The Impact of Transportation Infrastructure on Economic Growth: Evidence from China

Shuting Chen

University of Chicago

Thesis submitted in partial fulfillment of the requirements of the Master of Arts in Computational Social Science

Advised by Dr. Richard Evans

Contents

Abstract

This thesis explores the impact of having access to transportation infrastructure on regional economic growth, based on economic outcomes for non-metropolitan cities in China from 1997 to 2011. With the use of "straight line" identification strategy, this thesis finds that cities more distant to the constructed straight lines have less access to transportation networks. Further, results in empirical analysis suggest that there exists a positive effect of transportation infrastructure on economic growth, and the magnitude of the effect becomes larger when using night lights as the measure of economic outcomes.

1 Introduction

During the last two decades, developing countries have spent an enormous amount of investment on transportation infrastructure projects shaping their cities for decades to come. In 2007, about 20 percent of World Bank lending — which was larger than the total proportion of that on health, education, and social services — was allocated to transportation infrastructure projects (World Bank, 2007b). Particularly, transportation infrastructure has been treated as a key to facilitating economic growth and development. The historical fact that the construction of transportation infrastructure such as highways and railroads coincided with rapid economic growth in western countries provides appealing evidence for this argument (Banerjee et al., 2012). Nevertheless, when considering the causality of transportation infrastructure on economic growth, it is likely that transportation infrastructure is endogenous to the development process. For example, the huge construction of railroads could be the engine of economic growth or it could simply arise in response to development demand. Fogel (1962, 1964) thoroughly argued that the effect of innovations in transportation infrastructure on economic development was exaggerated in the United States, and that it actually was a result of government policies for boosting railroads.

Consequently, this paper explores the impact of having access to transportation infrastructure on regional economic growth, based on economic outcomes for non-metropolitan cities in China from 1997 to 2011. More Specifically, to address the problem of endogeneity on the placement of transportation networks, this paper follows the "straight line" identification strategy proposed by Banerjee et al. (2012). Instead of drawing a straight line between each historically important city and the nearest Treaty Port[1]/another historically important city, this paper adopts the strategy of drawing a straight line from one provincial capital city to the nearest capital city in another province/Treaty Port. This seems plausible as these lines likely capture most transportation networks and we only consider non-metropolitan cities. Using the distance to the nearest straight line as an instrument of measuring how much access to transportation infrastructure a city has, it is possible to examine the correlation or even causality between transportation networks and economic growth.

This paper conducts empirical analysis not only using prefecture cities' official economic data from Chinese Provincial Statistical Yearbooks but also using night lights data collected by U.S. Air Force Defense Meteorological Satellite Program (DMSP). Night lights remote sensing data has been widely used as a proper proxy for economic activity in recent literature (Harari, 2020; Henderson et al., 2012; Hodler and Raschky, 2014; Storeygard, 2016), which provides the potential to alleviate measurement errors in conventional data as well as concerns about Chinese official government statistics (Koch-Weser, 2013).

Under the assumptions of factor mobility (i.e. complete immobility of labor and capital is assumed to move at a cost), the underlying hypothesis is that non-metropolitan cities that are closer to the nearest straight lines more likely experi-

[1] Shanghai, Ningbo, Fuzhou and Guangzhou were four Treaty Ports chosen by the League of Eight Nations after Qing government's defeat in the First Opium War in 1842.

ence high economic growth than those far away. This seems reasonable since we use distance as an instrument for measuring the level of access to transportation infrastructure. The closer the cities are to the proposed straight line the more access they have to transportation networks. Based on our results in empirical analysis, more distant cities on average tend to have less access to transportation networks. Further, transportation infrastructure has a positive impact on economic growth and the magnitude of effect is larger when we use night lights data. This might attribute to fewer measurement errors embedded in night lights data than traditional Chinese government economic data.

The remainder of this thesis is structured as follows. The relevant literature is reviewed in Section 2. The multiple sources of data and adopted empirical methods are introduced in Section 3. Elaborated empirical analysis and results are discussed in Section 4. Section 5 concludes and discusses potential research directions.

2 Literature Review

This section elaborates recent literature relative to transportation infrastructure and economic growth. Particularly, this project emphasizes on three aspects: 1) How did previous researchers evaluate the impact of transportation infrastructure on economic outcomes? 2) What has been done by using night lights remote sensing data for economic activities, especially for economic growth? 3) What is the evolution for economic development and transportation construction in China over the past two decades?

2.1 Transportation infrastructure on economic outcomes

Since there are various reasons as to why strong transportation infrastructure may have positive impact on economic development (e.g. reducing trade costs and promoting market integration, a significant amount of research has been conducted in this field. In terms of identification strategies to evaluate the impact of transportation infrastructure, previous literature could be classified into research using "straight-line" instruments or not.

As previously mentioned, Banerjee et al. (2012) (who utilized the same country of interest) proposed an identification strategy by drawing a straight line connecting each historically important city to the nearest Treaty Port/other historically important city. This innovative exogenous source of networks construction relies on the fact that placement of networks in China tend to connect historical cities (Banerjee et al., 2012). Further, they constructed a theoretical trade model, which aims to understand the effects of distance on the mobility of goods as well as that of factors of production (i.e. labor and capital). One primary prediction of their theoretical model is that as long as the capital is more mobile than goods (with the assumption that labor is completely immobile), output per capita would always be higher and inequality lower in the regions closer to straight lines. This presumably provides a channel through which the construction of networks not only improves the efficiency of trade but also increases factor mobility. In fact, this idea has been demonstrated by Atack et al. (2010) finding that access to railroads has a significant positive impact on urbanization in the United States and Baum-Snow et al. (2017) investigating how urban transportation networks has shaped the urban form of Chinese cities since 1990. Similarly, Faber (2009) adopted both a straight-line instrument as well as estimates of transportation costs to investigate the influence of recently constructed highways in China on economic development.

Alternatively, a growing number of recent studies have proposed other appealing identification strategies to assess the effect of transportation infrastructure on economic performance (Hong et al., 2011; Yu et al., 2012; Beyzatlar et al., 2014; Sun and Cui, 2018). Michaels (2008) and Donaldson (2018) both employed a difference-in-difference (DD) approach to explore the effect of transportation networks; the former focused on the impacts of highway construction in the United States in the 1950s, while the latter studies the effects of railroad networks in 19th century India.

Moreover, Michaels (2008) also used an instrumental variables approach to extract the variation in access resulting from the fact that highways were more likely to be built in either a North-South direction or a East-West direction. Similarly, Keller and Shiue (2008) adopted instrumental variables approaches to study the opening up of railroads among fifteen states in Germany. Instead of from the point of view of factor mobility, these papers studied the effect of transportation infrastructure from the perspective of market integration, price convergence, and relative price of factors. The overall results show that transportation infrastructure enhances market integration and increases the degree of price convergence.

As mentioned in Section 1, this paper slightly modifies the "straight line" identification strategy developed by Banerjee et al. (2012) by drawing straight lines from one provincial capital city to the nearest capital city in another province or Treaty Port. This is not only because of their well-defined identification strategy and reliable empirical results, but also taking the similar scenario of research questions into account. Other than examining the effect of transportation networks on the level of per capita Gross Domestic Product (GDP), this paper evaluates the impact of transportation infrastructure on economic growth by taking the natural logarithm of per capita GDP. This can be treated as another contribution of this paper to previous literature since Banerjee et al. (2012) focused on examining the effect of transportation networks on the level of GDP rather than the growth. The details of empirical methods will be elaborated in Section 3.

2.2 Night lights data and economic growth

The second related literature is about the adoption of remote sensing data, especially night lights data, in evaluating economic performance. One might be curious about why we would like to use remotely sensed data and how this new source of data has been used in economics so far. As argued by Donaldson and Storeygard (2016), there are three main advantages of using such satellite data in economics. First, the advanced technologies of remote sensing make it feasible and affordable to collect a large scale of panel data, some of which cover a wide range of hard-to-measure characteristics. Second, comparing to traditional data, they are usually available at a substantially higher level of spatial resolution. Figure 1 in Bleakley and Lin (2012) represents a remarkable example of exploiting the spatial distribution of economic activity. The clearly visible pattern of night lights provides strong evidence to demonstrate "the causal effect of historical portage sites on the location of economic activity today". However, it seems impossible to investigate this causality by conventional US data sources. The third major advantage of remote sensing data is that data collected in the consistent manner with wide geographic coverage are accessible now. This eliminates the barriers for previous researches that were impeded due to the inconsistency of data within the same category from region to region.

