Natural City Growth in the People's Republic of China

PETER H. EGGER, GABRIEL LOUMEAU, AND NICOLE PÜSCHEL*

This paper analyzes the growth of Metropolitan Statistical Areas in the PRC between 1992 and 2013 by focusing on the night-light radiance—a measure of economic activity—of connected subcity places that we refer to as a natural city. This paper documents the rapid growth of natural cities in the PRC between 1992 and 2009 that was followed by a slight reduction in the size of some natural cities between 2010 and 2013 in the aftermath of the recent global financial crisis. Institutional factors—such as the location of places near Special Economic Zones, the ramifications of legal migration from rural to urban areas following reforms to the *hukou* (household registration) system, and infrastructure accessibility—are found to be important drivers of the integration of peripheral places into natural cities.

Keywords: city growth, Metropolitan Statistical Areas, People's Republic of China
JEL codes: O20, R12, R40

I. Introduction

With the increase in global population, the change in urbanization rates around the world is a dynamic phenomenon. While in 1994 only about 30% of the world's population lived in cities, as defined by national statistical offices, about 54% did in 2014.[1] In the People's Republic of China (PRC), which has been among the most dynamic economies in the world over the last quarter of a century, almost 25% of the population has moved to urban areas during the past 2 decades. The PRC's National New-Type Urbanization Plan, 2014–2020 targets an urbanization rate of 60% by 2020. While urbanization is often measured as the increase in the

[1]World Bank. 2015. *World Development Indicators*. http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?end=2015&locations=CN&start=1960&view=chart

population within the administrative boundaries of cities, urbanization in a broad sense is driven by three phenomena: (i) the increase in population density (and economic activity) within the administrative boundaries of existing urban zones, (ii) the increase in population density (and economic activity) in areas in the vicinity of administrative urban zones through the growth of Metropolitan Statistical Areas (MSAs), and (iii) (to a lesser extent) the physical growth of the administrative areas of cities.[2] This paper focuses on the first two phenomena, which are objects of interest in the theoretical and empirical urban economics literature focusing on city growth; urban sprawl, which goes hand in hand with the formation of densely populated urban subcenters; and the decentralization of economic activity (see, for example, Fujita and Ogawa 1980, 1982; Henderson and Mitra 1996; Glaeser and Kahn 2001, 2004; McMillen and Smith 2003; Burchfield et al. 2006; Garcia-López, Hémet, and Viladecans-Marsal 2016).

Unlike in many other countries, the recent growth of cities in the PRC has been governed by regulations. The country's one-child policy, which had been instituted in its most restrictive form between 1978 and 2015, led to a slump in overall population growth, reduced the growth rate of cities, and slowed the average urbanization rate. Furthermore, the *hukou* (household registration) system has restricted the internal migration of people to urban centers by limiting access to public goods such as health care, schools, universities, and official housing. Finally, the inception of Special Economic Zones (SEZs) has ensured the protection of the private property rights of foreign investors, alleviated taxes and tariffs, regulated the policy of land usage, and liberalized economic and labor laws in geographically confined zones. According to Wang (2013), most major cities in the PRC's 326 municipalities hosted some sort of SEZ by 2006. A consideration of these regulatory provisions—apart from factors capturing the economic attractiveness and amenities in cities—appears relevant as they may lead to a gap between actual and optimal city size in the PRC, thereby affecting the associated economies of scale and scope (see, for example, Au and Henderson 2006a, 2006b; Desmet and Rossi-Hansberg 2013), which can result in potentially significant output losses.

The PRC's extensive investments in transport infrastructure, particularly road and railway networks, have fundamentally reshaped the structure of its urban areas. In the early 1990s, the Government of the PRC began to renew and upgrade its transport infrastructure, which caused previously underdeveloped regions to grow faster as industries started to decentralize (Banerjee, Duflo, and Qian 2012; Faber 2014; Baum-Snow et al. 2016, 2017). For example, Baum-Snow et al. (2017) find that suburban ring roads have displaced an average of about 50% of central city

---

[2]The term MSA is mostly used in the context of the study of cities in the United States. In Europe, the literature primarily refers to a Functional Urban Area, which essentially describes the same concept of agglomerations measured by a minimum density of the population according to census data. In this paper, we utilize the term natural cities to indicate something similar, though it is based on the measurement of a city by remote-sensing (night-light radiance) data in conjunction with the City Clustering Algorithm.

industrial gross domestic product (GDP) to the outskirts of cities, while marginal radial railroads have displaced an additional 20%. Similarly, Baum-Snow et al. (2016) argue that expanded regional highway networks in the PRC have had a negative average effect on local population density, causing a reallocation of economic activity and altering the structure of the country's cities.

The focus of this paper is on the growth of natural cities, which are defined as connected places with a minimum level of night-light radiance as a measure of place- and time-specific economic activity (Henderson, Storeygard, and Weil 2012), and which are associated with the PRC's 300 largest administrative cities over the period 1992–2013. One major merit of using remote-sensing data to define cities is that such data are available at much higher frequency than population census data. Furthermore, the data collection itself is much more homogeneous in terms of timing and concept. The data suggest that the PRC's natural cities grew rapidly between 1992 and 2010 before shrinking to some extent in the last few years of the review period, which might be attributable to the detrimental effects of the recent global financial crisis. We document this phenomenon for all cities in terms of descriptive statistics and illustrate it exemplarily for two major agglomerations, Beijing and Shanghai. This paper explores these developments using econometric analysis and identifies institutional factors—as reflected in the proliferation of SEZs and the provisions of the *hukou* system—and infrastructure accessibility as being important determinants of natural city growth. We highlight the effects of road and railway accessibility, and illustrate that shocks to infrastructure can be expected to induce relatively rapid adjustments in natural city size over the next 20 years.

The remainder of the paper is organized as follows. Section II introduces the definition of a natural city employed in this paper and outlines the measurement thereof. The data and their descriptive statistics, empirical strategy, and results are presented in Section III. Section IV concludes.

II. Natural City Borders in the People's Republic of China, 1992–2013

In this paper, we employ a definition of city boundaries based on what we call natural borders. Natural city borders relate to the well-known concepts of MSAs and Functional Urban Areas (FUAs), which measure city size by activity rather than administrative boundaries (see, for example, Zipf 1949; Krugman 1996; Eaton and Eckstein 1997; Harris, Dobkins, and Ioannides 2001; Ioannides and Overman 2003; Eeckhout 2004; Rozenfeld et al. 2011). A general motivation to use a city definition based on either MSA or FUA is that they capture more accurately the extent of urban units, going beyond (and sometimes integrating several units with) administrative boundaries. When looking at emerging urban areas, especially in transition economies such as the PRC, the study of MSAs and FUAs follows an economic rather than an administrative logic. We define the boundaries of natural cities based on the City Clustering Algorithm (CCA) (Rozenfeld et al. 2008,

Figure 1. **Centroids of the 300 Biggest Administrative Cities in the People's Republic of China by Population, 2000**
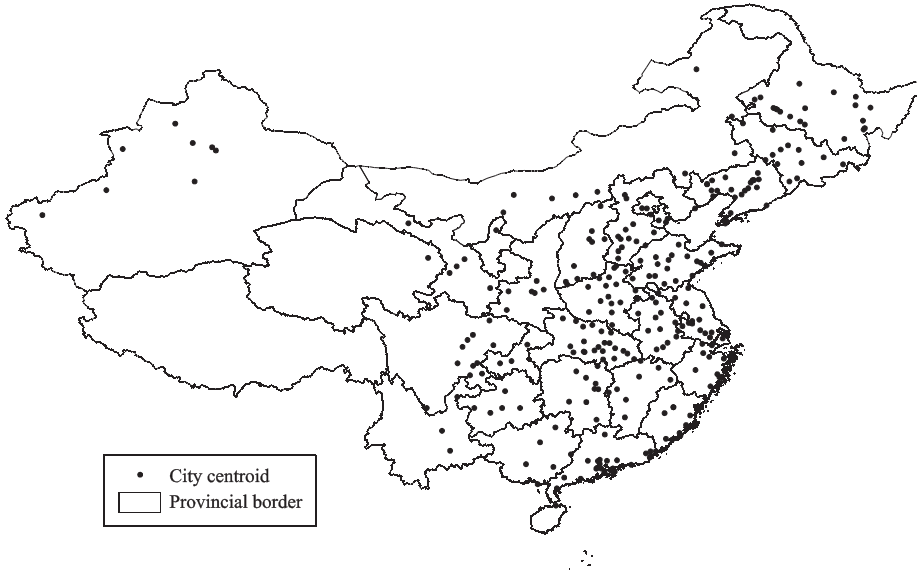
Source: Authors' illustration.

Rozenfeld et al. 2011), which we apply to remote-sensing (night-light radiance) data collected from satellites (Burchfield et al. 2006; Henderson, Storeygard, and Weil 2012). We measure the average night-light radiance in places that are 3 kilometers (km) in length by 3 km in width.[3] We are facing a trade-off between portraying and approximating the boundaries of small cities, especially in the early phases of the sample period, and the tractability of the data, particularly the application of the CCA.[4] The former requires sufficiently small places and the latter sufficiently few places. For those reasons, the consideration of 3 km × 3 km places was the finest-grained grid we could use given the time constraints. In general, one major advantage of using remote-sensing data to define natural cities is that annual data are available between 1992 and 2013, while MSA and FUA data are based on population censuses and therefore only available at lower frequency.

We consider the 300 biggest administrative cities in the PRC by population in the year 2000.[5] Figure 1 shows a map of the PRC and the location of the centroids of

---

[3] Individual places in the PRC bordering water or other boundaries may be smaller in size than 3 km × 3 km.

[4] The distribution of city sizes in the sample is presented in Figure 5.

[5] A list of the 300 biggest Chinese cities by population in 2000 is presented in Table A.1. Table A.2 includes a list of all natural cities by size in 2000. There are three different administrative levels of cities in the PRC's urban system: municipalities, prefecture-level cities, and county-level cities. With regard to the empirical analysis, we use administrative boundary information at the county level only. For further information on this point, please see section III.A.

all 300 cities covered. Very few cities are located in the western PRC, while there is a particularly high density in the vicinity of the coastal belt, which is not surprising provided the high degree of economic activity through international trade in that area.

The objects of interest in this study are the aforementioned 3 km × 3 km places. We define natural city borders on a uniform grid of such places for all cities in the sample. On this grid, we assign a place to a natural city in a year if (i) the average night-light radiance on the square exceeds a value of 40; and (ii) it is located near a cluster of places with average night-light radiance over 40, including the place that contains the city centroid (based on the CCA algorithm). We employ Version 4 of the Defense Meteorological Satellites Program–Operational Linescan System to measure night-light radiance at the pixel level (Croft 1978). The remote-sensing (night-light radiance) data therein take on values between 0 (no light) and 63 (maximum light). Night-light radiance data per pixel are available for all years between 1992 and 2013 based on pictures from six different satellites (F10, F12, F14, F15, F16, and F18), with some years covered by two satellites.[6] We chose the data such that the number of satellites they come from is minimized (F10 for 1992–1993, F12 for 1994–1999, F15 for 2000–2004, F16 for 2005–2009, and F18 for 2010–2013). The data comprise a raw-data version as well as a stable-data version, where the latter ensures that the data are not conflated by fire or firework incidents, or clouds or any other weather conditions. In this paper, we use the stable-light data version and compute the mean of radiance across all pixels within each place. In the final data set, we include all those places that were assigned to be in a natural city in any year between 1992 and 2013, and we track these places over the entire review period.

Figures 2 and 3 delineate the natural city with its city centroid (black dot) and administrative boundaries for Beijing and Shanghai for the years 1992, 1998, 2007, and 2013. In every panel, gray grids represent places that constitute the natural city in that particular year. Prefecture-level administrative city boundaries are indicated in black.[7] In the case of Beijing, we observe that its natural city size grew remarkably over the entire sample period. Especially from 1998 onward, the natural city of Beijing grew outward toward the northeast, which could be partly related to the 1993 opening of the Airport Expressway linking central Beijing to the Beijing Capital International Airport. Additional infrastructure investments to improve airport connectivity (e.g., Airport Express Subway) in preparation for the 2008 Olympic Games may have also contributed to the northeast developing more rapidly than other parts of Beijing.

