2021

THREE ESSAYS ON THE HIGH-SPEED RAIL NETWORK IN CHINA

Shiyu Cheng
University of Kentucky, shiyu.cheng@foxmail.com
Author ORCID Identifier:
https://orcid.org/0000-0003-0329-6559
Digital Object Identifier: https://doi.org/10.13023/etd.2021.322

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Shiyu Cheng, Student

Dr. William H. Hoyt, Major Professor

Dr. Carlos Lamarche, Director of Graduate Studies

THREE ESSAYS ON THE HIGH-SPEED RAIL NETWORK IN CHINA

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Gatton College of Business and Economics
at the University of Kentucky

By
Shiyu Cheng
Lexington, Kentucky
Director: Dr. William H. Hoyt, Professor of Economics and Public Policy
Lexington, Kentucky
2021

https://orcid.org/0000-0003-0329-6559

ABSTRACT OF DISSERTATION

THREE ESSAYS ON THE HIGH-SPEED RAIL NETWORK IN CHINA

My dissertation consists of three essays that study the economic consequences of China's high-speed rail (HSR) expansion.

In the first essay, I use the college admission cutoff scores to reveal students' college preferences under the enrollment quota. By exploiting the quasi-experimental variation in whether or not college cities are connected by the HSR network, I document a two-point increase in the cutoff scores following a HSR station opening in the college city using difference-in-difference (DD) approach. Colleges in the megacities experience a larger increase in cutoff scores after the station opening. These findings suggest that the HSR network stimulates "brain drain" from unconnected cities to connected cities, especially connected megacities.

The second essay examines the impact of better HSR accessibility on housing prices in Jiangsu Province. Using transaction data of new houses aggregated to the complex level, I compare the housing prices of properties close to the new HSR stations to those close to pre-existing HSR stations, before and after the new station openings. In a DD specification, I document that housing prices decrease by twenty percent in the areas where the station distance reduces due to the station opening outside the city.

The third essay investigates the impacts on household income. Using DD approach, I document that urban households experience a significant increase in total household income following the opening of HSR station in their city. While labor earnings increase, the probability of having business income decreases. Moreover, labor income of the households whose heads work in the manufacturing sector increases little, but for households whose heads work in the transport or communications sectors increases much more than other households, suggesting that the HSR network facilitates urban industry specialization.

KEYWORDS: High-Speed Rail, College Admission Score, Housing Price, Household Income

Author's signature: Shiyu Cheng

Date: August 1, 2021

THREE ESSAYS ON THE HIGH-SPEED RAIL NETWORK IN CHINA

By
Shiyu Cheng

Director of Dissertation: Dr. William H. Hoyt

Director of Graduate Studies: Dr. Carlos Lamarche

Date: August 1, 2021

ACKNOWLEDGMENTS

I am grateful for the help and support of many people. First, I would like to thank my advisor, Dr. William Hoyt, for his patient guidance and kind help throughout my life as a doctoral student at the University of Kentucky. I thank Dr. Carlos Lamarche, Dr. Josh Ederington, and Dr. Rajeev Darolia, the rest of my dissertation committee, for their time and feedback. I also thank Dr. Michael Samers for serving as the outside examiner to my dissertation defense. In addition, I am grateful to Dr. James Ziliak, my supervisor at the University of Kentucky Center for Poverty Research. My work benefited from many skills that I learned when I was a research assistant there. This dissertation would not be possible without them, and their passion on research inspires me to pursue my career in academia.

I also thank Dr. Wenbin Zang for continuous encouragement since my junior year at Southwestern University of Finance and Economics in China. I thank my friends, Sen Lu and Xuche Gong, who are working on their dissertations in different fields of economics, Wen Yang and Lishu Wang for always being there. My doctoral study would not have begun without their encouragement and support.

I want to acknowledge the support of my family. Due to the epidemic, I came back last winter and have lived with my parents since then. I treasure the half-year I spent with my family when I prepare the dissertation, and I am grateful for their support in my daily life. I will leave again this fall for work and take their love and sweet memory with me like always.

Finally, thank you, Shiyu, for not giving up.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

In the past two decades, China designed, planned, and built the world's largest high-speed rail (HSR) network. Its construction, at a remarkedly fast pace, obviously affected China's economy in numerous ways, particularly those localities that were connected to the HSR network and those localities that "lost out". My dissertation, which consists of three distinct essays, explores a few of the distribution impacts of this innovation in transportation infrastructure on the Chinese economy with a particular focus on its impact on the regional distribution of growth.

The HSR network serves two primary objectives. First, it reduces the heavy burden of the conventional rail system by providing additional passenger capacity. As a result, in the current rail system in China conventional slow rail mainly transports freight and the HSR mainly serves intercity passenger travel. The second goal of the Chinese government is to use the HSR network to connect the major cities in all provinces. By doing so, they expect to stimulate economic growth and regional rebalance, especially in economically backward regions.

In chapter 2, I provide evidence of an unintended consequence of the HSR network. Although the government wanted to reduce regional inequality in economic development, the HSR network may exacerbate it by facilitating the sorting of high-ability people into connected cities, especially connected megacities. In this chapter, I focus on a specific group of what might be considered high-ability people, college students. More specifically, I examine how the HSR affects the sorting of college students based on one measure of their ability, the scores on the National College Entrance Examination (NCEE). Under the enrollment quota, changes in college admission cutoff scores reflect the evolution of students' college preferences. By exploiting the quasi-experimental variation in whether or not college cities are connected by the HSR network, I document a two-points increase in the cutoff scores following a HSR station opening in the college city using difference-in-difference (DD) approach. However, if the students' home province has already connected to the HSR network, the effect of a following HSR route connecting the college city to the students' home province is statistically insignificant. As the HSR competes with other traffic modes in different distance ranges, the average treatment effects may be masked by the heterogenous effect across distance ranges. I find that station opening decreases the admission cutoff scores for in-province colleges while increases it for colleges out of students' home province. The positive effect is the greatest for out-of-province colleges that are close to the students' home province, i.e., within 150 km. The effects of route connection are still statistically insignificant no matter the college is in province or out of province, far or close. This suggests that the reduction in transportation cost may be not the main mechanism through which the HSR network changes students' college choices. To uncover the redistributional effects on college attendance. I further explore the heterogeneous impacts on the students from more developed provinces versus less developed provinces and find that the increases in the college cutoff scores for out-of-province colleges are mainly driven by the students from less developed provinces. As more than 96% of colleges in the sample are out of students' home province, I then focus on the out-of-province colleges to examine the most critical aspects that make out-of-province colleges more attractive after connecting to the HSR network. Colleges in the megacities experience a larger increase in cutoff scores after the station opening. HSR station in the college cities boosts cutoff scores for the elite colleges but decreases cutoff scores for the other colleges. At the same time, the HSR connection only has positive effects for colleges in the elite projects. These findings suggest that the HSR network stimulates "brain drain" from unconnected cities to connected cities, especially connected megacities.

In chapter 3, I analyze the effect of the HSR network on housing prices using Jiangsu Province as a case study. While a local government can decide where to place the new HSR station within its locality, it cannot influence the site selection process in other cities. Therefore, the reduction in distance to the closest HSR station caused by a new station opening outside of the city is exogeneous. Using transaction data of new houses aggregated to the complex level, I compare the housing prices of properties close to the new HSR stations to those close to pre-existing HSR stations, before and after the new station openings. In a DD specification, I document that housing prices decrease by twenty percent in the areas where the station distance is affected by the station opening in nearby cities. The effect of station opening in local city is statistically insignificant. A one-kilometer decrease in station distance is associated with a 0.5 percent decrease in housing prices. The impacts are negative and stable when I compare properties in different distance radius (106km, 75km, 50km, 25km, 15km) to the closest HSR station. Due to the lack of new complexes within 15km of the new station, I cannot exclude the possibility that the HSR station can have positive effects on housing prices within a closer distance. As suggested in the literature (Gibbons and Machin, 2005), the negative impact might be explained by a demand shift towards houses closer to the new HSR station.

The previous studies emphasize the distributional effects of the HSR rail network in human capital and housing market. In chapter 4, I employ household data from China

Household Income Project (CHIP) to investigate its impacts on household income. Using DD approach, I document that urban households experience a significant increase in total household income following the opening of a HSR station in their city. While labor earnings (wage/salary income) increase, the probability of having business income decreases. The likelihood of having property income does not change, on average. However, those effects are heterogeneous across industries and age cohorts. As the new transportation mode mainly serves the intercity travelers and decreases the face-to-face interaction costs between cities, HSR stimulates specialization towards industries that requires high communication skills and low manual tasks. My results reveal that labor income of the households whose heads work in the manufacturing sector increases little, but for households whose heads work in the transport or communications sectors increases much more than other households. Moreover, younger households aged between 25 to 34 gain little in total income following the HSR connection. At the same time, the rise in labor income is much smaller for the households with high-skilled heads. The changes in relative cohort sizes might explain these results.

The remainder of the dissertation consists of the aforementioned chapters.

CHAPTER 2. HIGH-SPEED RAIL NETWORK AND BRAIN DRAIN: EVIDENCE FROM COLLEGE ADMISSION SCORES IN CHINA

2.1 Introduction

Transportation infrastructure changes the spatial allocation of population (Redding and Turner, 2015). While the most apparent effects of changes in transportation infrastructure are changes in the populations of the affected regions, the characteristics of the people living in these regions may also change (Rhode and Strumpf, 2013). Though a vast literature finds evidence that transportation infrastructure affects the location of both employment and residence (Baum-Snow, 2007; Duranton and Turner, 2012; Baum-Snow et al., 2017; Baum-Snow, 2019), how it affects the spatial sorting of people of differing ability has been scarcely examined. In the case of this study, the measure of ability used is college admission scores.

Transportation infrastructure can contribute to the improvement of regional productivity through multiple channels. First, research in urban economics typically links transportation infrastructure to agglomeration effects that arise from an inflow of new workers (Duranton and Turner, 2012; Lin, 2017). Second, transportation infrastructure can boost productivity by increasing innovation even without any increase in density (Agrawal et al., 2017; Dong et al., 2020). Instead, it lowers the cost of knowledge flows, facilitates knowledge sharing across more distant regions. Third, transportation infrastructure may lead to sorting of workers that shifts the spatial distribution of workers' productivity. If workers who are inherently more productive choose to locate in cities with better transportation infrastructure, the cities will experience a growth in productivity. Less productive workers may be crowded out of the region. As a result, we expect to observe an inherently more able worker at the margin.

In this chapter, I test the third channel: whether the high-speed rail (HSR) network in China changes the college choices of high school graduates. While understanding decisions of where to attend college is of interest itself, the decisions made by college freshmen give insights into the locational choices of young workers entering the labor market. Students are influenced by many of the same factors that influence non-student migrants (McHugh and Morgan, 1984) because many students consider the place of college as a potential place of employment in the future. The fact that graduates often stay in the place of college after completing their education (Winters, 2011; Haapanen and Tervo, 2012; Ma and Pan, 2014) also leads credence to this argument.

The higher education system in China is heavily regulated. Every year colleges make enrollment plans, including quotas for the number of students admitted from each province. Once approved by the Ministry of Education (MOE), the number is unchangeable. For most students, scores on the National College Entrance Examination (NCEE, also known as Gaokao) are the only admission criteria. Holding enrollment constant, the change in college admission cutoff scores reflects the evolution of students' college preferences. If more high-scoring students apply to the college because of the HSR connection, its admission cutoff score will rise.

Building a stronger workforce is a crucial public policy issue, especially in China, where the Hukou (household registration) system still limits city sizes[1]. On the one hand,

---

[1] Under the Hukou system, people are attached to the locality of their residence registration. This registration confers specific local benefits, such as health care, public education, retirement pension, employment opportunities, etc. The system places barriers

local governments use preferential policies to attract high-skilled workers and usually award them Hukou status after working a sufficient number of years in the city. On the other hand, they support local institutes to train their raw talent who become high-skilled workers after graduation. Admission cutoff scores indicate the ability of incoming students and signal how successful the local government is in this competition for talent.

For colleges, admission cutoff scores show their popularity and signal their quality. Colleges with higher scores gain bargaining power and more resources from the government. However, any change in the cutoff scores caused by HSR construction is unintended -- cities are not connected to make certain colleges more attractive. Instead, one goal of expanding the HSR network is for China to rebalance its growth geographically (Lawrence et al., 2019).

The HSR network expands rapidly since 2008 and reaches 38,000 km (about 24,000 miles) in total length in 2020[2]. It reduces travel time for its users at speeds over 200 km/h (124 miles/h). The World Bank team (Lawrence et al., 2019) reports that the average cost for each kilometer of a HSR line is from 104 million CNY to 139 million CNY (24 – 32 million USD per mile). For such a huge investment, it is of great importance to fully understand its redistributional effect. If the HSR network encourages "brain drain" to the connected cities, particularly the megacities, it may exacerbate the regional inequality against its objective.

---

on migration and is used for population control of the cities. Au and Henderson (2006) consider the Hukou system a main reason for a large fraction of cities in China being undersized.

[2] Data source: "Railway Statistics Bulletin 2020", National Railway Administration of the People's Republic of China, 19 April 2021, http://www.nra.gov.cn/xxgkml/xxgk/xxgkml/202104/t20210419_147769.shtml (in Chinese). The number for 2013 is 11,000 km.

This chapter uses the expansion of the HSR network since 2008 as a quasi-experiment. With data on college admission scores in the Gaokao Web (http://college.gaokao.com/schpoint), the difference-in-difference (DD) results show that after the college city is connected to the HSR network (i.e., a HSR station opens in the city), the admission cutoff scores in the connected colleges increase by 2 points on average comparing to the scores in the unconnected colleges. However, if the students' home province has already connected to the HSR network, the effect of a following HSR route connecting the college city to the students' home province is statistically insignificant.

As the HSR competes with other transportation modes in different distance ranges, the average treatment effects may be masked by the heterogenous effect across distance ranges. I find that station opening decreases the admission cutoff scores for in-province colleges while increases it for colleges out of students' home province. The positive effect is the greatest for out-of-province colleges that are close to the students' home province, i.e., within 150 km. The effects of route connection are still statistically insignificant no matter the college is in province or out of province, far or close. This suggests that the reduction in transportation cost may be not the main mechanism through which the HSR network changes students' college choices.

To uncover the redistributional effects on college attendance, I further explore the heterogeneous impacts on the students from more developed provinces versus less developed provinces, and find that the increases in the college cutoff scores for out-of-province colleges are mainly driven by the students from less developed provinces. Moreover, the mobility of students from more developed provinces seems to increase more as the in-province colleges for those provinces experience a larger decrease following the HSR station opening.

I then focus on the out-of-province colleges to look into the most important aspects that make out-of-province colleges more attractive after connecting to the HSR network. Colleges in the megacities experience a larger increase in cutoff scores after the station opening. At the same time, the HSR connection only has positive effects for colleges in the elite projects. These findings suggest that the HSR network stimulates “brain drain” from unconnected cities to connected cities, and especially to the connected megacities.

This chapter contributes to the literature on high-skilled migration. In their review article, Kerr et al. (2017) recognize that a significant share of high-skilled migration occurs before the workers enter the labor market but during the skill-building process when they choose the location where they receive higher education. These migrants choose destinations based on the quality of available schools, the possibility of subsequently entering the labor market, and future professional opportunities (Rosenzweig, 2006; Kato and Sparber, 2013; Grogger and Hanson, 2011, 2015). While these studies focus on international migration, my research highlights the impacts of improving transportation technology on domestic migration within China.

This chapter also contributes to the literature on ability sorting under rigid labor demand (or restricted on resident population). Kleven et al. (2013) find evidence of sorting effects that low taxes attract high-ability football players who displace low-ability players and explain the results in a simple model of migration and taxation with rigid labor demand. Although with a very different driving force, my finding that the connected colleges experience an increase in the cutoff scores provides evidence of sorting effects under the enrollment quota. This has implications for the literature explaining the productivity advantages in big cities. Using Spanish data, De La Roca and Puga (2017) document that the distribution of workers' initial unobserved ability has no difference across city sizes. Hence, they conclude that sorting on unobservables is unimportant. This result can be true in a labor market with free entry and exit geographically. However, in a country like China, where the Hukou system is still working, and some cities have population caps, the sorting effects should have been paid more attention.

My study builds on the literature evaluating the economic impacts of the HSR network in China. Recent papers document that after the HSR connection, housing prices in the secondary cities increase (Zheng and Kahn, 2013), employment grows by 7 percent (Lin, 2017), researchers become more productive (Dong et al., 2020). My study sheds light on the redistributional effects. Qin (2016) finds that peripheral counties along the upgraded railway lines experienced reductions in GDP and GDP per capita following the upgrade because the economic activities divert from peripheral counties to the urban core. I also find evidence of redistributional effects that sharp human capital inequality across regions which is against the objective of rebalancing the regional economies by HSR construction.