Night lights data, which have been collected primarily by the Defense Meteorological Satellite Program–Operational Linescan System (DMSP-OLS) since 1992, have played an important role in assessing economic activity. This argument relies on the

Source: Bleakley and Lin 2012.
Notes: The figure shows night lights across part of the United States in 2003, illustrated with rivers (dotted lines) and the fall line (solid line). Substantial cities on the fall line, starting from the northeast, include Trenton, Philadelphia, Washington DC, Richmond, Augusta, Columbia, Little Rock, Fort Worth, Austin, and San Antonio.

fact that total visible light observed from Earth's surface at night has been widely treated as a proxy for local economic activity, with the assumption that lighting is a normal good. Several reasons might be accounted for the growing usage of night lights data in economics. The first reason is that these data are now much easier to access and process than decades ago. Another primary reason comes from a large demand of high-quality data. Deaton and Heston (2010) expressed strong concern and warning about the uncertainty surrounding many estimates of income after they conducted international comparisons using frequently-used global data from the Penn World Table (PWT) and the World Bank's World Development Indicators. Other researchers (Chen and Nordhaus, 2011; Johnson et al., 2013; Dawson et al., 2001) illustrated the relatively high degree of measurement error in the PWT by studying different revisions to the PWT data. Other than the issues of measurement error, another intractable problem is that most countries do not have consistent data for important macroeconomic variables such as GDP at the sub-national level.

In response to the issues of precisely measuring macroeconomic variables, especially for GDP, the amount of light observed from outer space has been commonly used as a proxy for economic activity (Doll et al., 2006; Elvidge et al., 1997; Henderson et al., 2012; Lee, 2018; Michalopoulos and Papaioannou, 2013a, 2013b). This accounts for the underlying reason why this paper plans to use night lights remote sensing data. The formal econometric scrutiny for the accuracy of lights as a proxy for economic growth has only been provided recently by Henderson et al. (2012). By considering various functional forms and controls for changes in dispersion of lights, the authors concentrated on the differences among GDP growth forecasts by using different sources of data (that is night lights data versus traditional GDP data or combined). The results suggest that the optimal estimate of GDP growth is the combination of equally weighting conventionally measured growth and growth measured by night lights data. Consequently, this paper follows their approaches, which can not only alleviate the impact of measurement errors and inconsistency in conventional

GDP data at sub-national level, but also offer the potential to find new insights for the impact of transportation infrastructure on economic growth with new data.

2.3 Economic growth and transportation infrastructure in China

Since China is the country of interest in this paper, the third related literature is about the evolution of economic growth and transportation infrastructure in China. In the pre-reform era (1949-1976), central planning policies made urban areas inherit relatively more endowment of human and physical capital than rural areas. As Chinese government has always paid strong attention to economic performance and industrialization, the proportion of government revenues spent on funding industrial development increased from around 30% in 1952 to around 60% in 1957 (Eckstein, 1977). However, restrictions on migration substantially inhibited the mobility of unskilled labor during the post-reform period and little financial development likely limited mobility of capital (West and Zhao, 2000). Although pre-reform China had experienced annual growth of 6 percent a year with some fluctuations, the post-reform China witnessed an more than 9 percent average real growth a year with fewer ups and downs (Hu and Khan, 1997; Cai, 2010). More importantly, labor and capital accumulation accounted for only about 58 percent of post-reform economic growth, while increases in productivity made up the rest. Currently, the annual growth target in China's 13th Five-Year Plan (2016 - 2020) is 6.5 percent, with the goal of balancing high speed of economic growth and the quality of growth (World Bank, 2018). In addition to achieve sustainable economic growth, China's central policymakers have focused more on energy safety and environmental protection (Song et al., 2018).

For the evolution of transportation infrastructure in China over the past two decades, one of the most impressive policies is China's National Trunk Highway System (NTHS). The objective of NTHS was to connect capital cities, cities with an urban population over 500,000, and border crossings on a single expressway network (World Bank, 2007a). This network construction originally planned to finish by 2020; however, it was completed ahead of schedule by the end of 2007 since highway construction became a source of the government's stimulus spending after the Asian financial crisis (Asian Development Bank, 2007). Faber (2009) explored this policy as a large scale natural experiment to understand whether the resulting trade cost reductions are a driver for the diffusion of industrial activity. It turns out that large scale inter-regional transportation network likely leads to a reduction in industrial and total output growth. Other than the policy of NTHS, the high-speed rail (HSR) network introduced in 2007 has significantly facilitated regional accessibility and connectivity (Wang and Duan, 2018).

3 Data and Methods

3.1 Data

Compared to previous literature in measuring the impact of transportation infrastructure on economic growth, one contribution of this paper is that it improves measurement accuracy by compiling two sources of data for measuring economic growth. The first source of data is prefecture cities' official economic data collected from Chinese Provincial Statistical Yearbooks and the second source of data is night lights data obtained by U.S. Air Force DMSP. To determine the geographical location of each prefecture city in China, this paper employs GIS Maps of 2010 China Prefecture Population Census Data as another source of data.

3.1.1 Night Lights Data

Night lights data have been collected by the United States Air Force DMSP using Operational Linescan System sensors[2] since the 1970s and they became openly digital available in 1992 (Henderson et al., 2012). This paper employs 26 satellite-year datasets (Average Visible, Stable Lights, & Cloud Free Coverages) from 1997 to 2011[3], including approximately 110 million satellite-year pixels for 178 prefecture cities in China. They are grid-based datasets and every 30 arc-second output pixel has been labeled by a digital number (an integer between 0 and 63), representing the

[2] Night lights data are publicly available at: `https://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html`

[3] Specifically, this paper uses data from satellite F12 for 1997-1999, F14 for 1997-2003, F15 for 2000-2007, F16 for 2004-2009, and F18 for 2010-2011.

intensity of lights (0 for no light) at nights for human settlements excluding sunlight, moonlight, aurorae, forest fires, and clouds.

As mentioned previously, night lights data have been frequently used in recent research for economic activity. Henderson et al. (2012) who provided the first formal econometric framework of using lights as a proxy for economic growth, demonstrated that the lights account for 20% of the variation in GDP growth, net of country and year fixed effect. Instead of focusing on country-level lights data, Harari (2020) and Storeygard (2016) used night lights data as a measure of city-level GDP for India and Sub-Saharan African countries respectively. Based on the verification that the strong relationship between light usage growth and income growth at the country level does hold at the city level, the authors converted the original pixel-level data into cities. Storeygard (2016) also illustrated that the elasticity of GDP with respect to light at the city level behaves in a similar manner as that at the national level in China. Therefore, this paper adopts light growth as a proxy for GDP growth at the city level. To convert the pixel-level lights data into cities, this paper follows the steps elaborated in Storeygard (2016), involving compiling 26 satellite-year data into a binary grid and converting ever-lit regions to polygons. Details about data conversion will be elaborated in Section 4.1.1.

3.1.2 Chinese Government Economic Data

The conventional local government data for economic activities are from Provincial Statistical Yearbooks published annually by each province in China. These panel data contain very comprehensive statistics from almost all social and economic aspects including but not limited to economic growth, population and employment, investment in fixed assets, prices, industry and energy. Data in a specific year contain more than 50 city-control variables for approximately 350 prefecture cities. In particular, this paper uses city-level data for GDP, the length of highway, the length of railway, and city-specific controls such as land area and resident population from 1997 to 2011[4]. The main issue of using these data is that they may have been measured using different techniques across provinces and over time. Hence, I conduct data cleaning work such as changing the units of measurement and modifying some variables with a consistent definition before implementing empirical analysis.

3.1.3 Spatial Data

To draw the straight lines from one provincial capital city to another provincial capital city/Treaty Port and measure the distance between each target city and the nearest straight line, this paper uses 2010 China Prefecture Population Census Data with GIS Maps in digital format obtained from the University of Michigan China

[4] Almost all Provincial Statistical Yearbooks are available up to 2017, however since night lights data are only publicly available up to 2013, this paper plans to consider 1997 - 2011 as the study period.