---

[6]The satellite identifiers correspond to those used by the Defense Meteorological Satellites Program. For further information, please see National Oceanic and Atmospheric Administration. Earth Observation Group. https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html

[7]Figures 2 and 3 show prefecture-level administrative boundaries. However, in the estimation, all variables that include information on administrative boundaries rely on county-level boundaries as those boundaries represent the city-size distribution in a better way than prefecture-level boundaries.

Figure 2. **Natural City and Administrative Borders over Time—Beijing**
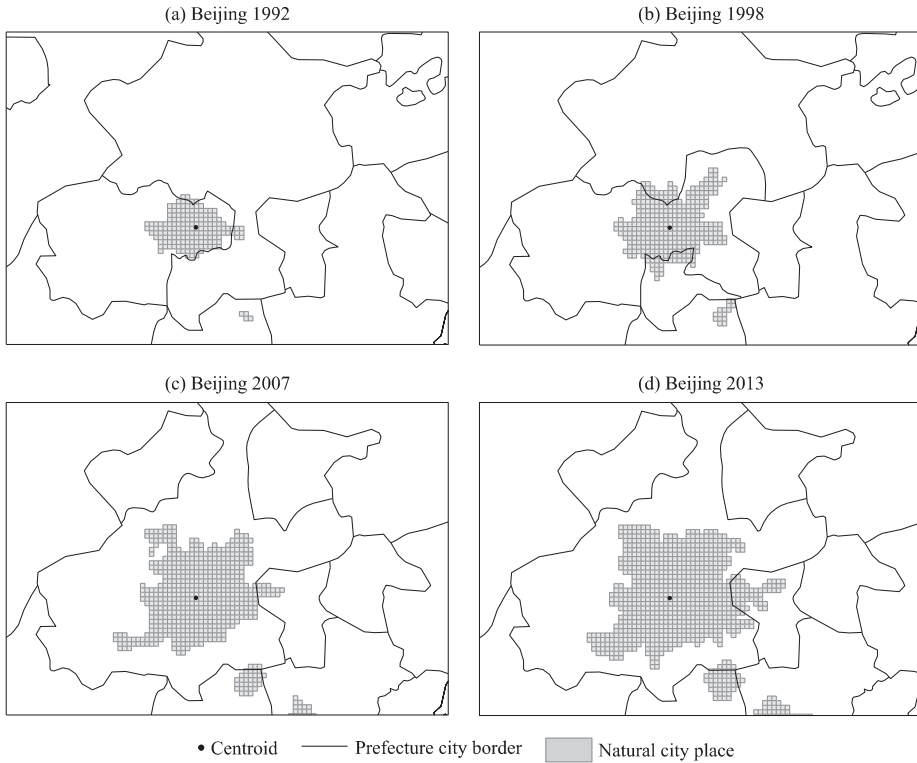
Source: Authors' illustration.

Similar to Beijing, Shanghai's natural city grew over the entire review period and mostly integrated urban areas along the downstream part of the Yangtze River. The example of Shanghai illustrates that, especially toward the end of the review period, several administrative cities merged into one natural supercity. The natural city of Shanghai in 1992 contained only one administrative centroid, while by 2013 it had incorporated a number of formerly distinct administrative and natural cities along the Yangtze River into one natural supercity. However, in spite of the general growth of natural cities through 2007–2013, many natural cities, including Beijing and Shanghai, shrank between 2010 and 2013, most likely as a consequence of the global financial crisis (Figure 4).

Table 1 reports average unconditional transition probabilities for natural city places for the whole sample of places considered. The table suggests that there is a high degree of persistence from 1 year to another: 92% of all natural city places keep their status, while about 90% of all places outside the natural city boundary remain outside that boundary from 1 year to another. The probability of acquiring natural city status amounts to 10%, while losing natural city status occurs in 7% of

Figure 3. **Natural City and Administrative Borders over Time—Shanghai**

(a) Shanghai 1992

(b) Shanghai 1998

(c) Shanghai 2007

(d) Shanghai 2013

• Centroid ⸻ Prefecture city border ▨ Natural city place
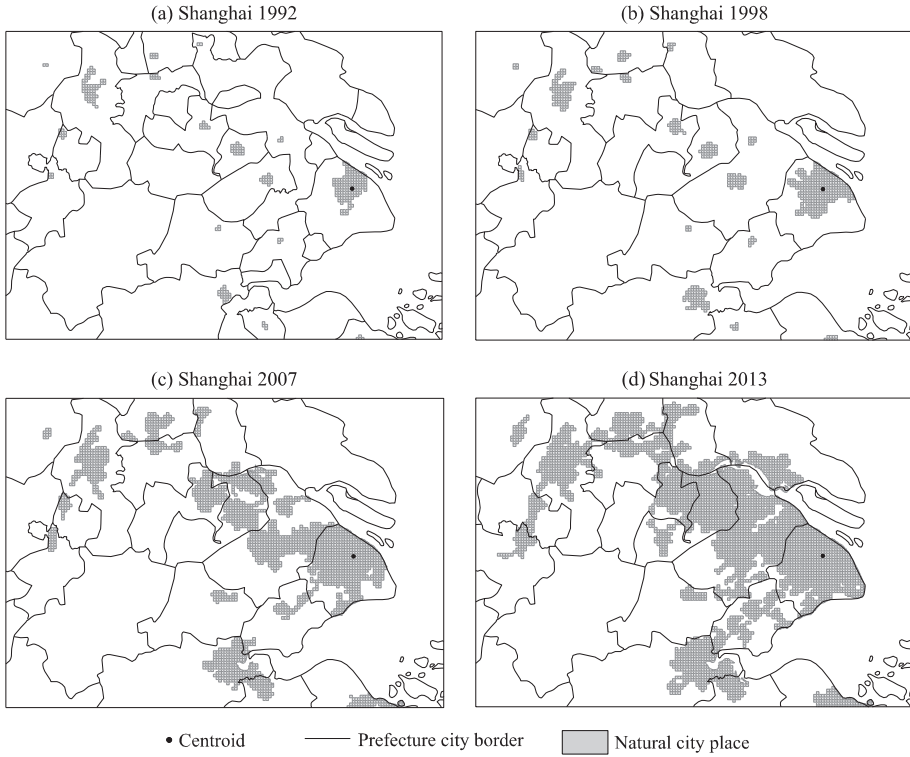
Source: Authors' illustration.

Figure 4. **Shrinking Natural Cities, 2010–2013—Beijing and Shanghai**

(a) Beijing

(b) Shanghai

• Centroid

Natural city place in:

☐ 2010 and 2013   ▨ 2010   ▨ 2013
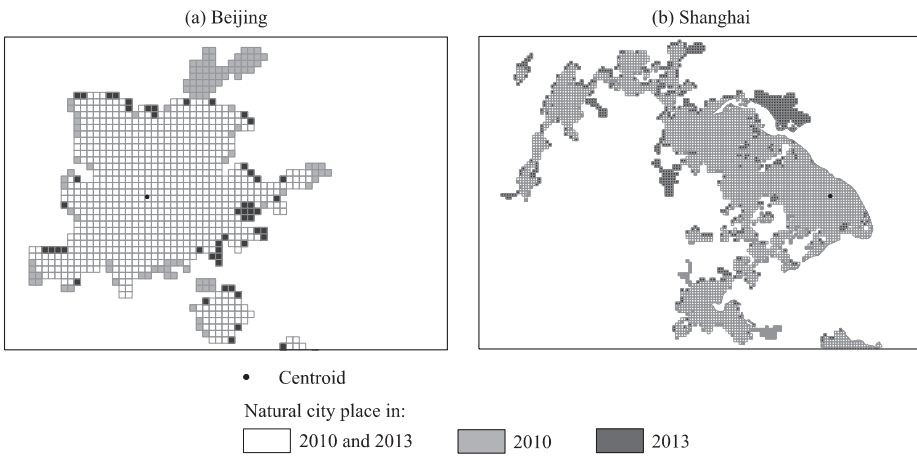
Source: Authors' illustration.

Table 1. **Transition Matrix**

		Target		
		Nat = 1	**Nat = 0**	**Total**
Origin	Nat = 0	10.06	89.94	100
	Nat = 1	92.63	7.37	100
	Total	39.65	60.35	100

Nat = natural city.
Source: Authors' calculations.

Figure 5. **Kernel Density Estimates of Natural City Size across Cities**
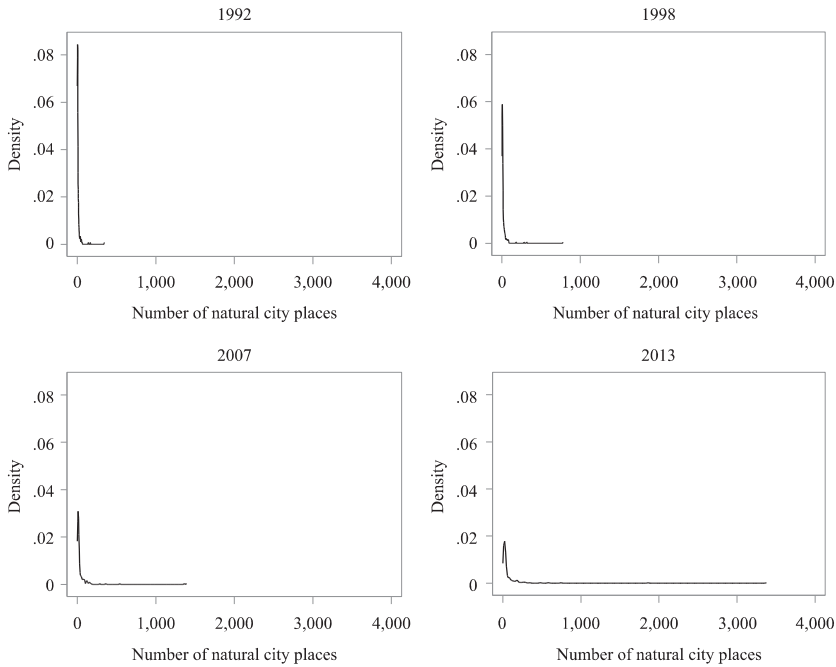
Source: Authors' calculations.

all cases from 1 year to another. The latter development is almost entirely driven by transitions during 2010–2013, reflecting the PRC's economic downturn in the aftermath of the global financial crisis.

Per Table 1, the average natural city size is expected to grow over the sample period. Figure 5 draws kernel density estimates of natural city sizes for the years 1992, 1998, 2007, and 2013. In each of the four panels of Figure 5, the horizontal axis shows the number of 3 km × 3 km places in a natural city. We observe that the average natural city size, reflected in the total number of places covered, increases remarkably with time. Especially in the beginning of the review period, the density mass is concentrated in the left tail of the distribution, indicating a great number of

relatively small natural cities and only a small number of very large supercities in the sample. Later in the review period, the degree of dispersion in terms of natural city size increases and the density mass in the left tail of the distribution gets smaller.

III. Drivers of Natural City Growth

In this section, we introduce all variables included in the subsequent empirical analysis.

A. Data

We use average (night-light) radiance data in a 3 km $\times$ 3 km place i at period t as the dependent variable to measure economic activity in that area. The variable $radiance_{it}$ is continuous and censored from below as well as from above, ranging from 0 (no light) to 63 (maximum light). Information on the source and the processing of the radiance data can be found in section II.

We identify three key categories of variables that drive natural city growth: geographical, climate, and institutional. The geographical variables include distance measures, some of which are time variant (indexed by both i and t) and others that are not (indexed by i only): distance to the administrative city center ($dist\ to\ center_i$), distance to the administrative city border ($dist\ to\ adborder_{it}$), distance to the nearest waterway ($dist\ to\ water_i$), distance to the ocean ($dist\ to\ ocean_i$), distance to the nearest road ($dist\ to\ road_i$), and distance to the nearest railway line ($dist\ to\ rail_i$). The geographical variables include a binary indicator that is unity if a place lies within the administrative boundary of the city centroid and zero otherwise ($within\ admin\ boundary_{it}$). Except for $dist\ to\ adborder_{it}$ and $within\ admin\ boundary_{it}$, which utilize annual data on administrative boundaries (at the county level) from the China Data Center at the University of Michigan, all distances are taken from OpenStreetMap using ArcGIS software.[8] Furthermore, we utilize topological information in the form of a measurement of altitude ($altitude_i$) from WorldClim Global Climate Data, and we control for the geographical location of each centroid by using information on its longitude and latitude from ArcGIS. The reason for including the latter two is that they relate to a place's accessibility. For instance, Chinese cities near the coast grew faster due to better accessibility to sea transport, which attracted foreign direct investment and was further stimulated by the formation of SEZs.

