The whiskers indicate uniform 95% sup-*t* confidence bands computed using the plug-in method ([Montiel Olea and Plagborg-Møller, 2019](#)).

FIGURE E-9
Selective migration: Within-city evidence (long window)



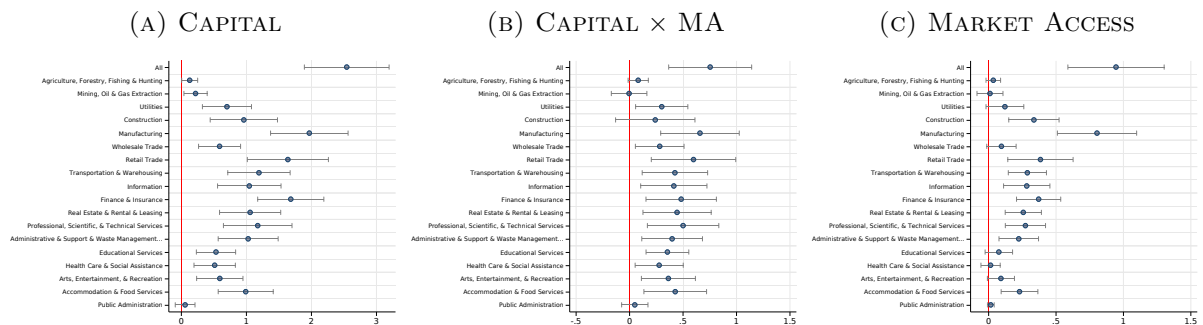
Notes: The figure illustrates event-study results from fixed effects regressions of the more than eight years of schooling dummy (blue circles) and log years of schooling (red triangles) on the binned sequence of treatment change dummies defined in the text. All specifications include the following individual-level controls: A gender dummy, a born-in-city dummy, age, and age squared. All specifications include city-year and cohort-at-move fixed effects as defined in the text. The gray error bars provide 95% confidence intervals based on standard errors clustered on the city level.

FIGURE E-10
Chinese aid by sector (2000-2014): Log commitments



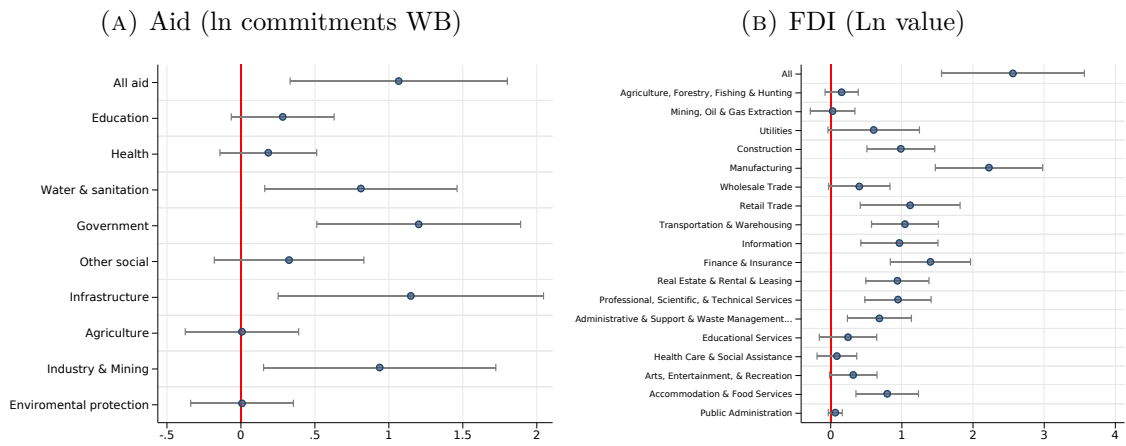
Notes: The figure plots estimates from regressions of Chinese development projects (2000–2014) in a particular sector on the fraction of years a city was a capital. Panel A reports estimates of the capital city effect at average levels of internal market access. Panel B reports results for the interaction of capital status with market access. Panel C reports results for the market access baseline effect. The definition of sectors follows the OECD’s Common Reporting Standard (see [Online Appendix A](#) for details). The gray error bars provide 95% confidence intervals based on standard errors clustered on initial regions.