The remainder of this chapter is organized as follows. Section 2.2 introduces the institutional background of China's higher education system and the HSR network. Section 2.3 describes the data, and Section 2.4 presents empirical strategies for quantitative analysis. Section 2.5 analyzes the empirical results. Section 2.6 concludes.

2.2 High-Speed Rail and College Admission in China

2.2.1 The High-Speed Rail (HSR) Network

In 2003, China built its first experimental high-speed rail (HSR) route, Qinhuangdao and Shenyang passenger-dedicated line, which is 405 km (250 miles) long. It was the only HSR route in China until 2008, when the HSR network began to experience rapid expansion, as shown in Figure 2.1[3].

The original objective of the HSR network was to provide additional capacity for the overloaded conventional rail network. The goal soon became improving regional and provincial connectivity to support economic development and urbanization with a focus on enhancing inclusiveness and reducing poverty in the economically backward regions. China is thus seeking to rebalance its growth geographically (Lawrence et al., 2019).

Given the objectives, the placement of HSR stations and routes is unlikely to be random. As per the original objective, more developed or fast-growing cities should be connected due to the urgent need for enhancing transportation capacity. Moreover, strong economic conditions allow cities to invest more in local universities or other amenities. To reach its objectives of enhancing inclusiveness and reducing poverty, the Ministry of Railway (MOR) should place HSR stations in disadvantaged cities, often lacking investment in higher education. We do not know which objective dominates when the MOR selects HSR cities -- thus, the sign of any potential bias in the OLS regressions is ambiguous.

[3] As the college data are from 2008 to 2013, I only include HSR maps between 2007 and 2013. The figures are from Lawrence et al. (2019).

As the HSR network is not designed to make certain colleges more attractive, there is little concern about the possibility of reverse causality. But it is still possible that some omitted variables (potential confounding factors), such as simultaneous investment, are correlated with the HSR status and students' college preferences. Difference-in-difference (DD) approaches are frequently used in the literature on HSR, including studies by, for example, Qin (2016), Lin (2017), Dong et al. (2020). Researchers have found similar pre-trends on numerous factors between the connected cities and unconnected cities[4]. Thus, the changes in these factors are more likely to be consequences of the HSR expansion than the determinants of the HSR placement. Assuming that any relative changes in amenities or college investment (potential confounding factors) following the HSR connection are caused by the HSR expansion, we estimate the "total effect" of HSR. If the parallel trend assumption is satisfied, the DD estimate of the total effect is unbiased.

2.2.2 College Admission

Scores on the National College Entrance Examination (NCEE, also known as Gaokao) are a good measure of students' inherent ability. Although the labor market does not reward high NCEE scores directly, employers do value college quality (Li et al., 2012) and undergraduate GPA (Li and Zhang, 2010). Both can be largely predicted by the NCEE scores.

To apply for colleges, students are required to take the exam in their home province (i.e., the province where their Hukou is). In 2017, the number of test-takers was

---

[4] Qin (2016) tests the pre-trends of GDP, GDP per capita, fixed asset investment. Lin (2017) tests railway ridership, road ridership, air ridership, employment, GDP, housing prices, fixed investments, retail sales, and total patents. Dong et al. (2020) test research productivity measured by research paper publications and citations.

approximately nine million. Around seventy percent of the test takers are finally admitted, and about fifteen percent can attend a first-tier university[5]. While closely monitored by the Ministry of Education (MOE), provinces have some flexibility in choosing exam questions and determining the details of the admission process. Within each province, the exam and the admission process are uniformly administrated. Therefore, the exam scores are only comparable within each province-year unit.

The typical exam and admission process can be described in five steps. First, in June, students take the exams. Second, based on the distribution of the scores and provincial admission quotas, each province announces tier cutoff scores. Students receive their exam scores and the provincial tier cutoff scores. Each tier has a separate admission process, which starts from the top tier and ends in the bottom tier. In the third step, students rank their college and major preferences and then submit the list of college-major pairs in the online admission system. In July, in the fourth step, students are assigned to colleges and majors one by one, from the highest score to the lowest score. The student with the highest score is assigned the first. Thus, they can be admitted wherever they want to attend the most. If the student's first choice college-major pair is filled by higher-scoring students, the system will search the next choice on the list until the student is assigned. If all the listed choices are full, the student will be lifted out of the first-round admission process. Finally, by the end of July, students will know to which university they are admitted.

[5] Colleges and universities in China can be divided into three broad tiers. The first tier is key universities that are typically administrated by the central government or the provincial governments. The second tier is regular universities, consisting of public universities administrated by the provincial governments and private universities. The third tier is three-year specialized colleges. In this chapter, I focus on the first-tier universities because the admission process makes the scores only comparable within each tier.

Students who miss the first round will be informed by the same online system about which universities still have spots open and repeat steps 3 and 4. If they still miss the second round, they can participate in the admission process (steps 3 and 4) of the next tier universities. Although some provinces let the students rank and submit their preferences before they know their scores, this will not affect our results since the policy applies to everybody within the province and the province-year fixed effects capture the differences in the admission process across provinces.

Admission cutoff scores may be interpreted as "prices" in the higher education market determined by the interaction between the students (demanders) and the colleges (suppliers). If more high-scoring students rank a particular college above others, the admission cutoff score will rise, and the low-scoring students who could have been admitted before will be crowded out. Therefore, the marginal student, who gets admitted at the lowest score, now has higher inherent ability.

2.3 Data

The HSR information is mainly from the Chinese High-Speed Rail and Airline Database (CRAD) of the Chinese Research Data Services (CNRDS) Platform, which includes station address, station opening date, route opening time, and main stops on the route[6].

University information and the scores between 2008 and 2013 are collected from the Gaokao website (http://college.gaokao.com/schpoint). For each university, I obtain its

[6] I confirmed the information on the stops of each route from the official railway service website (http://www.12306.cn).

address, status of Project 211/985[7], and enrollment from each province in a given year. The scores vary by college by province, by year, and by exam package[8]. Besides the cutoff scores, I also collect the average scores and the highest scores of the college.

For all the specifications, I control for a set of time-varying characteristics of the college cities, including GDP per capita, population, public green land area, and the number of hospital beds. These variables are collected from the China City Statistical Yearbook 2009 – 2014.

Table 2.1 provides summary statistics by treatment status. Panel A represents the sample used in the analysis of station opening effect and Panel B represents the sample used in the analysis of route connection effect. I drop the college cities with an HSR station before 2009 in the station sample and the city-province pairs connected by a HSR line before 2009 in the route sample, which leads to a slight difference in the sample sizes. To estimate the pure effects of a HSR route directly connecting to the students' home province, the students' home province must be connected to the HSR network before the route opening. Otherwise, the estimated effects of route will be mixed with the effects of station opening in home province.

The average number of students who take a certain exam package admitted by a college from a particular province in one year is 66 for the station sample. Enrollment in

---

7 Project 211 and Project 985 are two elite university projects established by the MOE. 211 stands for constructing 100 world-leading universities in the 21st century. There are 112 universities in Projects 211. 985 stands for the announcement date of the project – May in 1998. There are 39 universities in Project 985, and all the 39 universities are in Project 211 as well.

8 All test takers need to take three required subjects, Chinese, math, and a foreign language. Besides these, students can choose to take either the social-science package or the nature-science package. The social-science package includes history, political science, and geography. The nature-science package includes physics, chemistry, and biology.

the first-tier universities is stable over time. First, about 45 percent of our observations are for colleges under the central government's direct administration. Although the overall enrollment keeps growing, these colleges' enrollment only increases slightly (see Figure A 1). Second, first-tier universities offer top-quality education, which should limit their ability to expand.

Megacities are defined as Beijing, Shanghai, Guangzhou, and Shenzhen. Guangzhou and Shenzhou are located in Guangdong Province, so I consider students from Beijing, Shanghai, Guangdong Province as from mega-provinces. I also define smart cities that have at least two universities in Project 211/985[9]. Provinces that have a smart city are defined as smart provinces.

2.4 Empirical Model

When thinking of the HSR expansion, we are interested in the effects of a HSR station opening in the college city and the effects of a HSR route connecting the college city to the students' home province. In this section, I specify the empirical models that I estimate.

My empirical strategy has two dimensions: I estimate the effects of station opening and route connection separately using the difference-in-difference (DD) approach. Then I analyze the heterogeneous effects by adding interaction terms of dummies indicating

[9] There are nine smart cities, Beijing, Shanghai, Guangzhou, Nanjing, Wuhan, Tianjin, Chengdu, Changsha, Xi'an. The corresponding smart provinces are Beijing, Shanghai, Guangdong Province, Jiangsu Province, Hubei Province, Tianjin, Sichuan Province, Hunan Province, Shaanxi Province. They are provinces that have a smart city. All these nine cities belong to the "New First-Tier Cities" in China. In Dong et al. (2020), they are defined as the "megacities."

treatment status and dummies of college characteristics. I also split the sample into more developed and less developed home province subsamples.

2.4.1 Difference-in-Difference Specification

I first examine the relationship between college scores and HSR treatment status. The baseline estimation strategy is a DD specification of the form:

$$Score_{ipst} = \alpha + \beta Station_{i,t-1} + \gamma' X_{ct} + \rho Enroll_{ipst} + \delta_i + \theta_{spt} + \varepsilon_{ipst} \quad (2.1)$$

for the station effects, and

$$Score_{ipst} = \alpha + \beta Route_{ip,t-1} + \gamma' X_{ct} + \rho Enroll_{ipst} + \zeta_{cp} + \theta_{spt} + \varepsilon_{ipst} \quad (2.2)$$

for the route effects.

The dependent variable, $Score_{ipst}$, represents admission cutoff scores, average scores, or the highest scores, respectively, of college i for students choosing exam package s from home province p in year t. $Station_{i,t-1}$ indicates whether the city of college i has a HSR station in the previous year $t-1$. $Route_{ip,t-1}$ indicates whether a HSR line connects the college city to the students' home province in the previous year. The value of $Station_{i,t-1}$ (and $Route_{ip,t-1}$) equals zero for the year of the HSR station opening (route connection)[10] because the NCEE takes place in the middle of each year, and the HSR takes effect gradually.

I define HSR routes/lines as railway lines running at an average speed of 250 km/h or more (i.e., G-class passenger train service) and intercity lines running at an average speed of 200 km/h or more (i.e., C-class service)[11]. HSR connection between college city

---

[10] For example, suppose a station opens in 2012. For the observations in 2012 and before, the variable $Station_{i,t-1}$ equals 0. For the observations in the year 2013 and after, the variable $Station_{c_i,t-1}$ equals 1.

[11] This definition is consistent with Lin (2017).

and student province means students can take a HSR train from their home province capital to the college city directly without transfer in a third city.

The vector X_{ct} are the control variables for time-varying characteristics of the college city c, including GDP per capita, population, public green land area, the number of hospital beds. The variable $Enroll_{ipst}$ is the enrollment for the college-province pair for exam package s in year t. The station equation (2.1) includes university fixed effects (δ_i), and home province by year by exam package fixed effects (θ_{spt}). Standard errors are clustered at the college province level. The route equation (2.2) includes college-city and student-home-province pair fixed effects (ζ_{cp}). Other variables are defined as the same as in specification (2.1). Standard errors are clustered at the province pair level.

To test the parallel trend assumption for the DD specification, I estimate two separate regressions for the station effects, i.e., equation (2.3), and the route effects, i.e., equation (2.4). For the station opening effects, I regress scores on a set of dummies indicating one to four years before or after the HSR station opening. The year before the station opening is the base year thus omitted. I also include the same control variables and fixed effects as in equation (2.1). The $Station_{i,t}$ dummy equals one when the college city c_i has its own HSR station in operation. m and n represent its mth lag and nth lead.

$$Score_{ipst} = \alpha + \sum_{m=2}^{4} \beta_m Station_{i,t-m} + \sum_{n=0}^{4} \beta_n Station_{i,t+n} + \gamma' X_{ct} + \rho Enroll_{ipst} + \delta_i + \theta_{spt} + \varepsilon_{ipst} \tag{2.3}$$

For the route connection effects, I regress the scores on a set of dummies indicating one to four years before or after the HSR route connection. The specification is:

$$Score_{ipst} = \alpha + \sum_{m=2}^{4} \beta_m Route_{ip,t-m} + \sum_{n=0}^{4} \beta_n Route_{ip,t+n} \\ +\gamma' X_{ct} + \rho Enroll_{ipst} + \zeta_{cp} + \theta_{spt} + \varepsilon_{ipst} \tag{2.4}$$

Similarly, the year before the route connection is the base year and omitted. The same control variables and fixed effects of equation (2.2) are included.

2.4.2 Heterogeneous Analysis

As the HSR network may differently affect the colleges with specific characteristics, I then add interaction terms of dummies indicating treatment status and dummies of college characteristics into the regressions to conduct heterogeneous analysis. The specifications are

$$Score_{ipst} = \alpha + \beta_1 Station_{i,t-1} + \beta_2 D_i + \beta_3 Station_{i,t-1} * D_i \\ +\gamma' X_{ct} + \rho Enroll_{ipst} + \delta_i + \theta_{spt} + \varepsilon_{ipst} \tag{2.5}$$

$$Score_{ipst} = \alpha + \beta_1 Route_{ip,t-1} + \beta_2 D_i + \beta_2 Route_{ip,t-1} * D_i \\ +\gamma' X_{ct} + \rho Enroll_{ipst} + \zeta_{cp} + \theta_{spt} + \varepsilon_{ipst} \tag{2.6}$$

where D_i is a vector of dummies indicating college attributes. All other variables are the same as previously defined.

First, as the HSR competes with other transportation modes at different distance ranges, I let the vector D_i indicate the distance to students' home province. The omitted group is in-province colleges thus β_1 shows the effect of HSR connection on the in-province colleges. The effects on out-of-province colleges at different distance ranges are $\beta_1 + \beta_3$.

Second, students from provinces with different economic conditions may respond differently to HSR connection. To investigate the sorting effects, I then split the sample into more developed and less developed home province subsamples, i.e., mega-provinces versus secondary provinces, smart provinces versus other provinces (defined in section 2.3), and let D_i indicate whether the college locates in the megacity and whether it is in the elite project (211/985).

Results and detailed analysis are presented in section 2.5.2.

2.5 Results

2.5.1 Baseline Results

Figure 2.2 shows the event study results on the college cutoff scores. The HSR station/route has zero effects before its operation, which supports the parallel trend assumption. The dynamic impacts on the average scores and the highest scores are shown in Figure 2.3. Again, no evidence of different pre-trends is found.

Table 2.2 reports the DD results. Column (1) represents the impacts on the college cutoff scores, column (2) represents the impacts on average scores, and column (3) represents the impacts on the highest scores. The college admission cutoff scores in the connected colleges increase by 2 points on average following the HSR station opening. The effects are smaller and statistically insignificant for the average and highest scores. This is not surprising because, under the quota system of enrollment, we should observe the change of inherent ability on the marginal student, who gets admitted by the college with the lowest score. The average student and the top-scoring student have a larger set of feasible college choices, hence are less affected by the enrollment quota.

Note that the estimated effects are an average for all years following the HSR treatment. As the event study results show that these effects increase gradually over time, we may expect larger coefficients with additional years of data in the future. The system is far from equilibrium because more stations and routes are planned and under construction. The results here, then, should be interpreted as estimates of the short-run effects of the HSR network.

As can be seen in Table 2.2, Panel B, the coefficient on *Route* is statistically insignificant. Given that the sample is of home province already connected to the HSR network, this indicates that the effect of a subsequent HSR route connecting the college city to the students' home province is statistically insignificant. This might be because the primary mechanism through which the HSR network affects people's location choice may not be the reductions in transportation costs but increased economic opportunities created by the HSR connection in the destination city. These baseline findings are robust under alternative specifications using different definitions of the dependent variables and the independent variables (see Table A 1 and Table A 2).

However, these estimates of average treatment effects may be masked by the heterogenous effect across the students' home provinces and attributes of the college. Therefore, I examine how the effects of HSR may vary by a number of factors, including the distance between college and home province, in-province vs. out-of-province colleges, colleges in megacity vs. colleges not in megacity, and university programs. This analysis uses specifications (2.5) and (2.6) and divides the sample into the relevant subgroups.

2.5.2 Redistributional Effects and Heterogeneous Analysis

This section examines the redistributional effects of the HSR network with a focus on the cutoff scores. Not all colleges experience growth in admission cutoff scores after the connection, and students from different provinces respond differently to the HSR expansion.

2.5.2.1 Distance

HSR competes with highway and air for different ranges of distance. Highways dominate in the 0-150 km range, and HSR dominates between 150-800 km. In the range of 800-1,200 km, HSR competes with air. Air dominates if the distance is over 1,200 km[12] (See Figure A 2). In this section, I test if the impacts on colleges are different based on distance.