Data Center (CDC)[5]. As described in the data documentation, the GIS data is based on the national digital map (1:1 million scale, i.e. 1 cm on the map equals 10km on the ground) including rivers, roads, residential area and administrative boundaries; all city GIS maps are constructed for matching 2010 China population Census data covering all 31 municipalities, provinces and autonomous regions, except for Taiwan, Hong Kong and Macao. Specifically, 2010 China population Census data (city-level version) are cross-sectional data for 346 prefectures with 523 variables in short form and 351 variables in long form[6], respectively. A different version of similar GIS/map data were used in Banerjee et al. (2012)[7], which were from the University of Michigan China Data Center as well. Furthermore, all geographic measures such as defining centroids of cities and measuring the nearest distance from each centroid to the straight line have been constructed using R/QGIS software.

[5] CDC has no longer been available since September 2018 but previous data are still available at "Map Collection" in the Joseph Regenstein Library.

[6] These variables include family households, population and sex ratio by city, total employed population by industry and thrice industrial ratio by city, employed population by sex, medium-scale occupation and city and so on.

[7] The authors did not specify which GIS/map data they used, however, it is likely a similar one matching for the latest China population Census data at the time they wrote the paper.

3.2 Methods

The main issue of examining the impact of having access to transportation infrastructure on economic growth is about how to properly deal with endogeneity on the placement of transportation networks. One may argue that a region's high rate of economic growth likely results from its inherent properties such as it is a metropolitan city. Urban planners likely give priority to these regions when considering transportation infrastructure. Consequently, it is difficult to isolate the pure effect of having access to transportation networks on economic growth.

To address the endogeneity on the placement of transportation infrastructure, Banerjee et al. (2012) proposed a "straight line" identification strategy, drawing a straight line between each historically important city and the nearest Treaty Port/other historically important city in China. For each county in their samples, the nearest geographic distance from the county to a proposed straight line is the main source of exogenous variation for having access to transportation infrastructure. Since the terminal regions, which are historically important cities or Treaty Ports, are not chosen by considering existing transportation infrastructure and those proposed lines do well as proxy for transportation infrastructure, it is reasonable to treat the "straight line" as a good identification strategy. Hence, this paper follows their logic with the modification of drawing lines from one provincial capital city to the nearest Treat Port/another provincial capital city. This seems plausible because the construction of Chinese transportation infrastructure does not heavily depend on the connection among provincial capital cities.

First, to check whether the proposed straight lines work well as instruments for transportation infrastructure, this paper applies the following first stage equation:

$$TI_{cpt} = \alpha lnD_{cp} + \omega X_{ct} + \gamma_p + \delta_t + \varepsilon_{cpt} \tag{1}$$

where TI_{cpt} denotes transportation infrastructure of city c in province p in year t, D_{cp} denotes the distance to the nearest straight line between one provincial capital city and Treaty Port/another provincial capital city, X_{ct} denotes a vector of city-year fixed effects controlling for any city-level time-varying economic conditions constant throughout each city, γ_p represents province fixed effects, δ_t represents year fixed effects, and ε_{cpt} is the error term. Equation (1) are not in differences as we need to control for province fixed effects. In particular, this paper estimates the correlation between D_{cp} and TI_{cpt} with various specifications of transportation networks such as the length of highway, the length of railway and the overall measure as a function of both (i.e. different cases for dependent variable TI_{cpt}). The purpose of using different measures of transportation is to assess the robustness of D_{cp} as an instrument for transportation infrastructure.

The baseline specification for the second stage in empirical analysis is:

$$\triangle lny_{cpt} = \beta \triangle ln\hat{T}I_{cpt} + \omega \triangle X_{ct} + \triangle\delta_t + \triangle\varepsilon_{cpt} \tag{2}$$

where y_{cpt} denotes the economic outcome for city c in province p in year t, $\hat{T}I_{cpt}$ is the predicted length of transportation infrastructure obtained from equation (1), and other variables remain the same as described in equation (1). The standard errors are clustered at the city level. By taking the first difference of lny_{cpt}, we can evaluate the impact of transportation infrastructure on year to year economic growth. Specifically, this paper considers two cases of y_{cpt}: 1) y_{cpt} is the real per capita GDP as measured by data in Provincial Statistical Yearbooks; 2) y_{cpt} is the light output (i.e. the annual total night lights) for city c in province p in year t.

The identifying assumption for interpreting β as the causal effect of distance to the line on GDP growth is that the proximity to the line is the only difference between places close to the line and places far away. This means that we should not select the terminal cities so that the proposed straight lines only path through economically important regions. Hence, this paper draws lines among provincial capital cities and historical Treaty Ports, which cover most transportation networks in China. Moreover, the terminal cities are excluded in this study since their economic growth is usually spurred by extra support.

4 Analysis/Results

4.1 Data Analysis

4.1.1 Night Lights Data

The original 26 satellite-year datasets (Average Visible, Stable Lights, & Cloud Free Coverages, from 1997 to 2011) are stored in raster graphics images. To extract

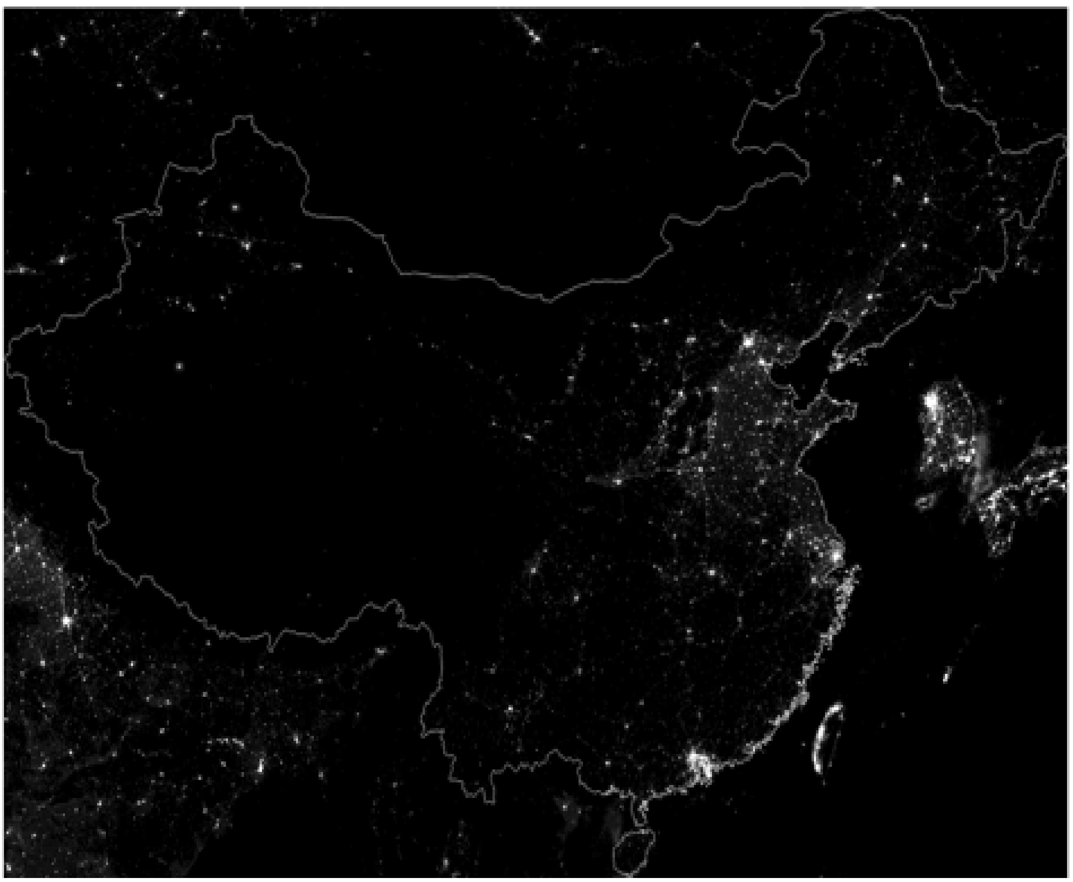


Figure 1: Lights at Night in China, 1997

Source: DMSP data collected by the United States Air Force.

these digital numbers and aggregate them into city level, I first crop images by the national-level GIS map of 2010 China Prefecture Population Data in QGIS3. Figure 1 and 2 display images of lights at night in China in 1997 and 2011, respectively.