FIGURE E-11
Capitals and FDI by industry (2003-2018): Ln Jobs



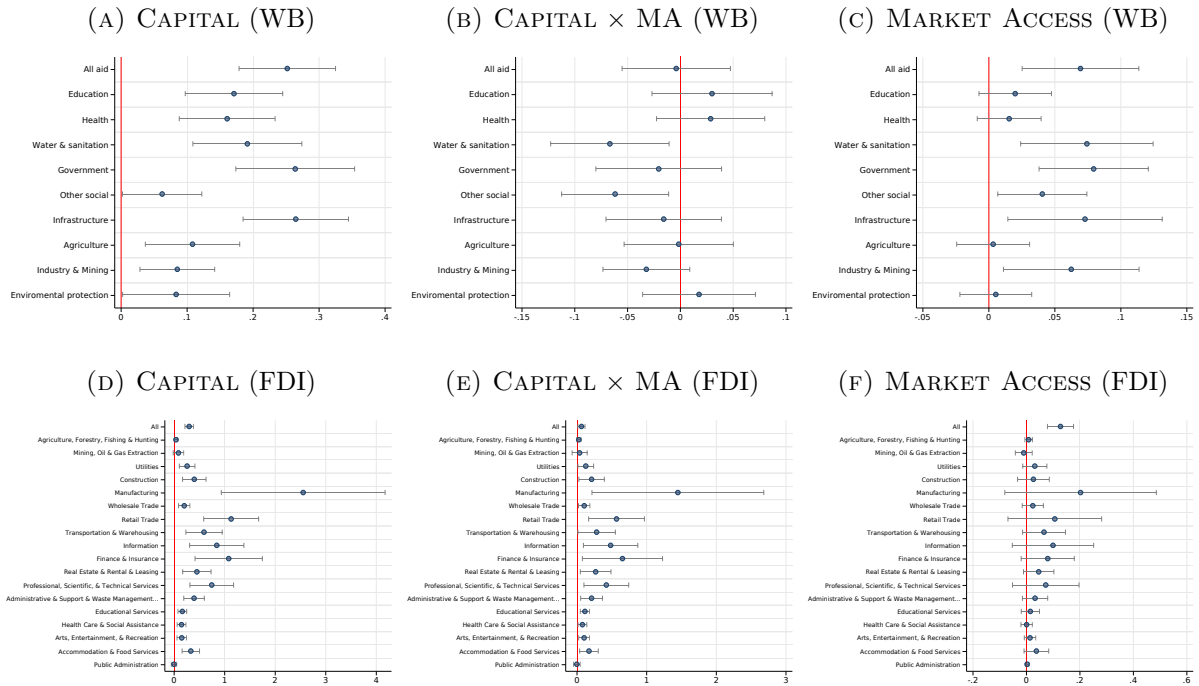
Notes: The figure plots estimates from regressions of log FDI created jobs + 1 in a particular sector on the fraction of years a city was a capital. Panel A reports estimates of the capital city effect at average levels of internal market access. Panel B reports results for the interaction of capital status with market access. Panel C reports results for the market access baseline effect. The definition of sectors follows the NAICS 2-digit sector classification (see [Online Appendix A](#) for details). The gray error bars provide 95% confidence intervals based on standard errors clustered on initial regions.

FIGURE E-12
Cities, aid and FDI: Market access effect



Notes: The figure plots estimates from regressions of the log of aid commitments +1 (WB) and log FDI projects values +1 on non-capital cities market access in 1990, the full set of fundamental controls and initial region fixed effects. FDI projects in a particular sector on the fraction of years a city was a capital. The FDI definition of sectors follows the NAICS 2-digit sector classification (see [Online Appendix A](#) for details). The gray error bars provide 95% confidence intervals based on standard errors clustered on initial regions.

FIGURE E-13
Aid and FDI: Extensive margin



Notes: The figure plots estimates from regressions of an indicator for the presence of at least one aid project (between 1994 and 2014) or FDI investment (between 2003 and 2018) on the fraction of years a city was a capital (in the respective period). Panel A (D) reports estimates of the capital city effect at average levels of internal market access. Panel B (E) reports results for the interaction of capital status with market access. Panel C (F) reports results for the market access baseline effect. The definition of sectors follows the OECD's Common Reporting Standard (see [Online Appendix A](#) for details). The definition of sectors follows the NAICS 2-digit sector classification (see [Online Appendix A](#) for details). The gray error bars provide 95% confidence intervals based on standard errors clustered on initial regions.

E-2. Additional tables

TABLE E-1
Identifying variation: Early and late developers

	<i>Late Developers</i>		<i>Early Developers</i>	
	Matched to city cores in 1990	City cores in 1990 with single changes	Matched to city cores in 1990	City cores in 1990 with single changes
<i>Panel A. Event-study period, 1987 – 2018</i>				
Always capitals	1,158	–	570	–
Gained status	300	248	35	32
Lost status	99	48	72	69
<i>Panel B. Diff-in-diff period, 1992 – 2013</i>				
Always capitals	1,222	–	583	–
Gained status	235	189	34	31
Lost status	72	35	54	51

Notes: The table shows summary statistics of the capital cities and urban clusters data for early and late developers following the urbanization in 1950 classification of [Henderson et al. \(2018\)](#). The urban clusters data in column 1 shows how many of these capital cities in late developers have been matched to cities that pass the detection thresholds of the city clustering algorithm. Column 2 shows the subset of these that experienced a single reform. Columns 3 and 4 repeat these summary statistics for early developers.

TABLE E-2
Baseline differences-in-differences

	<i>Dependent Variable: $\ln \text{LIGHTS}_{cit}$</i>					
	All Cities			Reformed Regions		
	(1)	(2)	(3)	(4)	(5)	(6)
CAPITAL	0.1043 (0.0279)	0.0852 (0.0273)	0.1049 (0.0283)	0.1351 (0.0297)	0.1085 (0.0303)	0.1096 (0.0323)
Fundamentals	–	✓	✓	–	✓	✓
City FE	✓	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	–	✓	✓	–
Ini. Region-Year FE	–	–	✓	–	–	✓
N	23909	23909	23909	8498	8498	8498
$N \times \bar{T}$	524867	524867	524867	186019	186019	186019

Notes: The table reports results from fixed effects regressions of the log of light intensity per square kilometer on capital city status. Standard errors clustered on initial regions are provided in parentheses.

