“Ghost Cities” versus Boom Towns: When Do China’s HSR New Towns Thrive?*

Lei Dong† Rui Du‡ Matthew Kahn§ Carlo Ratti¶ Siqi Zheng‖

Abstract

In China, local governments often build “new towns” far from the city center but close to new high-speed rail (HSR) stations. While some HSR new towns experience economic growth, others have been vacant for years and became “ghost towns.” This study explores the determinants of this heterogeneity. Using satellite imagery and online archives of government documents, we identify 180 HSR new towns. We use data on establishment growth to measure the local vibrancy of the new town at a fine spatial scale. Given that the placement of a new HSR station may reflect unobservable spatial attributes, we propose an instrumental variables strategy for the location of new HSR stations that builds on the recent economic geography literature. Our results show that the location and local market access are key determinants of the success of new towns.
JEL classification: R10, R11, Z20.

Keywords: New town creation, agglomeration, high-speed rail

---

*This paper has benefited from comments from Devin Michelle Bunten, Yannis M. Ioannides, Mark D. Partridge, Patricia Yanez-Pagans, Jeffrey Zabel, Junfu Zhang, and seminar participants at the Mini Urban Economics and Policy Workshop at MIT and research seminar series at the University of Oklahoma. We acknowledge research support from MIT Sustainable Urbanization Lab and are thankful to Yunhan Zheng, Jiantao Zhou, Dongxiao Niu, and Matteo Migliaccio for their superb research assistance. We also thank Kai Ying Lau for GIS database assistance. All errors are our own.

†MIT Sustainable Urbanization Lab; MIT Senseable City Lab; Email: leidong@mit.edu.

‡Department of Economics, Oklahoma State University; MIT Sustainable Urbanization Lab; MIT Center for Real Estate; Email: rui.du@okstate.edu.

§21st Century Cities Initiative, Johns Hopkins University, and NBER. Email: mkahn10@jhu.edu.

¶MIT Senseable City Lab, DUSP; Email: ratti@mit.edu.

‖MIT Sustainable Urbanization Lab, Center for Real Estate, and DUSP; Email: sqzheng@mit.edu.

1 Introduction

Transportation networks increase inter-city connectivity and reshape the geographical distribution of economic activity across and within cities. Recognizing this point, many Chinese city leaders are competing for high-speed rail (HSR) to promote local economic growth. In 2008, China built its first HSR line (Beijing-Tianjin), while by the middle of 2018 - only ten years after - China has 27,000 kilometers of HSR lines, accounting for approximately two-thirds of the world's total (Barrow, 2018).

A distinct feature of Chinese HSR stations is that these are newly-built (rather than redeveloped existing railway stations) located far away from the urban center. Given the unique land public finance scheme in China, city leaders have strong incentives to build "new towns" surrounding new HSR stations as a strategy to expand their cities and increase local government revenue from land auctions. Building a new town is also regarded as an effective industrial policy to attract new industries and population to boost the local economy. While some of those new towns have enjoyed economic growth, others have remained vacant for many years and become "ghost towns" (Woodworth and Wallace, 2017; Zhao and Ma, 2017).

What factors lead to the success or failure of these HSR new towns? We examine how station location choice and local economic demand shifts explain why we observe such variation in the urban economic vibrancy of HSR new towns. The literature on spatial equilibrium documents agglomeration and proximity to urban markets as important determinants of land rents (Koster and Rouwendal, 2013), which capture local economic growth trends. Recent research on the determinants of urban growth also emphasizes the importance of market access (Faber, 2014). Larger local demand for goods and services economic induces technological and innovation investment and human capital externalities (Acemoglu and Linn, 2004; Hanson, 2005).

In choosing a location for a new HSR station, the city government faces a key tradeoff between the costs of construction and demolition and the benefits of agglomeration economies. Since there are no official documents that define new towns and no data at the new town level ever exist, to measure the vibrancy created by a new town requires new data construction. We assemble satellite imagery data on all HSR stations and carefully compare them to online archives of government documents and different news sources to identify HSR new towns. Our final sample contains 180 HSR new towns. We then construct a panel dataset of new town vibrancy using new establishment data at a fine spatial resolution to examine the changes in very local economic growth. Specifically, we geocode the address attribute into coordinates with map geocoding services and compute bilateral great-circle distances between HSR stations, county centers, and city centers. We also collect county-level statistics, such as urban population, GDP, fiscal revenue, and the share of the agricultural sector to match with firm and distance data. Our granular panel dataset covers 2006-2015.

Using the least-cost path distance to instrument from the actual distance from the station the county center, we use 2SLS to estimate the impacts of new town's proximity to major spatial centers and its market access to test for these demand-side factors' roles in determining the success of a new town. We observe a strong attenuation of the agglomeration effect and a positive effect of the market access to existing urban markets.

We also use the potential new town location as a counterfactual location to estimate the impact of a newly-built HSR station on local urban vibrancy – whether it can support a new town around the station. Consistent with predictions of new economic geography theory, we show a strong, positive effect of building a new town on the local vibrancy compared to its matched counterfactual, conditioning on the distance to the host county center and market access to existing urban markets.

Our results document a negative competition effect where new towns close to the county centers receive no significant benefits from new town development. However, a new town is sustainable in a location further away from the host county center when the market access to existing urban markets is large enough. We track the economic activity in adjacent counties and find little evidence that the estimated positive treatment effect on local vibrancy is due to a reshuffle of economic activity over space.

We further examine the timing of new town vibrancy and show that locations close to the county center and with greater market access to existing urban markets are more likely to take off faster. A location experiences an economic boom four years earlier than it would have had the location not had an HSR station. This suggests that the "ghost city" phenomenon near HSR stations may represent a real option that city leaders offer to firms as they choose where to locate in the medium term. From the perspective of a city leader, if they expect a rise in local market potential in the medium term, then this also creates an option value justification for building an HSR new town in a given location even if they know in the short-term that the new town might not attract enough firms. Under this scenario, while the researcher observes a "ghost town" in the short run, this area may thrive in the medium-run.

Our paper is related to a growing body of literature examining the impact of HSR network on the local economy, the spatial distribution of economic activity and population (Sasaki et al., 1997; Xu, 2017; Li and Xu, 2018), employment and urban specialization patterns (Lin, 2017; Dong, 2018), and industry and product space (Gao et al., 2017; Zhu et al., 2019). The literature on the effect of the HSR network on local economic gains reaches mixed results. Some studies document a positive effect, possibly through the channel of market integration (Zheng and Kahn, 2013; Chen and Haynes, 2015; Ke et al., 2017; Ahlfeldt and Feddersen, 2017). Others find that the network connection leads to a reduction in local economic growth for non-targeted peripheral regions through a trade-based channel of decreasing transportation costs (Qin, 2017; Yu et al., 2019). Faber (2014) focuses on China's

National Trunk Highway System and documents similar negative effects on non-targeted peripheral counties, which appears to be driven by a trade-based channel in the light of falling trade costs.

We contribute to this literature by focusing on a unique urban development phenomenon in China — new towns built on undeveloped land. To the best of our knowledge, our research is the first empirical study to investigate the effect of new town development at the new town level. This unique spatial setting allows us to ask whether place-based strategies such as building completely new towns stimulates the growth of these emerging regions and what contributes to their successes or failures, which is important in understanding the consequences of new town development strategies.[1]

Our paper highlights the additional roles of agglomeration spillovers, industry location decisions, and market access, complementing the urban economics studies on spatial agglomeration, firm location choice, and endogenous urban growth (Palivos and Wang, 1996; Cohen and Paul, 2005). In particular, we document substantial benefits from building an HSR new town as well as a higher local impact of building a new town on firm growth for locations with a better local transport availability, adding another piece of evidence to the literature on industrial location behavior and economies of transport density (Mori and Nishikimi, 2002; Holl, 2004) as well as studies on the roles of transportation hubs and technology in determining the city location and its sustainability (Behrens, 2007).

Our research also connects to the literature using big data to measure urban vibrancy. Remote sensing data, especially nighttime light image, has been widely used by researchers for economic analysis at the global scale (Chen and Nordhaus, 2011; Henderson et al., 2012; Donaldson and Storeygard, 2016; Lu et al., 2018). Recently, with the rising availability of data based on smart-phones, new research has studied the geography of urban vibrancy (Chi et al., 2015; Dong et al., 2017; Huang et al., 2019). We contribute to this literature using micro-level firm distributions and satellite imagery data to describe the vibrancy dynamics at the new town level. In particular, the national-wide firm location data provide a novel measure of urban vibrancy.

The remainder of the paper proceeds as follows: Section 2 discusses the background and presents data details. We also present a case study of HSR new towns to motivate our analysis. Section 3 proposes our hypotheses to test and introduces our identification strategy. Section 4 presents the estimation results and discusses potential channels at work. Section 5 concludes.

[1] Diao et al. (2017) document the important role of intra-city travel cost in accessing the economic gains brought by the intercity HSR network connection.

2 Background and Data

2.1 High-speed rail

HSR has various definitions in different countries and literature. According to the official bulletin of China's Ministry of Railways, HSR is defined as "newly-built passenger-dedicated rail lines traveling at not less than 250 km/h (including lines with reserved capacity for an upgrade to the 250 km/h standard) on which initial service operate at not less than 200 km/h."[2] Under this definition, the Beijing-Tianjin inter-city line, which opened in 2008 for the Beijing Olympics, was the first passenger-dedicated high-speed rail line, and it connects the two largest cities in northern China with the top speed of 350 km/h. Since then, China has embarked on the fast track of HSR construction, and several important cross-regional lines were built to connect major cities, such as the Wuhan-Guangzhou line (opened in December 2009) and Beijing-Shanghai line (opened on June 2011).

The backbone of the HSR network in China was proposed in two "Mid-to-Long Term Railway Network Plan(s)". The first plan was initiated in 2004 and proposed a national grid system composing of eight high-speed rail corridors (also known as the 4+4 network).[3] This 4+4 national grid system was planned to be completed by 2020, but the government's stimulus accelerated the construction of the HSR lines. Especially after the 2008 Great Recession, the Chinese government launched a four trillion RMB package to stimulate infrastructure construction and domestic demands. As a result, China has 19,000 kilometers of high-speed rail by the end of 2015, far exceeding the target of 12,000 kilometers set in 2005. Thus, a new "Mid-to-Long Term Railway Network Plan" was proposed in 2016. This new plan extended the 4+4 high-speed rail grid to a larger 8+8 grid and expanded intercity lines to connect major metropolitan areas.[4] Specifically, the goal is to connect provincial capitals with other large and medium-sized cities with a population of more than 0.5 million and form a high-speed railway network that will cover the whole country. As of June 2018, ten years after the opening of the first HSR line, China's HSR network extends to about 27,000 km in total length and covers 29 provincial-level administrative divisions, becoming the most extensively uses HSR services in the world.[5]

Here, we digitize and geocode HSR stations and lines based on high-resolution transporta-

[2]See: http://www.gov.cn/flfg/2013-02/20/content_2334582.htm.

[3]Four lines run from north to south: Bejing-Shanghai, Bejing-Wuhan-Guangzhou-Shenzhen, Bejing-Shenyang-Harbin, and Hangzhou-Ningbo-Fuzhou-Shenzhen. Four lines go from east to west: Xuzhou-Zhengzhou-Lanzhou, Hangzhou-Nanchang-Changsha, Qingdao-Shijiazhuang-Taiyuan, and Nanjing-Wuhan-Chongqing-Chengdu. See: http://www.gov.cn/ztzl/2005-09/16/content_64413.htm

[4]Eight lines run from north to south and eight lines from east to west, see: http://www.gov.cn/xinwen/2016-07/20/content_5093165.htm.

[5]Chen et al. (2019) and Pan and Gao (2019) provide thorough information about the background of HSR development and its impacts from perspectives of regional economics, urban planning, public policy, and transportation.

tion maps published between 2017 and 2018. Besides geographical information, the source maps also have the attributes of opening year and level (national, regional, or intercity level) of each line, which provides additional control variables for our analysis. To obtain the accurate shape of the lines, we also collect the railway line data on OpenStreetMap[6]. By setting the OpenStreetMap data as the base layer, we draw a detailed map of the HSR lines on it. Since our analysis is conducted at the HSR new town level, we further collect the opening time of each station from Wikipedia and Baidu Baike,[7] and construct a new town panel dataset. Finally, we obtain 90 lines and 839 stations (Figure 1).

2.2 New towns in China

To benefit from substantial spillovers due to the HSR network connection, many Chinese cities have planned new towns around HSR stations (as a subcenter in the outskirts of the city, while still in the jurisdiction of the city), which are often called HSR new towns as there are many other different types of new towns in China. For brevity, "HSR new town" (or sometimes called "HSR new city" or "HSR new district") is referred to as "new town" in the rest of our paper. This "new town" development approach has been very popular in government planning in the HSR station location choice (Pan and Gao, 2019). This model promotes the development of new towns in regions far away from the urban center in the hope of vitalizing the vibrancy of new towns with spillovers received via the HSR network. Despite the prevalence of the new town model, the official documentation of new towns is scarce. Thus, we assemble information about new towns from various sources and seek to identify new towns around HSR stations.

We follow three steps to identify HSR new towns in China. First, we select the newly built HSR stations after the year 2008 (rather than the existing old stations). Second, we use Google Earth satellite imagery to identify stations with construction around them. Here, construction means that we can observe road networks, residential and industrial land development around the station after its opening, which are important criteria for new town development under our definition. Third, we collect online archives of government documents related to HSR new towns by querying Baidu searching engine two keywords: "HSR station name" and "new town". The second and third steps were performed by two research assistants with expertise in urban studies and GIS, and a third expert will make further confirmation when the conclusion is inconsistent. We successfully identify 180 HSR new towns. Figure 1 shows the national-wide geographical distribution of these new towns (orange dots). To further consider the geographical characteristics, we calculate the average elevation and slope

[6] http://www.openstreetmap

[7] Notice that not all stations of a line open at the same time, thus we searched the line names in Wikipedia and Baidu Baike to collect the station-level opening time information.