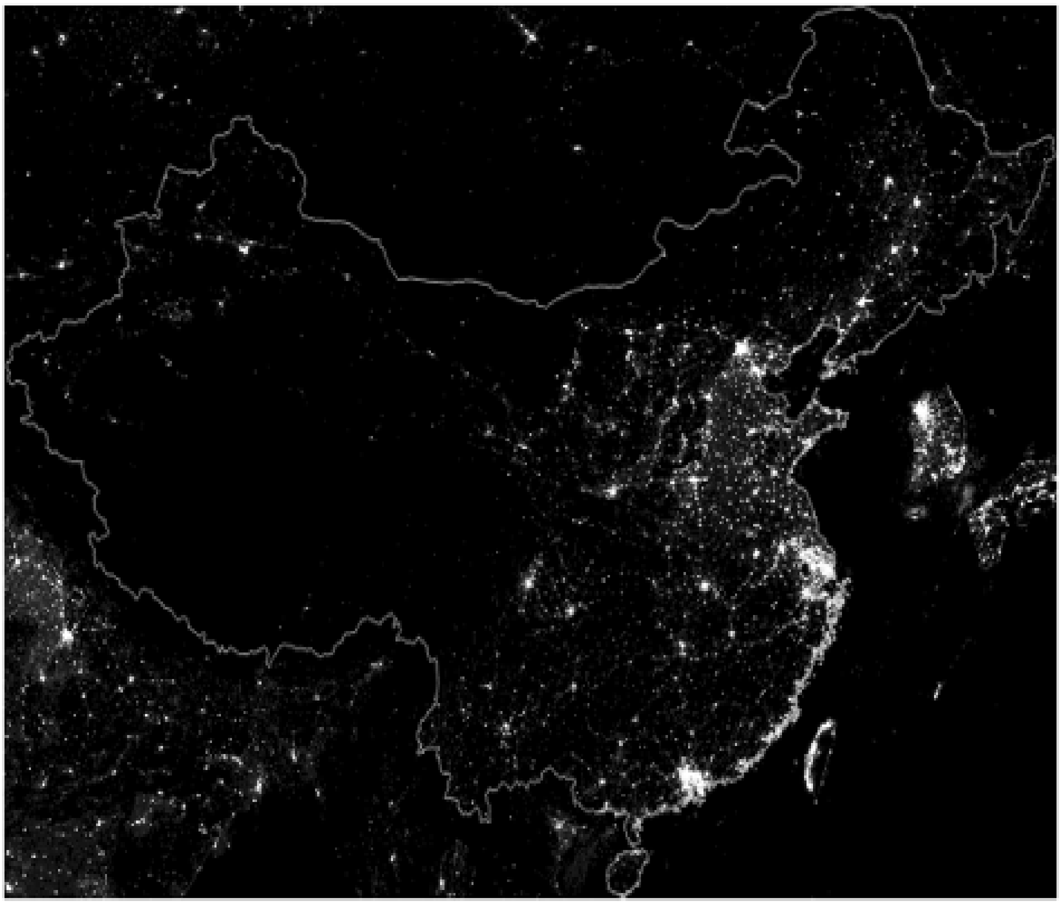


Figure 2: Lights at Night in China, 2011
Source: DMSP data collected by the United States Air Force.

North China and coastal area in the southeast China were most lit areas in both years and the intensity of light increased remarkably from 1997 to 2011 with a more dispersed drift in space. This coincides with the economic development in China from 1997 to 2011 when the whole country persistently experienced high-speed growth in economics, especially for regions in China's southeast coast and regions in north

China (around the capital city, Beijing). Nevertheless, most areas in Western China (including Xinjiang Uygur Autonomous Region, Tibet Autonomous Region, Qinghai Province, and Gansu Province) stayed in the dark due to their sparse populations.

Since this project targets at non-metropolitan prefecture cities, I extract digital numbers embedded in each pixel[8] and aggregate them into prefecture cities accordingly. Specifically, each prefecture city is treated as a polygon based on the

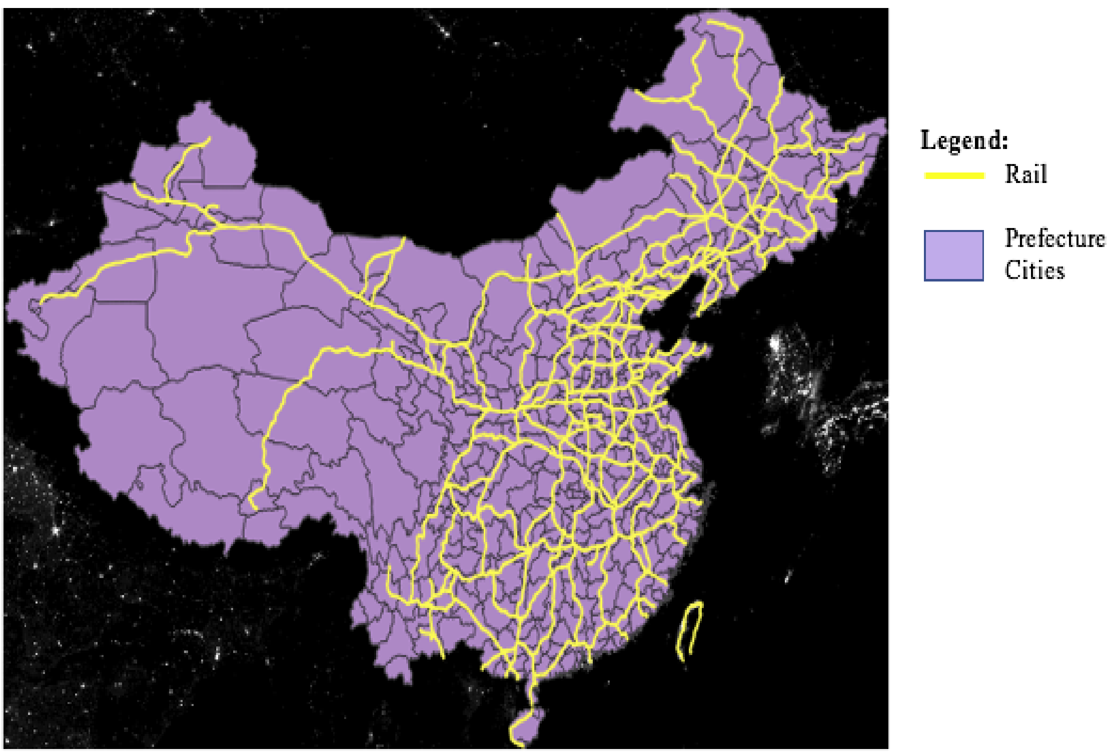


Figure 3: Lights at Night and Transportation Infrastructure in China, 2010

Source: DMSP data collected by the United States Air Force, GIS Map of 2010 China Prefecture Population Census Data, and Major railways in China

prefecture-level GIS map of 2010 China Prefecture Population Data. As shown in Figure 3, every purple polygon represents a prefecture city. After mapping each pixel

[8] Here, I use built-in function "extract" in R. All data processing for extracting night lights data has been done in R.

Table 1: Night Lights Data for Jiangsu Province, 1997

OBJECTID	Province	Prefecture City	Num of Pixels	Total NTL	Prop Zero	DN - Min	DN - Mean	DN - Max
171	Jiangsu	Wuxi	6291	84038	0.089334	0	13.358449	59
172	Jiangsu	Xuzhou	15844	110300	0.158167	0	6.961626	61
163	Jiangsu	Changzhou	6008	59961	0.082723	0	9.980193	59
169	Jiangsu	Suzhou	11547	140323	0.136053	0	12.152334	58
167	Jiangsu	Nantong	12264	102877	0.01484	0	8.388536	55
165	Jiangsu	Lianyungang	10471	63890	0.159679	0	6.101614	59
164	Jiangsu	Huai'an	13946	62674	0.340528	0	4.494048	56
173	Jiangsu	Yancheng	21434	99440	0.23682	0	4.639358	57
174	Jiangsu	Yangzhou	9173	73695	0.119481	0	8.033904	58
175	Jiangsu	Zhenjiang	5256	52588	0.046613	0	10.005327	56
170	Jiangsu	Taizhou	7983	63939	0.048603	0	8.009395	56
168	Jiangsu	Suqian	12018	47628	0.342153	0	3.963055	52

into a corresponding prefecture city, I compute the sum of digital numbers, representing the volume of lights used, within each prefecture city. The process of converting the pixel-level lights data into cities almost follows the steps elaborated in Storeygard (2016).