TABLE E-3
Alternate agglomerations and peripheries: Difference-in-differences

	<i>Dependent Variable: $\ln \text{LIGHTS}_{cit}$</i>					
	All Cities			Reformed Regions		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A. Growth of the larger agglomeration</i>						
CAPITAL	0.1313 (0.0312)	0.1022 (0.0300)	0.1278 (0.0310)	0.1644 (0.0329)	0.1252 (0.0332)	0.1294 (0.0354)
<i>Panel B. Growth in the periphery of the city</i>						
CAPITAL	0.1412 (0.0312)	0.1044 (0.0292)	0.1313 (0.0302)	0.1724 (0.0327)	0.1250 (0.0321)	0.1321 (0.0342)
N	23399	23399	23399	8362	8362	8362
$N \times \bar{T}$	513787	513787	513787	183142	183142	183142
Fundamentals	–	✓	✓	–	✓	✓
Agglomeration FE	✓	✓	✓	✓	✓	✓
Country-Year FE	✓	✓	–	✓	✓	–
Ini. Region-Year FE	–	–	✓	–	–	✓

Notes: The table reports results from fixed effects regressions of the log of light intensity per square kilometer on capital city status. Panel A reports results based on the larger agglomeration (based on a common growth rate for all cities in a country as in [Harari, 2020](#)). Panel B reports the results for the periphery (the buffer areas net of the initial core). Standard errors clustered on initial regions are provided in parentheses.

TABLE E-4
Different control groups: Child region and countrywide matches

	<i>Dependent Variable: ln LIGHTS_{cit}</i>					
	Light intensity in 1992			Population in 1990		
	Control city ranks within ... of treated city					
	± 2	± 3	± 4	± 2	± 3	± 4
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A. Control cities within ranks of treated cities in child district</i>						
CAPITAL	0.1011 (0.0334)	0.0942 (0.0328)	0.0832 (0.0324)	0.0892 (0.0323)	0.0946 (0.0316)	0.0903 (0.0314)
F-test pre-trends (<i>p</i> -val.)	0.191	0.237	0.335	0.191	0.237	0.335
<i>N</i>	545	644	720	524	627	703
<i>N</i> × $\bar{T}$	11822	13968	15613	11373	13603	15251
<i>Panel B. Control cities within ranks of treated cities in country</i>						
CAPITAL	0.0953 (0.0220)	0.0951 (0.0229)	0.0960 (0.0229)	0.0829 (0.0241)	0.0831 (0.0242)	0.0798 (0.0247)
F-test pre-trends (<i>p</i> -val.)	0.389	0.190	0.182	0.675	0.586	0.753
<i>N</i>	810	1034	1234	780	1005	1211
<i>N</i> × $\bar{T}$	17614	22478	26828	16988	21894	26386

Notes: The table reports results from fixed effects regressions of the log of light intensity per square kilometer on capital city status. Panel A matches treated cities to a varying number of control cities on the basis of their rank in terms of light intensity or population within the child district. Panel B matches treated cities to a varying number of control cities on the basis of their rank in terms of light intensity or population within the entire country. All regressions include city-fixed effects, child-region-by-year fixed effects (panel A) or country-year fixed effects (panel B), and time-varying coefficients on the fundamentals. We report an F-test for pre-trends tests for the null hypothesis that all leading terms in the equivalent event-study specification are jointly zero. Standard errors clustered on initial regions are provided in parentheses.

TABLE E-5
Fiscal decentralization: Difference-in-differences

	<i>Dependent Variable: ln LIGHTS_{cit}</i>			
	Revenue share (1)	Employment share (2)	(3)	(4)
CAPITAL	0.2019 (0.0678)	0.1200 (0.0652)	0.2968 (0.0627)	0.2645 (0.0675)
CAPITAL $\times$ FISCAL DECENTRALIZATION	0.1243 (0.0702)	0.0733 (0.0657)	0.2503 (0.0968)	0.2373 (0.0918)
Fundamentals	–	✓	–	✓
City FE	✓	✓	✓	✓
Ini. Region-Year FE	✓	✓	✓	✓
N	5683	5683	5861	5861
$N \times \bar{T}$	124785	124785	128812	128812

Notes: The table reports results from fixed effects regressions of the log of light intensity per square kilometer on capital city status and interactions of the status with the proxies for the degree of fiscal decentralization taken from Treisman (2008). Specifically, the subnational revenue share as a percentage of GDP (averaged 1994-2000), and the subnational government employment share (in 1997). The interactions of the capital city status with the proxies for the degree of fiscal decentralization ($\tilde{z}$) are standardized such that $\tilde{z} \equiv (z - \bar{z})/\sigma_z$. Standard errors clustered on initial regions are provided in parentheses.

TABLE E-6
Different light measures

	<i>Dependent Variable: ln LIGHTS_{cit}</i>				
	Stable lights raw (1)	Stable lights bottom fix (2)	Average lights raw (3)	Bluhm & Krause '18 raw (4)	Bluhm & Krause '18 bottom fix (5)
CAPITAL	0.0510 (0.0354)	0.0866 (0.0334)	0.0851 (0.0297)	0.0739 (0.0346)	0.1096 (0.0323)
Fundamentals	✓	✓	✓	✓	✓
City FE	✓	✓	✓	✓	✓
Ini. Region-Year	✓	✓	✓	✓	✓
N	8498	8498	8498	8498	8498
$N \times \bar{T}$	186019	186019	186019	186019	186019

Notes: The table reports results from fixed effects regressions of the log of light intensity per square kilometer using different light measures on capital city status. We add one before taking logs of lights per area in km in columns 1 and 4 to keep city-years with no observed light. The raw average lights data record a non-zero light intensity in every city-year. Standard errors clustered on initial regions are provided in parentheses.