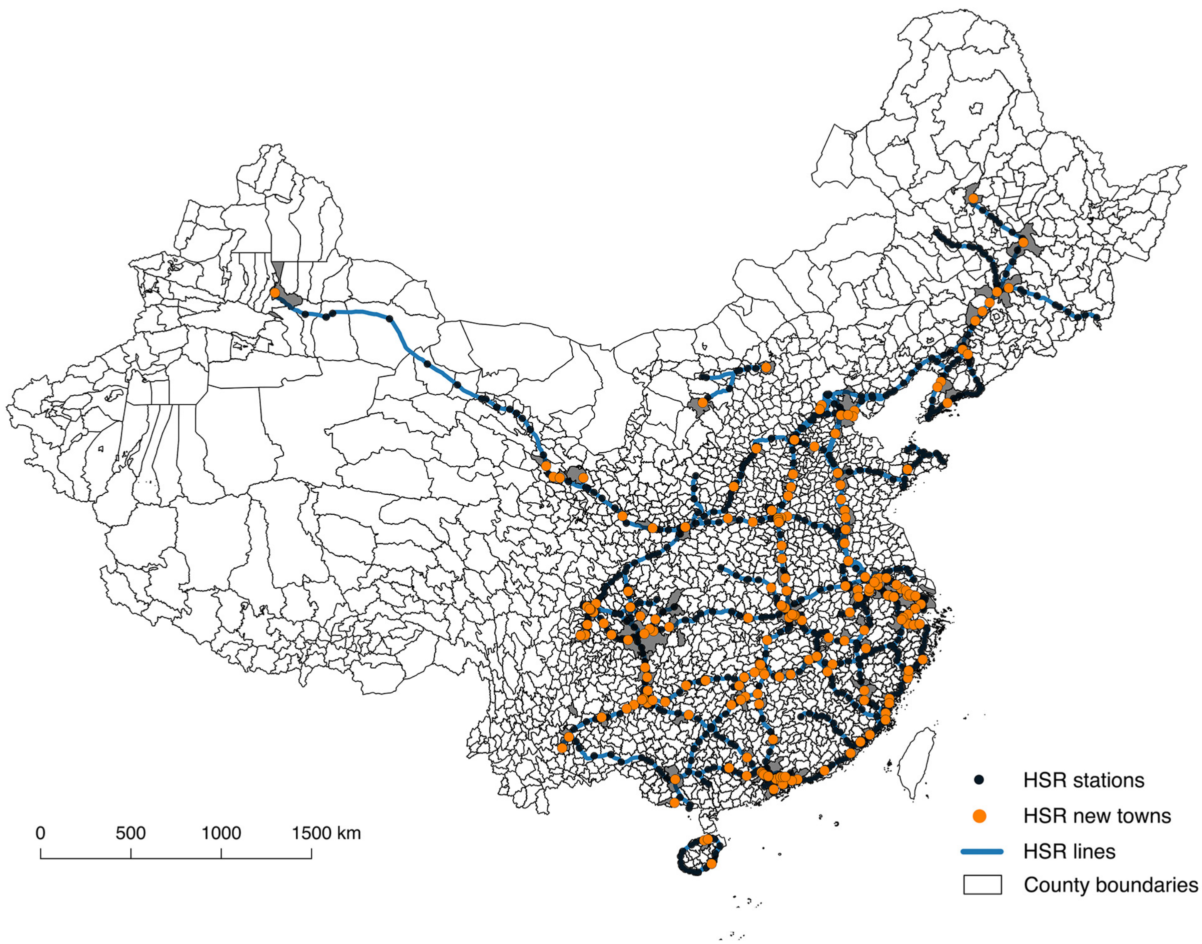


Figure 1: Mainland China's HSR new towns, stations, and lines.

at the new town level based on the digital elevation model (DEM) data.[8]

2.3 Firm data

To measure the urban vibrancy at a microscopic level, we collected firm registration record data from the registry database of the State Administration for Industry and Commercial Bureau of China (SAICBC). This dataset covers the registered information for all firms in China, with attributes including firm name, year established, address, operation start-tups, and so on. We geocode firm addresses into longitude and latitude using AMap API (https://lbs.amap.com/). Note that different cities may have streets or buildings with the same name, only querying the street name may return multiple coordinates. To reduce the possible error caused by address ambiguity, we also set the city name as a parameter for geocoding, and if the address of a firm record does not contain any description of the city

[8]The DEM dataset was collected from http://www.resdc.cn/data.aspx?DATAID=123.

name, we try to extract the city information from other fields such as the registration authority. Finally, we convert all geocoded coordinates into the WGS84 projection (AMap API returns coordinates with GCJ02 projection, a special projection system in China), and then aggregate firms into two different buffers – 3km and 5km – of HSR stations.[9] We use this buffer setting because we do not have exact boundaries of HSR new towns. According to some planning documents and relevant studies, the 3km and 5km buffers are basically in line with the development scale of the new town (Wang, 2016; Zheng et al., 2019). To investigate whether HSR new towns stimulate real economic growth rather than reshuffling economic activity across spaces, we also aggregate the firm data at the county level to build a county panel dataset of vibrancy measures. In Figure 2, we show the spatial distribution of newly established firms of the years 2010 and 2015 around Zhengdong HSR new town.

The limitations of the firm data should be noted: First, we only have registered firm addresses, which may not be the locations where actual economic activity takes place;[10] Second, firm size (e.g., the number of employees or the revenue) is unreported in the original dataset. Based on our observation and case studies, city officials use different strategies (subsidies and preferential policies) to first attract key players to locate to new town areas. These key players create industrial agglomeration which further impacts firm size, performance, and industrial location decisions (Li et al., 2012). We leave the analysis of these interactions for future research.

2.4 County socioeconomic data

Since the development of new towns usually happens at the county or district level, we conduct our analysis at the county level to capture intra-city location choice and agglomeration spillovers. Chinese administrative boundary data at the county level were collected from the Resource and Environment Data Cloud Platform at the Chinese Academy of Sciences. These county-level administrative units have been subdivided into county-level cities, counties, and districts (shixiaqu) of prefectural-level cities.[11] In our paper, if a prefectural-level city has multiple districts, they are merged as one county-level unit. For the sake of brevity, we refer to a prefectural-level city as a city for the remainder of this paper.

One important variable of interest is the distance between an HSR station and its corresponding county center. Here we define the county center as the area of the highest building

[9]We prefer to use a buffer of 5 km considering the possibility that some urban development does not occur in immediate proximity to the HSR station. For this reason, our main results are based on the number of new establishments within 5 km of the HSR station. The estimation results are robust to the size of the buffer. Results are available upon request.

[10]Some observations show that, to benefit from favorable policies (e.g., tax subsidies), firms may register the address in the industrial parks but do not operate there.

[11]More details about the administrative divisions of China can be found at https://en.wikipedia.org/wiki/Administrative_divisions_of_China.

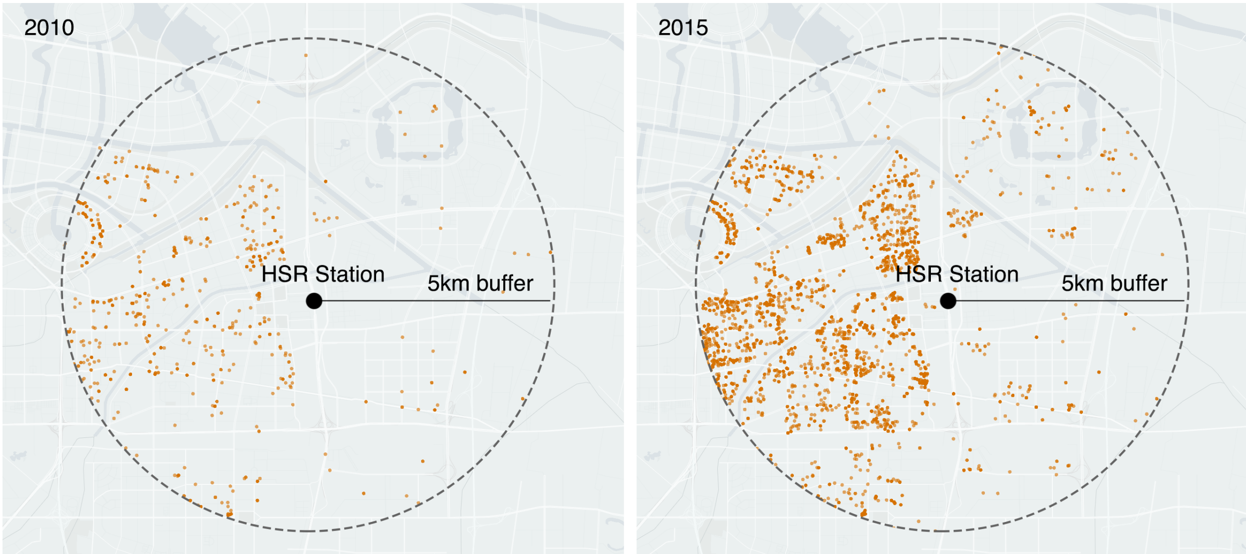


Figure 2: Spatial distribution of firms (orange points) around the Zhengdong HSR new town (Zhengzhou City, Henan Province).

density, which is identified by the satellite imagery. The distance between two geo-located points is then obtained by the Haversine formula, which calculates the great-circle distance between two coordinates on earth[12]. We also obtain the real travel distance and time between HSR station and its corresponding county center on the road networks by AMap API.

County-level statistics, such as GDP and population, are obtained from the County Economic Statistical Yearbooks from 2006-2015. We match the new-town-level data with county-level statistics to obtain the dataset for our panel analysis. Table 2 presents the summary descriptive statistics. Panel A presents the summary statistics for all years. Panel B reports 2006 descriptive statistics, while Panel C reports those for 2015. We choose this window to examine the dynamics in new town vibrancy as the window exactly coincides with two Five-Year Plans.

The average number of new establishments in a 5 km buffer is about three times of that in a 3 km buffer. This difference is larger in 2006 compared to 2015 (as measured by the ratio of firms within a 5 km buffer and a 3 km buffer), suggesting more establishments growth concentrated around HSR stations in the last decade. Some identified new towns have no firm activity in both 2006 and 2015, showing preliminary evidence of the "ghost towns" phenomenon in these locations despite observed real estate development from the satellite imagery.

HSR stations are often constructed in distant regions relative to the county center. The

[12] We use Haversine formula to compute the distance between points with longitude and latitude coordinates (ϕ_1, λ_1) and (ϕ_2, λ_2): $d = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2-\phi_1}{2}\right) + \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\lambda_2-\lambda_1}{2}\right)}\right)$, where R is the radius of the earth

average distance from the HSR station to the county center is 13 km. Our constructed least-cost distance is averaged at 20 km, implying that the actual location choice deviates from the least-cost choice so that trade costs are reduced and new towns can better access the agglomeration benefits from the county center. Within our expectation, Chinese cities have experienced tremendous growth in their local economies. The local economic size (measured by county government revenue) in 2015 is five times of that in 2006. The county-level population has increased from 2.11 million to 2.28 million. Agricultural employment share has risen to 0.9% from 2% during the estimation window, which is not what we expect. The proportion of counties that have a city status is roughly 80% and has remained almost unchanged despite some administrative division changes.[13]

2.5 A case study of a new town

To motivate our empirical analysis, we present a brief case study. Figure 3 presents the satellite imagery of two new towns developed around the HSR stations in the current year (post-HSR period): Zhengdong Station (in Zhengzhou city) and Chuzhou Station (in Chuzhou city). These two stations were opened around the same time. Zhengdong Station was opened in 2012 and Chuzhou Station began operation in 2011. The distance between these two new towns from the main city is relatively similar, both around 10 kilometers, but Zhengzhou has better local fundamentals (e.g., population, local government revenue, public transportation system, etc.). Satellite images show that Zhengdong has much more real estate development and seems more dynamic than Chuzhou.

We plot the trends of several variables using the data we assemble from various sources in Figures 4 and 5. In Figure 4, Panel A plots the trends of new establishments within a 5 km radius of the Zhengdong Station. Panel B plots the trends of satellite nighttime light intensity in the same area. Panel C plots the trends of market access to surrounding urban markets for the county where the station is located. We discuss the detail of the calculation of market access in our empirical strategy. In Panel D, we divide each variable by its initial value in our window to examine their growth patterns as opposed to their initial values.

Panel A of Figure 4 shows that the trend of establishment growth in the new town (the blue trend line) was relatively flat before the opening of Zhengdong Station but experienced a sharp increase after the station started operation. One may have a concern that the growth trajectory may simply pick up the time trends and time-specific policy shock. To net out year effects, we subtract the average number of new establishments in each year for all new towns from the number for Zhengdong. The red trend line depicts the trend of new establishments after adjusting for time trends. The trend line is slightly lower than the original trend but

[13] A county has a city status if there is a character "district" (Qu) or "City" (Shi) in its Chinese county name.