Table 1 displays aggregated night lights data for 12 prefecture cities in Jiangsu Province in 1997 [9]. For each city (rows in Table 1), there are six numerical features: number of pixels in the polygon of the city - Number of Pixels, total night lights - Total NTL (i.e. the sum of total DNs embedded in all pixels within a polygon), proportion of no lights - Prop Zero (i.e. DN = 0), minimum DN, average DN, and maximum DN values. The descriptive statistics of aggregated night lights data for 178 prefecture cities [10] in 1997 are shown in Table 2. One of the most notable findings is that the average proportion of zero DN is approximately 59%, which means that

[9] There were two satellites collecting data in 1997; Table 1 corresponds to the data of F121997.

[10] To make this project comparable with results in Banerjee et al. (2012), I only consider prefecture cities in the same fifteen provinces as they used. They elaborated why only considering these fifteen provinces (see p.21-22 in their paper) and I think it is reasonable to follow the same criteria in this project.

more than half of the whole nation did not have any economic activities at night in 1997. This coincides with what we have seen in Figure 1 and demonstrates that the economic state of China in 1997 was relatively poor.

Table 2: Descriptive Statistics for Night Lights Data - 178 Cities, 1997

	Num of Pixels	**Total NTL**	**Prop Zero**	**DN - Min**	**DN - Mean**	**DN - Max**
Count	178	178	178	178	178	178
Mean	23734.73	41707.93	0.59	0.08	4.27	50.21
Std	43307.77	39249.99	0.35	0.61	6.39	12.12
Min	1404	51	0	0	0.00	6
25%	9673.75	12224	0.23	0	0.72	43.25
50%	15755	26538	0.71	0	2.03	53
75%	24476.50	60186.75	0.90	0	6.21	60
Max	410009	165314	1.00	5	46.01	63

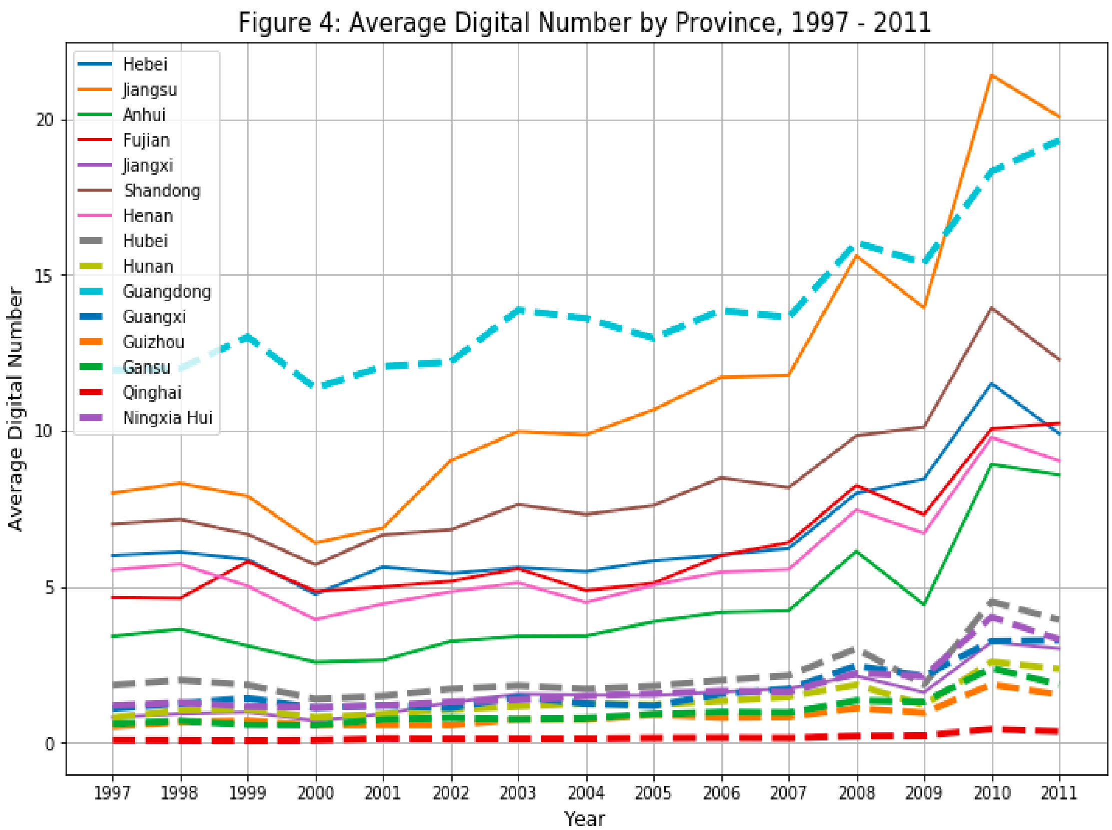
After applying the same data processing to each satellite-year dataset, I obtain 26 aggregated night lights datasets, each of which has the same features as Table 1 for 178 prefecture cities across 15 provinces. Since there were two satellites collecting night lights data in each year from 1997 to 2007, the total number of entries in 26 aggregated night lights datasets is $4,628$.

Table 3 summarizes descriptive statistics of aggregated night lights data for 178 prefecture cities from 1997 to 2011. Compared to those in 1997, the average number of total night lights for a single city increased from 41707.93 to 50670.72, while the average proportion of zero night lights dropped from 0.59 to 0.53. Since light growth is a proper proxy for economic growth at the country level (Henderson et al., 2012), this provides numerical evidence that the economic status of China has been improved a lot from 1997 to 2011.

Table 3: Descriptive Statistics for Night Lights Data, 1997 - 2011

	Num of Pixels	**Total NTL**	**Prop Zero**	**DN - Min**	**DN - Mean**	**DN - Max**
Count	4628	4628	4628	4628	4628	4628
Mean	23734.73	50670.72	0.53	0.10	5.07	52.76
Std	43190.62	48552.75	0.34	0.77	7.43	10.61
Min	1404	0	0	0	0	0
25%	9616.00	18561.25	0.18	0	0.95	47.75
50%	15755.00	33773.50	0.63	0	2.54	56
75%	24500.00	67153.50	0.85	0	6.59	61
Max	410009	472297	1	12	58.42	63

Additionally, it is necessary to examine the dynamic changes of light usage by province over the considered 15 years. Figure 4 displays the changes of average digital number in a pixel for each province from 1997 to 2011. As shown in Figure 4, there are two distinct segments where the light intensity of the upper part experienced a steadily increasing pattern with notable fluctuations while the light intensity of the lower part only increased a bit with relatively moderate fluctuations. These trends in light intensity are reasonable because the provinces corresponding to the upper segment are those in the central or eastern areas of China, while their counterparts in the lower segment are those in the western area of China. Economic growth of provinces in Central China or Eastern China has been spurred dramatically at higher speed since the Chinese Economic Reform. However, provinces in Western China have had relatively slow economic development due to the poor infrastructure, which has been gradually improved in the most recent years. The most significant fluctuation occurred around 2009, coinciding with the time period of global financial crisis starting in 2008. This somewhat demonstrates that light usage growth should



work well as a proxy for economic growth.

4.1.2 Chinese Government Economic Data

Since several Provincial Statistical Yearbooks in early years (i.e. before 2000) are not digital available online, I manually collected certain proportion of Chinese provincial government economic data with Provincial Statistical Yearbooks stored in the Joseph Regenstein Library.