We use the following time-invariant climate data: average annual rainfall during the period of observation ($rain_i$); average annual temperature ($temperature_i$);

---

[8]OpenStreetMap information is based on the most recent network information available only. Distance to the nearest road includes all types of different roads (e.g., private roads, lower-capacity highways, higher-capacity highways, and limited-access highways). Distance to the nearest railway line includes all types of railway lines (e.g., subway lines and interprovincial railway lines).

and average annual temperature variation ($sd\ temperature_i$, as measured by the standard deviation). Gridded climate data are available from WorldClim Global Climate Data.

The institutional variables represent two types of institutional changes that governed the PRC's urban growth: reforms in the *hukou* system and the formation of SEZs. Between the late 1970s and mid-2000s, a period which is referred to as the first wave of *hukou* reforms, restrictions on movement and work were eased, which led to a large inflow of rural workers into urban areas. In most provinces, the scale of reforms varied with city size. Generally, reforms have had little impact on institutions in the most attractive urban areas such as provincial capitals and large cities along the coastal belt. To capture the different effects, we introduce three binary indicators—$small_{it}$, $medium_{it}$, and $large_{it}$—which are unity if a province applied their latest *hukou* reforms to small, medium, or large cities, respectively, and zero otherwise. A combined effect of these reforms is captured in the binary indicator $hukou_{it}$, which is unity if any one of the three, two out of three, or all three city-size variables are unity, and zero otherwise. Time-variant information on the extent of the latest *hukou* reforms by province during the period 1998–2008 is available in Organisation for Economic Co-operation and Development (2013).

SEZs are geographic regions that are typically characterized by liberal economic policies designed to attract foreign investors and enhance economic activity. In this paper, we use the term SEZ as a generic term for all types of special economic zones and open areas, including Free Trade Zones, Economic and Technology Development Zones, and open coastal cities, among others. Wang (2013) characterizes four big waves in the formation of SEZs in the PRC (1979–1985, 1986–1990, 1991–1995, and 1996–2007) and lists the corresponding municipalities that were designated as SEZs in each of the first three waves. This allows us to code three different binary indicator variables ($firstwave_{it}$, $secondwave_{it}$, and $thirdwave_{it}$) of which the former two are time variant because of the time variation in administrative city boundaries. The third variable is time variant because in our coding there is no treatment of places and cities prior to 1995. We also include the combined effect of the three waves that is captured in the binary indicator SEZ_{it}, which is unity if any one of the three, two of the three, or all three SEZ wave indicator variables are unity, and zero otherwise. Since the information on SEZs provided in Wang (2013) pertains to the municipality level, and while data utilized here vary by place, we assume that all places within the treated municipalities were affected by SEZs in the same way.

As an additional control variable, we include the population density ($popdens_{i1990}$) in 1990.[9]

---

[9]Gridded population density data for 1990 by 2.5 arc-minute grid cells are available from the Socioeconomic Data and Applications Center.

Table 2.   **Summary Statistics—Inside and Outside Natural City and Total**

	Nat = 1		Nat = 0		Average	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**
Geography						
dist to road$_i$ (km)	0.23	0.32	0.37	0.44	0.32	0.40
dist to rail$_i$ (km)	2.50	3.26	4.05	4.68	3.46	4.26
dist to ocean$_i$ (km)	261.61	413.37	293.31	426.36	281.14	421.70
dist to water$_i$ (km)	1.65	1.94	2.17	2.45	2.00	2.28
dist to center$_i$ (km)	12.39	9.60	17.28	11.48	15.40	11.06
dist to adborder$_{it}$ (km)	3.15	3.07	3.76	3.35	3.52	3.26
within admin boundary$_{it}$	0.33	0.47	0.26	0.44	0.29	0.45
altitude$_i$ (m)	152.40	327.50	176.30	359.70	167.10	347.90
longitude$_i$	116.30	6.94	116.30	7.19	116.30	7.10
latitude$_i$	33.47	7.03	33.65	6.31	33.58	6.59
Climate						
rain$_i$ (mm)	95.37	47.32	92.17	41.83	93.40	44.04
temperature$_i$ (°C)	14.52	5.40	14.24	4.93	14.35	5.12
sd temperature$_i$ (°C)	9.13	2.72	9.23	2.48	9.19	2.57
hukou$_{it}$	0.69	0.46	0.38	0.48	0.50	0.50
small$_{it}$	0.65	0.48	0.37	0.48	0.48	0.50
medium$_{it}$	0.56	0.50	0.30	0.46	0.40	0.49
large$_{it}$	0.54	0.50	0.29	0.45	0.39	0.49
SEZ$_{it}$	0.69	0.46	0.54	0.50	0.60	0.49
firstwave$_{it}$	0.04	0.19	0.02	0.14	0.03	0.16
secondwave$_{it}$	0.46	0.50	0.37	0.48	0.41	0.49
thirdwave$_{it}$	0.39	0.49	0.27	0.45	0.32	0.47
popdens$_{i1990}$ (ppl/km^2)	1,209	1,837	812	775	964	1,305
radiance$_{i1992}$	29.67	17.79	13.02	8.93	19.41	15.37
radiance$_{it}$ (0–63)	53.93	6.95	21.78	11.58	34.12	18.59
Observations	266,613		166,061		432,674	

°C = degree Celsius, km = kilometer, m = meter, mm = millimeter, Nat = natural city, SD = standard deviation, ppl/km^2 = people per square kilometer, SEZ = Special Economic Zone.
Source: Authors' calculations.

B.   Descriptive Statistics

Table 2 summarizes the descriptive features of all variables by natural city status (within a natural city, Nat = 1; outside of a natural city, Nat = 0; Average), and reports the mean and standard deviation for each variable.

Table 2 indicates that places within a natural city are on average 1.4 times closer to the city centroid than places outside of a natural city. Similarly, places inside are 1.1 times closer to the coast, 1.3 times closer to waterways, 1.6 times closer to the nearest road, and 1.6 times closer to the nearest railway line. As expected, places are on average much closer to the nearest road (0.3 km) than to the nearest railway line (3.5 km). We also observe that places inside a natural city are closer to the nearest administrative border since administrative areas close to the

Table 3. **Summary Statistics—Averages for 1992, 1998, 2007, and 2013**

	1992		1998		2007		2013	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$hukou_{it}$	0.041	0.198	0.075	0.263	0.842	0.364	0.928	0.258
$small_{it}$	0.041	0.198	0.075	0.263	0.842	0.364	0.850	0.357
$medium_{it}$	0.000	0.000	0.000	0.000	0.724	0.447	0.773	0.419
$large_{it}$	0.000	0.000	0.000	0.000	0.698	0.459	0.742	0.437
SEZ_{it}	0.415	0.493	0.623	0.485	0.625	0.484	0.610	0.488
$firstwave_{it}$	0.026	0.160	0.027	0.161	0.027	0.161	0.027	0.161
$secondwave_{it}$	0.389	0.488	0.410	0.492	0.410	0.492	0.395	0.489
$thirdwave_{it}$	0.000	0.000	0.347	0.476	0.349	0.477	0.349	0.477
dist to $adborder_{it}$ (km)	3.610	3.350	3.540	3.290	3.500	3.230	3.520	3.220
within admin $boundary_{it}$	0.296	0.456	0.285	0.451	0.288	0.453	0.298	0.457
$radiance_{it}$ (0–63)	19.410	15.370	27.580	17.180	41.680	15.190	53.880	8.720
Observations	19,667		19,667		19,667		19,667	

km = kilometer, SD = standard deviation, SEZ = Special Economic Zone.
Source: Authors' calculations.

considered city centroids are smaller in the average year than areas outside of the considered administrative city centers. Places within and outside of natural cities do not differ in terms of their average location in terms of longitude and latitude, but they differ in terms of altitude: places inside natural cities have an average altitude 1.2 times lower than places outside. Only about 30% of all places in the data lie within the administrative boundaries of one of the 300 major city centroids in our sample. By comparison, 62% of all places are located inside natural cities in the average year. Finally, places inside and outside natural cities do not significantly differ in terms of average precipitation and temperature.

Table 2 further suggests that places inside natural cities are more densely populated and more luminous in the beginning of our study period (1.5 times and 2.3 times, respectively). Over the entire study period, both places inside and outside of natural cities have a higher radiance level than they did in 1992. Places inside of a natural city appear to experience a relatively stronger increase in radiance during the study period. These places are also an average of 2.5 times more luminous than places outside of a natural city over the entire study period.

Table 3 summarizes descriptive statistics (mean and standard deviation) for all time-variant variables by year (1992, 1998, 2007, and 2013). Table 3 suggests that the latest wave of *hukou* reforms (1996–2007) started impacting small cities—only 7.5% of all places in the sample were treated in 1998—before reaching medium-sized and large cities after 1998. Given that the *hukou* data are coded at the provincial level and that we consider the 300 biggest administrative cities in the PRC, it is not surprising that by 2013 almost 93% of all places in the sample had experienced some degree of *hukou* reform. Concerning the SEZ indicators, the first wave of reforms (1979–1985) included a relatively small number of places, with

only 2.6% of all places treated during this wave, whereas the second (1986–1990) and third (1991–1995) waves applied to more than one-third of all places in the sample. Consequently, about 62.3% of all places were assigned to an SEZ in 1995. Finally, Table 3 indicates that the average night-light radiance ($radiance_{it}$) increased from 19.4 in 1992 to 53.9 in 2013.

C.   Econometric Approach

In this subsection, we outline the econometric model used to estimate coefficients on the suspected determinants of the (night-light) luminosity of place i in year t, $radiance_{it}$. Two features of the dependent variable are worth mentioning: (i) it is censored from below at 0 and from above at 63, and (ii) it appears to be serially correlated.[10]

To respect both the double censoring and autocorrelation through equicorrelation (accruing to the repeated observation of places over time and the presence of place-specific effects) and through inertia, we postulate a dynamic Tobit model with double censoring and random effects. We account for dynamic adjustment by letting $radiance_{it}$ be a function of its first-, second-, and third-lagged values $R_{it} = (radiance_{it-1}, radiance_{it-2}, radiance_{it-3})$, respectively, and estimate it along the lines of Wooldridge (2005). Accordingly, the endogeneity of the lagged dependent variables on the right-hand side of the model—through the presence of time-invariant random shocks, μ_i, in the models—can be acknowledged by properly specifying the initial conditions of the process (Hsiao 2015).

Subsume all exogenous drivers of $radiance_{it}$ in the common vector X_{it} and let $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ be the unknown parameters on R_{it} and β be the unknown parameters on X_{it}. Furthermore, let ϵ_{it} be the (normalized) remainder disturbances in the processes. Then, we may introduce a latent, uncensored, normal counterpart to $radiance_{it}$, $radiance_{it}^*$, and relate the two of them as follows:

$$radiance_{it} = \begin{cases} 0 & if\ radiance_{it}^* \le 0 \\ radiance_{it}^* & if\ 0 < radiance_{it}^* \le 63 \\ 63 & if\ radiance_{it}^* > 63 \end{cases} \tag{1}$$

Moreover, we may specify the latent variable $radiance_{it}^*$ in a linear fashion as a function of the parameters of interest through

$$radiance_{it}^* = R_{it}\alpha + X_{it}\beta + \mu_i + \varepsilon_{it} \tag{2}$$

For the estimation of equation (2), we employ two alternative sets of initial conditions for R_{it}. One involves the observed radiance in the initial year of the data, $radiance_{i1992}$, and the other one additionally involves the time averages of all

---

[10]Even though the original night-light radiance data take on integer values only, the dependent variable used here is continuous over the entire range of the data as we take the average of the night-light radiance across the pixels within a 3 km $\times$ 3 km place.

time-variant variables in X_{it}. Since the functional form of the dynamic Tobit model with double censoring is nonlinear and X_{it} includes squared values of some of the determinants, we will report marginal effects only as is customary with nonlinear models.