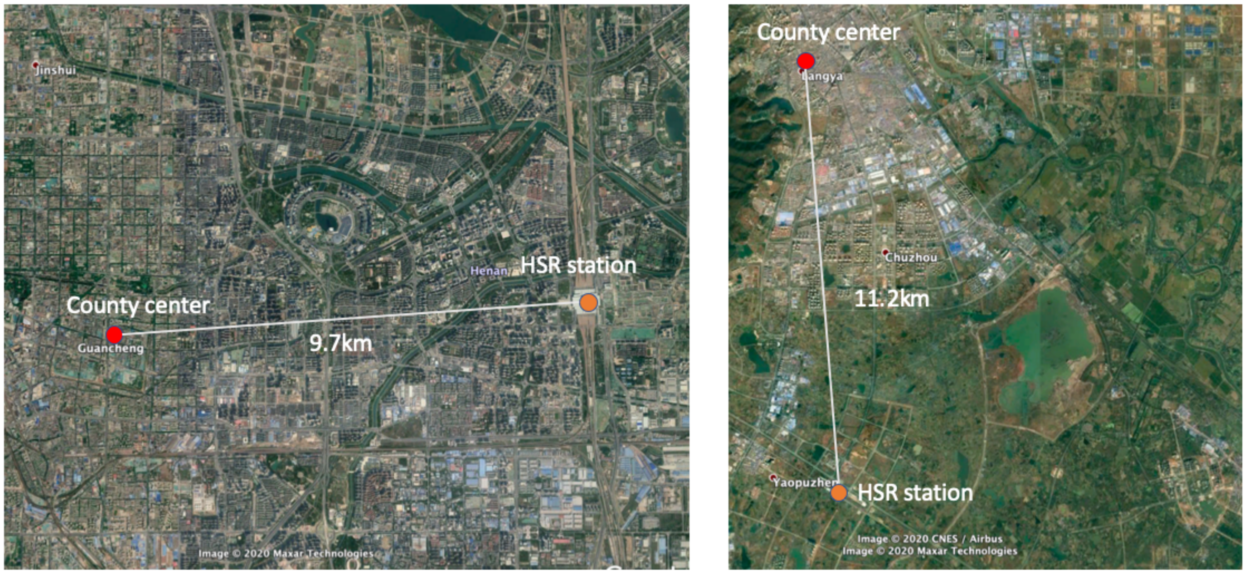
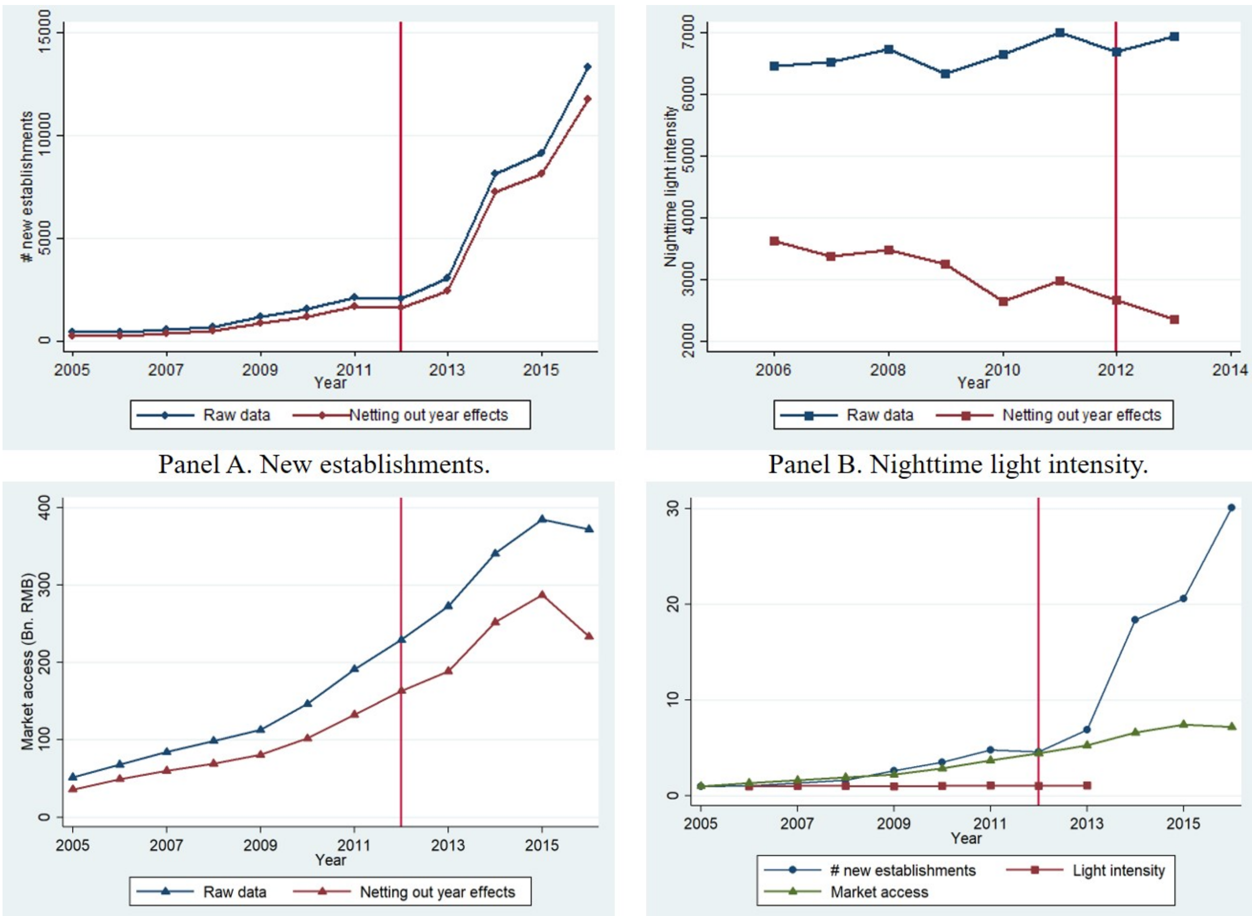


Figure 3: The satellite imagery of two typical new towns: Zhengdong Station of Zhengzhou ity, Henan Province (left); Chuzhou Station of Chuzhou City, Anhui Province (right). (Image Copyright: Google Earth)

resembles the overall pattern over the estimation window — Zhengdong started to boom after the opening of the HSR station.

In Panel B, the trend of satellite nighttime intensity remained flat throughout the estimation period. We document a downward trend after netting out year effects. These observations suggest that the upward trend in Zhengdong was not simply due to the real estate boom. Panel C suggests that a strong, growing local economy was one of the drivers for local economic growth. We document a constantly growing trend of local market access regardless of the consideration of year effects except for the period from 2015-2016. The normalized trends in Panel D demonstrate strong growth in new firm activity in the post-HSR period as well as an increasing trend of local market access, compared to their starting points. The satellite nighttime light intensity stood in the post-HSR period at a similar level as that in 2006.

We repeat our visual case analysis for Chuzhou Station in Figure 5. Panel A shows that the new establishment growth trend of the Chuzhou new town was flat using the raw data but declined sharply in the post-HSR period after netting out year effects. However, Panel B demonstrates a constant upward trend of nighttime light intensity over the estimation period. Since satellite light intensity may capture the construction of local infrastructure, the new town around the Chuzhou Station is likely a "ghost city", where we see a lot of real estate development but not much real economic activity (e.g., firm activity). Although the station is in good proximity to the county center, a relatively weak local economy as reflected in the stagnant market access growth in Panel C might undermine the urban success chance of the



Panel A. New establishments.

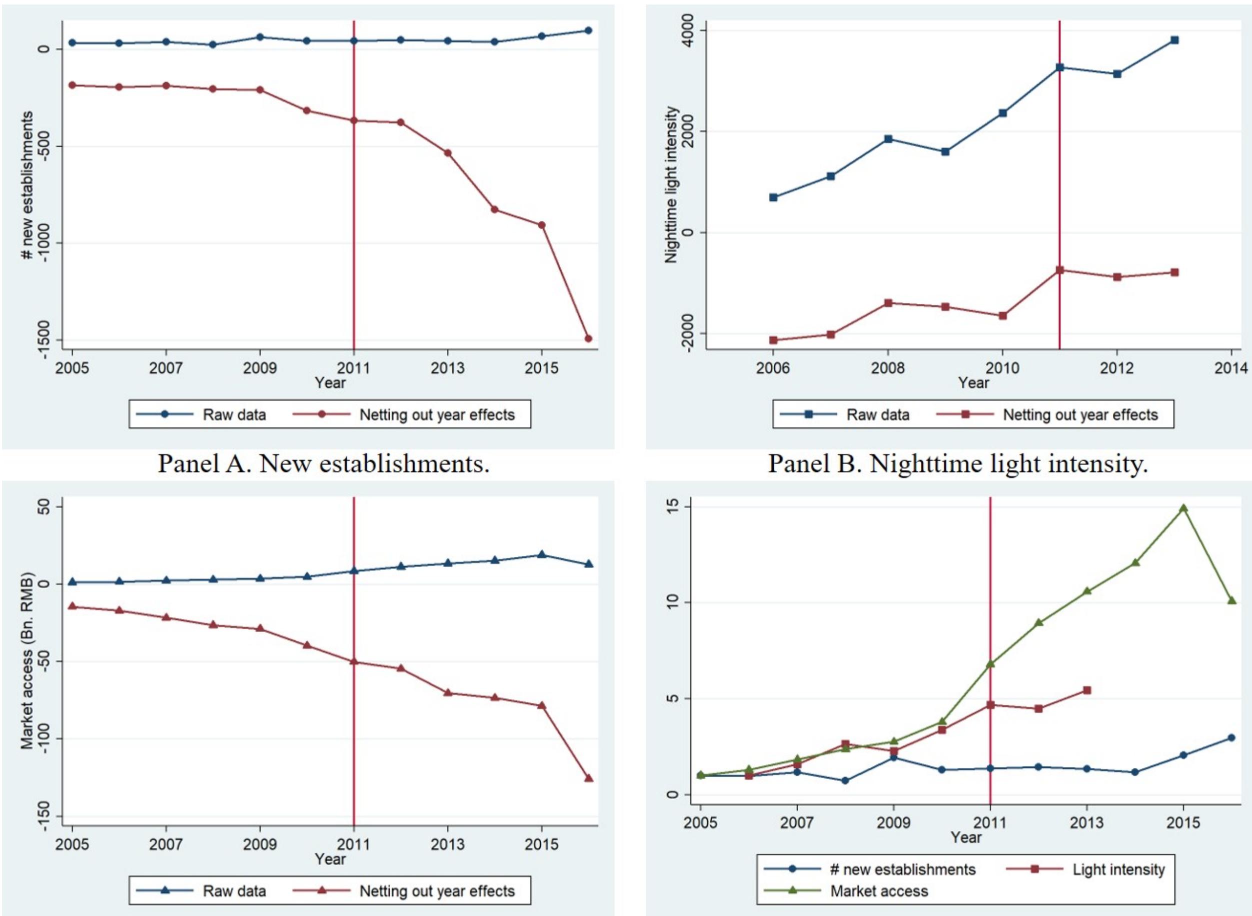
Panel B. Nighttime light intensity.

Panel C. Market access.

Panel D. Normalized trends.

Figure 4: Trends in the number of new establishments, nighttime light intensity (obtained from DMSP/OLS dataset (Donaldson and Storeygard, 2016)), and market access within a 5 km radius of Zhengdong Station.

new town area. Panel D affirms our observation. Note that the normalized trend of market access climbed up in the post-HSR period but was relatively weak compared to the average trend for other new towns (see the red trend line in Panel C).



Panel A. New establishments.

Panel B. Nighttime light intensity.

Panel C. Market access.

Panel D. Normalized trends.

Figure 5: Trends in the number of new establishments, nighttime light intensity (obtained from DMSP/OLS dataset), and market access within a 5 km radius of Chuzhou Station.

Given similar proximity to the county center, what other factors might lead the new town to different growth trajectories? To study these factors, we present pre-HSR and post-HSR summary statistics for Zhengdong and Chuzhou in Table 1. Table 1 documents some patterns we see in Figures 4 and 5. Both new towns have a similar distance to the county center and have a city status according to our definition. The average number of new establishments was much larger in Zhengdong and experienced substantial growth in the post-HSR period, while the average number of new establishments only stood at 57 after the Chuzhou Station began its operation. Zhengdong also had a much larger size of the local economy, with average market access of 196 billion RMB throughout the estimation window. In contrast, the average market access for Chuzhou is only about 8 billion RMB.

We then turn to the economic conditions of the county where the new town is located. The population size was larger in Zhengdong before the HSR station arrival. The average county population size in the pre-HSR window stood at 3.88 million and rose to 4.10 million in the post-HSR period. Chuzhou had a much smaller county population size to start with (0.53 million) and it barely grew in the post-HSR period. The county-level share of the agricultural sector was larger for Chuzhou than that for Zhengdong. Travel time by driving and public transit are lower for Zhengdong, suggesting a better road network and public transportation system. All of this preliminary evidence points to the importance of adjusting for differences in county-level characteristics in our analysis.

Our case study motivates many interesting questions about the growth trends of new towns in China. Given that an HSR station is built and a new town is developed in a location, what determines the growth trend of this new town? In our empirical strategy we present below, we focus on proximity to agglomeration economies and market access and adjust for differences in observed county characteristics. The growth trend of the local economy might be steeper if the new town is closer to the area's superstar city (or county subcenter) and if the local market access is high. We also ask, what is the causal impact of HSR new town development on local economic growth compared to what would have happened in a location if the HSR station had not been constructed? Are there differentials in local economic outcomes for new towns with different *ex-ante* county-level economic conditions? We proceed to answer these questions in the following sections of the paper.

3 The Empirical Strategy

Building upon the new economic geography theory and literature on new town development in China, we present several empirical hypotheses about the local economic growth of HSR new towns.

Hypothesis 1. HSR new town location choice hinges on construction and demolition costs, proximity to agglomeration economies and the size of the local economic shock. Given that a new town is built, the local vibrancy is determined by its location and market access to surrounding urban markets.

In the station location decision-making, local government officials consider the benefits and costs of various stakeholders. HSR network requires substantial construction and demolition costs as well as the cost of trading with nearby subcenters. To generate more land revenue, local officials need to build the station in a location with good access to nearby urban markets to make the new town attractive for businesses and population. A good location and a strong market potential can jointly make the new town model sustainable.

Hypothesis 2. The effect of proximity to the county center on the new town vibrancy is nonlinear as the new town location also determines its access to the urban markets in the other county subcenters.

The proximity to the center of the county where the station is located plays a crucial role in determining the local economic growth of the new town as the new town mostly obtain benefits from favorable policies instituted by the local government. However, through a trade-based channel, economic activity in other urban markets might also affect the chances of the new town success. This bell-shaped relationship between transport costs and agglomeration has been proposed by new economic geography theory as well as documented in empirical evidence (Teixeira, 2006).

Hypothesis 3. Building a new town generates local economic growth in the new town center. These benefits vary according to the *ex-ante* economic size, proximity to agglomeration economies, and transportation infrastructure availability.

Building a new town is indeed similar to designating an industrial park or high-tech zone to a place, which is a prominent place-based policy that Chinese local governments use to stimulate the local economy. One key factor for its success is the synergy between this top-down investment, and the location's fundamentals, such as the economic base, proximity to nearby agglomeration centers, and transportation infrastructure.

Hypothesis 4. Some new towns remain "ghost towns" in the early years but are likely to boom for years to come, depending on their locations and the size of market access.

Firms and individuals are likely to hold off their decisions to move into the new town even though key players (proximity to agglomeration economies, market access, favorable policies) are in place. The level of uncertainty about the local economic prospect and place-based policies could affect the timing of when a new town becomes vibrant. Building an HSR station in a good location and having strong market access could lower this uncertainty and accelerate the local economic growth.