Table 4 displays descriptive statistics of eight variables for all considered prefecture cities by regions. Specifically, the cities considered in this thesis can be divided into four regions: South China (Guangdong and Guangxi Provinces), North/Central

Table 4: Descriptive Statistics for Chinese Government Economic Data

	South China			North/Central China		
Variables	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Per Capita GDP (Yuan)	484	17384.92	16931.60	740	12117.10	9090.84
Land Area (sqkm)	484	11421.70	8537.40	765	13680.64	8158.58
Population (10k)	484	427.47	1259.61	752	489.08	235.35
Per Capita GDP Growth	451	0.12	0.11	676	0.15	0.43
Length of Railway (km)	30	2446.10	485.02	435	213.11	97.27
Length of Highway (km)	484	5744.66	4179.84	422	7925.02	6521.46
Length of Expressway (km)	15	1371.33	816.57	398	179.01	395.18
Length of Expressway and I - IV Highway (km)	300	5667.50	3899.05	398	9360.79	21318.24

	Eastern China			Western China		
Variables	Obs	Mean	Std. Dev.	Obs	Mean	Std. Dev.
Per Capita GDP (Yuan)	929	18784.29	18510.04	471	9588.75	13679.32
Land Area (sqkm)	929	11460.47	7495.41	474	40435.72	57530.78
Population (10k)	913	452.38	246.88	471	201.51	170.28
Per Capita GDP Growth	867	0.14	0.10	439	0.16	0.14
Length of Railway (km)	659	145.77	85.79	300	163.61	265.75
Length of Highway (km)	809	6840.82	5375.32	420	5890.06	4998.60
Length of Expressway (km)	520	217.17	533.10	211	81.36	90.52
Length of Expressway and I - IV Highway (km)	809	5963.19	4788.55	195	3248.36	2584.68

China (including Hebei, Hubei, Hunan, and Henan Provinces), Eastern China (including Jiangsu, Anhui, Fujian, Jiangxi, and Shangdong Provinces), and Western China (Guizhou, Qinghai, and Gansu provinces and Ningxia Hui Autonomous Region). Although the average per capita GDP growth (i.e. approximately 14%) is quite similar among cities from different regions, cities in Eastern China have the highest per capita GDP on average, which is almost double the value of per capita GDP for cities in Western China. The average population for cities in South, North/Central, and Eastern China are all over 4 million, with the highest average population of 4.89 million for cities in North/Central China. On the contrary, the average population for cities in Western China is only around 2 million, which is less than the half of that for cities in other three regions. In addition, the average land area for cities in Western China is about 40436 square kilometer (sqkm), more than three times of that for the remaining cities. These correctly reflect how the cities in Western China are different from those in other regions.

For different measures of transportation networks, the average and standard deviation of length of railway for cities in South China are much larger than those for other cities. This might attribute to the fact that very limited number of cities in South China have the data of length of railway. Further, variables *Length of Expressway* and *Length of Expressway and I - IV Highway* are sub-level measures of *Length of Highway*. Since the missing rate of *Length of Expressway* is relatively high, this thesis does not consider it as a specification for TI_{cpt} in empirical analysis.

4.1.3 "Straight Line" Strategy

To implement the "straight line" strategy described above, I compute the distances between each two provincial capital cities and/or Treaty Ports and draw lines from one provincial capital city to the nearest other provincial capital city and/or to the nearest Treaty Port. If there is more than one city/port where the difference in distances to the same city is less than 100km, I draw a line to all of them. Figure 5 displays a map of prefecture city boundaries (polygons with light-blue border), the constructed lines (in blue), provincial capital cities and four Treaty Ports, and major railways. Based on the constructed straight lines, I compute the nearest distance from each prefecture city to the straight line as an instrument for transportation

Figure 5: Straight Lines and Transportation Infrastructure

infrastructure. As shown in Table 5, cities in North/Central China and Eastern China are closest to the constructed lines on average, with the mean distance of 69.03 km and 84.53 km respectively. In terms of standard deviation, distance to line for cities in South China and Western China are more spread out than their counterparts in North/Central China and Eastern China. Since non-metropolitan cities in Eastern China and North/Central China have witnessed rapid economic growth, these statistics are consistent with the underlying hypothesis that cities closer

Table 5: Descriptive Statistics for Distance to Line - by Regions

	Num of cities	Mean	Std	Min	First Quartile	Median	Third Quartile	Max
North/Central China								
	Num of cities	Mean	Std	Min	First Quartile	Median	Third Quartile	Max
Distance to Line - km	765	69.03	69.38	2.09	17.80	39.51	117.07	285.34
South China								
	Num of cities	Mean	Std	Min	First Quartile	Median	Third Quartile	Max
Distance to Line - km	495	137.31	98.59	17.75	63.74	113.90	187.28	350.33
Eastern China								
	Num of cities	Mean	Std	Min	First Quartile	Median	Third Quartile	Max
Distance to Line - km	930	84.53	85.73	1.07	29.87	67.62	99.67	423.40
Western China								
	Num of cities	Mean	Std	Min	First Quartile	Median	Third Quartile	Max
Distance to Line - km	480	105.41	107.91	2.37	40.33	83.66	139.83	575.77

to the constructed lines are more likely to experience higher economic growth.

4.2 Results

4.2.1 Constructed Lines and Transportation Infrastructure

Table 6 shows the results for the first stage regression, corresponding to the model specification in equation (1). As mentioned previously, we consider four measures of transportation infrastructure as dependent variable TI_{cpt}: A. length of highway, B. length of railway, C. length of expressway and I - IV highway, D. average of length of

Table 6: Results for the First Stage

Dependent Variable: TI_{cpt}						
	A. Length of Highway			B. Length of Railway		
	(1)	(2)	(3)	(1)	(2)	(3)
ln Dist to Line	-0.3115	-0.2422	-0.2369	-0.0159	-0.0017	-0.2541
	(0.0803)	(0.0813)	(0.0794)	(0.0037)	(0.0033)	(3.1983)
Land Area		0.0153	0.0152		0.0024	0.0024
		(0.0034)	(0.0033)		(0.0001)	(0.0001)
Population			1.3368			0.1811
			(0.1418)			(0.0179)
Obs	2135	2135	2135	1411	1411	1411
Adj. R^2	0.065	0.075	0.118	0.769	0.819	0.833
	C. Length of Expressway and I - IV Highway			D. 0.5 Length of Highway + 0.5 Length of Railway		
	(1)	(2)	(3)	(1)	(2)	(3)
ln Dist to Line	-0.4218	-0.3482	-0.3031	-0.1614	-0.1010	0.0503
	(0.2441)	(0.2454)	(0.2453)	(0.0402)	(0.0408)	(0.0326)
Land Area		0.0474	0.0385		0.0101	0.0099
		(0.0186)	(0.0188)		(0.0016)	(0.0013)
Population			3.5847			4.6692
			(1.3018)			(0.1829)
Obs	1701	1701	1701	1411	1411	1411
Adj. R^2	0.027	0.031	0.036	0.129	0.156	0.363

Note: All regressions control for year and province fixed effects.
Standard errors shown in the parentheses are clustered at the city level.

highway and length of railway (i.e. the function of length of highway and length of railway). For each case of TI_{cpt}, we consider three scenarios: (1) baseline specification: only includes the logarithm of the distance to the nearest constructed lines (i.e. *ln Dist to Line*) as the explanatory variable; (2) includes both *ln Dist to Line* and city's land area; (3) further includes city's population as the third explanatory variable. Moreover, all regressions control for year and province fixed effects. To avoid losing too many degrees of freedom, we control for province fixed effects by classifying each province into one of four regions (i.e. North/Central, South, Eastern, or Western China).

Overall, almost all estimated coefficients of *ln Dist to Line* are negative except for the third scenario in Panel D whose estimate is positive and statistically insignificant even at 10% significance level. It is reasonable to have a negative correlation between transportation infrastructure and the distance to the nearest straight lines because under the assumption that the "straight line" identification strategy works well, more distant cities are supposed to have less access to transportation networks. This demonstrates that the proposed "straight line" strategy does represent how much access a city has to transportation in a proper manner. Additionally, the magnitude of *ln Dist to Line*'s estimate of the baseline model in Panel A is quite close to that in Banerjee et al. (2012), implying that the performance of "straight line" identification strategy has certain robustness to different levels of data (i.e. city-level data in this thesis and county-level data in Banerjee et al. (2012)).

Comparing regression results of the baseline specification for four measures of transportation infrastructure, the estimated coefficients of distance to the constructed lines are all statistically significant at 5% significance level, though the estimate of using *Length of Expressway and I - IV Highway* as dependent variable is least significant. In terms of adjusted R^2, approximately 77% of variation in length of railway could be explained by distance to the nearest straight lines, while only about 3% of variation in length of expressway and I - IV highway could be explained by the same variable. This might provide evidence that the constructed straight lines highly match the construction of railway, which seems logical not only by looking at the overlapped lines in Figure 5 but also considering the relative density of railway construction and highway construction in China.