Table 2.3 shows the results adding the interaction terms of HSR treatment status and the dummies indicating distance ranges. The omitted group is in-province colleges with zero in the distance. Following a HSR station opening in the city, in-province colleges experience an 8-points decrease in the cutoff scores, which implies that the HSR connection in home province encourages students to attend colleges out of their home province. Out-of-province colleges within 150 km range benefit more from the HSR connection comparing to farther colleges. The effects on colleges farther than 150 km are close, which may be because the distance is measured as the straight distance between the college city to the students' home province capital in this chapter. Lawrence et al. (2019) also acknowledge that the margins of the distance ranges are fuzzy because of the different price and speed assumptions on HSR lines.

The coefficients of HSR route at all ranges are statistically significant. This further confirms that travel costs play a limited role in changing students' college preferences.

---

[12] Lawrence et al. (2019) use the number of passengers and passenger-kilometers to define the dominant traffic mode in each distance band.

2.5.2.2 Choices between In-Province versus Out-of-Province Colleges

Table 2.4 reports heterogeneous impacts of HSR connection on in-province colleges and out-of-province colleges. Panel A shows the impacts of HSR station opening. Following HSR station opening in home province, the cutoff score for in-province colleges decreases by 7.8 points on average (column (1)). This suggests that the HSR connection increases students' mobility by encouraging more students to attend colleges out of their home provinces. Then I divide the sample into subgroups according to the students' home province. Columns (2) and (3) are the effects on students from mega-provinces or secondary provinces. Columns (4) and (5) are the effects on students from smart provinces or other provinces. In general, the economic conditions of mega-provinces are better than smart provinces and then other provinces. In-province colleges for mega-provinces experience the largest decrease following the HSR station opening. The decrease in smart provinces is also larger than it in other provinces. This implies that the mobility of students from more developed provinces seems to increase more due to HSR connection in an in-province city.

On average, HSR station opening in the college city out of home province will increase the cutoff scores by 2.39 (= -7.78 + 10.17) points. For the subgroups, the effects of HSR station opening in the out-of-province colleges are positive, except for the students from mega-provinces. The effect of HSR station in an out-of-province college city for students from mega-provinces is negative (-2.69 = -14 + 11.31). However, these effects are all statistically insignificant.

Panel B shows the impacts of HSR route connecting the students' home province to the college city. The coefficients have the same signs as in panel A but with smaller magnitudes. The only significant results are obtained for students from smart provinces in column (4). Being connected by a HSR route will decrease the cutoff scores in the in-province colleges by 3.7 points. This is mainly because when a HSR route connects the in-province college, the home provinces are also connected to somewhere out of the province, thus increasing students' mobility.

2.5.2.3 Colleges in Megacities versus Colleges in Secondary Cities

More than 96% of the colleges in the sample are out of students' home province. In Table 2.5, I focus on the out-of-province colleges and divide them by whether they locate in a megacity. Panel A shows the impacts of station opening. On average, the cutoff scores for colleges in mega-cities increase by 5.05 (= 1.212 + 3.838) points, while the increase for colleges in the secondary cities is statistically insignificant. Comparing the results for students from mega-provinces (column (2)) and other provinces (column (3)), we can find that the positive effects of HSR station for out-of-province colleges are mainly driven by students in the secondary provinces. Colleges in the megacities also become more popular with students from smart provinces after a HSR station opening in the college city (column (4)). Still, the impacts of HSR route are statistically insignificant in panel B.

2.5.2.4 College Quality Signaled by Project 211/985

Students with higher NCEE scores have a larger feasible set of college choices. If the high-scoring students change their college choices, the low-scoring students who would have been admitted may be crowded out to other places. Thus, I expect the effects of HSR connection is larger for better quality colleges. In China, Project 211 and Project 985 are two elite higher education projects. In general, people believe that education quality in the Project 985 colleges is higher than the rest of Project 211 colleges. Colleges in Project 211/985 are better than other colleges.

Table 2.6 shows the results adding the interaction term of HSR treatment status and the dummies indicating Project 211 college and Project 985 college. It is worth noting that when categorizing colleges by education quality, the coefficients on route connection in panel B become mostly significant. This suggests that for high school graduates, the most important determinant of college choice is college quality.

While colleges in Project 985 obtain the greatest increase in the cutoff scores, the scores in the rest of Project 211 colleges also increase due to the HSR connection. Colleges not in Project 211/985 losing their popularity after the connection. The HSR network encourages students (at least high-scoring students) to sort into the high-quality HSR connected colleges regardless of their home provinces. Colleges excluded by Project 211/985 but connected by the HSR seem to be the biggest losers. Their cutoff scores even decrease comparing to the unconnected colleges.

2.6 Conclusion

The urban economics literature generally asserts that the sorting of workers based on inherent ability is an unimportant source of productivity advantages in big cities (De La Roca and Puga, 2017). Here, I explore the sorting of college students caused by the HSR expansion in China. Colleges have enrollment quotas similar to migration barriers and population caps in big cities created by the Hukou system. My findings provide evidence that transportation infrastructure can change the spatial distribution of students' inherent ability under the enrollment quota, which suggests that we should pay more attention to the sorting effect when migration is restricted.

For my purposes here, student ability is measured by the NCEE scores. Holding enrollment constant, the change in college admission cutoff scores reflects the evolution of students' college preferences. The HSR network encourages high-scoring students to sort into high-quality colleges and colleges in the megacities. The "brain drain" effect exacerbates the regional inequality of human capital, which is against the objective of the HSR network -- to rebalance the regional development by HSR construction.

Tables

Table 2.1 Summary Statistics

Panel A: Sample for Station Opening Effect	Total	Control	Treat
	(1)	(2)	(3)
Dependent Variables:			
Cutoff Score	555.7	545.9	558.5
	(41.24)	(40.18)	(41.11)
Average Score	570.0	556.9	573.9
	(40.55)	(39.34)	(40.08)
Highest Score	590.0	575.1	594.4
	(42.16)	(41.41)	(41.37)
Cutoff Score for Tier-1 Colleges	532.1	531.9	532.2
	(36.26)	(38.66)	(35.53)
Treatment Status:			
Station	0.399	0	0.516
College Attributes:			
Enrollment for College-Province Pair by	66.14	63.12	67.02
Exam Package	(183.4)	(175.0)	(185.8)
Nature-Science Package	0.580	0.580	0.580
Project 211	0.387	0.333	0.403
Project 985	0.243	0.222	0.250
Under Central Govt.	0.458	0.256	0.517
Under MOE	0.406	0.247	0.452
Distance to College			
Out-of-Province College	0.964	0.958	0.965
Distance below 150 km	0.00310	0.000795	0.00377
Distance between 150 and 800 km	0.310	0.314	0.309
Distance between 800 and 1200 km	0.259	0.253	0.261
Distance greater than 1200 km	0.397	0.399	0.396
Characteristics of College Cities			
GDP per Capita	65,308	40,907	72,433
	(37,355)	(19,447)	(38,313)
Population	883.2	987.4	852.8
	(632.4)	(1,190)	(315.4)
Public Green Land Area	55.49	45.28	58.47
	(51.20)	(37.73)	(54.15)
Number of Hospital Beds for 10,000	55.10	48.88	56.92
Persons	(15.91)	(19.96)	(14.00)
# of observations	22,251	5,029	17,222

Table 2.1 Summary Statistics (Continued)

Panel B: Sample for Route Connection Effect	Total (1)	Control (2)	Treat (3)
Dependent Variables:			
Cutoff Score	554.8	552.2	569.9
	(42.37)	(42.84)	(35.99)
Average Score	568.9	566.2	584.3
	(41.79)	(42.35)	(34.60)
Highest Score	588.4	585.7	604.0
	(43.47)	(44.15)	(35.45)
Cutoff Score for Tier-1 Colleges	530.9	528.7	543.9
	(37.33)	(38.20)	(28.53)
Treatment Status:			
Route	0.0475	0	0.324
College Attributes:			
Enrollment for College-Province Pair by Exam Package	51.14	49.02	63.52
	(119.5)	(117.7)	(128.7)
Nature-Science Package	0.578	0.580	0.568
Project 211	0.377	0.373	0.401
Project 985	0.236	0.236	0.238
Under Central Govt.	0.446	0.438	0.492
Under MOE	0.388	0.378	0.447
Distance to College			
Out-of-Province College	0.987	0.990	0.974
Distance below 150 km	0.000684	0.000214	0.00342
Distance between 150 and 800 km	0.300	0.248	0.599
Distance between 800 and 1200 km	0.261	0.260	0.268
Distance greater than 1200 km	0.431	0.484	0.122
Characteristics of College Cities			
GDP per Capita	66,316	64,101	79,210
	(38,016)	(36,888)	(41,747)
Population	879.1	879.5	876.6
	(630.7)	(668.2)	(337.1)
Public Green Land Area	57.15	55.24	68.26
	(52.17)	(50.18)	(61.39)
Number of Hospital Beds for 10,000 Persons	55.29	54.45	60.23
	(16.06)	(16.20)	(14.27)
# of observations	21,945	18,728	3,217

Note: Standard deviation in the parentheses. Dummy variables only report the mean.

Table 2.2 Difference in Difference Results: Impacts of the HSR on Scores

Dependent Variable	(1) Score	(2) Avg. Score	(3) High. Score
Panel A: Impacts on Station Opening			
Station	2.086***	1.086	0.521
	(0.718)	(0.642)	(0.462)
Observations	22,251	22,251	22,251
R-squared	0.894	0.930	0.893
Fixed Effects	university, exam package by home province by year		
Panel B: Impacts on Route Connection			
Route	0.811	0.904	0.701
	(0.843)	(0.664)	(0.713)
# of Obs.	21,945	21,945	21,945
R-squared	0.792	0.794	0.774
Fixed Effects	college-city by student-home-province, exam package by home province by year		

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. All regressions include a constant, college enrollment, and characteristics in the college city -- GDP per capita, population, public green land area, the number of the hospital beds. Standard errors are clustered at the college province level for the station effects and at the province pair level for the route equation.

Table 2.3 Heterogeneous Effects by Distance between College City and Student Home Province

Dependent Variable: Cutoff Scores	(1) Station	(2) Route
HSR	-8.240***	1.113
	(2.081)	(8.051)
HSR *Dist.0_150	15.43***	2.497
	(1.337)	(8.295)
HSR *Dist.150_800	11.09***	0.439
	(1.943)	(8.057)
HSR *Dist.800_1200	9.349***	-2.182
	(1.731)	(8.136)
HSR *Dist.1200_	10.90***	-0.320
	(2.444)	(8.176)
Observations	22,251	21,945
R-squared	0.894	0.792
Fixed Effects	university, exam package by home province by year	college-city by student-home-province, exam package by home province by year

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. The omitted group is in-province colleges. All regressions include a constant, college enrollment, and characteristics in the college city -- GDP per capita, population, public green land area, the number of the hospital beds. Standard errors are clustered at the college province level for the station effects and at the province pair level for the route equation.

Table 2.4 Heterogeneous Effects on In-Province Colleges vs. Out-of-Province Colleges

Dependent Variable: Cutoff Scores	(1) All Students	(2) Students from Mega Prov	(3) Students from Other Prov	(4) Students from Smart Prov	(5) Students from Other Prov
Panel A: Impacts on Station Opening					
Station	-7.780***	-14.00***	-7.280***	-11.13***	-4.254
	(2.194)	(4.102)	(2.293)	(2.289)	(2.742)
Station * Out Univ	10.17***	11.31***	9.997***	12.45***	6.943**
	(2.112)	(3.796)	(2.167)	(2.241)	(2.738)
Observations	22,251	1,570	20,681	5,758	16,493
R-squared	0.894	0.940	0.886	0.932	0.872
Fixed Effects	university, exam package by home province by year				
Panel B: Impacts on Route Connection					
Route	-2.358		-2.268	-3.744***	-2.188
	(3.946)		(3.924)	(0.753)	(4.034)
Route * Out Univ	3.275	-1.246	3.425	4.096**	3.481
	(4.015)	(2.180)	(4.004)	(1.802)	(4.090)
Observations	21,945	1,319	20,626	5,379	16,566
R-squared	0.792	0.889	0.780	0.884	0.727
Fixed Effects	college-city by student-home-province, exam package by home province by year				

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. The omitted group is in-province colleges. All regressions include a constant, college enrollment, and characteristics in the college city -- GDP per capita, population, public green land area, the number of the hospital beds. Standard errors are clustered at the college province level for the station effects and at the province pair level for the route equation. Megacities include Beijing, Shanghai, Guangzhou, Shenzhen. Mega-provinces are Beijing, Shanghai, and Guangdong Province. Smart cities include Beijing, Shanghai, Guangzhou, Nanjing, Wuhan, Tianjin, Chengdu, Changsha, Xi’an. Smart provinces are Beijing, Shanghai, Guangdong Province, Jiangsu Province, Hubei Province, Tianjin, Sichuan Province, Hunan Province, Shaanxi Province.

Table 2.5 Heterogeneous Effects on Out-of-Province Colleges in Megacities vs. Secondary Cities

Dependent Variable: Cutoff Scores	(1) All Students	(2) Students from Mega Prov.	(3) Students from Other Prov.	(4) Students from Smart prov.	(5) Students from Other Prov.
Panel A: Impacts on Station Opening					
Station	1.212	-3.108**	1.483*	0.280	1.521*
	(0.778)	(1.423)	(0.821)	(0.789)	(0.847)
Station * Mega Univ	3.838***	-0.653	4.035***	4.512***	3.569***
	(0.711)	(2.575)	(0.696)	(1.140)	(0.715)
Observations	21,440	1,410	20,030	5,352	16,088
R-squared	0.893	0.943	0.886	0.933	0.872
Fixed Effects	university, exam package by home province by year				
Panel B: Impacts on Route Connection					
Route	0.184	-1.473	0.396	-0.852	0.972
	(0.786)	(3.269)	(0.775)	(1.369)	(0.861)
Route * Mega Univ	1.440	0.582	1.455	2.871	0.376
	(1.700)	(4.248)	(1.805)	(3.159)	(1.525)
Observations	21,670	1,315	20,355	5,343	16,327
R-squared	0.790	0.890	0.777	0.881	0.726
Fixed Effects	college-city by student-home-province, exam package by home province by year				

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. The omitted group is colleges in the secondary cities. All regressions include a constant, college enrollment, and characteristics in the college city -- GDP per capita, population, public green land area, the number of the hospital beds. Standard errors are clustered at the college province level for the station effects and at the province pair level for the route equation. Megacities include Beijing, Shanghai, Guangzhou, Shenzhen. Mega-provinces are Beijing, Shanghai, and Guangdong Province. Smart cities include Beijing, Shanghai, Guangzhou, Nanjing, Wuhan, Tianjin, Chengdu, Changsha, Xi’an. Smart provinces are Beijing, Shanghai, Guangdong Province, Jiangsu Province, Hubei Province, Tianjin, Sichuan Province, Hunan Province, Shaanxi Province.