TABLE E-7
Initial city size

	<i>Dependent Variable: $\ln \text{LIGHTS}_{cit}$</i>				
	Initial city size				
	30k	40k	50k	75k	100k
CAPITAL	0.1280 (0.0326)	0.1466 (0.0354)	0.1720 (0.0369)	0.1606 (0.0411)	0.1946 (0.0553)
Fundamentals	✓	✓	✓	✓	✓
City FE	✓	✓	✓	✓	✓
Ini. Region-Year	✓	✓	✓	✓	✓
N	5645	4101	3149	1932	1368
$N \times \bar{T}$	123552	89705	68848	42230	29908

Notes: The table reports results from fixed effects regressions of the log of light intensity per square kilometer on capital city status. Columns 1 to 5 restrict the estimation samples to cities with an initial population above 30 up to 100k inhabitants. Standard errors clustered on initial regions are provided in parentheses.

TABLE E-8
Ethnic diversity

	<i>Dependent Variable: $\ln \text{LIGHTS}_{cit}$</i>					
	All Cities			Reformed Regions		
	(1)	(2)	(3)	(4)	(5)	(6)
CAPITAL	0.1062 (0.0275)	0.0873 (0.0269)	0.1048 (0.0278)	0.1336 (0.0298)	0.1083 (0.0301)	0.1079 (0.0317)
CAPITAL $\times$ ELF	-0.0156 (0.0192)	-0.0150 (0.0182)	0.0012 (0.0195)	0.0051 (0.0205)	0.0007 (0.0204)	0.0090 (0.0210)
Fundamentals	—	✓	✓	—	✓	✓
City FE	✓	✓	✓	✓	✓	✓
Ini. Region-Year FE	✓	✓	✓	✓	✓	✓
N	23874	23874	23874	8465	8465	8465
$N \times \bar{T}$	524197	524197	524197	185392	185392	185392

Notes: The table reports results from fixed effects regressions of the log of light intensity per square kilometer on capital city status. The interactions of the capital city status with ethnic diversity ($\tilde{z}$) are standardized such that $\tilde{z} \equiv (z - \bar{z})/\sigma_z$. Standard errors clustered on initial regions are provided in parentheses.

TABLE E-9
Long differences (1992-2013): Larger agglomerations

	<i>Dependent Variables: Change in</i>			
	$\ln \text{LIGHTS}_{ci}$	$\ln \text{POP DENSITY}_{ci}$	$\ln \text{LIGHTS P.C.}_{ci}$	$\ln \text{URBAN INDEX}_{ci}$
	(1)	(2)	(3)	(4)
CAPITAL	0.3121 (0.0386)	0.2518 (0.0350)	0.0603 (0.0514)	0.0324 (0.0073)
Fundamentals	✓	✓	✓	✓
Initial-Region FE	✓	✓	✓	✓
<i>N</i>	6993	6993	6993	6000

Notes: The table reports results from long difference regressions of the change in log light density of a city over different epochs on the fraction of years in which a city is a capital (1992-2013). $\ln \text{LIGHTS}_{ci}$ is log light density, $\ln \text{POP DENSITY}_{ci}$ is log population density (where we take the closest population values 1990 and 2015), $\ln \text{LIGHTS P.C.}_{ci}$ is log light per capita and $\ln \text{URBAN INDEX}_{ci}$ is the remotely sensed urban index for built-up structures (re-scaled from -1 to 1 to 0 to 2). Standard errors clustered on initial regions are provided in parentheses.