3.1 The empirical strategy

To test Hypotheses 1 and 2, one immediate challenge to our identification is that station location choice is endogenous. There is a growing strand of literature concerned with rigorous identification strategies to address the non-random route placement issue. Many previous studies adopt an IV approach to address this issue. Earlier research uses the historical network to instrument the current network (Baum-Snow, 2007; Zheng and Kahn, 2013; Baum-Snow et al., 2017; Dong et al., 2020). Using the unique setting of China's HSR net- work, Zheng and Kahn (2013) and Dong et al. (2020) also use military troop deployments to instrument current network. Moreover, a growing body of empirical literature uses IV constructed using the least-cost path spanning tree networks (Banerjee et al., 2020; Faber, 2014; Yu et al., 2019).

We build upon this literature to construct a least-cost path distance IV considering decision-making about route and station choice, engineering, construction, and demolition costs, and market access to the existing urban markets involved in the process of new town development in China (Wang, 2016). This approach contributes to our understanding of how necessary information about benefits and costs are factored into the new town location choice decision-making. Moreover, this approach also provides plausible counterfactual new town locations which allow us to estimate the treatment effect of local transport infrastructure upgrades and new town development on the local vibrancy.

Our key identifying assumption is that counties across prefectural-level cities are similar conditional on local characteristics. This assumption allows us to exploit variations in the distance across counties to identify the economic impact of HSR new town location choice. Thus, the basic source of identification in our paper comes from the cross-county variations in the HSR station location choice and local economic shock. Conditioning on county-level characteristics, we compare new establishment growth of otherwise similar counties with different distances to their county centers and different prefectural-level economic shocks to identify the causal relationship.

In the first-stage regression of our 2SLS estimation, where we predict the actual location choice using the least-cost path instrument variable (IV), we model the decision- making process over the station location choice given the configuration of HSR lines and upstream/downstream stations. The decision-maker, either the local government or the Ministry of Transport, carefully weighs both the benefit side and cost side factors. The benefits come from market access effects and agglomeration spillovers which are captured by market access to other urban markets in the same prefectural-level city. The costs involve the trade costs in terms of trading with the main city as well as construction and demolition costs due to HSR route configuration, which is captured by our least-cost path distance IV. In this sense, we use both the benefits due to transportation technology advances and the costs of demolition and construction to explain location choice decision-making of new town development.

We then turn to estimate the treatment effect of receiving an HSR new town (an HSR station) on the vibrancy of the target location rather than examining urban vibrancy determinants given that the new town has already been built. It is also important to see if facts can be established about the booming and ghost towns in terms of variation in treatment effect for locations with different local characteristics.

To estimate the causal effect of building an HSR new station on the urban vibrancy in its vicinity, our main empirical challenge is to consider which locations might be alternative locations for building HSR stations. Our unique spatial unit provides us with a setting that is distinct from those in previous literature on the HSR network. These studies are usually conducted at a certain administrative level where counterfactual locations are locations that do not have HSR stations. In our case, however, counterfactual HSR new town locations could be any locations within the county jurisdiction but not in an existing urban area. Throwing darts on the map and constructing a buffer around the darts would give us plausible counterfactual HSR station locations compared to the actual new town center which is often built on a blank slate as well.

To overcome this challenge, one plausible counterfactual new town candidate could be the midpoint location of the straight line connecting the upstream and downstream stations for a given actual HSR station. This counterfactual location does not differ too much from the actual location in terms of access to existing urban markets and distance to the county center, as illustrated in our least-cost distance instrumental variable construction.[14] Although differences might exist in some first-nature geographical attributes, these differences can be controlled by calculating the spatial first difference. The average number of establishments within 5 km of the counterfactual station is 292.5 compared to 497.4 of the actual station. These numbers are 92.1 and 166.5 when we use a buffer of 3 km.

3.2 2SLS estimation

To address the concern about non-random HSR route placement, we use a minimum cost path distance IV approach. We apply this approach to the sample of HSR new towns (no counterfactuals are included). Our estimation strategy consists of two steps. First, we model the HSR station location choice from both the cost side and the benefit side. Specifically, we use the plausibly exogenous cross-county variations in the least-cost distance from the HSR station to the county center and market access to surrounding urban markets to predict the actual HSR station location choice using Eq.(1):

$$D_{icsp} = \alpha_1 + \beta_{11} d_{icsp} + \beta_{12} M_{csp,t} + \gamma_1 X_{csp,t} + \delta_{1p} + \lambda_{1t} + \epsilon_{1icsp,t} \tag{1}$$

D_{icsp} is the distance from the HSR station i to its county center c in prefectural-level city s of province p, which reflects the actual HSR station location choice. This distance variable is static with no variations across time. d_{icsp} is the least-cost distance from the HSR station i to its county center c in a prefectural-level city s of province p, which reflects the costs involved in the construction of a new HSR station given the HSR lines and two neighboring station locations. $M_{csp,t}$ represents the market access of county c in year t. $X_{csp,t}$ contains county-level local characteristics of county c in prefectural-level city s of province p in year t.

[14] Most matched midpoint locations are located in the same county where the actual HSR location is located. We exclude the matched pairs in which midpoint location is not in the same county. Another thing to note is that endpoints of HSR lines are also excluded since an upstream or a downstream station may not exist.

δ_{1p} represents province fixed effects and allows us to take regional time-invariant unobserved differences into account. λ_{1t} is year fixed effects and captures the time trend related to macroeconomic business cycles. $\epsilon_{1icsp,t}$ is the error term with mean zero.

Second, we use the predicted location choice of the HSR station form the first step and Bartik local economic shock at the prefectural-city level to explain the vibrancy dynamics of the HSR new town using Eq.(2):

$$y_{icsp,t} = \alpha_2 + \beta_{21}\hat{D}_{icsp} + \beta_{22}M_{csp,t} + \gamma_2 X_{csp,t} + \delta_{2p} + \lambda_{2t} + \epsilon_{2icsp,t} \tag{2}$$

y_{icspt} is the number of new establishments of the station (new town) i in county c, prefectural-level city s of province p in year t. $\hat{D}_{icsp}$ is the predicted distance from the station i to the center of county c in a prefectural-level city s of province p in year t. Similar to Eq.(1), $M_{csp,t}$ represents the market access of county c in year t. $X_{csp,t}$ contains county-level local characteristics of county c in prefectural-level city s of province p in year t. δ_{2p} represents province fixed effects, while λ_{2t} is year fixed effects. $\epsilon_{2icsp,t}$ is the error term.

Considering that our distance variables are static, we also estimate the model in a long-difference variant of Eq.(2). Specifically, we examine the vibrancy change at the HSR station location during the entire estimation window by choosing a base year t_0 and the most recent year t_1 in our sample. The specification is displayed in Eq.(3) below:

$$y_{icsp,t_1} - y_{icsp,t_0} = \alpha_3 + \beta_{31}\hat{D}_{icsp} + \beta_{32}M_{sp,t_0} + \gamma_2 X_{csp,t_0} + \delta_{3p} + \lambda_{3t_0} + \epsilon_{3icsp,t_0} \tag{3}$$

$y_{icsp,t_1} - y_{icsp,t_0}$ is the change in the number of new establishments from base year t_0 to a future year period t_1. t_0 is 2006 and t_1 is 2015 in our sample data. $\hat{D}_{icsp}$ is the predicted distance from the station i to the center of county c in a prefectural-level city s of province p in year t. Similar to Eq.(2), M_{csp,t_0} represents the market access of county c in base year t_0. X_{csp,t_0} contains county-level local characteristics of county c in prefectural-level city s of province p in base year t_0. δ_{3p} represents province fixed effects, while λ_{3t_0} is year fixed effects.

The coefficients β_{11} and β_{12} together pinpoint the mechanism of HSR new town location choice. To determine where to build the station and thus the new town, the decision-makers (either local government or the Ministry of Transport) consider the cost and the benefit. To minimize the construction cost, the actual station location cannot deviate too much from the least-cost location choice, implying a positive sign of β_{11}. On the benefit side, given a larger shock to the local economic development, the new town could potentially be located further away from the county center to access the same level of economic spillovers, implying a positive sign of β_{12}.

The key parameters to estimate in Eqs.(2) and (3) are β_{21}, β_{22}, β_{31}, and β_{32}. Both theoretical and empirical work in urban economics show the attenuation of agglomeration economies with distance, indicating that β_{21} and β_{31} are expected to be negative. We also expect a higher level of urban vibrancy in the HSR new town given larger economic spillovers at the prefectural-city level. Thus, β_{22} and β_{32} are likely to be positive. We fit our model to the data we have assembled to test these hypotheses.

Note that the two-step estimation strategy is naturally embedded in a 2SLS estimation strategy where we instrument the endogenous HSR station location choice with the least-cost distance from the HSR station to the county center. Therefore, Eq.(1) is the first-stage equation in our 2SLS estimation, while Eqs.(2) and (3) demonstrate the second-stage specifications.

We need two crucial baseline assumptions to identify the effect: First, counties are similar after controlling for local characteristics so cross-county variation can be used to estimate the parameter. Second, the counterfactual new town center along a least-cost path affects changes in new-town-level economic outcomes only through actual location choice, conditional on province fixed effects, city status, and county-level characteristics.

3.2.1 The least-cost distance IV

Inspired by Banerjee et al. (2020) and Faber (2014), we construct a least-cost IV to "predict" the possible location choice of the HSR station. As illustrated in Figure 6, we draw a straight line connecting the stations before and after (denoted as A and C, respectively) the HSR new town (denoted as B). The midpoint of the segment AC is regarded as the location that is most likely to build an HSR station. Therefore, the distance between the county center B_c and the midpoint is the IV for the real distance between the HSR new town and the county center. Note that there is a requirement for the IV here that station A and station C cannot be located in the county where B is located. If that happens, we continue to look up or down the HSR line to find the first station that meets the requirement.

We argue that this straight line IV is exogenous for the following reasons. First, a straight line connecting two neighboring stations is the least engineering cost solution if we neglect other geographical constraints. This assumption is particularly effective for HSR because the speed of HSR is very fast and the turning radius is large, so straightness is an important consideration for HSR construction. This is different from other transport infrastructures such as highways, which require more considerations into terrain conditions, e.g., slope and roughness (Faber, 2014) because the speed of cars is relatively low and the turning radius is small. Second, when we have a straight line connecting two neighboring stations, a natural consideration is that the midpoint of the line is the most appropriate point to build a station because it minimizes the travel distance of the residents on the line segment. Third, since the endpoints of the straight line (A and C in Figure 6) locate outside the county of the HSR new town, thus we assume that the construction of the IV is purely geometrical and

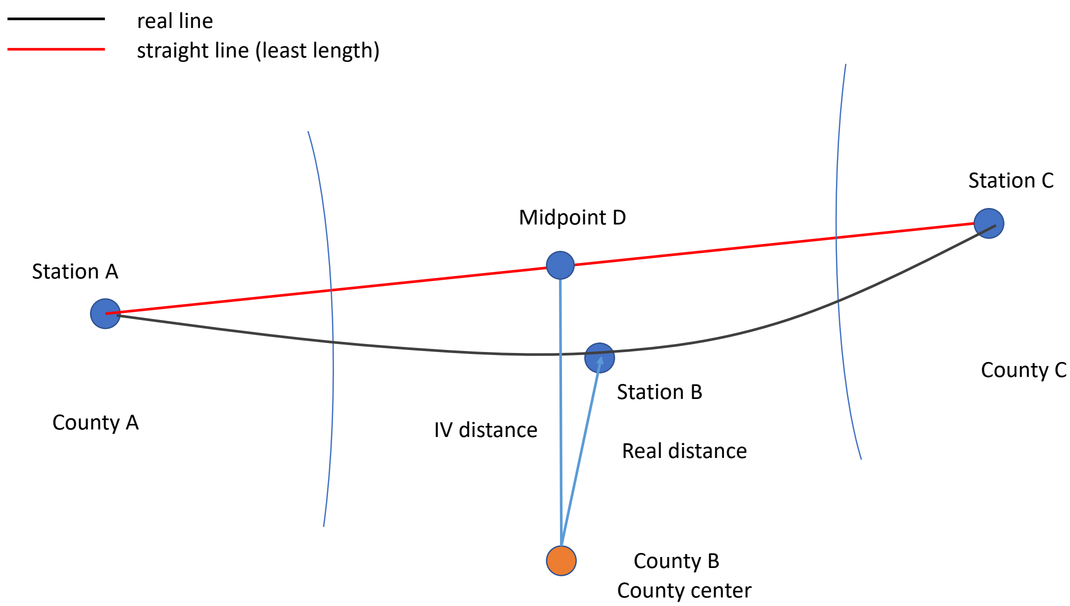


Figure 6: Illustration of the construction of the distance IV.

exogenous.

3.2.2 Market access

Following the literature (Hanson, 2005; Zheng and Kahn, 2013; Donaldson and Hornbeck, 2016), we measure the market access of a new town using the inverse-distance weighted sum of the urban markets around the host county where the new town is located.

$$MA_{is,t} = \sum_{j=1}^{N} Income_{js,t} \cdot e^{-\gamma d_{ijt}} \tag{4}$$

where $MA_{is,t}$ is the market access value of county i in year t. $Income_{js,t}$ is the market size of county j in the same city s in year t.[15] We use county government revenue as a proxy for the county market size as county GDP data contain many missing values in the county statistical yearbooks. d_{ijt} is the travel time (measured in minutes) by railway/HSR between county i and county j. γ is a spatial decay parameter that measures the size of travel cost and discounts urban market size in other counties based on bilateral travel time. N is the total number of counties in city s.

[15] Given the unique land finance scheme in China, it is likely that government officials try to gain the full market potential within the administrative boundary and reduce spillovers across boundaries. Thus, we only include existing economic mass in the same city

Following Zheng and Kahn (2013), we assume our counties are connected by railroads or highways. The spatial decay parameter γ is set to be 0.02, which is in the range of the spatial decay parameter estimates in the literature (Brakman et al., 2000; Hanson, 2005; Ahlfeldt and Feddersen, 2008, 2010). We assume the average travel speed to be 120 km/h before counties are connected by the HSR network.[16] The travel speed increases to 275 km/h after the HSR station is opened. The railroad distance between two subcenters is 1.2 times the great-circle distance between these two subcenters.