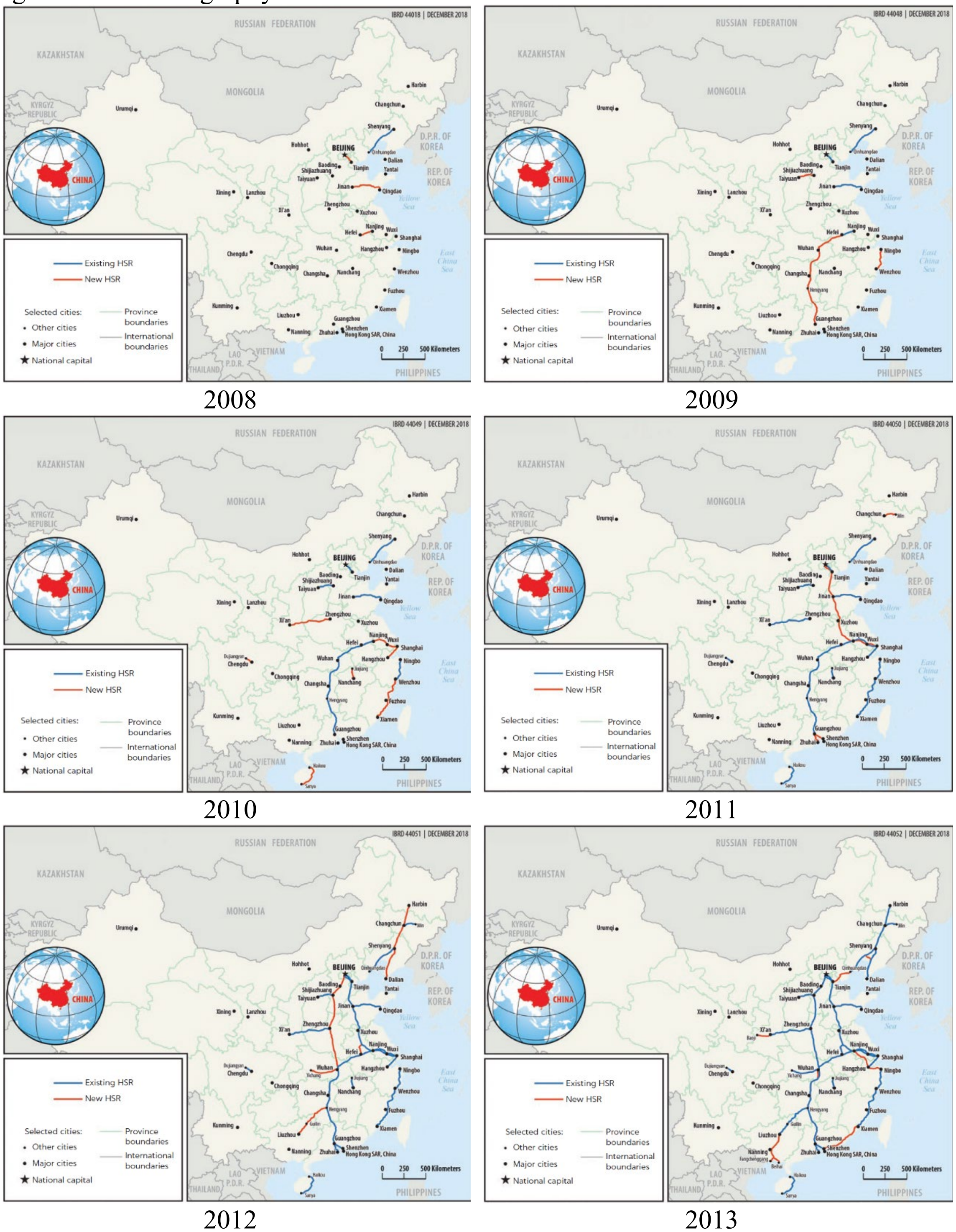
Table 2.6 Heterogeneous Effects on Out-of-Province Colleges by Elite Projects of Higher Education

Dependent Variable: Cutoff Scores	(1) All Students	(2) Students from Mega Prov.	(3) Other Prov.	(4) Students from Smart Prov	(5) Other Prov.
Panel A: Impacts on Station Opening					
Station	-2.365**	-7.796***	-2.055**	-2.416*	-2.257**
	(0.858)	(1.613)	(0.868)	(1.208)	(0.844)
Station *211Univ.	4.686***	4.248**	4.777***	3.245**	5.059***
	(0.996)	(1.521)	(1.005)	(1.420)	(0.907)
Station *985Univ.	10.83***	6.959***	11.23***	8.494***	11.53***
	(1.017)	(1.886)	(1.101)	(1.270)	(1.141)
Observations	21,440	1,410	20,030	5,352	16,088
R-squared	0.895	0.943	0.888	0.934	0.874
Fixed Effects	university, exam package by home province by year				
Panel B: Impacts on Route Connection					
Route	-3.743**	-8.686***	-3.459**	-7.945**	-2.308
	(1.552)	(2.592)	(1.592)	(3.101)	(1.465)
Route *211Univ.	4.285*	-0.432	5.242**	6.794*	4.893*
	(2.294)	(4.673)	(2.258)	(3.688)	(2.727)
Route *985Univ.	9.379***	18.05***	8.653**	18.47***	4.620
	(3.494)	(3.921)	(3.811)	(5.776)	(4.060)
Observations	21,670	1,315	20,355	5,343	16,327
R-squared	0.859	0.911	0.852	0.916	0.820
Fixed Effects	college-city by student-home-province, exam package by home province by year				

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. The omitted group is colleges that are not in the elite projects. All regressions include a constant, college enrollment, and characteristics in the college city -- GDP per capita, population, public green land area, the number of the hospital beds. Standard errors are clustered at the college province level for the station effects and at the province pair level for the route equation. Megacities include Beijing, Shanghai, Guangzhou, Shenzhen. Mega-provinces are Beijing, Shanghai, and Guangdong Province. Smart cities include Beijing, Shanghai, Guangzhou, Nanjing, Wuhan, Tianjin, Chengdu, Changsha, Xi’an. Smart provinces are Beijing, Shanghai, Guangdong Province, Jiangsu Province, Hubei Province, Tianjin, Sichuan Province, Hunan Province, Shaanxi Province.

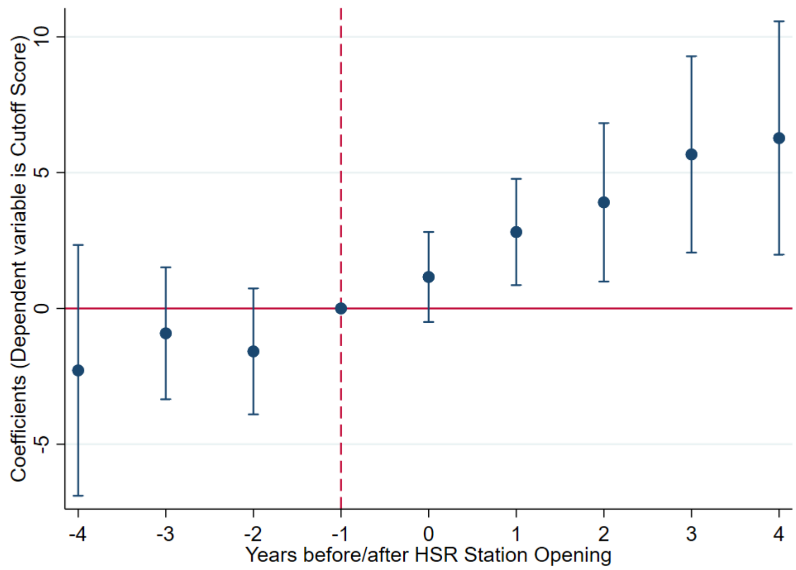
Figures

Figure 2.1 The Geography of HSR

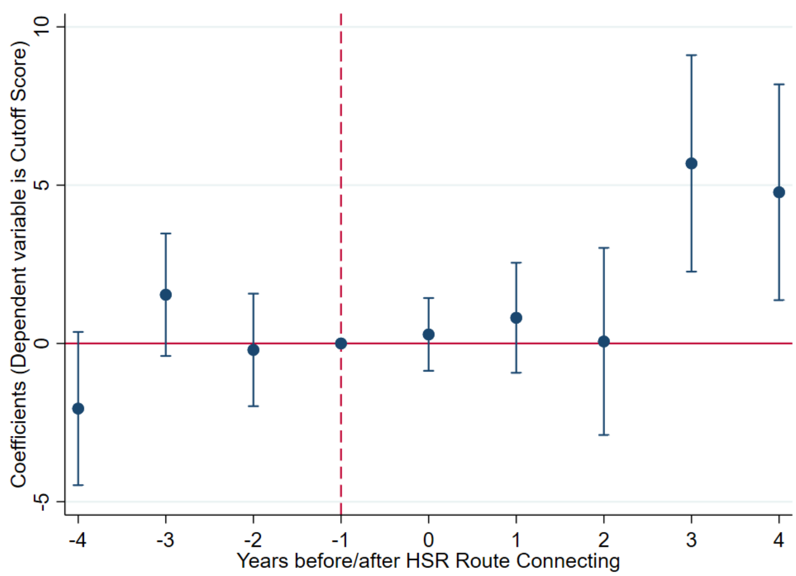


Source: Lawrence, M., Bullock, R. and Liu, Z., 2019. China's High-Speed Rail Development. International Development in Focus. Washington, DC: World Bank.

Figure 2.2 Event Study: The Dynamic Effect of HSR on College Admission Cutoff Scores

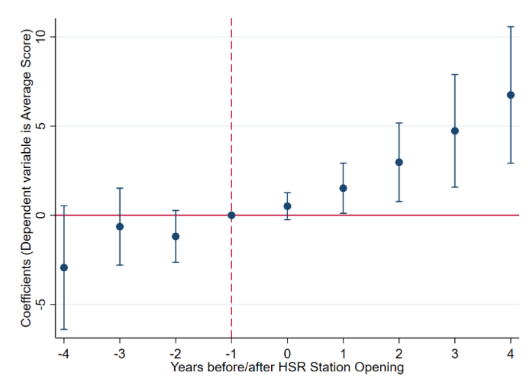


(A) Impact of HSR Station Opening

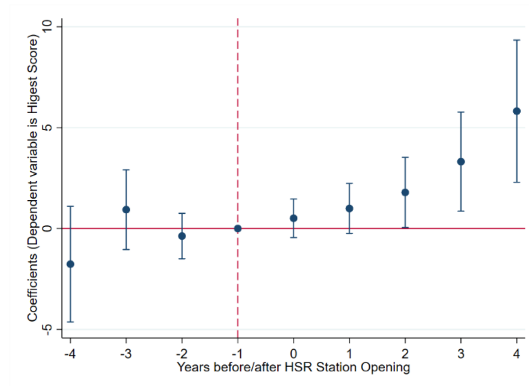


(B) Impact of HSR Route Connection

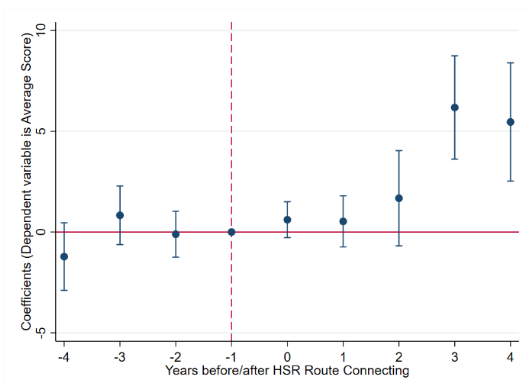
Figure 2.3 Event Study: The Dynamic Effect of HSR on Average Scores and Highest Scores



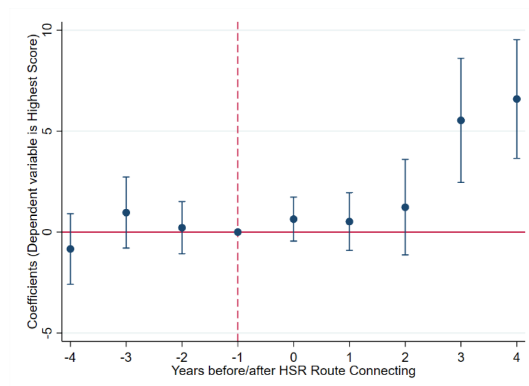
(A) Impact of HSR Station Opening On Average Scores



(B) Impact of HSR Station Opening On Highest Scores



(C) Impact of HSR Route Connection On Average Scores



(D) Impact of HSR Route Connection On Highest Scores

CHAPTER 3. HIGH-SPEED RAIL NETWORK AND HOUSING PRICES: EVIDENCE FROM JIANGSU PROVINCE

3.1 Introduction

Is better transport access always associated with increases in residential real estate prices? It may be true for intracity transportation infrastructure improvements, such as subways (Gibbons and Machin, 2005) and light rail lines (Billings, 2011), as people value the reduction in commuting time. However, intercity transportation improvements that have little impact on daily commuting costs may play a very different role in the housing market through geographical reorganization of economic activities.

In the past two decades, the Chinese government spends hundreds of billions of dollars to build the HSR network that mainly serves intercity passengers. As a result, HSR expansion in China is an ideal context to study the relationship between intercity transport access and housing prices. By the end of 2020, the total length of the HSR reached 38,000 km (about 24,000 miles)[13], the world's largest intercity transportation system.

Local governments in China also compete for HSR stations hoping they can boost their local economies and translate into higher future government revenues. While local governments can decide where to place new HSR stations within their cities, they cannot influence the site selection process in other cities. Therefore, using the reduction in distance to the closest HSR station caused by a new station opening outside of the city largely resolves the endogeneity concerns on the non-random placement of HSR stations.

[13] Data source: "Railway Statistics Bulletin 2020", National Railway Administration of the People's Republic of China, 19 April 2021, http://www.nra.gov.cn/xxgkml/xxgk/xxgkml/202104/t20210419_147769.shtml (in Chinese).

In this chapter, I compare the housing prices of properties close to the new HSR stations to those close to pre-existing HSR stations, before and after the new station openings. I combine unique data on transaction of first-time sold new houses that is aggregated at complex-level in Jiangsu Province with HSR station information to study the impacts of better HSR accessibility, defined as a reduction in station distance, on housing prices. In a difference-in-difference (DD) specification, I document that housing prices decrease by 20 percent in areas where the station distance is affected by the station opening outside the city. A one-kilometer decrease in station distance is associated with a 0.5 percent decrease in housing prices. The impacts are negative and stable when I compare properties in different distance radius to the closest HSR station. The effect of station opening within a city on housing prices is statistically insignificant.

This chapter contributes to the literature that estimates the impacts of transport access on residual real estate prices. Most researchers focus on the effects of intracity infrastructure improvements and using a relatively small radius to define the treatment and comparison groups. For example, Gibbons and Machin (2005) find that for properties within 2 km of a new subway station, a 1km reduction in station distance causes a 2 percent increase in housing prices. The comparison group is properties that are not affected by the new stations and within the 30km radium to the closest old station. Billings (2011) finds that housing prices within 1km of light rail transit stations increase by 4 percent for single-family properties and by 11 percent for condominiums following its announcement. The comparison group is properties within 1km of the proposed stations on the unselected lines. Relative to the literature, in this chapter, I study the variation in distance to the HSR stations. As the HSR routes mainly facilitate intercity traveling, it should affect properties at a longer distance. I do not restrict the station distance radius in the baseline estimation, and the subsequent analysis for properties within differing station distance radiuses confirms that the baseline estimation is robust.

This study also contributes to the literature evaluating the economic impacts of the HSR network in China. Recent papers document that the HSR network promotes economic growth in the connected cities (Ke et al., 2017), increases employment and facilitates urban industry specialization (Lin, 2017), stimulates knowledge spillovers (Dong et al., 2020) and innovation (Gao and Zheng, 2020). Zheng and Kahn (2013) find an increase in real estate prices in secondary cities after being connected to the megacity through a HSR line using city-specific real estate prices. This chapter uses housing data of a smaller unit to study the impacts of HSR stations on the property prices in the surrounding areas.

Li et al. (2020) use government land sale data to compare land within 3km of a HSR station and 3km of a conventional rail station. They find that residential land prices increase by 278 percent between 2008 and 2016 after the official announcement of station locations of the Beijing-Shanghai HSR line. The different signs of estimated coefficients between Li et al. (2020) and what I find here may be explained by the following. First, local governments may choose the area with the greatest development potential to place the HSR station. 3 km radium makes the treatment area most likely in the same city of the station. Thus, the effects of HSR stations may be overestimated. Second, land as an input of houses may have a quicker response to the HSR stations. As a result, they have enough land parcels sold within 3km radium even before the HSR stations are brought to use, while I do not have many new houses sold within the 15km radium after a new station opening. Finally, my results and Li et al. (2020) may reflect a demand shift towards houses closer to the new HSR station and cross the city border.

The remainder of the chapter is organized as follows. Section 3.2 describes the identification strategy, and Section 3.3 describes the data. Section 3.4 reports the results, and Section 3.5 concludes.

3.2 Identification Strategy

3.2.1 The Definition of Treated and Comparison Units

The goal of this chapter is to study the impact of HSR network on housing prices. Specifically, the opening of a new HSR station will increase the HSR accessibility for some areas by decreasing their distance to the closest HSR station, while it has no effects on the HSR accessibility for other areas whose nearest station does not change. Therefore, I compare the housing prices in the affected areas to those in the unaffected areas, before and after the new HSR station opening.

The definition of treated and comparison units is explained in Figure 3.1. Suppose in a Hotelling's linear city. There are two HSR stations at 0 and 1 at the beginning. For complexes locate between 0 and 1/2, station 0 is their closest HSR station. For complexes locate between 1/2 and 1, station 1 is their closest HSR station. Suppose there opens a new station at d, where d is greater than 1/2 and smaller than 1. While the new station changes the closest station for many complexes, some complexes are unaffected. Those complexes located in the unaffected areas (marked in gray) are defined as the comparison group. The yellow area between d/2 and 1/2 represents the area that is originally closer to station 0 but becomes closer to station d following its opening. The green area between 1/2 and (1-d)/2 represents the area that is originally closer to station 1 but becomes closer to station d following its opening.

3.2.2 Difference-in-Difference (DD) Specification

The starting point of my analysis is a simple DD model relating property values to the increase of HSR accessibility due to the opening of a closer HSR station:

$$\ln P_{it} = \alpha + \beta HSR_{it} + \rho' X_{it} + \gamma_s + \theta_c + \delta_t + \varepsilon_{it} \quad (3.1)$$

where $\ln P_{it}$ is the natural log of housing price, HSR_{it} is the interaction term of treatment status, i.e., whether the complex i locates between d/2 and (1-d)/2 in Figure 3.1, and the indicator of "post" periods, i.e., whether the station d has opened in period t. I use HSR_{it} instead of the classic $Treat_i * Post_t$ because the opening time of the new stations varies across areas. The vector X_{it} are complex-specific controls, including the average size of properties sold in the complex i and period t, and the distance between the complex i to its closest station s. Equation (3.1) also controls for the closest station fixed effects γ_s, city fixed effects θ_c, and time fixed effects δ_t, where s is the closest station to complex i at the end of studying periods, c is the city that the complex i locates in, and t is measured by month.