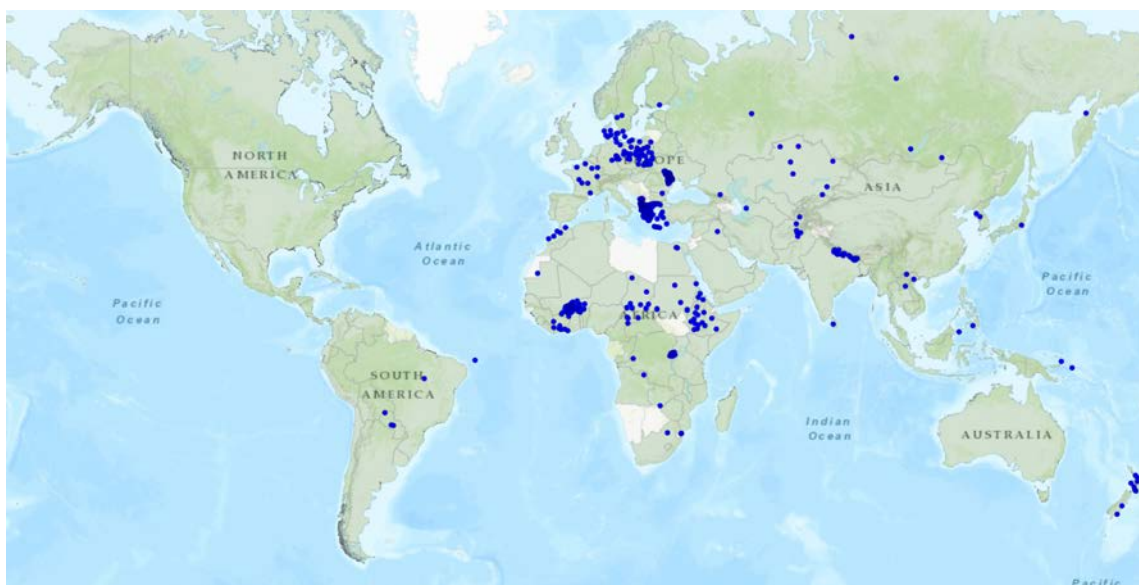
F. Former capitals and rump capitals

This appendix provides descriptive statistics on cities that lose their capital status, discusses pre-treatment trends, and discusses the appropriate comparison groups for these cities. We also report evidence on the performance of cities that lose capital status relative to their peers (cities that remain capitals).

F-1. Former capitals

Many cities across the globe have lost their status as capitals during the last three decades (see [Figure F-1](#)). About 63% of the observed 171 status losses in our sample occur during a centralization (mergers of two or more regions). In the other cases, a different city becomes the capital within the same region.

FIGURE F-1
Spatial distribution: Capital loss

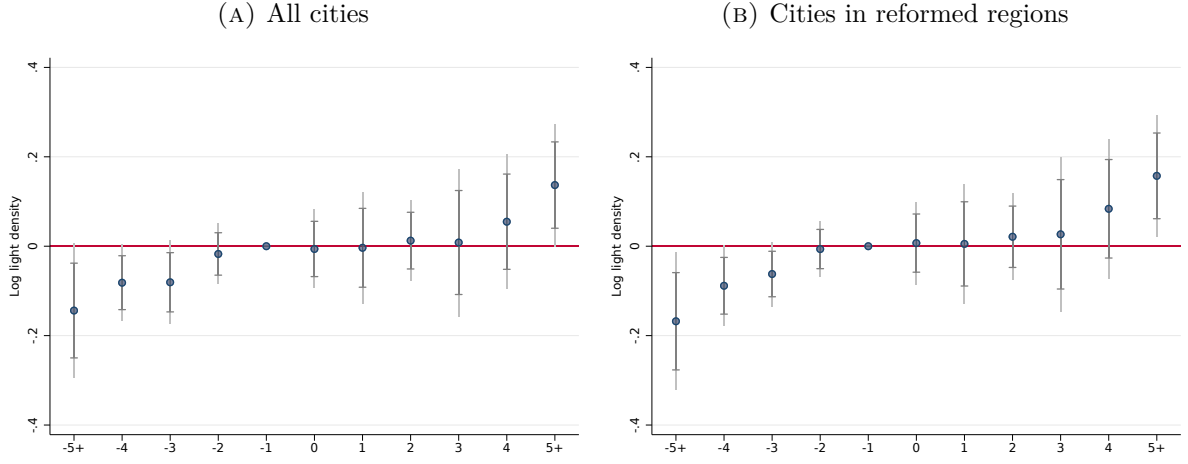


Notes: The figure shows all cities that lost their capital status during the 1987 to 2018 period in blue. Countries included in our sample are shaded green. The topographic base map is provided by Esri, using source material from Esri, TomTom, FAO, NOAA, and USGS.

We first turn to our baseline specification, which uses other non-capital cities as the control group. [Figure F-2](#) reports event-study estimates using our preferred specification with initial-region-by-year fixed effects and controls for locational fundamentals. There are significant and negative pre-trends. Capital cities that lose their status perform worse than non-capital cities before treatment. Regardless of why this occurs, identification is not feasible in our primary setting.

Of course, capitals that lose their status ought to be compared to cities that remain capitals. Unfortunately, this also implies that we now work with a drastically reduced