D.   Results

Table 4 summarizes the estimated effects of the lagged dependent variables associated with $\hat{\alpha}$, but only a subset of the effect estimates associated with $\hat{\beta}$.[11] For instance, we do not report the effects pertaining to variables used for the modeling of the initial condition with averages of the time-variant variables. Since the models are dynamic, the reported estimates should be interpreted as short-run effects materializing within a 3-year time window. Moreover, for the binary variables in X_{it} (e.g., the four variables each relating to either *hukou* or SEZ), we compare the average of the conditional mean when the variable takes on a value of unity for all places with the one when the variable takes on a value of zero for all places (Greene 2012). In Column (1) of Table 4, we model the initial condition as a function of the radiance in the initial year, $radiance_{i1992}$. In Column (2), the initial condition additionally includes the time averages of all time-variant variables. On a final note, the magnitudes of the total short-run effects of continuous variables in Table 4 should only be compared across such variables after normalization (e.g., by scaling them with the standard deviation of the respective variables in Table 3).

As the signs of significant effects do not differ qualitatively between Columns (1) and (2), and since the estimation of Column (2) is less efficient than for Column (1), we focus on the effects in Column (1). While we observe that the *hukou* and SEZ variables induce significant effects on $radiance_{it}$, we skip discussion of those effects here for the sake of brevity. Similarly, we forego discussion of the effects of geography and climate that are also reported in Table 4. In what follows, we focus on the effects of infrastructure, particularly roads and railways, near a place.

Two things stand out regarding these effects: (i) greater distance to transport infrastructure—such as roads, railway lines, and waterways—reduces the night-light radiance of a place; and (ii) the magnitude of the marginal effect of ln(*dist to road$_i$*) is around five times larger than that of ln(*dist to rail$_i$*). Clearly, these effects on $radiance_{it}$ reflect the importance of transport infrastructure, particularly roads, for local economic growth across all places in the sample.

---

[11]Table A.3 provides effects estimates akin to the dynamic Tobit model in Table 4 based on three alternative specifications that ignore censoring. These alternative models are linear models and always include satellite fixed effects. Apart from the infrastructure variables of interest, they are specified as follows: (i) the model in Column (1) does not include any other variables besides place fixed effects; (ii) the model in Column (2) is the same as in Column (1), but includes control variables; and (iii) the model in Column (3) is the same as in Column (2), but includes lags of the dependent variable and is an immediate linear counterpart to the dynamic Tobit model in Table 4. The results across these models and the dynamic Tobit model in Table 4 are very robust. As with the dynamic Tobit model, the fixed effects are parameterized in terms of averages of the time-variant explanatory variables.

Table 4. **Estimation Results for Dynamic Tobit**

	radiance$_{it}$ Full Sample (1)	radiance$_{it}$ Full Sample (2)	radiance$_{it}$ Nat = 1 (3)	radiance$_{it}$ Nat = 0 (4)	radiance$_{it}$ Admin = 1 (5)	radiance$_{it}$ Admin = 0 (6)
radiance$_{it-1}$	0.603***	0.602***	0.237***	0.524***	0.607***	0.598***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
radiance$_{it-2}$	0.194***	0.194***	0.0950***	0.149***	0.206***	0.189***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.002)
radiance$_{it-3}$	0.100***	0.100***	0.114***	0.028***	0.088***	0.106***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)
radiance$_{i1992}$	0.039***	0.039***	0.088***	0.190***	0.035***	0.044***
	(0.001)	(0.001)	(0.002)	(0.003)	(0.002)	(0.001)
ln(popdens$_{i1990}$)	0.149***	0.158***	0.098***	0.115***	0.110***	0.166***
	(0.011)	(0.011)	(0.019)	(0.023)	(0.024)	(0.013)
ln(dist to road$_i$)	−0.219***	−0.217***	−0.360***	−0.317***	−0.219***	−0.221***
	(0.007)	(0.007)	(0.013)	(0.014)	(0.013)	(0.009)
ln(dist to rail$_i$)	−0.041***	−0.038***	−0.066***	−0.091***	−0.003	−0.056***
	(0.008)	(0.008)	(0.014)	(0.015)	(0.015)	(0.009)
ln(dist to ocean$_i$)	−0.116***	−0.108***	−0.272***	−0.276***	−0.165***	−0.092***
	(0.010)	(0.010)	(0.018)	(0.018)	(0.021)	(0.011)
ln(dist to water$_i$)	−0.079***	−0.077***	−0.085***	−0.140***	−0.089***	−0.080***
	(0.008)	(0.008)	(0.015)	(0.015)	(0.015)	(0.010)
ln(dist to center$_i$)	−0.199***	−0.236***	−0.055*	−0.095***	−0.105***	−0.260***
	(0.016)	(0.016)	(0.028)	(0.031)	(0.027)	(0.021)
ln(dist to adborder$_i$)	−0.025***	−0.059	−0.015	0.056***	−0.022	−0.035***
	(0.009)	(0.044)	(0.017)	(0.017)	(0.017)	(0.012)
within admin boundary$_{it}$	−0.047*	−0.704***	−0.027	−0.065		
	(0.024)	(0.118)	(0.043)	(0.045)		
ln(altitude$_i$)	−0.010***	−0.085***	−0.306***	0.072**	−0.051*	−0.133***
	(0.016)	(0.016)	(0.032)	(0.029)	(0.028)	(0.020)
longitude$_i$	0.006*	−0.001	0.014**	−0.001	−0.018***	0.022***
	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.004)
latitude$_i$	0.054***	0.093***	0.087***	0.078***	0.119***	0.026**
	(0.011)	(0.012)	(0.02)	(0.020)	(0.022)	(0.013)
ln(rain$_i$)	−0.288***	−0.103*	−1.134***	−0.719***	0.089	−0.494***
	(0.055)	(0.057)	(0.107)	(0.099)	(0.102)	(0.067)
temperature$_i$	0.076***	0.079***	0.035	0.197***	0.077***	0.052***
	(0.012)	(0.012)	(0.023)	(0.021)	(0.020)	(0.016)
sd temperature$_i$	−0.003	−0.081***	−0.089***	−0.016	−0.082**	−0.033
	(0.018)	(0.019)	(0.033)	(0.032)	(0.035)	(0.022)
hukou$_{it}$	−0.750***	−1.316***	−0.036	−2.627***	0.596***	−0.963***
	(0.069)	(0.081)	(0.070)	(0.137)	(0.207)	(0.082)
small$_{it}$	1.128***	1.506***	0.592***	2.746***	−0.408**	1.556***
	(0.066)	(0.078)	(0.067)	(0.132)	(0.190)	(0.080)
medium$_{it}$	0.497***	1.093***	0.405***	1.226***	−0.441***	0.550***
	(0.054)	(0.074)	(0.062)	(0.095)	(0.165)	(0.064)
large$_{it}$	−0.696***	−0.927***	−0.809***	−0.809***	0.018	−0.785***
	(0.053)	(0.065)	(0.063)	(0.088)	(0.153)	(0.064)

Continued.

Table 4. *Continued.*

	radiance$_{it}$ Full Sample (1)	radiance$_{it}$ Full Sample (2)	radiance$_{it}$ Nat = 1 (3)	radiance$_{it}$ Nat = 0 (4)	radiance$_{it}$ Admin = 1 (5)	radiance$_{it}$ Admin = 0 (6)
SEZ$_{it}$	0.167***	0.977***	0.159**	0.561***	0.658***	0.048
	(0.040)	(0.185)	(0.074)	(0.074)	(0.103)	(0.047)
firstwave$_{it}$	−0.258***	−0.315***	0.011	−0.475***	−1.761***	0.079
	(0.071)	(0.075)	(0.125)	(0.142)	(0.179)	(0.081)
secondwave$_{it}$	−0.237***	−0.306***	−0.536***	−0.421***	−0.630***	−0.167***
	(0.035)	(0.041)	(0.064)	(0.064)	(0.101)	(0.039)
thirdwave$_{it}$	0.319***	−1.649***	0.467***	0.069	0.175**	0.421***
	(0.031)	(0.277)	(0.057)	(0.056)	(0.074)	(0.036)
Time averages		Yes				
Satellite effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	373,673	373,673	158,116	215,557	107,780	265,893
Places	19,667	19,667	19,656	16,756	6,144	14,187

Admin = within administrative boundary, Nat = natural city, SEZ = Special Economic Zone.
Notes: Reported coefficients are marginal effects. Standard errors are reported in parentheses. *** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$. All columns include squared terms for the following geography and climate variables: ln(dist to road$_i$), ln(dist to rail$_i$), ln(dist to ocean$_i$), ln(dist to water$_i$), ln(dist to adborder$_{it}$), ln(dist to center$_i$), ln(altitude$_i$), ln(rain$_i$), temperature$_i$, sd temperature$_i$. Column (2) includes time averages for all time-variant variables. All columns include satellite effects. All distance measures in the empirical estimation are in meters.
Source: Authors' calculations.

In Columns (3)–(6), we estimate the same model as in Column (1) for various subsamples of the data. Columns (3) and (4) divide the sample between places inside and outside of natural cities, while Columns (5) and (6) separate places inside and outside of the administrative borders of the major cities in our sample. Interestingly, the effect of ln(*dist to road$_i$*) in Column (3) is larger than in Column (4), while the opposite is observed for ln(*dist to rail$_i$*). Similarly, Column (6) shows a significant negative impact of ln(*dist to rail$_i$*), while the corresponding estimate in Column (5) is much smaller and not significant. The differences in the effects between Columns (3) and (4) on one hand and Columns (5) and (6) on the other—both in absolute terms and compared with Column (1)—reflect differences in the opportunity costs of certain types of transport infrastructure depending on the relative centrality or peripherality of places relative to the natural city or the administrative city center. In general, these results indicate that a marginal decline in distance to the road network leads places inside the natural city to grow relatively faster than places outside of it. However, a marginal decline in the distance to railway lines benefits peripheral areas more than central ones.[12]

Using the estimated effects from Column (1) in Table 4, we can predict the radiance level of all places from period to period and the change associated with an infrastructure improvement to the road or railway networks. We do so by reducing

___

[12]Table A.4 presents the results of the estimation of the baseline model (Column [1] in Table 4) on further subsamples, namely places in all four domains of the binary classification divided into administrative and natural city boundaries.

Figure 6. **Kernel Density Estimates—Observed versus Predicted Radiance Levels, 2007**
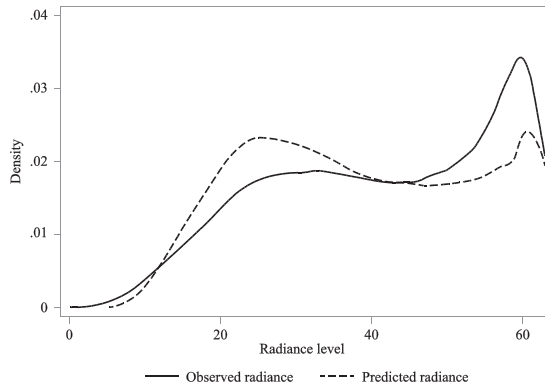
Figure 7. **Kernel Density Estimates—Observed versus Predicted Radiance Levels, All Years**
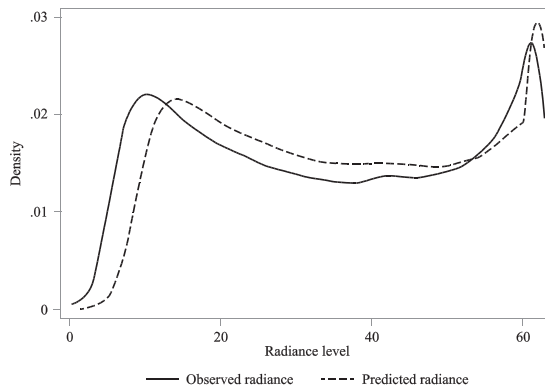
the distance to roads and railway lines by one standard deviation. We use 2007 as the benchmark year for this thought experiment since it is the year in which there are almost as many places outside (49%) as inside natural cities (51%). We predict the radiance level of all places in 2007 given the estimated coefficients associated with Column (1) of Table 4 and the variables in R_{it} and X_{it} as observed. We plot the kernel density estimates of observed and predicted radiance levels in 2007 in Figure 6.[13] Then, we shock ln(*dist to road$_i$*) and ln(*dist to rail$_i$*) alternatively by one standard deviation in 2007 and let the process run to see how such shocks impact radiance levels in the short and long term. Following the definition of a natural city used in this paper, we assume that any place will be part of a natural city in the counterfactual scenario if (i) its predicted radiance level amounts to at least 40, and

---

[13]The kernel density estimates of observed and predicted radiance levels for all years are plotted in Figure 7 and reflect a similar fit as the benchmark year (2007).