3.3 Treatment effect estimation

Using all identified HSR new towns and their matched counterfactual locations, we implement a difference-in-differences estimation approach to estimate the following baseline regression equation:

$$y_{icspt} = \alpha_4 + \beta_{40} \cdot treated_{icspt} + \beta_{41} D_{icsp} + \beta_{42} M_{cspt} + \gamma_4 X_{cspt} + \delta_{4i} + \lambda_{4t} + \epsilon_{4icspt} \quad (5)$$

where y_{icspt} is the number of new establishments within 5km of Station i in county c, prefectural-level city s, and province p in year t. $treated_{icspt}$ is a dummy variable that equals 1 if station i exists in year t. D_{icsp} is the great-circle distance of station i to the center of county c where the station is located. M_{cspt} is the market access of county c in year t. X_{cspt} is a vector of county-level observed characteristics of county c in year t. δ_{4i} is location/station fixed effects that capture the location-specific time-invariant unobserved heterogeneity. λ_{4t} controls for year fixed effects that consider unobserved time-related shocks to firm dynamics. ϵ_{4icspt} is a random error term with mean zero. β_{40} is the parameter of interest, which estimates the treatment effects of having an HSR station on the urban vibrancy of the surrounding area.

To examine the long-run changes in new town vibrancy, we also estimate Eq.(6), a long-difference specification of Eq.(5), as a robustness check.

$$y_{icspt_1} - y_{icspt_0} = \alpha_5 + \beta_{50} \cdot treated_{icspt_0} + \beta_{51} D_{icsp} + \beta_{52} M_{cspt_0} + \gamma_5 X_{cspt_0} + \delta_{5s} + \lambda_{5t_0} + \epsilon_{5icspt_0} \quad (6)$$

where $y_{icspt_1} - y_{icspt_0}$ is the change in the number of new establishments from base year t_0 to a future year period t_1. t_0 is 2006 and t_1 is 2015 in our sample data. All other variables are defined in similar ways as in Eq.(5) except that we control city fixed effects — δ_{5s} instead of location/station fixed effects to allow for enough within-group variation to identify the effect.

[16]The average speed of the conventional train is 120 km/h, which is also the common speed limit on highways.

The parameter of interest is β_{50}, which captures the long-run treatment effects of building an HSR new town on local economic growth.

Our identifying assumption is that the actual new town and the counterfactual new town tend to have similar trends in new firm formation before the HSR station is assigned to this county. This seems to be a reasonable assumption given that we define HSR new towns as the urban areas that witness little urban growth before the arrival of the stations from satellite imagery. Thus, their firm growth rates are likely to be similarly low before the peripheral land at the urban fringe is developed, adjusted for differences in county-level characteristics.

4 Results

4.1 Least-cost path 2SLS estimation

Table 3 presents our baseline panel estimation results of Eq.(2). Column (1) contains our two main explanatory variables: distance from the HSR station to the county center and access to the other urban markets in the city. We control for province fixed effects to take regional time-invariant heterogeneity into account, which allows us to use cross-city variation in the distance to the county center and market access to identify the vibrancy effects. We also include year fixed effects to capture any time trends and factors related to macroeconomic cycles. Column (2) adds log population size, the share of agricultural employment, and city status to Column (1). Standard errors are clustered at the city level to allow for arbitrary auto-correlation within cities.

We then use the least-cost path distance to instrument the actual distance from an HSR station to the county center and estimate the Eq.(2) using 2SLS estimation. Column (3) reports the second-stage results where we control for province fixed effects and year fixed effects but exclude other county-level covariates. Column (5) adds log population size, the share of agricultural employment, city status, a 4th order polynomial in local geographical characteristics to Column (3).

Column (4) presents the first stage results of Column (3), while Column (6) presents the first-stage results of Column (5). Note that the number of observations in Column (2) is smaller than that in Column (1) because of missing values in our covariates, as shown in the summary statistics table. Additionally, the sample size is smaller in Column (3), compared to Column (1), as we drop the cases where the actual HSR track and the least-cost track lie on different sides of the county center (see Figure 6).

Columns (1), (2), (3), and (5) of Table 3 reports a negative, significant coefficient on proximity to the county center across all specifications. The results are not sensitive to additional controls and 2SLS estimation. The elasticity of new establishments with respect to the distance from the station to the county center is roughly from -0.78 in Panel fixed effects estimation, while the size of the distance effect increases slightly to -0.83. The relative magnitudes of the coefficients in panel fixed effects estimation and 2SLS estimations suggest that the omitted variables in the panel fixed estimation bias our point estimates upward.

In Columns (1), (2), (3), and (5), the coefficient on the market access variable registers a positive sign and is statistically significant across specifications. These results are robust to the inclusion of additional controls. We find similar results when we use our 2SLS estimation strategy. In our preferred specification — Column (5), a 1% increase in market access leads to a 0.48% increase in new establishments, suggesting that better access to surrounding urban markets is more likely to sustain the new town development.

The first-stage results of 2SLS estimation highlight the location choice of HSR stations given benefits and costs of building an HSR station in the chosen location. The least-cost distance IV is valid according to the results of the weak identification test and underidentification test. A positive, significant coefficient on the distance IV suggests that construction and demolition cost, as well as proximity to agglomeration economies, are important considerations in the station location choice, conditional on market access to urban markets and other county-level characteristics in the city. A positive sign on market access in the first-stage regression suggests that HSR new town far away from the county center may be sustainable given a local economic shock that is large enough.

Table 4 repeats the analysis in Table 3 using the long-difference specification in Eq.(3).[17] The estimation window is from 2006-2015. We choose this window to cover two Five-Year Plan periods in China. As our distance variables are static, a long-difference setting could help uncover the main effects of distance and also allow us to investigate accumulated effects on local economic growth over the estimation window. To keep a reasonable sample size, we include the cases where the actual HSR line and the least-cost linear line are on different sides of the county center. However, our sample size in the long-difference setting still shrinks to less than 200 observations. Despite the small sample size, the results of Table 4 closely resemble those of Table 3 in all specifications, albeit slight differences in the size of coefficient estimates.

4.2 Nonlinearity in distance

An HSR station is likely to be a bypass station if it is too close to the main urban area. Jobs may still stay in the main urban area and the new town may not be able to gain benefits from the local economic shock. It is worth noting that our sample excludes those stations converted from existing rail stations and thus very close to the county center. An HSR new

[17] We can only estimate the model by controlling for a 3rd order polynomial in local geographical characteristics instead of a 4th order polynomial due to the reduced sample size in the long-difference setting.

town on average is about 13 km away from its county center, which is at the higher end of the close distance range in Bosker and Buringh (2017). Indeed, Chinese local governments tend to build new towns far away from their urban centers. Thus, we test for the positive effects of being at medium distance.

As the new town moves away from its county center, the new town may escape the urban shadow of the host county center and still have relatively good access to the urban market at the county center. We also consider the possibility that the neighboring county centers are likely to become more accessible to the new town as the station moves further away from its county center (or moves closer to other county subcenters). Both of these alternative explanations can lead to a reverse effect of proximity to agglomeration economies above medium distance. To test these hypotheses, we examine the nonlinearity in the effects of station location choice on the local economic growth of the new town.

Table 5 presents the results in a panel setting. The specifications are similar to those in Table 3. Column (1) reports the panel fixed effects estimation result for the regression controlling for only distance and market access variables. Instead of using the logarithmic distance to the county center, we include distance and its quadratic form to test the nonlinear impacts of station location choice. Column (2) adds county-level controls and a 4th order polynomial in geography characteristics to Column (1).

Column (3) renders the second-stage result of 2SLS estimation where we use the least-cost path distance and its quadratic term to instrument the actual distance and its quadratic term, respectively. Column (4) includes additional county-level characteristics and a 4th order polynomial in local geography characteristics based on the specification in Column (3). Province fixed effects and year fixed effects are included in all specifications to account for region-specific and time-specific unobserved heterogeneity correlated with both local economic characteristics and location choice. Standard errors are clustered at the city level to allow for arbitrary within-city auto-correlation.

Table 5 shows a significantly negative coefficient on distance and a significantly positive coefficient on its quadratic term, suggesting there exists a nonlinear relationship between proximity to the county center and its local economic growth. This nonlinear relationship is robust to additional controls and 2SLS estimation. Using the estimate in Column (4), the turning point is 31.50 km, which lies roughly in the middle of the medium distance range (20 km-50 km) in Bosker and Buringh (2017). Table 6 presents very similar results from a long-difference estimation. The turning point in Column (4) of Table 6 is 39.25 km.

The result supports our hypothesis that an HSR new town may capture economic spillovers from neighboring county centers as the station moves closer to them but further away from the center of the county where the station is built. In our sample dataset, the average radius of the counties containing HSR new towns is 27.4 km (sd. 12.4).[18] This average radius is slightly below the turning point revealed by the estimation results. In other words, given a strong local economic shock, local officials may choose to build the station close to the edge of the county boundary to amplify the economic spillover effects from nearby subcenters of other counties. An alternative explanation is that this observed location choice pattern is a result of competition for an HSR new town among county governments.

4.3 The treatment effect of new town development

We examine the treatment effect of new town development on the local economic growth by estimating Eqs.(5) and (6) with the data on HSR new towns and their matched counterfactual new town locations. Table 7 presents the estimation results. The first two columns report the baseline results estimating Eq.(5). Column (1) does not include any controls, while column (2) includes the log market access and county-level attributes such as population, the share of agricultural employment, and city status. We control for location/station and year fixed effects to consider station-specific and time-specific unobserved heterogeneity correlated with the station location choice. Note that distance to the county center and geography features are static at the station level and thus are absorbed by location/station fixed effects. Standard errors are clustered at locations to allow for arbitrary auto-correlation patterns within each station location.

Column (1) shows a strong positive treatment effect. Including county-level controls in column (2) does not substantially vary the result. Adjusted for differences in county-level characteristics, HSR new town gains 12.86% ($= e^{0.121} - 1$) more establishments than its counterfactual location where the HSR station and new town development are absent. These results are not surprising as building a new town using the HSR station as a hub is likely to boost up local economic activity through the channel of market integration and provision of favorable policies for an average location.

We estimate Eq.(6) to repeat our analysis in a long-difference setting. Columns (3) and (4) in Table 7 present estimation results. Column (3) includes city fixed effects and year fixed effects but does not adjust for differences in county-level covariates. Column (4) adds the log distance to the county center, log market access, county-level characteristics, and a 4th order polynomial in local geography characteristics. The long-difference specification documents a much larger treatment effect, implying about 277% ($= e^{1.327} - 1$) more establishments in the local area if the new town had been built. We interpret this large treatment effect as a medium-term impact of new town development on local economic growth over the decade (2006-2015).

[18] We calculate the area A of each county that has at least one new town. Then, we approximate each county as a circle and calculate the corresponding radius by $\sqrt{A}/\pi$.

This result is interesting as it could justify city officials' decisions to build a new town that might be a "ghost city" in the short term if they anticipated these medium-term effects and based their investment decision on the local economic prospect. Our analysis shows that a medium-term effect is larger than a short-run effect if an endogenous agglomeration (a new town) is created.

4.4 Heterogeneous treatment effects

Our observations show a sizable variation in the local vibrancy of new towns. Why have some new towns thrived while others have failed? According to the new economic geography theory, urban chances of new towns should depend on their locations relative to existing urban markets as well as the size of those markets. We test these theoretical predictions by documenting differential treatment effects across different levels of *ex-ante* economic conditions of the county where the new town is located. Due to the small sample size of the long-difference setting, we estimate Eq.(5) in all specifications concerning heterogeneous treatment effects, in which we control for location and year fixed effects and cluster the standard errors by location.

In Column (5) of Table 7, we interact the treatment indicator with four quartile dummy indicators of government revenue in 2005, where Q_i is the ith quartile ($i = 1, 2, 3, 4$).[19] Column (6) adds the log market access and county-level covariates to Column (5). Column (5) shows that the new town has the largest treatment effect (28.92% more new establishments than its counterfactual) if the HSR station is built in the county that lies above the Q_4 of *ex-ante* county-level government revenue (337.67 million RMB). This effect is even larger (31.26% more establishments) after adjusting for differences in county-level characteristics.

Column (7) of Table 7 regresses log new establishments on the interaction terms of the treatment dummy variable and quartile dummies of the 2005 county population. Column (8) adds the log market access and county-level covariates to Column (7). Results show that the treatment effect of new town development on the local vibrancy increases with the *ex-ante* population size at the county center. In Column (8), the treatment effect stands at 0.242, suggesting a 27.34% increase in the new establishments within a 5 km radius of the station given that the county population is above 2.11 million (Q_4 of 2005 county population). Endogenous city location theory (Fujita and Mori, 1996, 1997; Fujita et al., 1999; Behrens, 2007) predicts that a location's urban chances depend on its distance to existing urban centers. Locations too close to the existing urban center face competitions with the city, while locations far away from the existing urban center have poor access to existing urban markets. Bosker and Buringh (2017) document this nonlinear effect of distance on the existing

[19] The first HSR station was opened in 2008.

urban center on the new city's urban chances, i.e., the negative competition (urban shadow) effect and the positive effect of being at medium distance.

We test the predictions by interacting the treatment dummy variable with the quartile dummies of distance to the county center, travel time by driving, and travel time by public transit in Eq.(5). Results are reported in Table 8 where odd numbers of columns include location/station fixed effects and year fixed effects and even numbers of columns further control for market access and county-level covariates.

Columns (1) and (2) show that new towns close to the county center do not experience significant local economic growth. If any, the treatment effect has a negative sign negative if the new town is within a 5.35 km radius of the county center, suggesting the presence of an urban shadow effect. New towns at medium distance (5.42 km-15.58 km) witness the largest gains from developing a new town at the HSR station. The coefficient on $treated \times Q_4$ is significant in Column (1) but becomes statistically indifferent from zero once we adjust for differences in county-level characteristics. Although the new town may capture more cross-county-border spillovers by locating far away from the center of the county where it is located (thus closer to other county subcenters), this effect vanishes when we consider market access which may absorb cross-county-border agglomerative externalities.