FIGURE F-2
Former capitals vs. all cities

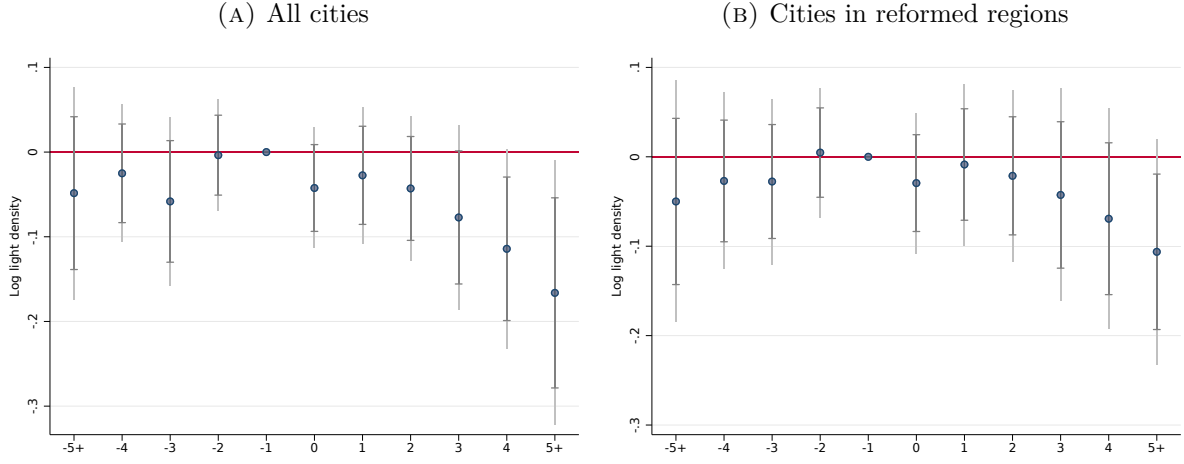


Notes: The figure illustrates results from fixed effects regressions of the log of light intensity per square kilometer on the binned sequence of treatment change dummies (capital loss) defined in the text. Panel A shows estimates for all ever capital cities based on a specification with country-year effects. Panel B shows estimates for ever-capital cities in reformed regions based on a specification with final-region-by-year fixed effects. All regressions include city fixed effects. The gray error bars provide 95% confidence intervals based on standard errors clustered on final regions. The whiskers indicate uniform 95% sup- t confidence bands computed using the plug-in method ([Montiel Olea and Plagborg-Møller, 2019](#)).

sample size (434 capital cities in the sample of reformed regions) and a design that more closely resembles a staggered event study with a small control group. Moreover, we do not have enough degrees of freedom to allow for time-varying coefficients on the locational fundamentals. In [Figure F-3](#), we run event studies on the set of ever capitals using again binned treatment change indicators for city loss. Note that we excluded cities that became capitals during our sample. Hence, the comparison groups differ a lot from our standard approach. The identifying variation in panel A is based on the difference between cities that are always capitals within the country compared to capitals that lose that status sometime during our sample period. The identifying variation in panel B is restricted to mergers of administrative regions in which one city loses its status, and the other city becomes the capital of the whole region. Note that focusing on mergers also has implications for the fixed effects we can include. Instead of initial-region-by-year fixed effects, we now use final-region-by-year fixed effects. This allows us to compare cities within the merging region and control for unobserved trends in the constituent parts before their merger.

The results show a clear pattern. We find no evidence suggesting the presence of pre-trends. Hence, capitals that will subsequently lose their capital status are not declining relative to always capitals before treatment. After the capital status is removed, we observe a steady loss of economic activity that takes longer to materialize than our main result but suggests a decline of similar magnitude in the medium run.

FIGURE F-3
Former capitals vs. always capitals



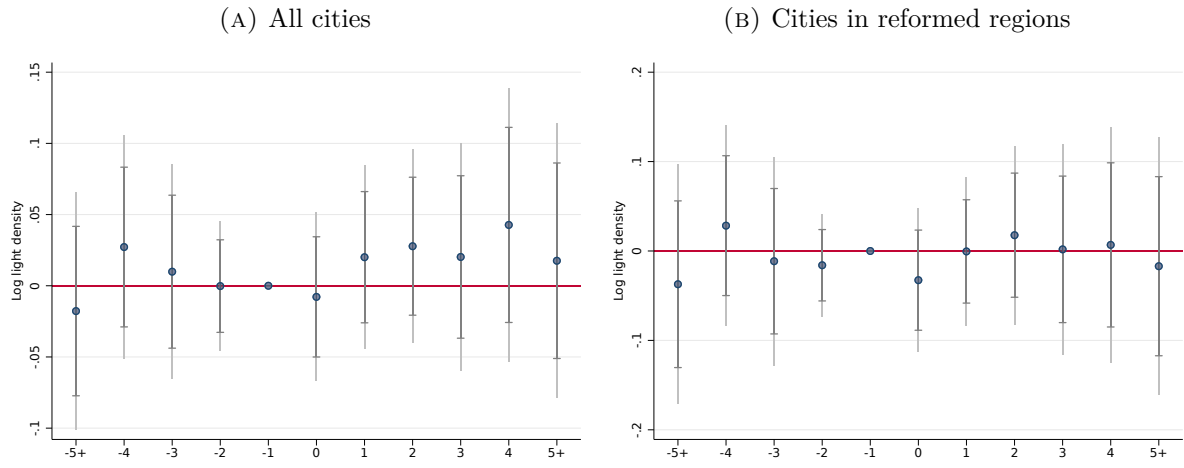
Notes: The figure illustrates results from fixed effects regressions of the log of light intensity per square kilometer on the binned sequence of treatment change dummies (capital loss) defined in the text. Panel A shows estimates for all ever-capital cities based on a specification with country-year fixed effects. Panel B shows estimates for ever-capital cities in reformed regions based on a specification with final-region-by-year fixed effects. All regressions include city fixed effects. The gray error bars provide 95% confidence intervals based on standard errors clustered on final regions. The whiskers indicate uniform 95% sup- t confidence bands computed using the plug-in method (Montiel Olea and Plagborg-Møller, 2019).

F-2. “Rump” capitals

A related issue to the loss of a political premium is the effect of decentralization on existing capitals that lose part of their territory. We refer to these cities as “rump capitals,” i.e., capitals that rule over a smaller jurisdiction after a decentralization reform that creates new additional capitals in the initial ‘parent’ region.