Table 5. **Transition Matrix of Counterfactual ln(dist to road$_i$)**
Definition: Nat = 1 if radiance > 40, Nat = 0 if radiance < 40

		Counterfactual: ln(dist to road$_i$)		
		Nat = 1	**Nat = 0**	**Total**
Baseline	Nat = 1	99.95$_{(2007)}$	0.05$_{(2007)}$	100
		99.72$_{(2012)}$	0.28$_{(2012)}$	100
		99.80$_{(2017)}$	0.20$_{(2017)}$	100
		99.95$_{(2022)}$	0.05$_{(2022)}$	100
		99.99$_{(2027)}$	0.01$_{(2027)}$	100
	Nat = 0	0.66$_{(2007)}$	99.34$_{(2007)}$	100
		6.18$_{(2012)}$	93.82$_{(2012)}$	100
		30.15$_{(2017)}$	69.85$_{(2017)}$	100
		51.85$_{(2022)}$	48.15$_{(2022)}$	100
		83.39$_{(2027)}$	16.61$_{(2027)}$	100
	Total	43.50$_{(2007)}$	56.50$_{(2007)}$	100
		61.43$_{(2012)}$	38.57$_{(2012)}$	100
		85.57$_{(2017)}$	14.43$_{(2017)}$	100
		96.91$_{(2022)}$	3.09$_{(2022)}$	100
		99.73$_{(2027)}$	0.27$_{(2027)}$	100

Nat = natural city.
Note: In addition to the radiance threshold, the City Clustering Algorithm condition is a necessary condition for a place to be assigned as a natural city (Nat = 1). The City Clustering Algorithm condition implies that a place is near a cluster of places with average radiance greater or equal to the threshold.
Source: Authors' calculations.

(ii) it is connected to other places in the natural city with a radiance level of at least 40.

In Tables 5 and 6, we report effects of these shocks on radiance levels in 2007—as well as after 5, 10, 15, and 20 years—compared with the baseline predictions. Table 5 shows the effect of a shock on road infrastructure. Most places predicted to lie inside a natural city in the baseline case remain inside it in the counterfactual scenario after 5 years (99.7% in 2012) and after 20 years (100% in 2027). However, the share of places in the sample that are predicted to lie outside of the natural city in the baseline but inside of it in the counterfactual scenario steadily increases over time in response to the shock from 0.7% in 2007 to 30.2% in 2017 and to 83.4% in 2027. The magnitude of the effect is relatively high because the actual number of places not in a natural city after 2017 is relatively small by construction of the data set.[14] Figures 8 and 9 illustrate the examples of Beijing and Shanghai, respectively.[15]

---

[14]The data set includes only those places that were in a natural city at some point between 1992 and 2013. This implies that all places that are not yet in a natural city in 2007 have a high probability of becoming part of a natural city within a few years.

[15]The sample includes only those places that are part of a natural city in any of the years covered.

Table 6. **Transition Matrix of Counterfactual ln(dist to rail$_i$)**
Definition: $Nat = 1$ if radiance > 40, $Nat = 0$ if radiance < 40

		Counterfactual: ln(dist to rail$_i$)		
		Nat = 1	Nat = 0	Total
Baseline	Nat = 1	100.00$_{(2007)}$	0.00$_{(2007)}$	100
		99.92$_{(2012)}$	0.08$_{(2012)}$	100
		99.91$_{(2017)}$	0.09$_{(2017)}$	100
		99.96$_{(2022)}$	0.04$_{(2022)}$	100
		99.98$_{(2027)}$	0.02$_{(2027)}$	100
	Nat = 0	0.13$_{(2007)}$	99.87$_{(2007)}$	100
		1.02$_{(2012)}$	98.98$_{(2012)}$	100
		4.18$_{(2017)}$	95.82$_{(2017)}$	100
		10.47$_{(2022)}$	89.53$_{(2022)}$	100
		15.34$_{(2027)}$	84.66$_{(2027)}$	100
	Total	43.22$_{(2007)}$	56.78$_{(2007)}$	100
		59.43$_{(2012)}$	40.57$_{(2012)}$	100
		80.36$_{(2017)}$	19.94$_{(2017)}$	100
		94.31$_{(2022)}$	5.69$_{(2022)}$	100
		98.64$_{(2027)}$	1.36$_{(2027)}$	100

Nat = natural city.
Note: In addition to the radiance threshold, the City Clustering Algorithm condition is a necessary condition for a place to be assigned as a natural city ($Nat = 1$). The City Clustering Algorithm condition implies that a place is near a cluster of places with average radiance greater or equal to the threshold.
Source: Authors' calculations.

The picture is similar, albeit of a smaller magnitude, when looking at the effect of a shock on rail infrastructure as shown in Table 6. The extreme majority of places predicted inside a natural city in the baseline are also predicted to lie inside in the counterfactual analysis after 5 years (99.9% in 2012) and (100% in 2027). The share of places predicted to lie outside in the baseline but inside in the counterfactual ($Nat = 0$ in baseline, $Nat = 1$ in counterfactual) is also increasing over time from 0.1% in 2007 to 4.2% in 2017 to 15.2% in 2027. The smaller magnitude of the effect reflects the smaller magnitude of the coefficient of ln(*dist to rail$_i$*) compared to the coefficient of ln(*dist to road$_i$*) estimated in Column (1) of Table 4.[16]

Our finding that transport infrastructure has a positive effect on local economic activity is well aligned with the findings in Banerjee, Duflo, and Qian (2012). They indicate that transport networks lead to higher levels of GDP per capita, even though the effect reported is small in magnitude. In line with

---

[16]In Tables A.5 and A.6, we report the dynamic responses to an infrastructure shock when fixing the night-light-radiance threshold for a place to be inside a natural city to 50 instead of 40. As expected, the main message of the results holds, even though the share of places predicted to lie outside the natural city in the baseline and inside in both counterfactuals is lower than in Tables 5 and 6. This simply reflects the distribution of night-light radiance across places as shown in Figure 7.

Figure 8. **Counterfactual Road and Rail—Beijing over Time**

Counterfactual Road	Counterfactual Rail
(a) Beijing 2007	(b) Beijing 2007
(c) Beijing 2012	(d) Beijing 2012
(e) Beijing 2017	(f) Beijing 2017

• Centroid

Natural city place in:

☐ Both   ▨ Baseline   ■ Counterfactual

Source: Authors' illustration.

Figure 9. **Counterfactual Road and Rail—Shanghai over Time**
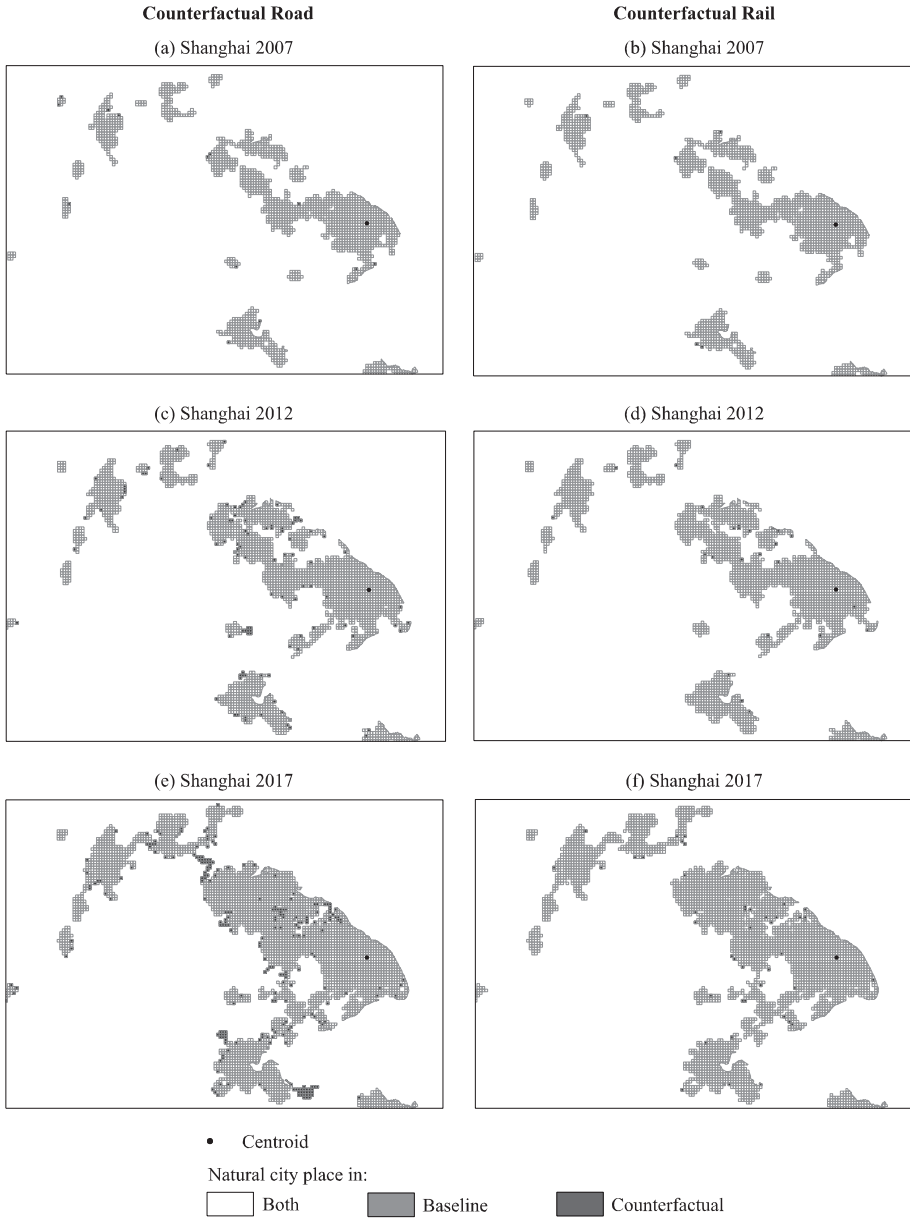
Source: Authors' illustration.

Baum-Snow (2007) and Baum-Snow et al. (2016), the results in our paper also suggest that better transport connectivity increases local economic activity in suburban areas. Considering the population density in city centers versus suburban areas, Baum-Snow et al. (2016, 2) suggest that "each additional radial highway displaced about 4% of [the] central city population to suburban regions and that the existence of some ring road capacity in a city reduced city population by about 20%." Contrary to these findings, we observe a positive effect of transport infrastructure on natural city growth, with a positive effect on both central and more peripheral areas of an average natural city. These results contrast with Faber (2014), who, looking at peripheral counties outside the commuting zones of metropolitan areas, finds that highway network connections have led to lower GDP growth among peripheral counties. This difference in findings suggests that transport networks have different effects on economic activity in remote areas than in metropolitan areas.

IV. Conclusions

This paper documents patterns in the size and growth of natural cities in the PRC for the 300 largest urban entities between 1992 and 2013. Rather than using administrative data on economic outcomes and their determinants, the paper identifies the boundaries of a natural city, which is related more closely to the notion of MSAs or Functional Urban Zones, in terms of the night-light radiance of connected places that measure 3 km × 3 km. Ultimately, the boundaries of natural cities are determined by applying the CCA to remote-sensing data for those places during the review period.