Since a city's density of transportation is likely correlated with its land area, we consider land area as an important control for each city. The second column in each panel of Table 6 displays coefficient estimates when land area is included in equation (1). As expected, land area is positively correlated with transportation networks no matter which measure of TI_{cpt} is used. Nevertheless, the coefficient for the log distance to the straight lines becomes insignificant in Panels B and C when including land area. This might suggest that there exists strong correlation between distance to lines and land area and land area probably has more explanation power for length of railway and length of expressway and I - IV highway.

Further, when we include population as another control for each city, the coefficient for the log distance to the nearest lines is only significant when the dependent variable is the length of highway. Although the values of adjusted R^2 in Panel A are much smaller than those in Panel B, the coefficients of *ln Dist to Line* in Panel A are all statistically significant regardless of model specification. This implies that distance to the nearest constructed lines is likely robust to various model specifications when transportation infrastructure is represented by the length of highway.

4.2.2 The Effect on Economic Growth

To illustrate the effects of transportation infrastructure on economic growth, we

Table 7: Results for the Second Stage: Fitted Highway and Fitted Railway

Dependent Variable: $\triangle$ lny_{cpt}						
	ln Per Capita GDP			ln Night Lights		
	(1)	(2)	(3)	(1)	(2)	(3)
$\triangle$ ln Fitted Highway	0.2335	0.2335	0.2358	0.4694	0.4323	0.4061
	(0.0244)	(0.0243)	(0.0242)	(0.0329)	(0.0325)	(0.0317)
$\triangle$ Land Area		$4.95 * 10^{-7}$	$4.78 * 10^{-7}$		$-3.63 * 10\ 10^{-6}$	$-3.51 * 10^{-6}$
		$(5.08 * 10^{-7})$	$(5.08 * 10^{-7})$		$(6.81 * 10^{-7})$	$(6.66 * 10^{-7})$
$\triangle$ Population			$-6.00 * 10^{-5}$			$2.66 * 10^{-4}$
			$(2.18 * 10^{-5})$			$(2.85 * 10^{-5})$
Obs	2135	2135	2134	2135	2135	2134
Adj. R^2	0.098	0.097	0.100	0.254	0.263	0.296
	ln Per Capita GDP			ln Night Lights		
	(1)	(2)	(3)	(1)	(2)	(3)
$\triangle$ ln Fitted Railway	0.1081	0.0690	0.0778	0.2298	0.1465	0.0981
	(0.0304)	(0.0254)	(0.0231)	(0.0457)	(0.0378)	(0.0261)
$\triangle$ Land Area		$-4.89 * 10^{-7}$	$-3.19 * 10^{-7}$		$-4.12 * 10^{-6}$	$-1.99 * 10^{-6}$
		$(5.80 * 10^{-7})$	$(5.84 * 10^{-7})$		$(8.64 * 10^{-7})$	$(6.59 * 10^{-7})$
$\triangle$ Population			$2.53 * 10^{-4}$			$2.61 * 10^{-3}$
			$(7.90 * 10^{-5})$			$(8.92 * 10^{-5})$
Obs	1409	1400	1391	1409	1400	1391
Adj. R^2	0.099	0.096	0.107	0.167	0.184	0.534

Note: All regressions control for year and province fixed effects.
Standard errors are shown in the parentheses.

estimate the effects of predicted transportation networks on the log of per capita GDP and the log of night lights respectively. Table 7 and Table 8 display the results for the second stage in two-stage least squares (2SLS) regression based on equation (2), starting with a specification that only controls for year and province fixed effects. The underlying independent variable is the predicted value of transportation infrastructure obtained from regression results in the first stage.

As shown in the column (1) of each panel in Table 7 and Table 8, the coefficients for the log of predicted transportation networks are positive and statistically significant at 1% level, implying a positive correlation between transportation infrastructure and per capita GDP growth/night lights growth. Moreover, the predicted transportation networks seem to have more explanatory power on night lights growth, in terms of both the magnitude of coefficients and the values of adjusted R^2. Specifically, as shown in column (1) of the upper two panels in Table 7, the elasticity between the predicted length of highway and night lights is 0.4694, while the elasticity between the predicted length of highway and per capita GDP is 0.2335. This means that a 1% increase in the fitted length of highway induces a 0.47% increase in the intensity of night lights and a 0.23% increase in per capita GDP, on average. Additionally, the predicted length of highway accounts for 25.4% of variation in the intensity of night lights while it only accounts for 9.8% of variation in per capita GDP. Although the differences of adjusted R^2 between night lights and per capita GDP for other measures of transportation networks are not as large as those for the length of highway, all four

Table 8: Results for the Second Stage: Fitted Expressway and I - IV Highway and Fitted Highway and Railway

Dependent Variable: △ lny_{cpt}						
	ln Per Capita GDP			In Night Lights		
	(1)	(2)	(3)	(1)	(2)	(3)
△ ln Fitted Expressway and I - IV Highway	0.0807	0.0461	0.0501	0.2076	0.1825	0.0781
	(0.0223)	(0.0208)	(0.0218)	(0.0274)	(0.0261)	(0.0203)
△ Land Area		-6.65 * 10^{-6}	-6.90 * 10^{-6}		-2.04 * 10^{-6}	-2.48 * 10^{-6}
		(1.00 * 10^{-6})	(1.02 * 10^{-6})		(1.26 * 10^{-6})	(9.47* 10^{-7})
△ Population			-4.29 * 10^{-5}			2.19 * 10^{-3}
			(7.37 * 10^{-5})			(6.85 * 10^{-5})
Obs	1626	1619	1603	1626	1619	1603
Adj. R^2	0.072	0.098	0.096	0.092	0.085	0.493
	ln Per Capita GDP			In Night Lights		
	(1)	(2)	(3)	(1)	(2)	(3)
△ ln Fitted Highway and Railway	0.3738	0.3433	0.2218	0.6432	0.5613	0.2137
	(0.0260)	(0.0249)	(0.0207)	(0.0383)	(0.0366)	(0.0238)
△ Land Area		7.68 * 10^{-7}	3.63 * 10^{-7}		-2.10 * 10^{-6}	-1.3151 * 10^{-6}
		(5.41 * 10^{-7})	(5.51 * 10^{-7})		(7.96 * 10^{-7})	(6.34 * 10^{-7})
△ Population			9.04 * 10^{-5}			2.45 * 10^{-3}
			(7.67 * 10^{-5})			(8.82 * 10^{-5})
Obs	1411	1409	1388	1411	1409	1388
Adj. R^2	0.227	0.217	0.182	0.316	0.309	0.560

Note: All regressions control for year and province fixed effects.
Standard errors are shown in the parentheses.

measures of transportation infrastructure consistently fit the intensity of night lights better than per capita GDP. Presumably, the relatively lower quality of data for per capita GDP makes it harder to be accurately estimated by the proposed independent variables.

Columns (2) and (3) in Table 7 and Table 8 represent estimated coefficients after controlling for land area or both land area and population for each city. The magnitude of coefficients for land area and population is very small and close to zero, though most of them are statistically significant. One notable finding is that the adjusted R^2 becomes larger when we control for more variables in the cases of Table 7 while it fluctuates without a consistent pattern in the cases of Table 8. To some extent, this coincides with the finding in the first stage that the constructed straight lines perform better when we represent transportation infrastructure by the length of highway or the length of railway.

For the magnitude of coefficients of TI_{cpt}, corresponding to the first row of each panel in Table 7 and Table 8, it decreases dramatically as the number of controls increases except for scenarios using the predicted length of highway (i.e. cases in the upper panels of Table 7). The coefficients for the log of predicted length of highway are quite similar with different model specifications as well as different measures of economic outcomes. This likely suggests that the length of highway has a more persistent impact on economic growth over other measures of transportation networks.

5 Conclusion

To explore the impact of transportation infrastructure on economic growth, especially at the sub-national level for developing countries, this thesis uses data of 178 non-metropolitan prefecture cities in China from 1997 to 2011. Based on the reliable and acceptable performance of "straight line" identification strategy in previous literature (Banerjee et al., 2012; Faber, 2009), this thesis adopts similar identification strategy to mitigate or even eliminate the endogeneity of transportation networks. Instead of drawing straight lines between one historically important city to the nearest historically important city/Treaty Port (Banerjee et al., 2012), this thesis draws straight lines from one provincial capital city to the nearest provincial capital city/Treaty Port. The nearest geographic distance from each prefecture city to a constructed straight line serves as exogenous variation for measuring how much access a city has to transportation infrastructure.