In Columns (3)-(6) of Table 8, we measure the proximity to the county center using travel time instead to better capture real trade costs involved in transporting workers, goods, and services between the new town center and the county center. These travel time measures also contain important information about the capacity of local transportation infrastructure. In particular, driving time may reflect the road network conditions, while travel time by public transit may reflect the availability of the public transportation system (bus routes, subway systems, rapid trains, etc.). Note that travel time is retrieved in the current period. Thus, the cross-section variation in travel time may not be the same as that in *ex-ante* travel time but it is a good proxy.

Results in columns (3) and (4) closely resemble those in Columns (1) and (2). Due to negative competition effects, the treatment effect on the new towns close to the county center is not statistically different from zero. Conditional on county-level covariates, new towns at medium distance witness substantial new establishment creation, while the positive effect diminishes as the distance of the new town to the county center exceeds a certain threshold (about 30.87 minutes by driving). A new town would see a 32.98% ($= e^{0.285} - 1$) increase in new establishments if it takes about 20-30 minutes to drive to the county center from the new town center.

Columns (5) and (6) document the positive treatment effects on new towns at medium distance as well. Column (6) shows that a new town would see a 19.12% ($= e^{0.175} - 1$) increase in new establishments if it takes about 40-60 minutes to travel to the county center from the new town center by public transit. However, we do not observe any urban shadow effect. One possible reason might be that there is no public transit for some new towns to travel to the county center. Travel time retrieved from the API is then the travel time by walking, which distorts information about real travel costs in this circumstance.

4.5 Does this place-based policy merely reshuffle economic activity?

For any place-based policy, it is crucial to examine whether the policy stimulates real economic growth rather than just reshuffling economic activity between the target area and neighboring non-target areas. To do so, we also collect panel data on new establishments of counties adjacent to the treated locations to track neighboring economic activity over time. Pure growth in the new town center might show no significant drop in new business creation.

Table 9 presents the estimation results of the treatment effect on local economic activity in adjacent subcenters. Column (1) includes only the treatment group (HSR new towns) and regresses the log total new establishments in all neighboring counties on a dummy variable *Post* that equals 1 if the HSR station is in operation in the current year. Column (2) adds county-level controls (the log market access, county population, the share of the agricultural sector, and city status) to Column (1). Location/station fixed effects and year effects are included to absorb location-specific and time-specific shocks that might bias the estimates.

Column (3) expands the sample to include all matched midpoint counterfactual locations and estimates Eq.(5) except that the outcome variable is replaced with the log total new establishments in neighboring counties. In Column (4), we control for county-level covariates. Robust standard errors are clustered by stations to allow for arbitrary auto-correlation within each new town location. This investigation is to compare the economic activity in the adjacent counties of the treated location to that of the counterfactual location in the absence of the HSR arrival and new town development.

The coefficients on *Post* and *Treated* are all negative but statistically insignificant at the 10% level, providing strong evidence of real economic growth benefiting from HSR new town development rather than the reshuffle of economic activity over space. New establishments in the new town are more likely to stem from new business activities instead of the relocation of existing firms from neighboring counties. However, due to the lack of detailed firm-level information, we are not able to tell if the establishment is a new firm or a facility/branch expansion. Thus, we cannot completely rule out the possibility that new establishments burgeoning in the new town are those relocate from counties that are not adjacent to the county where a new HSR station opens. We leave this for future research.

4.6 Do some "ghost towns" prosper later?

While some HSR towns may be "ghost towns" in the short run, there remains the possibility that they will prosper later. For each new town, the number of years it takes for local economic growth to materialize can vary substantially due to the uncertainty about the timing of economic booms and downturns. In other words, panel data in our sample may be right-truncated so that some potential boom towns have not taken off by 2016.

To examine the timing of new town success and failures, we convert our panel data (including counterfactual locations) to survival data. We define our "failure" event (becoming vibrant), a bit counterintuitively, as the occurrence that a location has experienced positive firm growth for at least four consecutive years during our estimation window.[20] The time each location needs to thrive is recorded in the data. In our sample, about 38% of locations have thrived by the end of the estimation window. The average time a location needs to witness constant local economic growth is about 10 years with the minimum time being 5 years. This preliminary examination of our survival data could add evidence to the prevailing "ghost city" phenomenon in China.

Table 10 presents results from three survival models. Column (1) presents results estimating a Cox proportional hazards model. Column (3) reports the estimation results of a parametric survival model. Column (5) renders results estimating treatment effects for survival data using the inverse-probability-weighted regression adjustment method. Columns (2), (4) and (6) add additional controls (log county population, the share of the agricultural sector, and city status) to Columns (1), (3), and (5), respectively. We include the log distance to the county center and log market access in all specifications.

The Cox proportional hazards model and parametric survival model show very similar results. As the station moves away from the county center, the hazard ratio of thriving decreases, albeit the insignificant coefficient on the distance when we adjust for differences in county-level characteristics. Larger market access raises the hazard ratio of new town vibrancy. These results highlight the role of location and market access in accelerating local economic success in the new town development.

We combine the survival analysis and treatment-effects estimation in Columns (5) and (6) The results show that, if all locations upgrade to the HSR network and build a new town, the average time to the local vibrancy would be 3.4-4.4 years less than the counterfactual that no location in the population of interest builds an HSR station. That is to say, the average treatment effect of HSR new town development is to hasten the time to local vibrancy by about 4 years. The estimated average time to local economic vibrancy when no location in the population of interest has a station is 14.8-16.1 years.

[20] We do not use the years it takes to reach vibrancy after the HSR opens as a majority of stations are constructed after 2012, which makes it impossible to observe the vibrancy outcome by our definition.

Figures 7 and 8 provide a visual analysis of survival model estimates. Figure 7 presents Kaplan-Meier failure estimates by treated and controlled locations. It shows that locations that receive an HSR station tend to have a higher likelihood of realizing local vibrancy than their matched counterfactual locations. Figure 8 plots smoothed hazard estimates across time, showing similar patterns where treated locations are always more likely to become vibrant compared to controlled locations although the hazard estimates for two groups seem to converge over time.

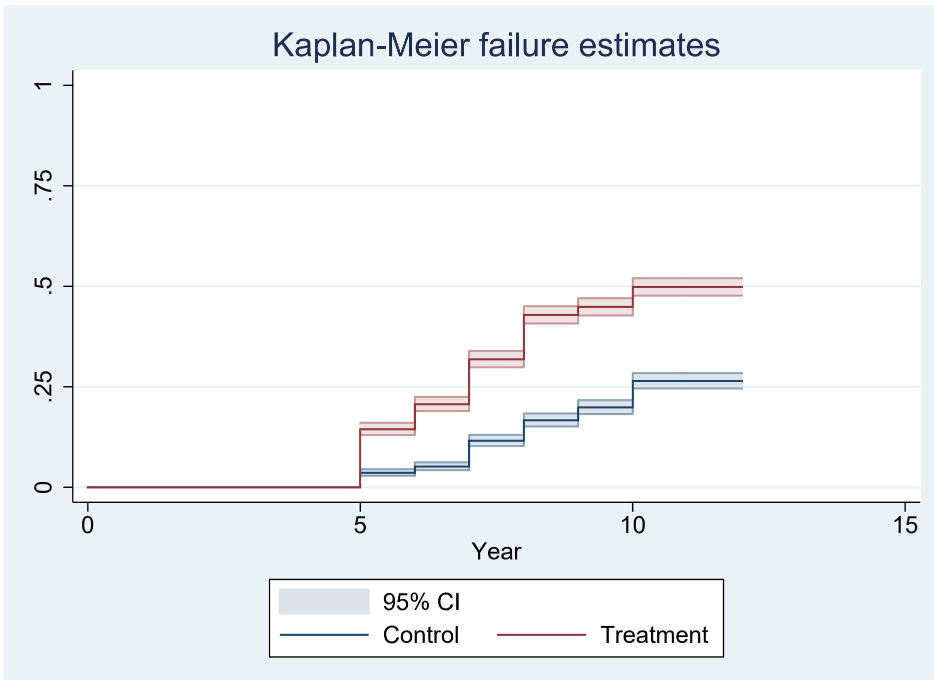


Figure 7: Kaplan-Meier failure estimates: Treated vs. controlled.

A back-of-the-envelope calculation suggests that building an HSR new town would reduce the average time needed for steady economic growth to 11.4-12.7 years. Note that our estimation window covers 12 years (2005-2016). Our survival analysis complements our previous analysis to provide a plausible explanation for a large number of "ghost cities" in China. To put it simply, some locations fail due to its undesirable location and weak local economic shock. Some others are currently "ghost cities" but may thrive in the future given a good location and good market access.

These results raise another interesting question: The HSR station has been built but economic activity has not co-located in the new town yet. In fact, on average, it takes more than 10 years to reach a steady growth trend. Why does economic activity not gravitate to the new town area sooner? One possible explanation is the uncertainty about the local economic prospect. Dixit and Pindyck (1994) discuss a theoretical approach to the capital investment decisions of firms, stressing the uncertainty of the economic environment in which

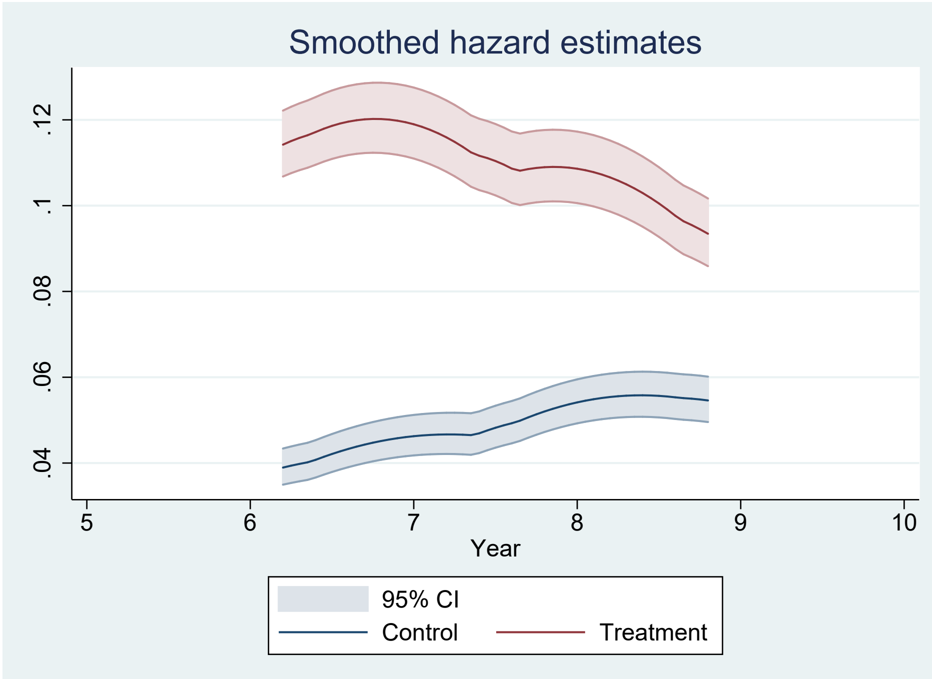


Figure 8: Smoothed hazard estimates: Treated vs. controlled.

these decisions are made.

Government officials and firms may embrace a real option valuations approach to guide their decision-making. On one hand, city officials hold an option to build a new town, although we do not model this process in this study. On the other hand, firms have the option to open businesses if the new town thrives. An option has a value when uncertainty ends so that firms and individuals are more certain about the future economic trends of the new town. Higher option values are associated with higher risks, suggesting that cities and firms that execute the option when high uncertainty disappears can realize large benefits of holding such a real option.

If the uncertainty level is high, the wise investment strategy for firms is to delay to minimize any possible incurred costs and wait until the vicinity booms. Localized economic shock (e.g., market access, favorable policies, improved transportation infrastructure, etc.) that is large enough may reduce the risk and alter the timing of the economic boom. In our case, the uncertainty could be an economic risk or a political one. For example, if a major regional project targeted at a nearby area is announced by the central government to start in the next couple of years, the uncertainty about short term growth can be reduced by the policy signal and the new town is likely to boom at that point.

4.7 City-level Bartik shock

To deal with the endogeneity issue with the local economic shock, we include the predicted local gross domestic product (GDP) instead of the actual local GDP. The predicted local GDP is obtained using the Bartik shift-share approach (Bartik, 1991), which provides plausibly exogenous variation in local GDP.[21]

China City Statistic Yearbooks started to report employment for 19 industry sectors in 2002. Following the literature (Détang-Dessendre et al., 2016; Chen et al., 2017; Zheng and Du, 2020), we regress the GDP growth between 2002 and subsequent years T on 2002 employment shares of 19 industries, prefectural-level control variables, and provincial fixed effects in prefectural-level city c.

$$\%\triangle y_{2002-T} = \alpha_0 + \sum_{i=1}^{20} \alpha_i * shr_{ic} + \beta \boldsymbol{X_{cT}} + \gamma_p + e_{cT} \tag{7}$$

where $\%\triangle y_{2002-T}$ is the growth rate of GDP or market integration from the initial year 2002 to year T. shr_{ic} is the 2002 employment share of industry sector i in city c as the total employment in industry i. $\boldsymbol{X_{cT}}$ represents the control variables at prefectural-level city c in a subsequent year T. γ_p is the provincial fixed effects. e_{cT} is the idiosyncratic error for city c in year T.

We then use the fitted values (predicted GDP growth rate $\%\triangle \hat{y}_{2002-T}$) from Eq.(7) and the GDP in the initial year 2002 to predict the local GDP level in year T.

$$y_{cT} = (1 + \%\triangle \hat{y}_{2002-T}) y_{c2002} \tag{8}$$

This Bartik shift-share type local economic shock provides plausibly exogenous variations since it is based on OLS using initial prefectural-level industry employment shares and does not constrain to national growth rates. We run the first-stage regressions of 2SLS estimation where we use the Bartik local economic shock to instrument the actual GDP. Our Bartik local economic shock strongly predicts the local GDP level in the first stage.[22] Results are available in the appendix on our website.

[21] Although we could include both the endogenous location choice and the endogenous local economy in the model and instrument them using least-cost distance and Bartik local economic shock, we focus on modeling station location choice in the first step and directly control for the exogenous economic shock. The coefficient on the Bartik local economic shock in the second-stage analysis is more of a reduced-form coefficient.