To study how the reduction in station distance affect the housing prices, I then generalize the specification (3.1) by replacing the dummy indicating treatment status to the continuous changes in station distance:

$$\ln P_{it} = \alpha + \beta \Delta dist_{it} + \rho' X_{it} + \gamma_s + \theta_c + \delta_t + \varepsilon_{it} \quad (3.2)$$

where $\Delta dist_{it}$ is the interaction term of the reduction in distance, i.e., the station distance at t=0 minus the station distance at t=T[14], and the indicator of "post" periods. Hence, $\Delta dist_{it}$ is always equal to zero for the comparison group.

The quasi-experiment nature of the new HSR station opening employed in this chapter eliminates the concerns of non-random placement of the treatment group. First, I compare the housing prices of properties close to the new HSR stations to those close to pre-existing HSR stations, before and after the new station openings. Second, while local governments can decide where to place new HSR station within their cities, they cannot influence the site selection process in other cities. Therefore, the reduction in distance to the closest HSR station caused by a new station opening outside of the city should be exogenous to them.

The key assumption in DD estimation is that the housing prices in the comparison and treatment groups should have similar growth patterns before the new HSR station opening. To test for the common trend assumption, I use an event study model of this form:

$$\ln P_{it} = \alpha + \sum_{m=2}^{5} \beta_m HSR_{it+m} + \sum_{n=0}^{9} \beta_n HSR_{it-n} + \rho' X_{it} + \gamma_s + \theta_c + \delta_t + \varepsilon_{it} \quad (3.3)$$

where HSR_{it} indicates the quarter in which the station distance changes. It equals 1 only if the new HSR station opens in the quarter t and the complex i's HSR accessibility is affected by the newly opened HSR station. The variable HSR_{it+m} is its mth lead and HSR_{it-n} is its nth lag. The quarter right before the new station opening is omitted as the base period. All other variables are previously defined.

---

14 $t \in [0, T]$.

3.2.3 Heterogeneous Analysis

First, I divide complexes into 1km grids and add interaction terms of dummies indicating treatment status and dummies indicating distances into the regressions to estimate the heterogeneous effects of better HSR accessibility on housing prices across distance:

$$\ln P_{it} = \alpha + \beta_0 HSR_{it} + \sum_k \beta_k' HSR_{it} * R_{ik} + \rho' X_{it} + \gamma_s + \theta_c + \delta_t + \varepsilon_{it} \quad (3.4)$$

where $R_{ik} = I(k \leq dist_{s_i} < k + 1)$ represents each 1km distance range to its closest HSR station (k is positive integers). The dummy indicating properties within the 1km radius of the HSR station ($k = 0$) is omitted. That is, the effect on properties within the 1km radius of the HSR station is captured by β_0 and the effect of better HSR accessibility for houses located within the k+1th km ranges is $\beta_0 + \beta_k$.

Second, as the two original stations at 0 and 1 may be very far away from each other and the effect of HSR station may vary across station distance, I further restrict the sample by the distances to the closest station. In other words, I compare properties that locate in different distance radius (106km, 75km, 50km, 25km, 15km) to the closest HSR station.

3.3 Data and Study Area

3.3.1 Data Source

Housing data used in this chapter are collected from the China Real Estate Information Corporation (CRIC) data system (http://www.cricchina.com). Real estate enterprises and researchers widely use it to study the dynamics of the housing market in China. I obtain its transaction data for new houses that first-time sold in Jiangsu Province from 2015 to 2018. The transaction data is aggregated to the complex level for each month, including complex address, complex average housing prices (in CNY), complex average housing sizes (in square meters), and the month in which the transaction happens.

HSR information is from the Chinese High-Speed Rail and Airline Database (CRAD) of the Chinese Research Data Services (CNRDS) Platform, which includes station address and station opening date.

After gathering geographic information on HSR stations and housing complexes from Baidu Maps, I calculate the distances between complexes to the HSR stations using the *geodist* Stata command.

3.3.2 Jiangsu Province

I use Jiangsu Province as a case to estimate the effect of improving HSR accessibility on housing prices. Jiangsu Province is one of the most developed provinces in China, which locates in the east coast region and is adjacent to Shanghai City. Its GDP per capita in 2018 is 115,930 CNY[15], which is 75% higher than the national average. Due to the data availability, I exclude the city of Nanjing (the province capital), Lianyungang, Yangzhou, Suqian in the analysis of this chapter, which are marked by gray in Figure 3.2.

Jiangsu Province is an ideal context for the study of HSR accessibility. The HSR network already connects many cities in Jiangsu by the end of 2014 (marked in dark blue in Figure 3.2) as the rest all get connected by 2020. Even the pre-connected cities build new HSR stations between 2015 and 2018. This provides a great variation in the treatment status. Not only the areas in the cities that are not connected by 2014 experience a reduction

[15] This is about 17,565 USD, estimated using the average exchange rate in 2018, i.e., 1 USD = 6.6 CNY

in station distance but also some areas at the edge of the pre-connected cities are affected by the opening of new HSR stations. In addition, cities in Jiangsu Province are irregularly shaped like strips, which ensures a large body of comparison group that is not affected by the opening of new HSR stations.

3.3.3 Summary Statistics

Table 3.1 presents the summary statistics. The average property price is about 1 million CNY, and the average size is about 125.7 square meters. The reduction in station distance is caused by new station openings in May 2016 or in September 2016. In Table 3.1, I divide the comparison groups into "before" and "after" periods by the first treatment time. That is, January 2015 to April 2016 is defined as the pre-treatment period, and May 2016 to December 2018 is defined as the post-treatment period for the comparison group in this table. The housing prices of the comparison group are about 165,000 CNY higher than the treatment group in the pre-treatment period and experience a larger growth after the opening of new stations.

Figure 3.3 shows the changes in the distribution of station distance for the comparison group before and after the first treatment time as previously defined. The distribution does not change much during the study periods from 2015 to 2018, suggesting that housing supply across distances does not change in the short run. Therefore, changes in the distribution of station distance for the treatment group are caused by the shock of new stations opening rather than the change in housing supply.

Figure 3.4 shows the distribution of station distance for the treatment group. The left column represents the observations whose closest HSR station is inside the city, while the right column represents the observations whose closest HSR station is outside the city.

We can see very different patterns in the reduction in station distance caused by a new station opening inside the city versus outside the city. After the new station opens in the city, station distances for the affected areas are below 35km. The station opening outside the city also decreases the station distance for the affected area, but the station distances still vary between 15km to 155km.

3.4 Results

3.4.1 Baseline Results of Better HSR Accessibility

Figure 3.5 shows the results of the event study on housing prices. The new HSR station has zero effects before its opening, no matter the station is in the city (Figure 3.5 (B)) or out of the city (Figure 3.5 (C)). The housing prices in the affected areas decrease following the HSR station opening outside the city, and the gap in housing prices in the affected area and the unaffected area becomes larger over time.

Table 3.2 reports the baseline results estimated by the DD specification (3.1). Column (1) shows that better access to the HSR station, defined as the reduction in station distance due to a new station opening, is associated with a 16 percent decrease in the affected area following the new station opening. This effect is mainly driven by the new station opening outside of the city, as the effect of a new station opening inside the city is statistically insignificant (Column (2)). Column (3) analyzes the impact of a new station opening outside the city. As discussed in the previous sections, I assume local governments can decide the location of HSR stations inside their cities but have little influence on the location of HSR stations outside of their city. Therefore, the reduction in station distance caused by a station opening outside of the city is exogenous. It will decrease the housing prices of the affected areas by about 20 percent following the opening of station. As shown in Figure 3.5, the gap between affected area and unaffected area is widen overtime so the 20 percent decrease can be seen as a short-run effect.

3.4.2 Effects of Each Additional Kilometer Reduction in Station Distance

Table 3.3 reports the regression estimates of the specification (3.2). Column (1) shows that a 1km reduction in station distance is associated with a 0.453 percent decrease in housing prices of the affected areas. In the last two columns of Table 3.1, we can find that the average reduction in station distance for the treatment group is about 30 km (126.5 – 96.62 and 118.2 – 87.06). Thus, the results in Table 3.3 are in line with the DD estimation in Table 3.2, in which the average price decrease is 16.2 percent (0.453 * 30 = 13.59). Column (3) shows that housing prices decrease in areas affected by the new station opening outside of the city by 0.5 percent for each kilometer reduction in station distance. The effects of a new in-city station (Column (2)) are smaller comparing to the effects of a new out-of-city station.

As the location of HSR stations outside of the city is exogenous, further analysis will focus on the effects of new HSR stations outside of the city. Table 3.4 shows the regression results controlling for additional time-varying city-level characteristics (Column (2)) and city-by-year fixed effects (Column (3)). The magnitudes decrease but are still negative and statistically significant. Column (4) uses unaffected areas whose closest HSR station is outside the city as the comparison group, and the results are very close to the baseline results in Column (1).

3.4.3 Heterogeneous Effects across Distance to the Closest Stations

Figure 3.6 plots the effects of better HSR accessibility for houses regarding the distance to the closest station at the end of the studying periods estimated by specification (3.4). The y-axis represents the estimated $\beta_0 + \beta_k$. The red line is obtained by local cubic smoothing, and the gray area shows its 95% confidence interval. Figure 3.6 (B) shows that the effect of in-city HSR station opening is very close to zero, while the impact of the out-of-city opening is significantly negative. The area within 75km of the new HSR station in the nearby cities experiences the largest decrease in housing prices following its opening.

3.4.4 Robustness Check Using Alternative Radius to the Closest Station

As shown in Figure 3.3 and Figure 3.4, the station distance for the comparison groups is concentrated below 50 km, but it is over 50 km for the treatment group that is closer to an out-of-city station. To ensure that the baseline result is not driven by the different trends in housing prices for two groups that has very different station distances, I further compare properties that locate in different distance radius (106km, 75km, 50km, 25km, 15km) to the closest HSR station.

The results are reported in Table 3.5. The impacts are negative and stable when using different distance radius. Due to the lack of new complexes within 15km of the new station, I cannot exclude the possibility that the HSR station can have positive effects on housing prices within a closer distance. As suggested in the literature (Gibbons and Machin, 2005), the negative impact might be explained by a demand shift towards houses closer to the new HSR station.

3.4.5 Pre-Existing HSR Station Can Not Offset the Negative Impacts

Table 3.6 presents how alternative transportation modes in the city offset the negative effects of the new HSR station opening outside the city. The negative impact is smaller for the affected areas in a city with a passenger-dedicated waterway (columns (1) and (2)) or an airport (columns (3) and (4)). However, having a pre-existing HSR station inside the city in 2014 does not offset the negative effects (columns (5) and (6)), which suggests that the impacts of HSR stations are localized. The distance to the HSR station matters. The HSR connection may reorganize the economic activities across cities as well as localities within a city.

3.5 Conclusion

I analyze the effect of the HSR network on housing prices using Jiangsu Province as a case study. While a local government can decide where to place the new HSR station within its locality, it cannot influence the site selection process in other cities. Therefore, the reduction in distance to the closest HSR station caused by a new station opening outside the city is exogeneous. Using transaction data of new houses aggregated to the complex level, I compare the housing prices of properties close to the new HSR stations to those close to pre-existing HSR stations, before and after the new station openings. In a DD specification, I document that housing prices decrease by twenty percent in the areas where the station distance is affected by the station opening in nearby cities. The effect of station opening inside the city is statistically insignificant. A one-kilometer decrease in station distance is associated with a 0.5 percent decrease in housing prices. The impacts are negative and stable when I compare properties in different distance radius (106km, 75km, 50km, 25km, 15km) to the closest HSR station. Due to the lack of new complexes within 15km of the new station, I cannot exclude the possibility that the HSR station can have positive effects on housing prices within a closer distance. As suggested in the literature (Gibbons and Machin, 2005), the negative effect might be explained by a demand shift towards houses closer to the new HSR station.

Tables

Table 3.1 Summary Statistics

VARIABLES	Total	COMPARISON		TREAT	
		Before	After	Before	After
Housing prices	1.007e+06	886,598	1.208e+06	721,271	795,990
	(734,579)	(546,889)	(882,130)	(403,159)	(488,506)
Housing sizes	125.7	125.6	130.2	117.9	117.6
	(42.82)	(39.57)	(50.09)	(28.94)	(28.07)
Better HSR accessibility	0.177	0	0	0	1
	(0.382)	(0)	(0)	(0)	(0)
Closest HSR is out of	0.908	0.950	0.954	0.711	0.823
city	(0.289)	(0.218)	(0.210)	(0.453)	(0.382)
Station distance by 2014	42.34	10.25	14.57	118.2	126.5
	(56.59)	(9.673)	(16.61)	(52.87)	(47.90)
Station distance by 2018	34.31	10.25	14.57	87.06	96.62
	(45.62)	(9.673)	(16.61)	(52.39)	(48.58)
# of observations	59,744	15,833	28,093	5,228	10,590

Note: Standard deviation in the parentheses.

Table 3.2 Better HSR Accessibility and Housing Prices

	(1) all/all	(2) in/all	(3) out/all
HSR	-0.162**	-0.0139	-0.210***
	(0.0558)	(0.0764)	(0.0515)
Observations	59,744	47,313	56,357
R-squared	0.643	0.655	0.642
Closest Station FE	YES	YES	YES
City FE	YES	YES	YES
Year by Month FE	YES	YES	YES

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. All regressions include a constant, average housing sizes in squared meters, and distance to the closest HSR station by the end of 2018. Standard errors are clustered at the city level.

Table 3.3 Reduction in Station Distance and Housing Prices

	(1) all/all	(2) in/all	(3) out/all
Reduction in Distance	-0.00453***	-0.00180*	-0.00523***
	(0.00105)	(0.000924)	(0.00111)
Observations	59,744	47,313	56,357
R-squared	0.644	0.656	0.642
Closest Station FE	YES	YES	YES
City FE	YES	YES	YES
Year by Month FE	YES	YES	YES

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. All regressions include a constant, average housing sizes in squared meters, and distance to the closest HSR station by the end of 2018. Standard errors are clustered at the city level.

Table 3.4 Robustness Check

VARIABLES	(1) out/all	(2) out/all	(3) out/all	(4) out/out
Panel A: Impacts of Better HSR Accessibility				
HSR	-0.210***	-0.157**	-0.0578*	-0.200***
	(0.0515)	(0.0572)	(0.0288)	(0.0506)
Observations	56,357	56,357	56,357	54,258
R-squared	0.642	0.647	0.652	0.646
Panel B: Impacts of 1km Reduction in Distance to the Closest HSR Station				
Reduction in Distance	-0.00523***	-0.00329***	-0.00147*	-0.00502***
	(0.00111)	(0.000802)	(0.000740)	(0.00106)
Observations	56,357	56,357	56,357	54,258
R-squared	0.642	0.647	0.652	0.645
Closest Station FE	YES	YES	YES	YES
City FE	YES	YES		YES
Year by Month FE	YES	YES	YES	YES
City Controls		YES		
City by Year FE			YES	

Table 3.5 Robustness Check Using Different Distance Radius

	(1) out	(2) < 106 km	(3) < 75 km	(4) < 50 km	(5) < 25 km	(6) <15 km
Panel A: comparing areas close to a new station outside of the city to all unaffected areas						
HSR	-0.210***	-0.292***	-0.200**	-0.206**	-0.151*	-0.293***
	(0.0515)	(0.0521)	(0.0661)	(0.0633)	(0.0675)	(0.0288)
Observations	56,357	48,196	44,521	43,295	38,466	30,427
R-squared	0.642	0.644	0.657	0.659	0.670	0.711
Panel B: comparing areas close to a new station outside of the city to the unaffected areas close to a pre-existing station outside of the city						
HSR	-0.200***	-0.280***	-0.182**	-0.189**	-0.139*	-0.285***
	(0.0506)	(0.0518)	(0.0589)	(0.0554)	(0.0663)	(0.0331)
Observations	54,258	46,097	42,422	41,196	36,677	28,812
R-squared	0.646	0.648	0.662	0.666	0.673	0.715
Closest Station FE	YES	YES	YES	YES	YES	YES
City FE	YES	YES	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES	YES	YES

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. All regressions include a constant, average housing sizes in squared meters, and distance to the closest HSR station by the end of 2018. Standard errors are clustered at the city level.