We specify the corresponding event for capitals that experience a reduction in their jurisdiction and estimate event studies comparing their performance to the set of always capitals. Figure F-4 presents the results. We find no evidence in favor of pre-treatment trends or any change in activity after a city becomes a “rump capital.” The economic gains of new capital cities appear not to come at the cost of the old ones, at least not in the short to medium run.

FIGURE F-4
Rump capitals vs. always capitals



Notes: The figure illustrates results from fixed effects regressions of the log of light intensity per square kilometer on the binned sequence of treatment change dummies (for rump capitals). Panel A shows estimates comparing rump capitals to always capitals based on a specification with country-year fixed effects. Panel B shows estimates comparing rump capitals to always capitals in reformed regions based on a specification with initial-region-by-year fixed effects. All regressions include city fixed effects. The gray error bars provide 95% confidence intervals based on standard errors clustered on initial regions. The whiskers indicate uniform 95% sup- t confidence bands computed using the plug-in method ([Montiel Olea and Plagborg-Møller, 2019](#)).

Additional references

- AidData.** 2017. “WorldBank.GeocodedResearchRelease_Level1_v1.4.2 geocoded dataset [dataset].” Aid Data Williamsburg, VA and Washington, DC. AidData [distributor]. <http://aiddata.org/research-datasets> (accessed September 2, 2020).
- Bai, Ying, and Ruixue Jia.** 2023. “The Economic Consequences of Political Hierarchy: Evidence from Regime Changes in China, 1000-2000 C.E.” *Review of Economics and Statistics*, 105(3): 626–645.
- Bardhan, Pranab K., and Dilip Mookherjee.** 2000. “Capture and governance at local and national levels.” *American Economic Review*, 90(2): 135–139.
- Bluhm, Richard, and Melanie Krause.** 2022. “Top lights: Bright cities and their contribution to economic development.” *Journal of Development Economics*, 157: 102880.
- Bluhm, Richard, and Melanie Krause.** 2025. “Data Storage for Bluhm and Krause (2022): Top lights: Bright cities and their contribution to economic development.” Zenodo [distributor]. <https://doi.org/10.5281/zenodo.15533718>.
- Bluhm, Richard, Axel Dreher, Andreas Fuchs, Bradley C. Parks, Austin M. Strange, and Michael J. Tierney.** 2024. “Replication data for: Connective Financing: Chinese Infrastructure Projects and the Diffusion of Economic Activity in Developing Countries.” Mendeley Data [distributor]. <https://doi.org/10.17632/235v7ksk8y.1>.
- Bluhm, Richard, Axel Dreher, Andreas Fuchs, Bradley C. Parks, Austin M. Strange, and Michael J. Tierney.** 2025. “Connective Financing: Chinese Infrastructure Projects and the Diffusion of Economic Activity in Developing Countries.” *Journal of Urban Economics*, 145: 103730.
- Brinkhoff, Thomas.** 2020. “City Population [dataset].” citypopulation.de [distributor]. <https://www.citypopulation.de> (accessed June 12, 2020).
- Campante, Filipe R., and Quoc-Anh Do.** 2014. “Isolated Capital Cities, Accountability, and Corruption: Evidence from US States.” *American Economic Review*, 104(8): 2456–81.
- Eberle, Ulrich J., J. Vernon Henderson, Dominic Rohner, and Kurt Schmidheiny.** 2020. “Ethnolinguistic diversity and urban agglomeration.” *Proceedings of the National Academy of Sciences*, 117(28): 16250–16257.
- FAO.** 2015. “Global Administrative Unit Layers (GAUL) [dataset].” FAO, Rome, Italy [distributor]. <https://data.apps.fao.org/catalog/dataset/global-administrative-unit-layers-gaul-2015> (accessed June 22, 2018).
- Financial Times Ltd.** 2020. “fDi Markets: Global Investment Database.” Financial Times Ltd., London, United Kingdom [distributor]. <https://www.fdimarkets.com/> (accessed August 5, 2020).
- GADM.** 2018–2022. “Global Administrative Areas (GADM) Versions 3.4 to 4.1 [dataset].” Center for Spatial Sciences at the University of California, Davis [distributor]. <http://www.gadm.org> (accessed August 16, 2019).
- Harari, Mariaflavia.** 2020. “Cities in bad shape: Urban geometry in India.” *American Economic Review*, 110(8): 2377–2421.
- Henderson, J. Vernon, Tim Squires, Adam Storeygard, and David Weil.** 2017. “Replication Data for: The Global Distribution of Economic Activity: Nature, History, and the Role of Trade.” Harvard Dataverse [distributor]. <https://doi.org/10.>

7910/DVN/MO6RJT.