[22] As Chen et al. (2017) point out, this modified version of Bartik IV should be the best linear unbiased estimator (BLUE) and addresses two issues with the traditional Bartik IV in the first stage. First, it constrains growth rates for all industries in each city to equal their respective national growth rates rather than the actual industry patterns, often leading to a weak IV issue. Second, the Bartik IV is ineffective when the national trends are relatively weak compared to idiosyncratic city-level shocks, resulting in a weak first stage.

5 Discussion and Conclusion

Many Chinese cities build new towns to boost up local economic growth and increase fiscal revenues in their jurisdictions. While the development of new towns in some cities is successful, many other new towns are underdeveloped and have become "ghost towns." We focus on one specific type of new towns — HSR new town in this paper. Our goal is to examine how the HSR station location choice interacts with existing urban markets to determine the urban vibrancy of HSR new towns. We combine information from satellite images, local new sources, and nighttime light intensity to manually identify 180 HSR new towns out of 839 HSR stations. We then build the first-of-its-kind urban vibrancy measures using information from firm establishments at the new town level.

The local vibrancy of new towns connects to the ultimate question about the origins of cities: What leads to the emergence of new cities? Economic geography theory predicts that a location's physical geography and its position relative to already-existing cities (2nd nature geography) matters most for its urban chances. Empirical evidence shows that a location's physical geography characteristics are the dominant determinants of its urban chances (Bosker and Buringh, 2017). Our results suggest that the geography and the physical location play critical roles in determining the urban growth of planned new towns as well. We document the attenuation of within-city agglomeration spillovers with the distance to the urban center. In this respect, distance to the urban center reflects the transport cost of accessing the economic activity in the core-peripheral spatial structure.

The nonlinear effect of the distance to the urban center further points to a sub-center story where multiple neighboring sub-centers produce agglomeration spillovers to which the new town has access. Inter-regional market and trade integration goes beyond the administrative boundaries. The finding is consistent with the channels of trade integration in the increasing returns to scale trade theory and economic geography (Helpman and Krugman, 1985; Fujita et al., 2001). Our results also highlight the presence of the urban shadow effect, making new towns close to the county center less likely to succeed if the existing urban market size is large. These results are consistent with both theoretical work (Fujita and Mori, 1996, 1997; Fujita et al., 1999; Behrens, 2007) and empirical exercises (Bosker and Buringh, 2017; Cuberes et al., 2019).

Our study places more emphasis on 2nd nature geography rather than on 1st nature geography. We make some clarifications for this point: First, 1st nature geography is well captured in the HSR route placement as decision-makers consider construction and demolition costs which reflect preferences for certain favorable 1st nature geography attributes, such as terrain ruggedness, inland waters, existing transport infrastructure, etc. Second, new town development is mainly pushed forward by the administrative power of local government, providing sufficient capital investment and favorable policies to promote growth in the target location. For this reason, 2nd nature geography characteristics are likely to be the dominant determinants of the local vibrancy. Third, the relative importance of 1st nature geography diminishes due to advances in transport technology (e.g., HSR connections) and lower trade barriers that substantially reduce trade costs.

The market size effect is another potential channel, which highlights the role of a higher demand for local goods and services due to local economic shock (Acemoglu and Linn, 2004). These demand linkages between regions are strong and growing over time (Hanson, 2005; Zheng and Kahn, 2013; Donaldson and Hornbeck, 2016). Moreover, our results are consistent with recent empirical evidence documenting strong causal links between firm activity and market access to surrounding urban markets in China, both across and within cities (Zheng and Du, 2020; Du and Zheng, 2020).

Note that there is much empirical evidence showing that the core-periphery mechanism often leads to a reduction in local GDP growth among peripheral non-targeted regions connected by the HSR network due to *ex-ante* asymmetric market sizes and local fundamentals (Faber, 2014; Qin, 2017; Yu et al., 2019). Nevertheless, the new town development model examined in our paper is distinct from the traditional core-periphery setting in that new towns are targeted regions with favorable policies and capital influx. Falling trade costs, agglomeration forces from the urban center, existing market size, and policy favoritism are all at work sustaining the growth of new towns, although our data do not allow us to examine the effects of policy incentives.

To evaluate the policy, another main question to address is whether the local vibrancy gains in new towns are the net creation of jobs or merely a reshuffle of economic activity between new towns and neighboring metropolitan centers. Our data do not allow us to trace the distributional changes of individual firm activity. By tracking firm growth patterns in adjacent counties, our results suggest no economic activity reshuffling between the new town center and neighboring urban centers. There are several plausible explanations for our findings: First, local transport infrastructure upgrades and favorable policies along with the new town development lead to strong economic growth that stems from the new creation of business opportunities because of the reduced fixed cost of entrepreneurship. Second, new establishments may come from other cities in light of a competition channel where local governments build new cities to compete for financial resources, industries, and population. Third, the HSR network extends the scope of the general equilibrium from a within-city one to an intercity one due to reduced commuting cost, allowing urban residents to work in the new town but live in another location. This channel aligns with a suburban job growth story in light of a polycentric model (Glaeser and Kahn, 2004).

It is important to point out the vibrancy outcomes observed might be temporary and subject to change in the long run. It may take a few years to accumulate capital and a talented workforce to sustain a vibrant new town. As most stations were built after 2012, our data are likely to be right-truncated. Our survival analysis highlights a perspective of real options evaluation to help us understand decisions made by local government officials and firms. Results show that upgrading to the HSR network and building a new town hasten the time needed to spur robust local economic growth. It takes roughly 15 years on average for a location without an HSR station to thrive, while an HSR new town takes about 4 years less to witness steady economic growth. This suggests that many "ghost towns" that have an HSR station in our sample may thrive in the next decade or so, conditional its location and the size of existing nearby urban markets.

One may argue that this is quite a long period to see the benefits of building a new town. The term of office for local government officials is about 3-5 years. So why does the local government leader still build new towns even if it is difficult to realize the payoff in the short run? While our study cannot provide direct answers to this question, the literature highlights a political explanation — phantom urbanization: Local governments tend to favor the aesthetic urban form over economic, demographic or environmental repercussions (Sorace and Hurst, 2016).

Our study has implications for the place-based policies targeting at promoting local economic development by building HSR new towns. These policies are instituted to take advantage of the market integration on the massive HSR network but sometimes fail to boost up local economic activity. Can an HSR station sustain a new town built from scratch? Our results suggest that the success of a new town requires careful long-term planning. The exact odds depend mainly on policymakers' careful location choice decision- making and benefit-cost analysis concerning important factors such as the construction cost, existing local urban markets, coordination among nearby county centers, and local economic and political uncertainty.

References

Daron Acemoglu and Joshua Linn. Market size in innovation: theory and evidence from the pharmaceutical industry. *The Quarterly Journal of Economics*, 119(3):1049–1090, 2004.

Gabriel Ahlfeldt and Arne Feddersen. Determinants of spatial weights in spatial wage equations: A sensitivity analysis within a consistent framework. *Working Papers*, 022, 2008.

Gabriel Ahlfeldt and Arne Feddersen. From periphery to core: Economic adjustments to high speed rail. *ERSA Conference Papers*, ersa11p545, 2010.

Gabriel M Ahlfeldt and Arne Feddersen. From periphery to core: measuring agglomeration effects using high-speed rail. *Journal of Economic Geography*, 18(2):355–390, 2017.

Abhijit Banerjee, Esther Duflo, and Nancy Qian. On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, 145: 102442, 2020.

Keith Barrow. Ten years, 27,000km: China celebrates a decade of high-speed. *International Railway Journal*, 2018.

Timothy J Bartik. Who benefits from state and local economic development policies? *WE Upjohn Institute for Employment Research*, 1991.

Nathaniel Baum-Snow. Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805, 2007.

Nathaniel Baum-Snow, Loren Brandt, J Vernon Henderson, Matthew A Turner, and Qinghua Zhang. Roads, railroads, and decentralization of Chinese cities. *Review of Economics and Statistics*, 99(3):435–448, 2017.

Kristian Behrens. On the location and lock-in of cities: geography vs. transportation technology. *Regional Science and Urban Economics*, 37:22–45, 2007.

Maarten Bosker and Eltjo Buringh. City seeds: Geography and the origins of the European city system. *Journal of Urban Economics*, 98:139–157, 2017.

Steven Brakman, Harry Garretsen, and Marc Schramm. The empirical relevance of the new economic geography: Testing for a spatial wage structure in Germany. *CESifo Working Paper Series*, 395, 2000.

Anping Chen, Tianshi Dai, and Mark Partridge. Agglomeration and firm wage inequality: Evidence from China. *Working Paper*, 2017.

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

Zhenhua Chen and Kingsley E Haynes. Impact of high speed rail on housing values: An observation from the Beijing-shanghai line. *Journal of Transport Geography*, 43:91–100, 2015.

Zhenhua Chen, Kingsley E. Haynes, Yulong Zhou, and Zhaoxin Dai. *High Speed Rail and China's New Economic Geography: Impact Assessment from the Regional Science Perspective*. Edward Elgar, 2019.

Guanghua Chi, Yu Liu, Zhengwei Wu, and Haishan Wu. Ghost cities analysis based on positioning data in China. *arXiv preprint arXiv:1510.08505*, 2015.

Jeffrey P Cohen and Catherine J Morrison Paul. Agglomeration economies and industry location decisions: the impacts of spatial and industrial spillovers. *Regional Science and Urban Economics*, 35(3):215–237, 2005.

David Cuberes, Klaus Desmet, and Jordan Rappaport. Urban growth shadows. *Federal Reserve Bank of Kansas City, Research Working Paper*, 19-08, 2019.

Cécile Détang-Dessendre, Mark D Partridge, and Virginie Piguet. Local labor market flexibility in a perceived low migration country: The case of french labor markets. *Regional Science and Urban Economics*, 58:89–103, 2016.

Mi Diao, Yi Zhu, and Jiren Zhu. Intra-city access to inter-city transport nodes: The implications of high-speed-rail station locations for the urban development of Chinese cities. *Urban Studies*, 54(10):2249–2267, 2017.

Robert K. Dixit and Robert S. Pindyck. *Investment under Uncertainty*. Princeton University Press, 1994.

Dave Donaldson and Richard Hornbeck. Railroads and American economic growth: A "market access" approach. *Quarterly Journal of Economics*, 131(2):799–858, 2016.

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

Lei Dong, Sicong Chen, Yunsheng Cheng, Zhengwei Wu, Chao Li, and Haishan Wu. Measuring economic activity in China with mobile big data. *EPJ Data Science*, 6(1):29, 2017.

Xiaofang Dong. High-speed railway and urban sectoral employment in China. *Transportation Research Part A*, 116, 2018.

Xiaofang Dong, Siqi Zheng, and Matthew E. Kahn. The role of transportation speed in facilitating high skilled teamwork across cities. *Journal of Urban Economics*, 115:103212, 2020.

Rui Du and Siqi Zheng. Agglomeration, housing affordability, and new firm formation: The role of subway network in Beijing. *Journal of Housing Economics*, 48:101668, 2020.

Benjamin Faber. Trade integration, market size, and industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies*, 81(3):1046–1070, 2014.

Masahisa Fujita and Tomoya Mori. The role of ports in the making of major cities: self-agglomeration and hub-effect. *Journal of Development Economics*, 49:93–120, 1996.

Masahisa Fujita and Tomoya Mori. Structural stability and evolution of urban systems. *Regional Science and Urban Economics*, 27(4-5):399–442, 1997.

Masahisa Fujita, Paul Krugman, and Tomoya Mori. On the evolution of hierarchical urban systems. *European Economic Review*, 43(2):209–251, 1999.

Masahisa Fujita, Paul R Krugman, and Anthony J Venables. *The spatial economy: Cities, regions, and international trade.* MIT Press, 2001.

Jian Gao, Bogang Jun, Alex Pentland, Tao Zhou, Cesar A Hidalgo, et al. Collective learning in China's regional economic development. *arXiv preprint arXiv:1703.01369*, 2017.

Edward L. Glaeser and Matthew E. Kahn. *Handbook of Regional and Urban Economics. Chapter 56 - Sprawl and Urban Growth.* North Holland, 2004.

Gordon H. Hanson. Market potential, increasing returns and geographic concentration. *Journal of International Economics*, 67(1):1–24, 2005.

Elhanan Helpman and Paul R Krugman. *Market structure and foreign trade: Increasing returns, imperfect competition, and the international economy.* MIT Press, 1985.

J Vernon Henderson, Adam Storeygard, and David N Weil. Measuring economic growth from outer space. *American Economic Review*, 102(2):994–1028, 2012.

Adelheid Holl. Manufacturing location and impacts of road transport infrastructure: empirical evidence from spain. *Regional Science and Urban Economics*, 34(3):341–363, 2004.

Bo Huang, Yulun Zhou, Zhigang Li, Yimeng Song, Jixuan Cai, and Wei Tu. Evaluating and characterizing urban vibrancy using spatial big data: Shanghai as a case study. *Environment and Planning B: Urban Analytics and City Science*, 0(0):1–17, 2019.

Xiao Ke, Haiqiang Chen, Yongmiao Hong, and Cheng Hsiao. Do China's high-speed-rail projects promote local economy? — new evidence from a panel data approach. *China Economic Review*, 44:203–226, 2017.

Hans RA Koster and Jan Rouwendal. Agglomeration, commuting costs, and the internal structure of cities. *Regional Science and Urban Economics*, 43(2):352–366, 2013.

Dongya Li, Yi Lu, and Mingqin Wu. Industrial agglomeration and firm size: Evidence from china. *Regional Science and Urban Economics*, 42(1-2):135–143, 2012.

Zhigang Li and Hangtian Xu. High-speed railroads and economic geography: Evidence from Japan. *Journal of Regional Science*, 58(4):705–727, 2018.

Yatang Lin. Travel costs and urban specialization patterns: Evidence from China's high speed railway system. *Journal of Urban Economics*, 98:98–123, 2017.

Heli Lu, Chuanrong Zhang, Guifang Liu, Xinyue Ye, and Changhong Miao. Mapping China's ghost cities through the combination of nighttime satellite data and daytime satellite data. *Remote Sensing*, 10(7), 2018.