The key results of our analysis include the following. First, the number of distinct natural city centers decreased during the review period due to the absorption of some natural cities by others. This was particularly the case for larger cities, such as Shanghai, that formed natural supercities during the review period. Second, we detected rapid growth for the average natural city, which is in accordance with population census data that are only available at less frequent time intervals than night-light data, and adheres to the PRC's goal of fostering the rate of urbanization. The results suggest that natural cities grew considerably beyond the administrative boundaries of cities, which calls into question policies that target urbanization rates and other related development objectives based on administrative city boundaries. Third, the global financial crisis at the end of the last decade left its marks on natural city growth as some Chinese natural cities in our sample shrank between 2010 and 2013. Fourth, infrastructure improvements to the road and railway networks benefit agglomerations, although railway network improvements are expected to mainly benefit peripheral areas of cities more so than road improvements.

In future work, we plan to focus more explicitly on the difference between time-variant administrative versus natural city boundaries, and shed further light

on the dynamic process of responses to the two exogenous shocks. While we mainly used institutional variables related to the *hukou* system and SEZs as control variables, we will scrutinize their effects more closely after having coded them at a greater level of detail in order to understand these effects with greater precision and a broader scope than was possible in the current paper. Finally, we will investigate the effects of changes in the PRC's infrastructure networks, which was not possible with the data at hand, to better identify the associated effects on economic and other outcomes.

References

Au, Chun–Chung, and J. Vernon Henderson. 2006a. "Are Chinese Cities Too Small?" *Review of Economic Studies* 73 (3): 549–76.

_____. 2006b. "How Migration Restrictions Limit Agglomeration and Productivity in China." *Journal of Development Economics* 80 (2): 350–88.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian. 2012. "On the Road: Access to Transportation Infrastructure and Economic Growth in China." NBER Working Paper No. 17897.

Baum-Snow, Nathaniel. 2007. "Did Highways Cause Suburbanization?" *The Quarterly Journal of Economics* 122 (2): 775–805.

Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang. 2016. "Highways, Market Access, and Urban Growth in China." SERC Discussion Papers 0200.

_____. 2017. "Roads, Railroads, and Decentralization of Chinese Cities." *Review of Economics and Statistics*. http://www.mitpressjournals.org/doi/abs/10.1162/REST_a_00660#.WMK1-7YrJE4

Burchfield, Marcy, Henry G. Overman, Diego Puga, and Matthew A. Turner. 2006. "Causes of Sprawl: A Portrait from Space." *The Quarterly Journal of Economics* 121 (2): 587–633.

Croft, Thomas A. 1978. "Nighttime Images of the Earth from Space." *Scientific American* 239 (1): 86–98.

Desmet, Klaus, and Esteban Rossi-Hansberg. 2013. "Urban Accounting and Welfare." *American Economic Review* 103 (6): 2296–327.

Eaton, Jonathan, and Zvi Eckstein. 1997. "Cities and Growth: Theory and Evidence from France and Japan." *Regional Science and Urban Economics* 27 (4–5): 443–74.

Eeckhout, Jan. 2004. "Gibrat's Law for (All) Cities." *American Economic Review* 94 (5): 1429–51.

Faber, Benjamin. 2014. "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System." *Review of Economic Studies* 81 (3): 1046–70.

Fujita, Masahisa, and Hideaki Ogawa. 1980. "Equilibrium Land Use Patterns in a Non-Monocentric City." *Journal of Regional Science* 20 (4): 455–75.

_____. 1982. "Multiple Equilibria and Structural Transition of Non-monocentric Urban Configurations." *Regional Science and Urban Economics* 12 (2): 161–96.

Garcia-López, Miquel-Àngel, Camille Hémet, and Elisabet Viladecans-Marsal. 2016. "Job Decentralization, Subcenter Formation and Public Transportation." https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=SAEE40&paper_id=148

Glaeser, Edward L., and Matthew E. Kahn. 2001. "Decentralized Employment and the Transformation of the American City." NBER Working Paper No. 8117.

_____. 2004. "Sprawl and Urban Growth." In *Handbook of Regional and Urban Economics*, edited by J. Vernon Henderson and Jacques-François Thisse, 2481–527. Amsterdam: Elsevier.

Greene, William H. 2012. *Econometric Analysis*, *Pearson International Edition*. London: Pearson Education, Limited.

Harris-Dobkins, Linda, and Yannis M. Ioannides. 2001. "Spatial Interactions among US Cities: 1900–1990." *Regional Science and Urban Economics* 31 (6): 701–31.

Henderson, J. Vernon, and Arindam Mitra. 1996. "The New Urban Landscape: Developers and Edge Cities." *Regional Science and Urban Economics* 26 (6): 613–43.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.

Hsiao, Cheng. 2015. *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Ioannides, Yannis M., and Henry G. Overman. 2003. "Zipf's Law for Cities: An Empirical Examination." *Regional Science and Urban Economics* 33 (2): 127–37.

Krugman, Paul. 1996. "Confronting the Mystery of Urban Hierarchy." *Journal of the Japanese and International Economies* 10 (4): 399–418.

McMillen, Daniel P., and Stefani C. Smith. 2003. "The Number of Subcenters in Large Urban Areas." *Journal of Urban Economics* 53 (3): 321–38.

National Oceanic and Atmospheric Administration. Earth Observation Group. https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html (accessed August 11, 2016).

OpenStreetMap. https://www.openstreetmap.org/#map=5/51.500/-0.100 (accessed August 11, 2016).

Organisation for Economic Co-operation and Development (OECD). 2013. *OECD Economic Surveys: China 2013*. Paris: OECD Publishing.

Rozenfeld, Hernán D., Diego Rybski, José S. Andrade, Michael Batty, H. Eugene Stanley, and Hernán A. Makse. 2008. "Laws of Population Growth." *Proceedings of the National Academy of Sciences* 105 (48): 18702–7.

Rozenfeld, Hernán D., Diego Rybski, Xavier Gabaix, and Hernán A. Makse. 2011. "The Area and Population of Cities: New Insights from a Different Perspective on Cities." *American Economic Review* 101 (5): 2205–25.

Socioeconomic Data and Applications Center. http://dx.doi.org/10.7927/H4XK8CG2 (accessed August 11, 2016).

Wang, Jin. 2013. "The Economic Impact of Special Economic Zones: Evidence from Chinese Municipalities." *Journal of Development Economics* 101: 133–47.

Wooldridge, Jeffrey M. 2005. "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity." *Journal of Applied Econometrics* 20 (1): 39–54.

World Bank. 2015. World Development Indicators. http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?end=2015&locations=CN&start=1960&view=chart

Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Boston: Addison–Wesley.

Appendix

Table A.1. List of the 300 Biggest Administrative Cities in the People's Republic of China by Population, 2000

Rank	Name	Population	Rank	Name	Population	Rank	Name	Population	Rank	Name	Population
1	Shanghai	13,278,500	76	Shatian	648,400	151	Maoming	346,500	226	Zouxian	196,500
2	Beijing	7,209,900	77	Mudanjiang	646,700	152	Jixi	345,200	227	Chenzhou	193,300
3	Wuhan	4,104,300	78	Zhenjiang	643,000	153	Zhoukou	345,000	228	Badaojiang	192,900
4	Chengdu	4,064,700	79	Yancheng	638,800	154	Jinxi	337,300	229	Wanxian	191,700
5	Tianjin	3,945,900	80	Shaoyang	636,700	155	Shiongshui	336,300	230	Shangqiu	191,300
6	Shenyang	3,527,800	81	Jinzhou	630,800	156	Zhongshan	334,100	231	Haibowan	190,000
7	Xian	3,480,600	82	Taizhou	622,400	157	Zhaoqing	331,400	232	Ezhou	189,300
8	Chongqing	3,378,900	83	Taian	618,900	158	Dongchang	328,500	233	Quanzhou	188,200
9	Guangzhou	3,244,900	84	Dandong	604,600	159	Tieling	324,100	234	Zaoyang	187,900
10	Harbin	3,129,300	85	Luancheng	598,500	160	Daipo	322,300	235	Kaiyuan	186,200
11	Nanjing	2,870,200	86	Panjin	582,200	161	Xinpu	322,200	236	Dongling	184,600
12	Taiyuan	2,690,500	87	Guilin	580,200	162	Dezhou	317,900	237	Zaozhuang	184,000
13	Changchun	2,337,000	88	Kaifeng	577,400	163	Luzhou	317,200	238	Jiutai	183,800
14	Zhengzhou	2,052,100	89	Zhangjiakou	573,800	164	Jingdezhen	314,400	239	Baiyin	183,500
15	Jiulong	2,040,200	90	Yingkou	571,400	165	Chifeng	314,200	240	Yizheng	183,100
16	Jinan	1,961,500	91	Haikou	568,900	166	Shishi	308,800	241	Jingzhou	181,200
17	Dalian	1,925,200	92	Zhanjiang	568,000	167	Tongling	307,600	242	Qincheng	176,700
18	Changsha	1,891,000	93	Huaiyin	564,400	168	Dongying	303,800	243	Nanping	176,300
19	Hangzhou	1,881,500	94	Neijiang	562,700	169	Yanji	302,900	244	Hangu	175,600
20	Shijiazhuang	1,683,800	95	Puyang	561,300	170	Suizhou	302,700	245	Laiyang	174,900
21	Nanchang	1,657,900	96	Bengbu	561,200	171	Xuanhua	302,600	246	Luan	174,700
22	Jilin	1,625,700	97	Shihezi	553,400	172	Baicheng	292,300	247	Xianning	173,900
23	Tangshan	1,600,600	98	Yangzhou	548,600	173	Hulan,ergi	283,200	248	Jingmen	170,900
24	Qingdao	1,584,300	99	Jiamusi	548,400	174	Anshun	283,100	249	Weinan	170,800
25	Urumqi	1,424,300	100	Yueyang	542,800	175	Quanwan	277,700	250	Jiaozhou	170,200
26	Luoyang	1,417,200	101	Maanshan	536,700	176	Sanmenxia	274,100	251	Liupanshui	169,600
27	Xinyang	1,412,300	102	Xiamen	535,600	177	Linyi	268,500	252	Jian	169,400

Continued.

Table A.1. Continued.

Rank	Name	Population	Rank	Name	Population	Rank	Name	Population	Rank	Name	Population
28	Lanzhou	1,409,200	103	Shaoguan	529,300	178	Jiujiang	268,200	253	Bozhou	169,300
29	Fushun	1,409,000	104	Zhangzhou	519,700	179	Huizhou	265,700	254	Fuling	168,900
30	Hefei	1,362,900	105	Wuhu	512,500	180	Nanyang	265,500	255	Honghu	168,500
31	Xianggangdao	1,345,800	106	Xingtai	512,500	181	Wuzhou	264,200	256	Huanggang	168,300
32	Baotou	1,226,600	107	Zhunmen	499,600	182	Yingcheng	263,600	257	Zhaodong	167,800
33	Anshan	1,226,200	108	Jiaozuo	497,800	183	Aksu	263,500	258	Yuhong	167,600
34	Shantou	1,221,300	109	Foshan	496,300	184	Mianyang	261,800	259	Beipiao	166,900
35	Guiyang	1,160,700	110	Yuanlong	494,800	185	Jincheng	259,600	260	Hengshan	166,700
36	Suzhou	1,150,200	111	Tanggu	490,000	186	Wafangdian	258,400	261	Wulanhaote	166,500
37	Handan	1,129,800	112	Zhangye	489,600	187	Shangrao	258,300	262	Linhe	166,400
38	Fuzhou	1,120,200	113	Siping	486,800	188	Tongliao	257,500	263	Huzhou	165,500
39	Xuzhou	1,099,200	114	Kuiqing	480,100	189	Suihua	256,700	264	Fuyang	165,300
40	Datong	1,053,600	115	Yichang	478,200	190	Heze	256,600	265	Mentougou	165,200
41	Wuxi	1,010,500	116	Panzhihua	473,900	191	Sucheng	254,200	266	Longfeng	164,200
42	Xianyang	976,200	117	Xiangfan	471,500	192	Jining	248,400	267	Deyang	163,600
43	Kunming	967,900	118	Jiaojiang	470,500	193	Laohekou	248,300	268	Xiaogan	163,400
44	Benxi	918,900	119	Cangzhou	470,400	194	Rizhao	246,800	269	Yulin	162,600
45	Changzhou	917,400	120	Liaoyuan	470,400	195	Yibin	246,400	270	Zhicheng	162,300
46	Pingdingshan	906,700	121	Jiaxing	466,900	196	Kashi	240,400	271	Nanpiao	161,300
47	Baoding	888,100	122	Yinchuan	465,900	197	Yining	240,000	272	Sujiatun	159,500
48	Nanning	880,400	123	Zhuhai	460,600	198	Hebi	239,300	273	Shanwei	158,600
49	Wenzhou	867,200	124	Changde	457,600	199	Mianchang	238,900	274	Korla	158,400
50	Qiqihar	860,000	125	Jiangmen	457,500	200	Beihai	238,200	275	Beian	158,300
51	Huainan	859,300	126	Aomen	453,300	201	Ganzhou	236,500	276	Yichun	157,800
52	Huaibei	808,100	127	Shashi	446,100	202	Xiantao	234,800	277	Acheng	157,300
53	Xining	782,700	128	Chengde	442,300	203	Jinzhou	231,900	278	Daliang	156,700
54	Hengyang	780,700	129	Hengshui	432,500	204	Ranghulu	229,800	279	Boshan	156,600
55	Hohhot	762,700	130	Luqiao	428,600	205	Chuzhou	228,500	280	Dunhua	156,600
56	Anyang	759,900	131	Baoji	415,200	206	Linfen	226,000	281	Qianjiang	156,000

Continued.