Since the terminal cities of constructed straight lines are not under consideration, the proposed straight lines are supposed to work in a similar manner as those used in Banerjee et al. (2012). Particularly, this has been verified based on the empirical results for the first stage regression. Given the significantly negative correlation between the log of distance to the nearest constructed lines and transportation infrastructure, it is appropriate to conclude that the nearest distance to a constructed line serves as a proper instrument for access to transportation.

By employing four different representations of transportation networks, I propose a channel through which we can demonstrate whether the proposed instrument is robust to different measures of transportation with various model specifications. It turns out that the estimated coefficients for the log of distance to the nearest straight lines change not only as the measure of TI_{cpt} changes but also as the model specification changes by including more controls within the same measure of TI_{cpt}. The coefficients for the log distance to the straight lines only remain statistically significant for different model specifications when we use length of highway as TI_{cpt}. Further, the log distance to the straight lines can explain nearly 80% of variation in the length of railway. Therefore, although the estimated coefficients for the log distance to the constructed lines seem not robust to different measures of transportation networks, it is still reasonable to conclude that this proposed instrument does work properly to provide exogenous variation to length of highway or length of railway. More importantly, this thesis provides another empirical evidence to the validity of "straight line" identification strategy by using city-level data.

In terms of the significantly positive correlation between the log predicted amount of transportation infrastructure and log per capita GDP/night lights, this thesis finds positive effect of transportation infrastructure on economic growth. However, the effect is weakened when we consider more city-level controls and is not robust under different measures of transportation networks. One contribution of this thesis is that it uses the intensity of night lights as a measure of economic outcomes. Our results suggest that the predicted measure of transportation infrastructure has more explanatory power on the growth of night lights. Although there might exist the concern that using night lights data would amplify the effect, it is still worth trying to employ night lights data due to its relatively higher quality than the conventional government data, especially for regions which have poor data collection system. Moreover, one could use other measures of transportation infrastructure such as the distance to navigable river or the distance to coastline and control for more city-level fixed effects such as education level and unemployment rate. This might provide more comprehensive and accurate analysis on the effect of transportation infrastructure on economic growth.

References

Atack, J., Bateman, F., Haines, M., & Margo, R. A. (2010). Did railroads induce or follow economic growth?: urbanization and population growth in the American Midwest, 1850–1860. *Social Science History*, *34*(2), 171–197.

Banerjee, A., Duflo, E., & Qian, N. (2012). *On the road: Access to transportation infrastructure and economic growth in China* (Working Paper No. 17897). National Bureau of Economic Research.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. (2017). Roads, railroads, and decentralization of Chinese cities. *Review of Economics and Statistics*, *99*(3), 435–448.

Beyzatlar, M. A., Karacal, M., & Yetkiner, H. (2014). Granger-causality between transportation and gdp: A panel data approach. *Transportation Research Part A: Policy and Practice*, *63*, 43–55.

Bleakley, H., & Lin, J. (2012). Portage and path dependence. *The Quarterly Journal of Economics*, *127*(2), 587–644.

Cai, F. (2010). Chinese rural reform in 30 years: An analysis in terms of institutional economics. *Frontiers of Economics in China*, *5*(1), 150–168.

Chen, X., & Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, *108*(21), 8589–8594.

Dawson, J. W., DeJuan, J. P., Seater, J. J., & Stephenson, E. F. (2001). Economic information versus quality variation in cross-country data. *Canadian Journal of Economics*, *34*(4), 988–1009.

Deaton, A., & Heston, A. (2010). Understanding PPPs and PPP-based national accounts. *American Economic Journal: Macroeconomics*, *2*(4), 1–35.

Doll, C. N., Muller, J.-P., & Morley, J. G. (2006). Mapping regional economic activity from night-time light satellite imagery. *Ecological Economics*, *57*(1), 75–92.

Donaldson, D. (2018). Railroads of the Raj: Estimating the impact of transportation infrastructure. *American Economic Review*, *108*(4-5), 899-934.

Donaldson, D., & Storeygard, A. (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives*, *30*(4), 171–98.

Duncan, T. (2007). Retrospective analysis of the road sector, 1997-2005. *Operations Evaluation Department, Asian Development Bank*.

Eckstein, A., & Eckstein, F. E. (1977). *China's economic revolution*. CUP Archive.

Faber, B. (2009). Integration and the periphery: The unintended effects of new highways in a developing country. In *Urban Economics Sessions at the NARSC 2009 conference* (pp. 18–21).

Fogel, R. W. (1962). A quantitative approach to the study of railroads in American economic growth: a report of some preliminary findings. *The Journal of Economic History*, *22*(2), 163–197.

Fogel, R. W. (1964). *Railroads and American economic growth: essays in econometric history* (Vol. 296). Johns Hopkins Press Baltimore.

Harari, M. (2020). Cities in bad shape: Urban geometry in India. *American Economic Review*, *110*(8), 2377-2421.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, *102*(2), 994–1028.

Hodler, R., & Raschky, P. A. (2014). Regional favoritism. *The Quarterly Journal of Economics*, *129*(2), 995–1033.

Hong, J., Chu, Z., & Wang, Q. (2011). Transport infrastructure and regional economic growth: evidence from China. *Transportation*, *38*(5), 737–752.

Hu, Z. F., & Khan, M. S. (1997). Why is China growing so fast? *Staff Papers*, *44*(1), 103–131.

Imhoff, M. L., Lawrence, W. T., Stutzer, D. C., & Elvidge, C. D. (1997). A technique for using composite DMSP/OLS "city lights" satellite data to map urban area. *Remote Sensing of Environment*, *61*(3), 361–370.

Johnson, S., Larson, W., Papageorgiou, C., & Subramanian, A. (2013). Is newer better? Penn World Table revisions and their impact on growth estimates. *Journal of Monetary Economics*, *60*(2), 255–274.

Keller, W., & Shiue, C. H. (2008). *Institutions, technology, and trade* (Working Paper No. 13913). National Bureau of Economic Research.

Koch-Weser, I. N. (2013). The reliability of China's economic data: An analysis of national output. *US-China Economic and Security Review Commission Staff Research Project*, *4*.

Lee, Y. S. (2018). International isolation and regional inequality: Evidence from sanctions on North Korea. *Journal of Urban Economics*, *103*, 34–51.

Michaels, G. (2008). The effect of trade on the demand for skill: Evidence from the interstate highway system. *The Review of Economics and Statistics*, *90*(4), 683–701.

Michalopoulos, S., & Papaioannou, E. (2013a). National institutions and subnational development in Africa. *The Quarterly Journal of Economics*, *129*(1), 151–213.

Michalopoulos, S., & Papaioannou, E. (2013b). Pre-colonial ethnic institutions and contemporary African development. *Econometrica*, *81*(1), 113–152.

Song, M., Peng, J., Wang, J., & Zhao, J. (2018). Environmental efficiency and economic growth of China: a Ray slack-based model analysis. *European Journal of Operational Research*, *269*(1), 51–63.

Storeygard, A. (2016). Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa. *The Review of Economic Studies*, *83*(3), 1263–1295.

Sun, Y., & Cui, Y. (2018). Evaluating the coordinated development of economic, social and environmental benefits of urban public transportation infrastructure: Case study of four Chinese autonomous municipalities. *Transport Policy*, *66*, 116–126.

Wang, L., & Duan, X. (2018). High-speed rail network development and winner and loser cities in megaregions: The case study of Yangtze River Delta, China.

Cities.

West, L. A., & Zhao, Y. (2000). *Rural labor flows in China* (No. 42). University of California Inst of East.

World Bank. (2007a). *Domestic trade impacts of the expansion of the national expressway network in China*. The World Bank.

World Bank. (2007b). *Evaluation of World Bank support to transportation infrastructure*. The World Bank.

World Bank. (2018). *The World Bank in China: Overview*. The World Bank.

Yu, N., De Jong, M., Storm, S., & Mi, J. (2012). Transport infrastructure, spatial clusters and regional economic growth in China. *Transport Reviews*, *32*(1), 3–28.