Tomoya Mori and Koji Nishikimi. Economies of transport density and industrial agglomeration. *Regional Science and Urban Economics*, 32(2):167–200, 2002.

Theodore Palivos and Ping Wang. Spatial agglomeration and endogenous growth. *Regional Science and Urban Economics*, 26(6):645–669, 1996.

Haixiao Pan and Ya Gao. Development of high-speed rail in the people's republic of China. *ADBI Working Paper Series*, 959, 2019.

Yu Qin. No county left behind? The distributional impact of high-speed rail upgrades in China. *Journal of Economic Geography*, 17(3):489–520, 2017.

Komei Sasaki, Tadahiro Ohashi, and Asao Ando. High-speed rail transit impact on regional systems: does the shinkansen contribute to dispersion? *The Annals of Regional Science*, 31(1):77–98, 1997.

Christian Sorace and William Hurst. China's phantom urbanisation and the pathology of ghost cities. *Journal of Contemporary Asia*, 46(2), 2016.

Lan Wang. *Studies on HSR New Town Planning and Development*. Tongji University Press, 2016.

Max D. Woodworth and Jeremy L. Wallace. Seeing ghosts: parsing China's "ghost city" controversy. *Urban Geography*, 38(8), 2017.

Mingzhi Xu. Riding on the new silk road: Quantifying the welfare gains from high-speed railways. *University of California, Davis: Working Paper*, 2017.

Feng Yu, Faqin Lin, Yihong Tang, and Changbiao Zhong. High-speed railway to success? The effects of high-speed rail connection on regional economic development in China. *Journal of Regional Science*, forthcoming:1–20, 2019.

Shenghuan Zhao and Dongqing Ma. Ghost city phenomenon along China's high-speed railway grid. *International Journal of Sustainable Society*, 9(3):210–225, 2017.

Longfei Zheng, Fenjie Long, Zheng Chang, and Jingsong Ye. Ghost town or city of hope? the spatial spillover effects of high-speed railway stations in China. *Transport Policy*, 81: 230–241, 2019.

Siqi Zheng and Rui Du. How does urban agglomeration integration promote entrepreneurship in China? evidence from regional human capital spillovers and market integration. *Cities*, 97:102529, 2020.

Siqi Zheng and Matthew E Kahn. China's bullet trains facilitate market integration and mitigate the cost of megacity growth. *Proceedings of the National Academy of Sciences*, 110(14):E1248–E1253, 2013.

Shengjun Zhu, Chong Wang, and Canfei He. High-speed rail network and changing industrial dynamics in Chinese regions. *International Regional Science Review*, pages 495–518, 2019.

Table 1: A case study: Zhengdong Station vs. Chuzhou Station

Panel A: Zhengdong Station	Full window	Pre-HSR	Post-HSR
# new establishments (5 km)	3550.75	1342.22	10176.33
Nighttime light intensity (5 km)	7517.67	6660.00	14379
Dist to the county ctr	9.68	9.68	9.68
Got'v revenue (Bn. RMB)	38.76	27.95	71.19
Market access (Bn. RMB)	195.98	139.42	365.65
Population (Million people)	3.94	3.88	4.10
Agr. share	0.63	0.75	0.02
Travel time by driving	-	-	22.86
Travel time by public transit	-	-	40.13
City status	Y	Y	Y
Panel B: Chuzhou Station	**Full window**	**Pre-HSR**	**Post-HSR**
# new establishments (5 km)	48.08	39.17	57.00
Nighttime light intensity (5 km)	2233.38	1527.60	3409.67
Dist to the county ctr	11.24	11.24	11.24
Got'v revenue (Bn. RMB)	2.81	1.09	4.53
Market access (Bn. RMB)	8.05	2.75	13.36
Population (Million people)	0.53	0.53	0.54
Agr. share	6.16	9.60	2.72
Travel time by driving	-	-	26.57
Travel time by public transit	-	-	62.60
City status	Y	Y	Y

Table 2: Descriptive statistics

	Obs.	Mean	Std. Dev.	Min.	Max.
Panel A: All years					
# establishments (5 km)	3,982	394.93	1216.17	0	34830
# establishments (3 km)	3,982	129.31	491.83	0	15236
Distance to county center (km)	3,982	13.25	12.43	2.17	71.46
Least-cost distance (km)	3,600	19.65	14.06	0.70	79.99
Gov't revenue ($\times 10^4$ RMB)	3,980	1.90×10^5	4.63×10^6	3021	6.41×10^7
Population ($\times 10^4$ people)	3,968	225.42	318.11	10	2449
Agri. employment share (%)	2,516	3.52	5.43	0.01	37.22
City Status	3,982	0.81	0.40	0	1
Panel B: 2006					
# establishments (5 km)	344	183.62	389.73	0	3439
# establishments (3 km)	344	54.2	131.42	0	1540
Distance to county center (km)	344	13.46	12.85	2.17	71.46
Least-cost distance (km)	312	19.85	14.36	0.70	79.99
Gov't revenue ($\times 10^4$ RMB)	344	6.65×10^5	1.56×10^6	3290	1.58×10^7
Population ($\times 10^4$ people)	344	211.26	305.53	10	2126.99
Agri. employment share (%)	220	5.55	5.61	0.45	25.20
City Status	344	0.80	0.40	0	1
Panel C: 2015					
# establishments (5 km)	358	750.09	1921.37	0	26260
# establishments (3 km)	358	262.73	816.86	0	11760
Distance to county center (km)	358	13.46	12.83	2.17	71.46
Least-cost distance (km)	324	19.90	14.34	0.70	79.99
Gov't revenue ($\times 10^4$ RMB)	358	3.19×10^6	6.86×10^6	27734	5.52×10^7
Population ($\times 10^4$ people)	358	228.04	337.19	12	2432.09
Agri. employment share (%)	256	0.86	1.89	0.01	12.05
City Status	358	0.80	0.40	0	1

Table 3: The short-run determinants of new town economic growth.

DV: Ln(# new est.+1)	Panel FE (1)	Panel FE (2)	2SLS (3)	2SLS (4)	2SLS (5)	2SLS (6)
Ln(distance)	-0.749***	-0.778***	-0.670***		-0.829***	
	(0.116)	(0.143)	(0.220)		(0.240)	
Ln(distance IV)				0.454***		0.501***
				(0.071)		(0.083)
Ln(market access)	0.554***	0.422***	0.548***	0.093***	0.488***	0.093
	(0.056)	(0.086)	(0.062)	(0.032)	(0.099)	(0.069)
Ln(population)		0.147			0.045	-0.040
		(0.133)			(0.135)	(0.088)
Agr. share		0.001			0.001	-0.0004
		(0.011)			(0.013)	(0.006)
City status		0.344*			0.509***	0.090
		(0.183)			(0.176)	(0.113)
Underidentification test						
Kleibergen-Paap rk LM statistic				27.06***		22.21***
Weak identification test						
Kleibergen-Paap Wald rk F statistic				40.90		36.45
10% maximal IV size				16.38		16.38
Province FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
County-level controls	N	Y	N	N	Y	Y
4th order polynomial in geography	N	Y	N	N	Y	Y
Obs.	1920	1214	1510		963	
Adj. R^2	0.608	0.608	0.660		0.664	

Standard errors in parentheses are clustered at the city level

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4: The medium-run determinants of new town economic growth.

DV: ΔLn(# new est.+1)	Panel FE		2SLS			
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(distance)	-0.681***	-0.806***	-0.608***		-0.851***	
	(0.152)	(0.220)	(0.216)		(0.243)	
Ln(distance IV)				0.438***		0.429***
				(0.102)		(0.108)
Ln(market access)	0.589***	0.450***	0.581***	0.075	0.469***	0.124
	(0.081)	(0.139)	(0.072)	(0.047)	(0.117)	(0.103)
Ln(population)		0.115			0.078	-0.099
		(0.210)			(0.183)	(0.128)
Agr. share		0.018			0.009	-0.025
		(0.023)			(0.019)	(0.017)
City status		0.307			0.476***	0.193
		(0.244)			(0.184)	(0.160)
Underidentification test						
Kleibergen-Paap rk LM statistic				21.12***		19.71***
Weak identification test						
Kleibergen-Paap Wald rk F statistic				18.53		15.89
10% maximal IV size				16.38		16.38
Province FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
County-level controls	N	Y	N	N	Y	Y
3rd order polynomial in geography	N	Y	N	N	Y	Y
Obs.	166	106	150		96	
Adj. R^2	0.541	0.492	0.608		0.524	

Standard errors in parentheses are clustered at the city level

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Nonlinear vibrancy effects of distance.

DV: Ln(# new est.+1)	Panel FE		2SLS	
	(1)	(2)	(3)	(4)
Distance	-0.147***	-0.144***	-0.174***	-0.189***
	(0.021)	(0.026)	(0.044)	(0.046)
Distance^2	0.002***	0.002***	0.003***	0.003***
	(0.000)	(0.000)	(0.001)	(0.001)
Ln(market access)	0.587***	0.440***	0.618***	0.541***
	(0.057)	(0.091)	(0.067)	(0.115)
Ln(population)		0.182		0.068
		(0.141)		(0.135)
Agr. share		0.004		0.004
		(0.011)		(0.014)
City status		0.364**		0.455**
		(0.173)		(0.181)
1st-stage F-Stat (IV)			57.16	51.81
1st-stage F-Stat (IV^2)			63.58	45.38
Province FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
County-level controls	N	Y	N	Y
4th order polynomial in geography	N	Y	N	Y
Obs.	1920	1214	1510	963
Adj. R^2	0.627	0.619	0.659	0.634

Standard errors in parentheses are clustered at the city level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: The medium-term nonlinear vibrancy effects of distance.

DV: $\triangle$Ln(# new est.+1)	Panel FE		2SLS	
	(1)	(2)	(3)	(4)
Distance	-0.135***	-0.167***	-0.159***	-0.157***
	(0.029)	(0.037)	(0.048)	(0.039)
Distance^2	0.002***	0.003***	0.003***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Ln(market access)	0.623***	0.423***	0.648***	0.468***
	(0.085)	(0.142)	(0.080)	(0.125)
Ln(population)		0.319*		0.175
		(0.174)		(0.156)
Agr. share		0.016		0.004
		(0.023)		(0.021)
City status		0.375		0.429**
		(0.235)		(0.190)
1st-stage F-stats (IV)			26.15	41.35
1st-stage F-stats (IV^2)			22.79	66.51
Province FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
County-level controls	N	Y	N	Y
3rd order polynomial in geography	N	Y	N	Y
Obs.	166	106	150	96
Adj. R^2	0.560	0.547	0.603	0.543

Standard errors in parentheses are clustered at the city level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Treatment effects of HSR new town development by *ex-ante* economic size.

DV: Ln(# new est.+1)	DID		Long-difference		*ex-ante* gov't revenue		*ex-ante* population	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated	0.137***	0.121**	2.159***	1.327***				
	(0.046)	(0.053)	(0.193)	(0.175)				
Treated×Q1					0.087	0.129	0.128	0.155*
					(0.075)	(0.080)	(0.082)	(0.092)
Treated×Q2					0.228**	0.250**	-0.032	-0.006
					(0.104)	(0.112)	(0.074)	(0.078)
Treated×Q3					0.005	-0.058	0.216**	0.160
					(0.074)	(0.085)	(0.104)	(0.115)
Treated×Q4					0.254***	0.272***	0.220***	0.242**
					(0.079)	(0.103)	(0.066)	(0.095)
Location FE	Y	Y	N	N	Y	Y	Y	Y
City FE	N	N	Y	Y	N	N	N	N
Year FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y	N	Y
Obs.	3982	2522	344	223	3982	2522	3982	2522
Adj. R^2	0.966	0.960	0.521	0.623	0.966	0.960	0.966	0.960

Standard errors in parentheses are clustered at the location level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Treatment effects of HSR new town development by *ex-ante* travel cost.

DV: Ln(# new est.+1)	Distance to county center		Travel time by driving		Travel time by public transit	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated×Q1	-0.048	-0.041	0.025	0.063	0.199**	0.192*
	(0.059)	(0.069)	(0.071)	(0.080)	(0.094)	(0.108)
Treated×Q2	0.218***	0.198**	0.215***	0.285***	0.169**	0.175**
	(0.064)	(0.094)	(0.081)	(0.097)	(0.069)	(0.085)
Treated×Q3	0.151*	0.150*	0.211***	0.083	0.211*	0.197
	(0.088)	(0.089)	(0.077)	(0.084)	(0.108)	(0.128)
Treated×Q4	0.224**	0.208	0.096	0.095	0.046	0.035
	(0.111)	(0.142)	(0.111)	(0.111)	(0.082)	(0.096)
Location FE	Y	Y	N	N	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	N	Y	N	Y
Obs.	3982	2522	3840	2522	3840	2522
Adj. R^2	0.966	0.960	0.965	0.960	0.965	0.960

Standard errors in parentheses are clustered at the location level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Economic activity in adjacent subcenters.

DV:	Only treated locations		Full sample	
Ln(# new est. in adj. counties)	(1)	(2)	(3)	(4)
Post	-0.010	-0.015		
	(0.034)	(0.043)		
Treated			-0.006	-0.009
			(0.031)	(0.033)
Location FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
Controls	N	Y	N	Y
Obs.	1872	1227	3744	2454
Adj. R^2	0.974	0.977	0.974	0.977

Standard errors in parentheses are clustered at the location level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Treatment effects estimation with survival-time data.

DV: Hazard ratio	Cox		Parametric		ATE	
	(1)	(2)	(3)	(4)	(5)	(6)
Ln(dist to county ctr)	-0.074*	-0.063	-0.103**	-0.075		
	(0.038)	(0.046)	(0.043)	(0.052)		
Ln(market access)	0.118***	0.271***	0.118***	0.331***		
	(0.017)	(0.031)	(0.019)	(0.036)		
ATE (treated 1 vs. 0)					-4.370***	-3.397***
					(0.403)	(0.467)
POmean (treated = 0)					16.112***	14.837***
					(0.352)	(0.406)
Controls	N	Y	N	Y	N	Y
Obs.	3972	2402	3972	2402	3972	2402

Robust standard errors in parentheses.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$