Table A.1. *Continued.*

Rank	Name	Population	Rank	Name	Population	Rank	Name	Population	Rank	Name	Population
57	Xinxiang	757,800	132	Yangquan	402,100	207	Liaocheng	225,700	282	Leshan	155,800
58	Shenzhen	752,200	133	Zunyi	401,200	208	Saertu	225,700	283	Gaomi	155,500
59	Liuzhou	748,200	134	Jining	399,900	209	Tongchuan	224,500	284	Linhai	154,600
60	Zhaotong	742,400	135	Shiyan	398,000	210	Xintai	223,600	285	Guangshui	154,300
61	Zhuzhou	729,000	136	Dongguan	380,700	211	Guangyuan	223,100	286	Kaili	153,200
62	Hegang	724,700	137	Pingxiang	379,900	212	Yuci	221,700	287	Xinzhou	152,900
63	Langfang	721,800	138	Weifang	376,700	213	Tianmen	220,100	288	Yichun	152,700
64	Ningbo	721,100	139	Putian	371,700	214	Nanchong	219,700	289	Dingzhou	152,000
65	Zigong	709,900	140	Xuchang	371,200	215	Wuxue	219,100	290	Karamay	150,200
66	Qinhuangdao	707,200	141	Saigong	369,900	216	Yiyang	218,800	291	Xinji	149,400
67	Xiangtan	692,900	142	Xigong	365,600	217	Shuangyashan	209,300	292	Gongzhuling	147,900
68	Fuxin	687,800	143	Changji	365,400	218	Zhumadian	209,000	293	Huangyan	147,000
69	Changzhi	683,600	144	Yangjiang	365,000	219	Haicheng	206,400	294	Huadian	145,900
70	Zhangdian	679,200	145	Qitaihe	356,500	220	Sanming	205,500	295	Jinhua	145,800
71	Nantong	677,200	146	Chaozhou	354,200	221	Linxia	203,200	296	Hanzhong	145,700
72	Huangshi	671,400	147	Luohe	350,900	222	Suzhou	199,500	297	Yushan	145,700
73	Rongcheng	669,500	148	Shaoxing	350,700	223	Yuncheng	197,500	298	Fuyu	145,500
74	Yantai	652,200	149	Chaoyang	350,100	224	Hailar	197,300	299	Huicheng	145,300
75	Liaoyang	650,100	150	Anqing	346,600	225	Anda	197,000	300	Chizhou	144,300

Source: Tageo. People's Republic of China. http://www.tageo.com/index-e-ch-cities-CN.htm (accessed July 12, 2016).

Table A.2. List of Natural Cities in the People's Republic of China by Size, 2000

Rank	Name	Merged	Area (km²)	Rank	Name	Merged	Area (km²)	Rank	Name	Merged	Area (km²)
1	Guangzhou	8	8,473.02	73	Xingtai		117.00	145	Yichang		54.00
2	Beijing	1	3,042.00	74	Guiyang		108.00	146	Yining		54.00
3	Shanghai		3,027.42	75	Jinxi		108.00	147	Yizheng		54.00
4	Tianjin		1,125.00	76	Mianyang		108.00	148	Yuci		54.00
5	Ranghulu	2	1,044.00	77	Siping		108.00	149	Zaozhuang		54.00
6	Shenyang	3	990.00	78	Xiangfan		108.00	150	Acheng		45.00
7	Harbin		756.00	79	Xianyang		108.00	151	Anqing		45.00
8	Nanjing		738.00	80	Xinxiang		108.00	152	Badaojiang		45.00
9	Wuhan		711.00	81	Haikou		100.68	153	Haibowan		45.00
10	Changchun		684.00	82	Cangzhou		99.00	154	Huainan		45.00
11	Qingdao		640.41	83	Huaiyin		99.00	155	Laiyang		45.00
12	Dalian	1	638.59	84	Korla		99.00	156	Qitaihe		45.00
13	Hangzhou		621.42	85	Nanchang		99.00	157	Shaoguan		45.00
14	Urumqi		594.00	86	Pingdingshan		99.00	158	Xiangtan		45.00
15	Xian		558.00	87	Shihezi		99.00	159	Xinyang		45.00
16	Jinan		477.00	88	Xinpu		99.00	160	Yangquan		45.00
17	Taiyuan		459.00	89	Yingkou		99.00	161	Yueyang		45.00
18	Chengdu		450.00	90	Zhaoqing		99.00	162	Yulin		45.00
19	Shijiazhuang		423.00	91	Zhanjiang		92.56	163	Yuncheng		45.00
20	Shishi	1	404.31	92	Benxi		90.00	164	Zhuzhou		45.00
21	Baotou		396.00	93	Liaocheng		90.00	165	Aksu		36.00
22	Zhengzhou		387.00	94	Maanshan		90.00	166	Baiyin		36.00
23	Kunming		369.00	95	Zhangzhou		90.00	167	Boshan		36.00
24	Suzhou (JS)		360.00	96	Rizhao		86.51	168	Chuzhou		36.00
25	Chongqing		351.00	97	Bengbu		81.00	169	Fuyang		36.00
26	Shantou		346.83	98	Dandong		81.00	170	Gaomi		36.00
27	Jilin		342.00	99	Fuxin		81.00	171	Gongzhuling		36.00
28	Zhuhai	1	337.01	100	Hailar		81.00	172	Putian		36.00

Continued.

Table A.2. *Continued.*

Rank	Name	Merged	Area (km²)	Rank	Name	Merged	Area (km²)	Rank	Name	Merged	Area (km²)
29	Hohhot		324.00	101	Huzhou		81.00	173	Suihua		36.00
30	Anshan		315.00	102	Jiaozuo		81.00	174	Xinzhou		36.00
31	Yantai		307.56	103	Jining (SD)		81.00	175	Xuchang		36.00
32	Dongying		306.00	104	Panzhihua		81.00	176	Zhaodong		36.00
33	Fuzhou		306.00	105	Yancheng		81.00	177	Zhoukou		36.00
34	Wuxi		306.00	106	Yushan		81.00	178	Zhumadian		36.00
35	Hefei		297.00	107	Zhenjiang		81.00	179	Anda		27.00
36	Changsha		279.00	108	Changji		72.00	180	Dunhua		27.00
37	Ningbo		270.00	109	Chaozhou		72.00	181	Huangyan		27.00
38	Fushun		243.00	110	Guilin		72.00	182	Jingmen		27.00
39	Wenzhou		238.53	111	Huaibei		72.00	183	Jining (NM)		27.00
40	Tanggu		233.44	112	Jixi		72.00	184	Kaiyuan		27.00
41	Changzhou		225.00	113	Kashi		72.00	185	Linhai		27.00
42	Datong		225.00	114	Linfen		72.00	186	Suzhou (AH)		27.00
43	Xuzhou		225.00	115	Nanyang		72.00	187	Tongling		27.00
44	Baoding		216.00	116	Shangqiu		72.00	188	Wafangdian		27.00
45	Handan		207.00	117	Shaoxing		72.00	189	Weinan		27.00
46	Nanning		207.00	118	Taizhou		72.00	190	Wuzhou		27.00
47	Zhangdian		180.00	119	Yanji		72.00	191	Yichun (HL)		27.00
48	Huizhou		171.00	120	Baicheng		63.00	192	Bozhou		18.00
49	Liaoyang		171.00	121	Baoji		63.00	193	Deyang		18.00
50	Linyi		171.00	122	Changde		63.00	194	Fuling		18.00
51	Luoyang		171.00	123	Chifeng		63.00	195	Hangu		18.00
52	Panjin		171.00	124	Dongchang		63.00	196	Huadian		18.00
53	Weifang		171.00	125	Hengshui		63.00	197	Jiujiang		18.00
54	Yinchuan		171.00	126	Huangshi		63.00	198	Maoming		18.00
55	Jiaojiang	1	169.39	127	Hulan		63.00	199	Mianchang		18.00
56	Langfang		162.00	128	Jinhua		63.00	200	Sucheng		18.00

Continued.

Table A.2. *Continued.*

Rank	Name	Merged	Area (km²)	Rank	Name	Merged	Area (km²)	Rank	Name	Merged	Area (km²)
57	Qinhuangdao		159.86	129	Karamay		63.00	201	Xiaogan		18.00
58	Fuyu		153.00	130	Luohe		63.00	202	Xinji		18.00
59	Anyang		144.00	131	Tieling		63.00	203	Zunyi		18.00
60	Puyang		144.00	132	Tongliao		63.00	204	Shanwei		16.93
61	Qiqihar		144.00	133	Zouxian		63.00	205	Chenzhou		9.00
62	Xiamen		140.68	134	Beihai		62.86	206	Dingzhou		9.00
63	Xining		135.00	135	Yangjiang		61.26	207	Huanggang		9.00
64	Yangzhou		135.00	136	Chengde	1	54.00	208	Jiutai		9.00
65	Dezhou		126.00	137	Haicheng		54.00	209	Luancheng		9.00
66	Jiamusi		126.00	138	Heze		54.00	210	Suizhou		9.00
67	Jinzhou (LN)		126.00	139	Jiaozhou		54.00	211	Wanxian		9.00
68	Mudanjiang		126.00	140	Jincheng		54.00	212	Xintai		9.00
69	Jiaxing		117.00	141	Jingzhou	1	54.00	213	Zhangye		9.00
70	Rongcheng		117.00	142	Liaoyuan		54.00				
71	Taian		117.00	143	Wulanhaote		54.00				
72	Wuhu		117.00	144	Xuanhua		54.00				

AH = Anhui Province, HL = Heilongjiang Province, JS = Jiangsu Province, km² = square kilometer, LN = Liaoning Province, NM = Inner Mongolia Autonomous Region, and SD = Shandong Province.

Notes: "Merged" reports the total number of cities, if any, that merged into the respective city listed under "Name." Merged cities in rank order are as follows: 1. Guangzhou: Daliang, Dongguan, Foshan, Huicheng, Jiangmen, Shenzhen, Shiongshui, and Zhongshan; 2. Beijing: Mentougou; 5. Ranghulu: Longfeng and Saertu; 6. Shenyang: Dongling, Sujiatun, and Yuhong; 12. Dalian: Jinzhou (Dalian); 20. Shishi: Quanzhou; 28. Zhuhai: Aomen; 55. Jiaojiang: Luqiao; 136. Chengde: Chaoyang; 141. Jingzhou: Sashi. The total number of natural cities and merged natural cities may be less than 300 as some cities did not meet the radiance threshold (40) in 2000. The area of the natural cities is calculated using the number of places per city, where one grid has a maximum size of 3 kilometers × 3 kilometers (and is less if located at the country's border).