Table 3.6 Heterogeneous Effects for Cities with Alternative Transportation Modes for Passengers

VARIABLES	(1) out/all	(2) out/out	(3) out/all	(4) out/out	(5) out/all	(6) out/out
HSR	-0.277***	-0.267***	-0.403***	-0.354***	-0.211***	-0.201***
	(0.0513)	(0.0508)	(0.0599)	(0.0549)	(0.0511)	(0.0503)
HSR * Waterway	0.100***	0.101***				
	(0.0209)	(0.0211)				
HSR * Airport			0.195**	0.155**		
			(0.0711)	(0.0574)		
HSR * HSR by 2014					0.0140	0.0232
					(0.0657)	(0.0590)
Constant	7.387***	7.361***	7.387***	7.361***	8.124***	7.951***
	(0.194)	(0.215)	(0.194)	(0.215)	(0.226)	(0.237)
Observations	56,357	54,258	56,357	54,258	56,357	54,258
R-squared	0.643	0.646	0.642	0.646	0.642	0.646
Closest Station FE	YES	YES	YES	YES	YES	YES
City FE	YES	YES	YES	YES	YES	YES
Year by Month FE	YES	YES	YES	YES	YES	YES

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. The omitted group is cities that do not have the specific transport mode. All regressions include a constant, average housing sizes in squared meters, and distance to the closest HSR station by the end of 2018. Standard errors are clustered at the city level.

Figures

Figure 3.1 Definition of Treated Units and Comparison Units

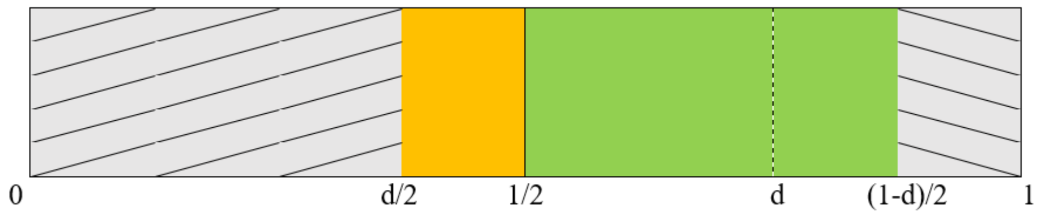


Figure 3.2 Jiangsu Province

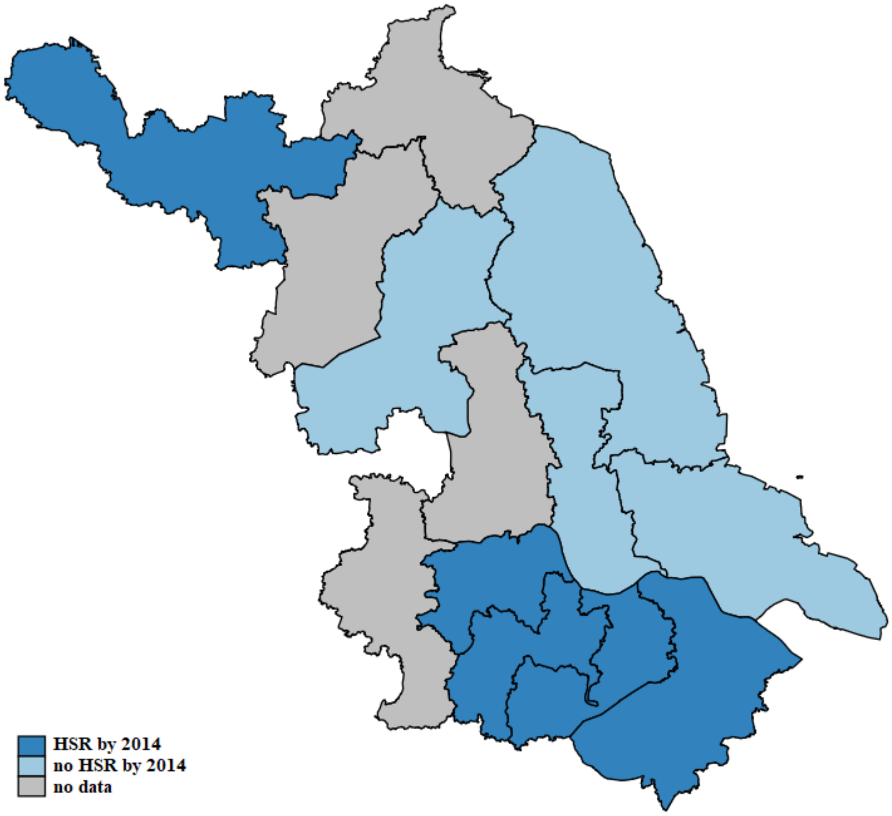


Figure 3.3 Distance to the Closest HSR Station for the Comparison Group

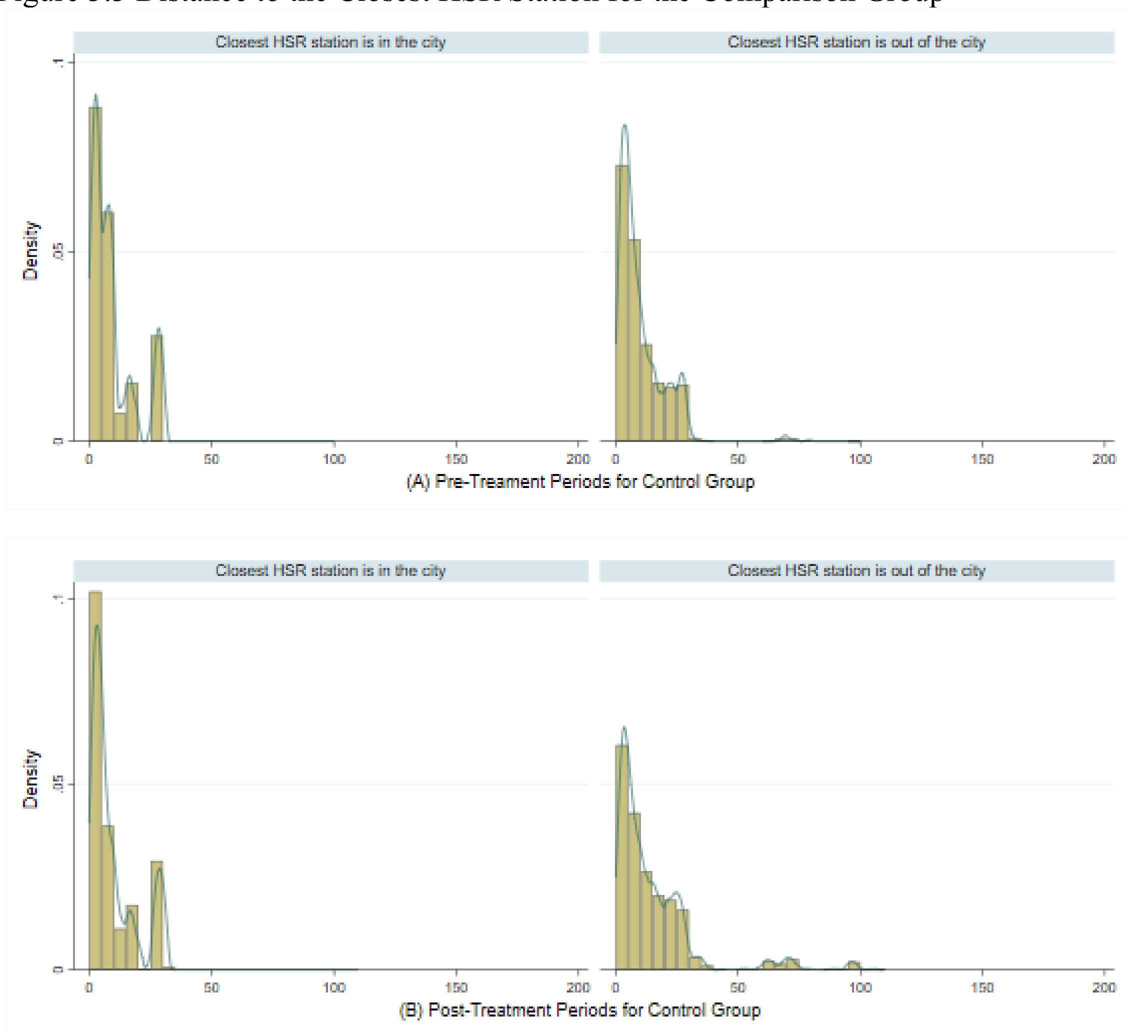


Figure 3.4 Distance to the Closest HSR Station for the Treatment Group

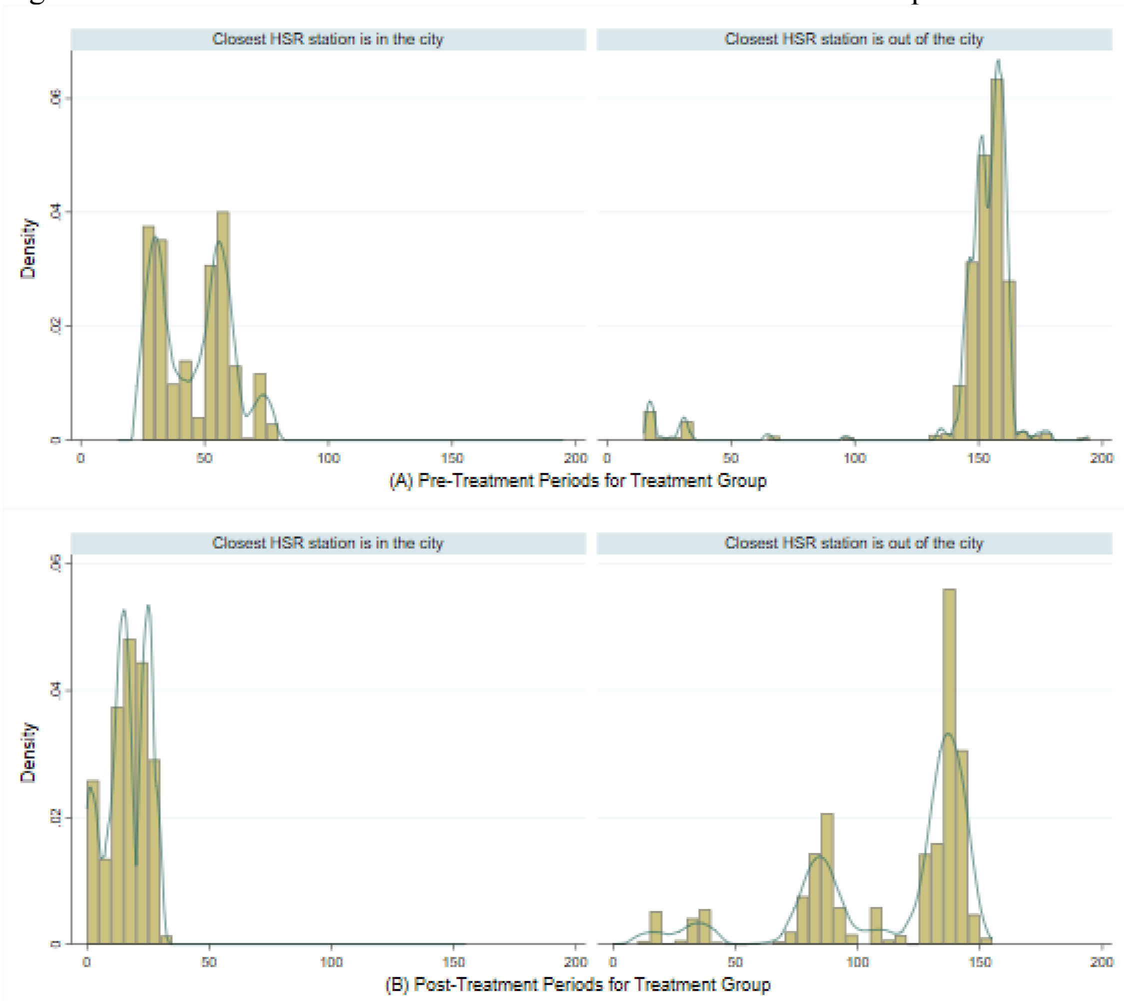
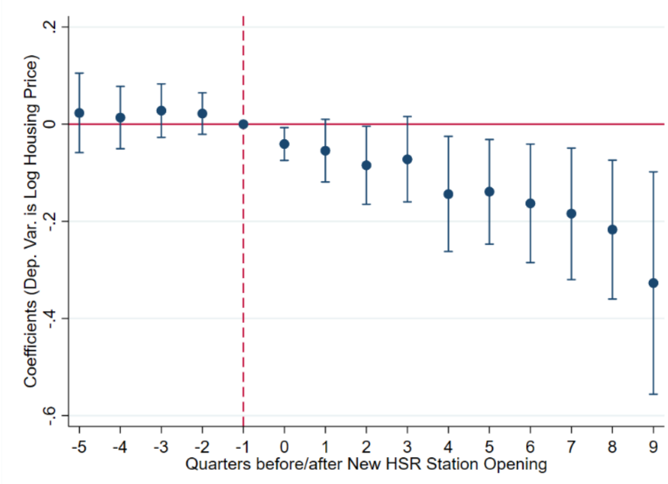
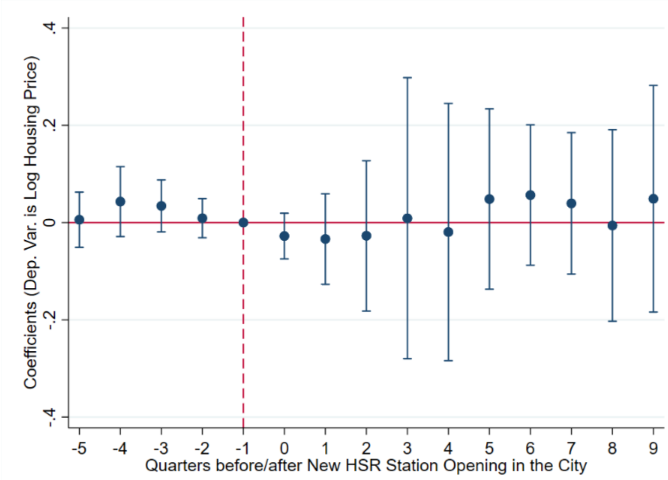


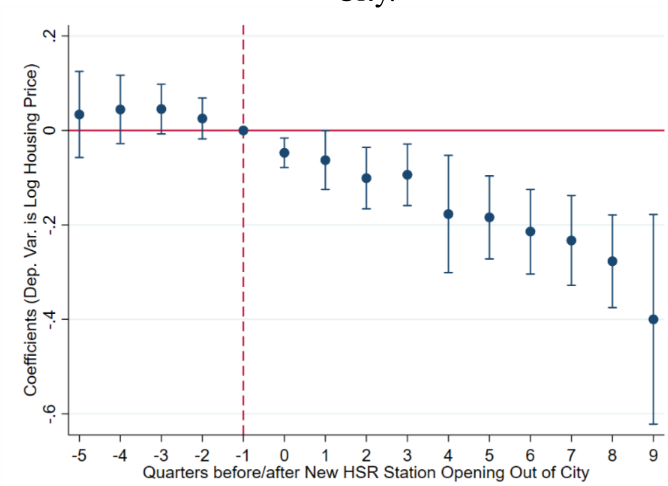
Figure 3.5 Event Study: The Dynamic Effect on Housing Prices



(A) Impact of Reduction in Station Distance.



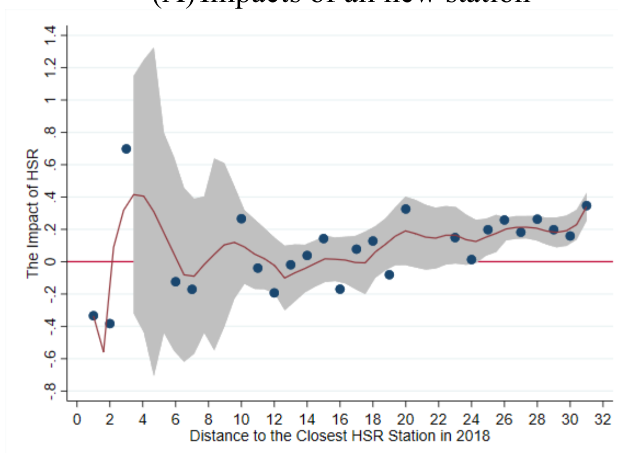
(B) Impact of Reduction in Station Distance Caused by a New HSR Station within the City.



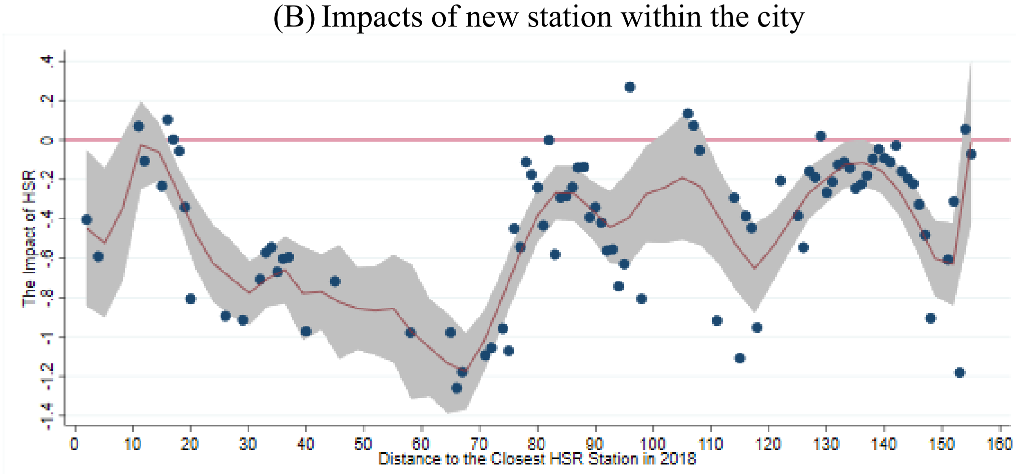
(C) Impact of Reduction in Station Distance Caused by a New HSR Station Outside of the City.