- Henderson, J. Vernon, Tim Squires, Adam Storeygard, and David Weil.** 2018. “The Global Distribution of Economic Activity: Nature, History, and the Role of Trade.” *Quarterly Journal of Economics*, 133(1): 357–406.
- Henderson, J. Vernon, Vivian Liu, Cong Peng, and Adam Storeygard.** 2020. “Demographic and health outcomes by Degree of Urbanisation: Perspectives from a new classification of urban areas.” European Commission, Directorate-General for Regional and Urban Policy.
- ICF.** 1986-2019. “Demographic and Health Surveys (various) [dataset].” Rockville, Maryland: ICF [distributor]. <https://dhsprogram.com/data/> (accessed May 1, 2020).
- IIASA/FAO.** 2012. “Global Agro-ecological Zones (GAEZ v3.0) [dataset].” IIASA, Laxenburg, Austria and FAO, Rome, Italy [distributor]. <https://www.gaez.iiasa.ac.at> (accessed July 22, 2017).
- Jarvis, Andrew, Edward Guevara, Hannes I. Reuter, and Andy Nelson.** 2008. “Hole-filled SRTM for the globe Version 4 [dataset].” CGIAR-CSI SRTM 90m Database. International Center for Tropical Agriculture (CIAT) [distributor]. <https://csidotinfo.wordpress.com/data/srtm-90m-digital-elevationdatabase-v4-1> (accessed August 22, 2019).
- Kiszewski, Anthony, Andrew Mellinger, Andrew Spielman, Pia Malaney, Sonia Ehrlich Sachs, and Jeffrey Sachs.** 2004a. “A global index representing the stability of malaria transmission.” *American Journal of Tropical Medicine and Hygiene*, 70(5): 486–498.
- Kiszewski, Anthony, Andrew Mellinger, Andrew Spielman, Pia Malaney, Sonia Ehrlich Sachs, and Jeffrey Sachs.** 2004b. “A global index representing the stability of malaria transmission [dataset].” Gordon C. McCord [distributor]. https://www.dropbox.com/scl/fi/t4je3iqo795gdu18v4swr/ME_raster.zip?rlkey=921dnfkahuxh4g4jy0uh8mmdl (accessed August 22, 2019).
- Law, Gwillim.** 2010. *Administrative subdivisions of countries*. Jefferson, NC:McFarland & Company.
- Law, Gwillim.** 2020. “Statoids.com [dataset].” Statoids.com [distributor]. <http://www.statoids.com/statoids.html> (accessed January 15, 2020).
- Lewis, M. Paul,** ed. 2009. *Ethnologue: Languages of the World, 17th ed.* Dallas, TX:SIL International.
- Marshall, Monty G., and Ted Robert Gurr.** 2020. “Polity5: Political Regime Characteristics and Transitions, 1800-2018 [dataset].” Center for Systemic Peace [publisher]. <https://www.systemicpeace.org/inscrdata.html> (accessed September 25, 2020).
- Montiel Olea, José Luis, and Mikkel Plagborg-Møller.** 2019. “Simultaneous confidence bands: Theory, implementation, and an application to SVARs.” *Journal of Applied Econometrics*, 34(1): 1–17.
- National Geospatial-Intelligence Agency.** 2010. “World Port Index [dataset].” National Geospatial-Intelligence Agency Maritime Safety Information [distributor]. <https://msi.nga.mil/Publications/WPI> (accessed June 11, 2019).
- Natural Earth.** 2018. “1:10m physical vectors, version 4.1.0 [dataset].” Natural Earth [distributor]. <https://www.naturalearthdata.com/downloads/10m-physical-vectors> (accessed April 24, 2019).
- Runfola, Daniel, et al.** 2020a. “geoBoundaries: A global database of political

- administrative boundaries.” *PloS one*, 15(4): e0231866.
- Runfola, Daniel, et al.** 2020*b*. “geoBoundaries: A global database of political administrative boundaries [dataset].” William & Mary geoLab [distributor]. <https://www.geoboundaries.org> (accessed April 1, 2020).
- Schiavina, Marcello, Sergio Freire, and Kytt MacManus.** 2019. “GHS-POP R2019A - GHS Population Grid Multitemporal (1975, 1990, 2000, 2015) [dataset].” European Commission, Joint Research Centre (JRC) [distributor]. <https://doi.org/10.2905/0C6B9751-A71F-4062-830B-43C9F432370F>.
- Storeygard, Adam.** 2016. “Farther on down the road: Transport costs, trade and urban growth in sub-Saharan Africa.” *Review of Economic Studies*, 83(3): 1263–1295.
- Sun, Liyang, and Sarah Abraham.** 2021. “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects.” *Journal of Econometrics*, 225(2): 175–199.
- Treisman, Daniel.** 2008. “Decentralization Dataset [dataset].” www.sscnet.ucla.edu/polisci/faculty/treisman/Papers/Decentralization.xls (accessed February 21, 2023).
- University of Delaware.** 2015. “Terrestrial Precipitation and Air Temperature: 1900–2014 Gridded Monthly Time Series (V 4.01) [dataset].” NOAA Physical Sciences Laboratory [distributor]. https://psl.noaa.gov/data/gridded/data.UDel_AirT_Precip.html (accessed August 22, 2019).
- U.S. Air Force Weather Agency.** 1992–2013. “DMSP Operational Linescan System (OLS) Nighttime Lights Time Series.” NOAA National Centers for Environmental Information (NCEI). <https://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html> (accessed August 6, 2019).
- U.S. Geological Survey.** 1984–2012. “Landsat 5 TM Collection 1 Tier 1 Raw Scenes [dataset].” Processed using Google Earth Engine. https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LT05_C01_T1 (accessed July 21, 2020).
- U.S. Geological Survey.** 1999–2021. “Landsat 7 Collection 1 Tier 1 Raw Scenes [dataset].” Processed using Google Earth Engine. https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LE07_C01_T1 (accessed July 21, 2020).