Source: Authors' calculations.

Table A.3. **Estimation Results for Alternative Non-Tobit Models**

	OLS radiance$_{it}$ (1)	OLS radiance$_{it}$ (2)	OLS radiance$_{it}$ (3)
radiance$_{it-1}$			0.609***
			(0.002)
radiance$_{it-2}$			0.194***
			(0.002)
radiance$_{it-3}$			0.100***
			(0.002)
radiance$_{i1992}$			0.019***
			(0.001)
ln(popdens$_{i1990}$)			0.098***
			(0.008)
ln(dist to road$_i$)	−2.851***	−2.195***	−0.185***
	(0.066)	(0.063)	(0.006)
ln(dist to rail$_i$)	−2.915***	−2.405	−0.044***
	(0.068)	(0.067)	(0.007)
ln(dist to ocean$_i$)		−1.434***	−0.090***
		(0.092)	(0.008)
ln(dist to water$_i$)		−1.417***	−0.080***
		(0.072)	(0.007)
ln(dist to center$_i$)		−7.130***	−0.159***
		(0.143)	(0.013)
ln(dist to adborder$_i$)		−0.404***	−0.094**
		(0.138)	(0.046)
ln(altitude$_i$)		−0.261*	−0.066***
		(0.145)	(0.013)
longitude$_i$		−0.026	−0.004
		(0.034)	(0.003)
latitude$_i$		1.183***	0.044***
		(0.131)	(0.011)
ln(rain$_i$)		−2.387***	−0.160***
		(0.561)	(0.047)
temperature$_i$		1.183***	0.060***
		(0.120)	(0.011)
sd temperature$_i$		−1.633***	−0.066
		(0.186)	(0.013)
hukou$_{it}$		−1.966***	−1.502***
		(0.253)	(0.079)
small$_{it}$		3.059***	1.667***
		(0.257)	(0.079)
medium$_{it}$		2.094***	1.093***
		(0.231)	(0.071)
large$_{it}$		−0.480**	−0.807***
SEZ$_{it}$		0.520***	0.769***
		(0.179)	(0.118)
firstwave$_{it}$		−2.192***	−0.245***
		(0.311)	(0.059)

Continued.

Table A.3.   *Continued.*

	OLS radiance$_{it}$ (1)	OLS radiance$_{it}$ (2)	OLS radiance$_{it}$ (3)
secondwave$_{it}$		0.003	−0.214***
		(0.220)	(0.032)
thirdwave$_{it}$		0.401***	−1.468***
		(0.144)	(0.222)
Constant	64.250***	140.600***	3.105***
	(1.652)	(10.220)	(0.684)
Place fixed effects	Yes	Yes	Yes
Satellite effects	Yes	Yes	Yes
Control variables		Yes	Yes
Lagged variables			Yes
Observations	432,674	432,674	373,673
Places	19,667	19,667	19,667

OLS = ordinary least squares, sd = standard deviation, SEZ = Special Economic Zone.
Notes: Reported coefficients are marginal effects. Standard errors are reported in parentheses. *** = p < 0.01, ** = p < 0.05, and * = p < 0.1. Squared terms are included for the following geography and climate variables: ln(dist to road$_i$), ln(dist to rail$_i$), ln(dist to ocean$_i$), ln(dist to water$_i$), ln(dist to adborder$_{it}$), ln(dist to center$_i$), ln(altitude$_i$), ln(rain$_i$), temperature$_i$, and sd temperature$_i$. All distance measures in the empirical estimation are in meters.
Source: Authors' calculations.

Table A.4.   **Estimation Results for Dynamic Tobit by Categories**

	(1) radiance$_{it}$ Nat = 1, Admin = 1	(2) radiance$_{it}$ Nat = 1, Admin = 1	(3) radiance$_{it}$ Nat = 1, Admin = 1	(4) radiance$_{it}$ Nat = 1, Admin = 1
radiance$_{it-1}$	0.487***	0.530***	0.263***	0.226***
	(0.005)	(0.003)	(0.003)	(0.002)
radiance$_{it-2}$	0.167***	0.142***	0.106***	0.090***
	(0.005)	(0.003)	(0.003)	(0.002)
radiance$_{it-3}$	0.024***	0.030***	0.090***	0.124***
	(0.005)	(0.003)	(0.003)	(0.002)
radiance$_{i1992}$	0.161***	0.207***	0.079***	0.092***
	(0.006)	(0.004)	(0.003)	(0.002)
ln(popdens$_{i1990}$)	0.180***	0.084***	0.103**	0.094***
	(0.044)	(0.027)	(0.040)	(0.022)
ln(dist to road$_i$)	−0.324***	−0.315***	−0.294***	−0.387***
	(0.025)	(0.016)	(0.022)	(0.016)
ln(dist to rail$_i$)	−0.004	−0.131***	−0.083***	−0.046***
	(0.028)	(0.018)	(0.025)	(0.017)
ln(dist to ocean$_i$)	−0.342***	−0.245***	−0.320***	−0.236***
	(0.040)	(0.021)	(0.035)	(0.022)
ln(dist to water$_i$)	−0.118***	−0.140***	−0.113***	−0.061***
	(0.027)	(0.018)	(0.026)	(0.018)

Continued.

Table A.4.   *Continued.*

	(1) radiance$_{it}$ Nat = 1, Admin = 1	(2) radiance$_{it}$ Nat = 1, Admin = 1	(3) radiance$_{it}$ Nat = 1, Admin = 1	(4) radiance$_{it}$ Nat = 1, Admin = 1
ln(dist to center$_i$)	−0.268***	0.123***	0.219***	−0.290***
	(0.055)	(0.040)	(0.044)	(0.038)
ln(dist to adborder$_i$)	−0.003	0.036	0.025	−0.026
	(0.031)	(0.022)	(0.029)	(0.021)
ln(altitude$_i$)	0.159***	0.039	−0.188***	−0.364***
	(0.048)	(0.037)	(0.052)	(0.039)
longitude$_i$	−0.040***	0.014*	−0.012	0.025***
	(0.010)	(0.008)	(0.010)	(0.008)
latitude$_i$	0.050	0.090***	0.275***	−0.009
	(0.037)	(0.024)	(0.038)	(0.025)
ln(rain$_i$)	−0.561***	−0.778***	−0.636***	−1.503***
	(0.173)	(0.122)	(0.189)	(0.130)
temperature$_i$	0.147***	0.220***	0.151***	−0.101***
	(0.034)	(0.029)	(0.035)	(0.031)
sd temperature$_i$	0.031	−0.056	−0.272***	−0.159***
	(0.060)	(0.040)	(0.062)	(0.043)
hukou$_{it}$	−1.068***	−2.844***	1.398***	−0.122
	(0.302)	(0.165)	(0.272)	(0.090)
small$_{it}$	0.985***	3.073***	−0.559**	0.822***
	(0.281)	(0.162)	(0.252)	(0.085)
medium$_{it}$	0.156	1.573***	−0.708***	0.331***
	(0.223)	(0.112)	(0.243)	(0.077)
large$_{it}$	−0.188	−1.034***	0.190	−0.930***
	(0.203)	(0.102)	(0.227)	(0.078)
SEZ$_{it}$	1.323***	0.476***	1.076***	−0.244***
	(0.192)	(0.086)	(0.173)	(0.089)
firstwave$_{it}$	−3.040***	−0.198	−2.047***	0.835***
	(0.388)	(0.159)	(0.293)	(0.142)
secondwave$_{it}$	−0.810***	−0.471***	−1.362***	−0.362***
	(0.185)	(0.072)	(0.169)	(0.072)
thirdwave$_{it}$	−0.380***	0.224***	0.534***	0.688***
	(0.135)	(0.066)	(0.120)	(0.069)
Constant	11.90***	21.32***	25.09***	15.14***
	(3.227)	(3.970)	(2.600)	(2.921)
Satellite effects	Yes	Yes	Yes	Yes
Observations	51,995	55,785	159,772	106,121
Places	5,957	4,781	12,330	14,040

Nat = natural city, Admin = within administrative boundary, sd = standard deviation, SEZ = Special Economic Zone.

Notes: Reported coefficients are marginal effects. Standard errors are reported in parentheses. *** = p < 0.01, ** = p < 0.05, and * = p < 0.1. Squared terms are included for the following geography and climate variables: ln(dist to road$_i$), ln(dist to rail$_i$), ln(dist to ocean$_i$), ln(dist to water$_i$), ln(dist to adborder$_{it}$), ln(dist to center$_i$), ln(altitude$_i$), ln(rain$_i$), temperature$_i$, and sd temperature$_i$. All columns include satellite effects. All distance measures in the empirical estimation are in meters.

Source: Authors' calculations.

Table A.5.   **Robustness—Transition Matrix of Counterfactual ln(dist to road$_i$)**

Definition: Nat $= 1$ if radiance > 50, Nat $= 0$ if radiance < 50

| | | Counterfactual: ln(dist to road$_i$) | | |
		Nat $= 1$	Nat $= 0$	Total
Baseline	Nat $= 1$	$99.87_{(2007)}$	$0.13_{(2007)}$	100
		$99.70_{(2012)}$	$0.30_{(2012)}$	100
		$99.51_{(2017)}$	$0.49_{(2017)}$	100
		$99.48_{(2022)}$	$0.52_{(2022)}$	100
		$99.50_{(2027)}$	$0.50_{(2027)}$	100
	Nat $= 0$	$0.49_{(2007)}$	$99.51_{(2007)}$	100
		$2.77_{(2012)}$	$97.23_{(2012)}$	100
		$7.94_{(2017)}$	$92.06_{(2017)}$	100
		$18.58_{(2022)}$	$81.42_{(2022)}$	100
		$35.03_{(2027)}$	$64.97_{(2027)}$	100
	Total	$28.71_{(2007)}$	$71.29_{(2007)}$	100
		$39.02_{(2012)}$	$60.98_{(2012)}$	100
		$51.50_{(2017)}$	$48.50_{(2017)}$	100
		$67.11_{(2022)}$	$32.89_{(2022)}$	100
		$82.16_{(2027)}$	$17.84_{(2027)}$	100

Nat $=$ natural city.
Note: In addition to the radiance threshold, the City Clustering Algorithm condition is a necessary condition for a place to be assigned as a natural city (Nat $= 1$). The City Clustering Algorithm condition implies that a place is near a cluster of places with average radiance greater or equal to the threshold.
Source: Authors' calculations.

Table A.6. **Robustness—Transition Matrix of Counterfactual**
ln(dist to rail$_i$)

Definition: Nat $= 1$ if radiance > 50, Nat $= 0$ if radiance < 50

		Counterfactual: ln(dist to rail$_i$)		
		Nat $= 1$	**Nat $= 0$**	**Total**
Baseline	Nat $= 1$	$100.00_{(2007)}$	$0.00_{(2007)}$	100
		$99.93_{(2012)}$	$0.07_{(2012)}$	100
		$99.88_{(2017)}$	$0.12_{(2017)}$	100
		$99.82_{(2022)}$	$0.18_{(2022)}$	100
		$99.84_{(2027)}$	$0.16_{(2027)}$	100
	Nat $= 0$	$0.06_{(2007)}$	$99.94_{(2007)}$	100
		$0.52_{(2012)}$	$99.48_{(2012)}$	100
		$1.54_{(2017)}$	$98.46_{(2017)}$	100
		$3.65_{(2022)}$	$96.35_{(2022)}$	100
		$6.31_{(2027)}$	$93.69_{(2027)}$	100
	Total	$28.43_{(2007)}$	$71.57_{(2007)}$	100
		$37.70_{(2012)}$	$62.30_{(2012)}$	100
		$48.32_{(2017)}$	$51.68_{(2017)}$	100
		$61.34_{(2022)}$	$38.66_{(2022)}$	100
		$74.68_{(2027)}$	$25.32_{(2027)}$	100

Nat $=$ natural city.

Note: In addition to the radiance threshold, the City Clustering Algorithm condition is a necessary condition for a place to be assigned as a natural city (Nat $= 1$). The City Clustering Algorithm condition implies that a place is near a cluster of places with average radiance greater or equal to the threshold.

Source: Authors' calculations.