Figure 3.6 Housing Price Gradient, with Distance to the Closest Station in 2018

(A) Impacts of all new station



(B) Impacts of new station within the city



(C) Impacts of new station outside of the city

CHAPTER 4. HIGH-SPEED RAIL NETWORK AND URBAN HOUSEHOLD INCOME IN CHINA

4.1 Introduction

As a natural consequence of regional economic development and the spatial redistribution of economic activities, transportation infrastructure improvements may result in uneven effects on earnings across industries, skill levels, and demographic groups. Yet, while abundant literature emphasizes the relationship between transportation costs and population/employment (Baum-Snow, 2007; Duranton and Turner, 2012; Sheard, 2014; Baum-Snow and Turner, 2017), land/real estate prices (Gibbons and Machin, 2005; Donaldson and Hornbeck, 2016; Bogart, 2009), trade (Duranton et al., 2014; Michael, 2008), industry composition (Chandra and Thompson, 2000), and economic growth (Banerjee et al., 2012; Farber, 2014; Storeygard, 2016), little attention has been paid to the effects of transportation infrastructure on income, especially at the household level.

Here I use China's recent construction of its high-speed rail (HSR) network as a natural experiment to investigate the effects of the large-scale transportation project on household income and the mechanisms determining it. China currently has the world's largest HSR network, which began to connect its major cities in 2008. As of 2020, the total in-operation length reached 38,000 km[16]. It provides fast, comfortable, and affordable transport services to inter-city travelers. The ambitious rail plan is to connect about 230 Chinese cities in all provinces by the end of 2030. Based on the Chinese experience, I

[16] Data source: "Railway Statistics Bulletin 2020", National Railway Administration of the People's Republic of China, 19 April 2021, http://www.nra.gov.cn/xxgkml/xxgk/xxgkml/202104/t20210419_147769.shtml (in Chinese). The number for 2013 is 11,000 km.

study whether being connected to the HSR network affects the total income of urban households and income from different sources.

To evaluate the effects of HSR on income distribution in China, I employ household data from China Household Income Project (CHIP) combined with city HSR connection information. Using difference-in-difference (DD) approach, I document that urban households experience a significant increase in total household income following the opening of a HSR station in their city. While labor earnings (wage/salary income) increase, the probability of having business income decreases. The likelihood of having property income does not change, on average. However, those effects are heterogeneous across industries and age cohorts.

Numerous mechanisms lead to these results. First, the HSR network facilitates urban industry specialization (Lin, 2017). As the new transportation mode mainly serves the inter-city travelers and decreases the face-to-face interaction costs between cities, HSR stimulates specialization towards industries that requires high communication skills and low manual tasks. My results reveal that labor income of the households whose heads work in the manufacturing sector increases little, but for households whose heads work in the transport or communications sectors increases much more than other households.

The second mechanism I analyze is the changes in the characteristics of urban households that follow the HSR connection of a city. My results show that the share of local households[17] in the population decreases after the HSR station opens in the city, suggesting that the HSR network encourages more migrants to flow into the connected

[17] Local household is defined as a household whose head has a local "hukou", where "hukou" refers to the household registration system. Having local hukou is very important for the household to enjoy local social benefits and public services.

cities. After the connection, the average age decreases, and the average education level increases for the household heads living in the connected cities. As a consequence, income patterns differ across age and education groups. Younger households aged between 25 to 34 gain little in total income following the HSR connection. Although their labor income increases, the likelihood that they have property income decrease dramatically. In contrast, older households experience increases in both labor income and the likelihood of having property income. At the same time, the rise in labor income is much smaller for the households with high-skilled heads[18].

This study contributes to the literature on estimating the impact of transportation infrastructure projects on income. Chandra and Thompson (2000) find that counties in the U.S. experience growth in total earnings and earnings in manufacturing, retail trade, services, transportation, and public utility industries after the interstate highway directly passes through them. Meanwhile, the untreated adjacent counties experience increased manufacturing earnings and decreased total earnings, retail trade, and government earnings. Michaels (2008) finds that the interstate highway increases trucking income in rural counties they cross relative to other rural counties. He also finds that highways increase the relative wage of high-skilled workers to low-skilled workers in high-skill counties and reduce it in low-skill counties. Banerjee et al. (2012) aggregate rural household income to the county and year level and find that the distance to a "line"[19] does not affect the income

---

[18] High-skilled worker is defined as individuals who have a four-year college degree. The increase in labor income for households whose heads have a three-year college degree is similar to it for whose heads do not.

[19] The "line" refers to the nearest straight line connecting two historical cities in China. As they predict the placement of transportation infrastructure, their estimation can be considered as an intent-to-treat effect.

growth. Relative to the literature, this chapter uses household-level information, tests for a new transportation infrastructure project, and focus on the urban area.

My study builds on the literature evaluating the economic impacts of the HSR network in China. Recent papers document that the HSR network promotes economic growth in the connected cities (Ke et al., 2017), increases employment and facilitates urban industry specialization (Lin, 2017), boosts real estate prices (Zheng and Kahn, 2013), increases government revenue by raising land prices (Li et al., 2020), stimulates knowledge spillovers (Dong et al., 2020) and innovation (Gao and Zheng, 2020).

This chapter and Kong et al. (2021) both use data from CHIP to study the impacts of HSR connections on income. Kong et al. (2021) choose the rural-urban migrant sample and focus on the migrants' wage evolution. I, instead, use the urban household sample to study the urban household income dynamics. Kong and co-authors find a decrease in wages for low-skilled migrants in the connected cities and attribute it to the loss of market power because more low-skilled migrants move into the connected cities. I, in contrast, find increases in wage income for both low-skilled and high-skilled households. Moreover, subsequent analysis reveals that the residents in the connected cities become more educated on average after the connection. Therefore, we have no consensus regarding the mechanism through which that HSR connection affects labor income. As the urban household sample does not exclude rural-urban migrants, it should have better representativeness of all urban residents. Finally, the cities surveyed in CHIP vary across waves. To guarantee the accuracy of the DD estimate, I only use data on households in cities that appear in every wave of the survey. As Kong et al. (2021) use all cities regardless of their absence in some waves, their results possibly reflect changes in average income caused by the changing composition of cities.

The rest of the chapter is organized as follows. Section 4.2 describes the data. Section 4.3 presents empirical strategies and identification assumptions. Section 4.4 reports the baseline results, and Section 4.5 analyzes the heterogeneous effects and the mechanisms. Section 4.6 concludes.

4.2 Data

4.2.1 Data Sources

Household information is obtained from China Household Income Project (CHIP). Seven waves of repeating cross-sectional surveys have been conducted in 1988, 1995, 1999, 2002, 2007, 2008, and 2013, which are widely used by researchers interested in the dynamics of income distribution in China. To better serve my objectives in this research, I only use the last four waves and keep only cities that appear in every wave of CHIP 2002, 2007, 2008, and 2013. Additionally, several factors make it preferable to further exclude the 2008 CHIP wave from the analysis. First, at the household level, CHIP 2008 lacks information on detailed income sources. Second, both CHIP 2007 and 2008 belong to the larger Rural-Urban Migrants in China (RUMiC) survey project. They survey the same households in both years; thus, including CHIP 2008 could potentially introduce downward bias caused by smaller variation in income between 2007 and 2008. In addition, the HSR network expansion started in 2008. Therefore, including CHIP 2008 brings ambiguity to the definition of treatment groups and the interpretation of results.

The HSR station opening information is from the Chinese High-Speed Rail and Airline Database (CRAD) of the Chinese Research Data Services (CNRDS) Platform, which is consistent with the released route opening information on the National Railway Administration (NRA) official website[20]. Opening time in this chapter is defined as the date that the first HSR station opens in the city. The exact opening time for the cities in the sample is listed in Table B 1. Although Chongqing has stations that open by the end of 2013, it is in the comparison group for several reasons. First, literature on the economic impacts of HSR connection (Lin, 2017; Dong et al., 2020) finds that HSR connection does not have significant effects in the opening year, let alone connection time at the end of the year. Second, following the literature (Lin, 2017; Lawrence et al., 2019), I define HSR routes/lines as railway lines running at an average speed of 250 km/h or more (i.e., G-class passenger train service) and inter-city lines running at an average speed of 200 km/h or more (i.e., C-class service). The designed maximum speed for the only HSR route passing through Chongqing at that time is 200 km/h, and it connects to a route with the designed maximum speed of 160 km/h in Hubei Province. The speed of this line is below the threshold of HSR route definition. Thus, it is not considered as a HSR route in this chapter.

Figure 4.1 shows the development of HSR network between 2008 and 2013. The red lines show the new HSR routes that open in current year, and the blue lines show HSR routes that open in previous years. Nanjing and Hefei connect to each other in 2008. The route that connects Chengdu City is still isolated from the other part of the network by the

---

20 "High Speed Rail in China (by Opening time, 20151001)", National Railway Administration of the People's Republic of China, 22 October 2015, http://www.nra.gov.cn/ztzl/hyjc/gstl_/zggstL/gtxl/201602/t20160216_21088.html.

end of 2013. In the map of 2013, it is clear that Chongqing City is not connected by any HSR routes.

4.2.2 Summary Statistics

City-level socioeconomic variables are collected from China Statistical Yearbook for Regional Economy 2004, 2008, 2009, and 2014, including GDP, GDP growth rate, population, fixed investment, the number of hospital beds per 10,000 people, and public libraries.

Table 4.1 provides the summary statistics of the key variables. The outcome variables are households' total income, wage/salary income, the dummies for whether having business income and property income. At the household level, I control for family size and the characteristics of the household heads, including gender, age, education level, self-reported health status, and urban hukou status. Self-reported health is discrete between one to five, where one stands for very unhealthy and five stands for very healthy. The total number of households in the sample is 7070 and decreases to 5006 when excluding CHIP 2008.

4.3 Empirical Strategy

4.3.1 Difference-in-Difference Estimation

The primary goal of this chapter is to estimate the effect of HSR connection on household income for people living in the connected city. This section talks about the empirical strategies used in this chapter.

The baseline estimation uses the difference-in-difference (DD) approach as:

$$Y_{it} = \alpha_0 + \alpha_1 CONNECT_{it} + \beta_1{}' H_{it} + \beta_2{}' X_{ct} + \gamma_c + \delta_t + \varepsilon_{it} \quad (4.1)$$

The dependent variable, Y_{it}, represents the natural log of total income for household i in year t in the baseline estimation. The dummy variable $CONNECT_{it}$ indicates whether the city in which the household i resides has an in-operation HSR station by the end of last year[21]. The vector H_{it} includes control variables for household characteristics, including family size, the dummy for female head, head's age, years of education, self-reported health status, dummies for urban hukou and local hukou. The vector X_{ct} includes control variables for city characteristics, including log GDP, GDP growth rate, log population, log fixed investment, log hospital beds per 10,000 people, and the number of public libraries. All specifications include city fixed effects, γ_c, and year fixed effects, δ_t. The error term, ε_{it}, is clustered at the city level.

To test for the mechanisms through which HSR connection affects household income, I examine three different income sources: wage/salary income[22], whether the household has business income, and whether it has property income. Besides income, I also examine family size and other characteristics of the household heads.

4.3.2 Instrumental Variable Estimation

Another empirical concern is that the placement of HSR stations and routes is unlikely to be random due to the objectives of the central government. In this chapter, the ten cities that are all selected to being connected to the HSR network ultimately in the initial version of Mid-to-long Term Railway Development Plan made by the NRA in 2004.

---

[21] The value of $CONNECT_{it}$ equals zero for the year of the HSR station opening. For example, Nanjing and Hefei are connected by a HSR route in 2008 but $CONNECT_{it}$ equals zero for these two cities in 2008 and equals one in 2013. As a result, only year 2013 is defined as the "post" period in this chapter.

[22] In log form.

Thus, the endogeneity issue is mainly caused by the omitted variables which are both correlated with when a city is connected to the HSR network and the household income in that city.

To address these endogeneity concerns, I employ an instrument variable (IV) approach. Following the literature, for example, Baum-Snow et al. (2017), Baum-Snow and Turner (2017), and Dong et al. (2020), I use a historical railroad map in 1962[23] to construct the IVs, whether the city is on a historical rail line in 1962 and whether the city has a junction where a railroad line crosses another. They are strong predictors of road placement today but are built too long ago to affect the current economy. Equation (4.3) shows the first stage specification, and Equation (4) shows the second stage specification:

$$CONNECT_{it} = \theta_0 + \theta_1 Pass1962_{it} + \theta_2 Xing1962_{it} + \beta_1^{1\prime} H_{it} + \beta_2^{1\prime} X_{ct} + \gamma_c^1 + \delta_t^1 + \nu_{it} \quad (4.3)$$

$$Y_{it} = \alpha_0 + \alpha_1 \widehat{CONNECT}_{it} + \beta_1^{2\prime} H_{it} + \beta_2^{2\prime} X_{ct} + \gamma_c^2 + \delta_t^2 + \varepsilon_{it} \quad (4.4)$$

where $Pass1962_{it}$ equals to one if the city is on a historical rail line in 1962 and the year is 2013, and $Xing1962_{it}$ equals to one if the city has a rail junction in 1962 and the year is 2013.

4.3.3 Triple-Difference Estimation

I employ triple-difference (DDD) approach to study the heterogeneous effects of HSR connection on different groups and control for confounding trends that are potentially omitted in the baseline estimation. The specification is

---

23 Baum-Snow et al. (2017) and Baum-Snow and Turner (2017) use 1962 transport networks as instruments for modern transportation networks to study the effects of roads on regional economic growth. Dong et al. (2020) also use the railway connection status in 1962 as one of their IVs to predict the effects of HSR on research productivity.

$$Y_{it} = \alpha_0 + \alpha_1 CONNECT_{it} + \alpha_2 CONNECT_{it} * D_i + \beta_1{}' H_{it} + \beta_2{}' X_{ct}$$
$$+\gamma_c * D_i + \delta_t * D_i + \gamma_c + \delta_t + \varepsilon_{it} \quad (4.5)$$

where D_i is a vector of dummies indicating households characteristics, such as industry, age cohort, and skill level of the household head. All other variables are the same as previously defined. In addition to the city fixed effects, γ_c, and year fixed effects, δ_t, I control for the city by group fixed effects, $\gamma_c * D_i$, and the year by group fixed effects, $\delta_t * D_i$.

First, Lin (2017) points out that the HSR network facilitates industry specialization. As the HSR network mainly serves inter-city travelers, it decreases face-to-face communication costs and shifts the specialization pattern of connected cities towards skilled and communication-intensive sectors. She finds that the employment increases more for the industries with a higher reliance on non-routine cognitive skills, for example, retail and wholesale industry, finance industry. I let D_i indicate different industries to investigate the effects of HSR connection on income across industries. Due to the small number of observations in other industries, I can only compare the effects on households whose heads working in the following industries: (1) manufacturing, (2) construction, (3) retail, hotel and catering services, (4) transportation, warehousing and post, (5) government. The effect of HSR connection on all other industries is α_1 while the effects on these five specific industries are $\alpha_1 + \alpha_2$.

Second, the HSR connection may have different effects across age cohorts and skill levels. I then let D_i indicate different age cohorts and educational groups.

4.3.4 Placebo Test

To confirm that the HSR station does not affect household income before its existence and at the year of opening, I further conduct a placebo test by adding interaction terms of connection dummies and year dummies.

$$Y_{it} = \alpha_0 + \sum_m (\alpha_{1m} CONNECT1_i * year_m + \alpha_{2m} CONNECT2_i * year_m) + \beta_1' H_{it} + \beta_2' X_{ct} + \gamma_c + \delta_t + \varepsilon_{it} \quad (4.6)$$

I divide the treatment groups into cities connected by 2008 and cities connected after 2008. The dummy variable $CONNECT1_i$ indicates cities with stations opening in 2008, and $CONNECT2_i$ indicates cities with stations opening between 2009 and 2013; $year_m$ is the year dummy indicating the year 2002, 2008, and 2013, where $m \in \{2002, 2008, 2013\}$. The year 2007 is omitted as the base year.

4.4 Baseline Results for Total Income

4.4.1 DD Results

Figure 4.2 provides evidence on whether the parallel trend assumption for the DD approach is satisfied. Figure 4.2 (A) and (B) show the dynamic effects of the HSR connection using an event study model, which includes leads and lags of the initial connection dummy. However, the household data used in this chapter only have four years (2002, 2007, 2008, 2013) and ten cities. Therefore, the coefficient of each lead or lag reports the weighted average differences in income changes between cities connected in a specific year and the cities in the comparison group[24]. Moreover, since no city first connects to the HSR network in 2007 or 2012, the coefficient of the first lag is unidentified. As a result, the coefficients do not properly reflect the evolution of income patterns.

I then conduct a placebo test using specification (4.6), and plot the coefficients of the interaction terms, α_{1m} and α_{2m}, with respect to the survey year. Figure 4.2 (C) and (D) show that between 2002 and 2007, when there are no real HSR connections, the change of differences in income between the two treatment groups and the comparison group is not statistically significant. Furthermore, although two cities connect to the HSR network in 2008, the figures show no significant effect in the year when the HSR first connects the city. This supports my data choice of deleting CHIP 2008. The test for common trend assumption for data without CHIP 2008 is plotted in Figure B 1. Similarly, flat trends in income differences between groups before the HSR connection are found.

Table 4.2 presents the baseline estimates of the effect of HSR connection on overall income for urban households based on specification (4.1). The dependent variable is the natural log of total income at the household level. When excluding CHIP 2008 for all households (Column (3)), the estimated effect of HSR connection in the connected cities is a 29.9 percent increase in income relative to households in the unconnected cities, which is equivalent to a 4.5 percent annual increase between 2007 and 2013. The estimate is slightly larger than the one including CHIP 2008 in column (1). About 97 percent of the households in the sample have local hukou, suggesting they are not recent migrants. Columns (2) and (4) look at the impacts of HSR connection on local households only. The

---

[24] For example, the coefficient for the station opening year shows the difference in income changes between 2007 and 2008 for the cities where station opens in 2008 and other cities in the comparison group.

results show that households with local hukou, who are more likely to be natives or longer-term residents, experience a larger gain in total income.

4.4.2 IV Results

Although all cities in the sample are planned to have HSR connection by the central government in 2004, it is still possible that the regression model omits some variables that affect the HSR connection by 2013 and the household income. To address these concerns, I use historical railroad information to construct IVs in the 2SLS specifications (4.3) and (4.4). Table 4.3 presents the first stage results that regress HSR connection dummy on the IVs and controls. It shows that the IVs are strong predictors of the HSR connection status and pass various identification tests.

Table 4.4 reports the impacts of the HSR connection on household income using IV approach. These effects are larger than the DD estimation in Table 4.2, indicating that the endogeneity issue does exist. One possible explanation is that as the initial goal of the HSR network is to release the burden of the overloaded conventional rail network, more developed regions are connected first. The urgent demand for better transportation infrastructure increases the possibility of being connected to the HSR network by 2013 and is negatively related to income growth. This would lead to a downward bias in the OLS estimation.

4.5 Mechanisms and Heterogeneous Analysis

4.5.1 Income Sources

To analyze the mechanisms through which the HSR connection affects household income, I start by examining the effects on different income sources. Table 4.5 column (1) replicates the effects on total income in Table 4.2. Panel A of Table 4.5 shows the results for all households. For households with positive labor earnings, wage increases by 46.1 percent after the HSR connection in the connected cities, i.e., a 6.5 percent annual increase. However, the possibility of having business income decreases by 14.3 percent. This may suggest that wage income and business income are complementary for households in urban China. The effect of HSR connection on the likelihood of having property income is statistically insignificant from zero. The results for households with local hukou in Panel B show a similar pattern. Table B 2 reports the IV estimates for different income sources. Still, positive effects on wage income, negative effects on the possibility of nonzero income, and no effects on the possibility of nonzero property income are found.

4.5.2 Heterogeneous Effects of HSR Connection

4.5.2.1 Heterogeneous Effects by Industries

Table 4.6 reports the estimates of α_1 and α_2 in specification (4.5). Compared to households in other industries, households with heads in the manufacturing sector have less gain in wage income after the HSR connection. The growth in wage income for the households with heads in the transportation, warehousing, and post sector doubles relative to other industries. However, the possibility of having business income increases for households in the manufacturing industry and decreases for households in the transportation sector. It supports the argument that wage income and business income are somewhat complementary.

Households working in the transportation and government sectors both experience greater increases in total income but through very different sources. The additional increase in wage income is the driving force of total income growth for households in the transportation sector. In contrast, the households in the government sector are the only group that is more likely to have property income after the HSR connection.

4.5.2.2 Heterogeneous Effects by Age Cohort

Table 4.7 shows the heterogenous effects across age cohorts. The youngest generation aged between 25 to 34 has little gain in total income from the HSR connection. Wage income increases but the probability of having other income sources decreases. The older generations experience increases in total income and the possibility of having property income. Surprisingly, the oldest generation aged between 55 to 65 has the largest increase in wage income.

4.5.2.3 Heterogeneous Effects by College Degree

Table 4.8 reports the heterogenous effects across skill levels. Column (2) shows that the growth in wage income caused by HSR connection is similar for households whose heads have or do not have a three-year college degree (associate degree). However, when using different definitions of college degrees, we can find that the rise in wage income is much smaller for the households whose heads have a four-year college degree (bachelor's degree).

4.5.3 HSR Changes the Composition of Residents

One potential mechanism behind the heterogeneous impacts of the HSR connection on household income across age cohorts and educational groups is the changes in the composition of people living in the connected cities caused by the HSR connection. Table 4.9 reports the DD results that regress household characteristics on the HSR connection dummy and control for the full set of city characteristics, city fixed effects, and year fixed effects. Column (1) shows that the share of local households decreases after the HSR station opens in the city. This suggests that the HSR network encourages more people to move into connected cities. Moreover, after the connection, the average age decreases, and the average education level increases for the household heads living in the connected cities. These results suggest that the HSR network facilitates the sorting of people based on age and education.

4.6 Conclusion

In this chapter, I study the effect of the HSR network on household income in urban China using information on a primary household income survey between 2002 and 2013. I first examine how city-level HSR connection status affects total income. The baseline effect of HSR connection for all urban households in the connected cities is a 29.9 percent increase in income relative to households in the unconnected cities. This is equivalent to a 4.5 percent annual increase between 2007 and 2013. Wage income and business income are complementary to each other. While the HSR connection increases wage income, it decreases the probability of having business income. I then analyze the heterogeneous effects across industries, age cohorts, and education groups. Wage income increases little for households whose heads work in the manufacturing sector and increases much more for households in the transportation sector. Younger households in the connected cities are less likely to have property income after the connection, making the effects of HSR connection on their total income statistically insignificant from zero. In addition, the rise in wage income is smaller for the households whose heads have a four-year college degree. Finally, HSR connection increases the share of nonlocal households in the connected cities. The average age of household heads decreases, and education attainment increases because of the HSR connection. These results suggest that the HSR network facilitates the sorting of people based on age and education, which is a potential mechanism through which HSR affects network affects household income.

Tables

Table 4.1 Summary Statistics

VARIABLES	2002 Control (1)	2002 Treat (2)	2007 Control (3)	2007 Treat (4)	2008 Control (5)	2008 Treat (6)	2013 Control (7)	2013 Treat (8)	2002-13 Total (9)
Panel A: Household Income and Characteristics									
Total Income	21,820	27,973	45,109	54,715	41,657	52,478	73,152	105,862	52,666
	(12,988)	(18,434)	(40,337)	(46,698)	(34,289)	(51,875)	(39,970)	(65,682)	(49,153)
Wage/Salary Income	16,673	21,630	31,506	39,718	39,274	47,945	49,018	79,017	41,218
	(11,413)	(18,003)	(38,868)	(43,621)	(24,912)	(38,246)	(38,070)	(61,901)	(41,901)
Business Income ≠ 0	0.0812	0.0777	0.0943	0.0950			0.275	0.190	0.116
Property Income ≠ 0	0.0714	0.165	0.0989	0.154			0.987	0.877	0.302
CONNECT	0	0	0	0	0	0	0	1	0.0895
Family Size	2.899	3.023	3.115	2.948	3.272	3.047	2.992	2.976	3.013
	(0.713)	(0.699)	(0.958)	(0.828)	(0.877)	(0.761)	(1.043)	(1.010)	(0.834)
Female Household Head	0.500	0.412	0.499	0.333	0.493	0.323	0.394	0.302	0.370
Age of Household Head	45.94	47.20	48.01	46.72	47.30	46.08	48.11	46.14	46.74
	(8.475)	(8.795)	(9.859)	(9.924)	(9.659)	(9.263)	(10.04)	(9.931)	(9.533)
Years of Education	10.88	10.99	9.800	10.09	10.03	10.27	10.39	11.61	10.45
	(3.027)	(2.922)	(2.193)	(2.136)	(2.055)	(1.847)	(3.442)	(3.383)	(2.531)
3-Year College or Above	0.279	0.246	0.290	0.334	0.507	0.513	0.264	0.365	0.365
4-Year College or Above	0.0909	0.0555	0.106	0.151	0.325	0.307	0.114	0.185	0.177
Self-Reported Health	3.682	3.662	3.582	3.803	3.765	3.780	3.881	4.019	3.777
Status (1 to 5)	(0.856)	(0.807)	(0.880)	(0.771)	(0.835)	(0.718)	(0.851)	(0.827)	(0.795)
Urban Hukou	0.994	0.990	0.968	0.968	0.980	0.975	0.938	0.951	0.972
Local Hukou	0.990	0.993	0.979	0.969	0.991	0.976	0.984	0.930	0.975
Panel B: CITY Characteristics									
GDP	1,727	1,488	3,372	3,191	4,229	3,867	12,269	9,121	4,098
	(615.8)	(797.8)	(1,482)	(1,787)	(1,786)	(2,104)	(2,081)	(3,891)	(3,403)

GDP Growth Rate	10.5	12.3	15.5	15.7	14.3	13.3	12.2	10.6	13.6
	(0.481)	(0.917)	(0.162)	(1.162)	(0.0788)	(1.983)	(0.343)	(0.734)	(2.179)
Population	2,736	666.8	2,646	682.5	2,702	694.0	2,882	799.9	1,122
	(951.1)	(199.9)	(1,164)	(213.9)	(1,144)	(214.3)	(472.2)	(220.3)	(962.3)
Fixed Investment	868.4	558.4	2,556	1,616	3,320	2,012	10,110	4,917	2,405
	(321.0)	(270.6)	(1,195)	(521.2)	(1,492)	(602.2)	(1,743)	(1,371)	(2,335)
Hospital Beds per 10,000 people	20.04	38.02	22.12	41.34	25.12	46.14	33.34	65.10	40.74
	(1.849)	(8.075)	(0.993)	(9.230)	(2.465)	(8.943)	(0.141)	(12.07)	(13.61)
NO. of Public Libraries	38.95	12.92	36.25	12.82	36.69	13.35	41.92	15.25	18.52
	(12.72)	(5.208)	(13.33)	(5.498)	(13.00)	(5.338)	(5.780)	(4.891)	(12.49)
N	308	1,171	435	2,073	345	1,719	386	633	7,070

Note: Standard deviation in the parentheses. For the dummy variables, only sample means are reported.

Table 4.2 Impacts of HSR Connection on Household Income

	All Waves		Without CHIP 2008	
	All	Local Hukou	All	Local Hukou
	(1)	(2)	(3)	(4)
CONNECT	0.264**	0.313**	0.299*	0.343*
	(0.110)	(0.106)	(0.154)	(0.152)
Observations	7,070	6,891	5,006	4,871
R-squared	0.424	0.426	0.514	0.517
City FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. The dependent variables are log total income. All regressions include a constant, household characteristics (i.e., family size, dummy for female head, head's age, years of education, self-reported health status, dummy for urban hukou and local hukou) and city characteristic (i.e., log GDP, GDP growth rate, log population, log fixed investment, log hospital beds per 10,000 people, the number of public libraries). Standard errors are clustered at the city level.

Table 4.3 The First-Stage Regressions of Household Income on HSR Connection

Dependent Variable:	All Waves		Without CHIP 2008	
CONNECT	All	Local Hukou	All	Local Hukou
	(1)	(2)	(3)	(4)
IV1: pass1962 *y2013	0.359*	0.361*	0.415**	0.417**
	(0.165)	(0.165)	(0.174)	(0.174)
IV2: xing1962 *y2013	0.939***	0.937***	0.944***	0.942***
	(0.0625)	(0.0638)	(0.0548)	(0.0561)
Observations	7,070	6,891	5,006	4,871
R-squared	0.961	0.959	0.969	0.968
City FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Anderson canon. corr. LM statistic				
	4820.680	4658.739	3433.267	3316.238
P-value	0.000	0.000	0.000	0.000
Cragg-Donald Wald F stat.				
	7546.11	7162.601	5434.566	5167.085
Critical Value	19.93	19.93	19.93	19.93
Sargan statistic				
	2.646	2.029	2.694	1.961
P-value	0.1038	0.1543	0.1007	0.1614

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. The dependent variables are the dummy of HSR connection in a city by the end of the year. All regressions include a constant, household characteristics (i.e., family size, dummy for female head, head's age, years of education, self-reported health status, dummy for urban hukou and local hukou) and city characteristic (i.e., log GDP, GDP growth rate, log population, log fixed investment, log hospital beds per 10,000 people, the number of public libraries). Standard errors are clustered at the city level.

Table 4.4 The Results from Instrumental Variable (2SLS) Estimation

	All Waves		Without CHIP 2008	
	All	Local Hukou	All	Local Hukou
	(1)	(2)	(3)	(4)
CONNECT	0.336**	0.398***	0.417**	0.474***
	(0.135)	(0.133)	(0.169)	(0.168)
Observations	7,070	6,891	5,006	4,871
R-squared	0.424	0.426	0.514	0.517
City FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. The dependent variables are log total income. All regressions include a constant, household characteristics (i.e., family size, dummy for female head, head's age, years of education, self-reported health status, dummy for urban hukou and local hukou) and city characteristic (i.e., log GDP, GDP growth rate, log population, log fixed investment, log hospital beds per 10,000 people, the number of public libraries). Standard errors are clustered at the city level.

Table 4.5 Impacts of HSR Connection by Income Sources (DD Estimation)

	Log Total Income	Log Wage Income	Having Business Income	Having Property Income
	(1)	(2)	(3)	(4)
Panel A: All Households				
CONNECT	0.299*	0.461**	-0.143**	-0.00666
	(0.154)	(0.143)	(0.0559)	(0.0633)
Observations	5,006	4,476	5,006	5,006
R-squared	0.514	0.378	0.072	0.555
Panel B: Households with Local Hukou				
CONNECT	0.343*	0.492***	-0.138**	0.00508
	(0.152)	(0.137)	(0.0572)	(0.0624)
Observations	4,871	4,364	4,871	4,871
R-squared	0.517	0.377	0.065	0.561
City FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. Household data is from CHIP 2002, 2007, 2013. The corresponding dependent variable for each column is log total income, log wage income, the dummy of whether having business income, or property income, respectively. Columns (3) and (4) are estimated using linear probability model. The regression of wage income only uses households with positive wage income. Panel A reports the results for all households, and Panel B reports the results for households with local hukou. All regressions include a constant, household characteristics (i.e., family size, dummy for female head, head's age, years of education, self-reported health status, dummy for urban hukou and local hukou) and city characteristic (i.e., log GDP, GDP growth rate, log population, log fixed investment, log hospital beds per 10,000 people, the number of public libraries). Standard errors are clustered at the city level.

Table 4.6 Heterogeneous Impacts by Industry

	Log Total Income	Log Wage Income	Having Business Income	Having Property Income
	(1)	(2)	(3)	(4)
CONNECT	0.308	0.430**	-0.0997	0.0193
	(0.173)	(0.186)	(0.0568)	(0.0602)
CONNECT * Manufacture	-0.111	-0.324*	0.0823**	-0.0678
	(0.108)	(0.159)	(0.0351)	(0.0545)
CONNECT * Construct	0.158	-0.195	0.0544	-0.0537
	(0.108)	(0.128)	(0.105)	(0.122)
CONNECT * Retail&Hotel/Rest.	-0.108	-0.0520	0.0148	-0.155***
	(0.0845)	(0.141)	(0.0867)	(0.0275)
CONNECT * Transport&Comm.	0.241***	0.346**	-0.104*	0.0269
	(0.0509)	(0.130)	(0.0516)	(0.0483)
CONNECT * Government	0.322***	0.0918	-0.0989	0.128***
	(0.0865)	(0.0725)	(0.0544)	(0.0325)
Observations	5,006	4,476	5,006	5,006
R-squared	0.527	0.393	0.126	0.563
City FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Industry by City FE	YES	YES	YES	YES
Industry by Year FE	YES	YES	YES	YES

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. Household data is from CHIP 2002, 2007, 2013. The corresponding dependent variable for each column is log total income, log wage income, the dummy of whether having business income, or property income, respectively. Columns (3) and (4) are estimated using linear probability model. The regression of wage income only uses households with positive wage income. All regressions include a constant, household characteristics (i.e., family size, dummy for female head, head's age, years of education, self-reported health status, dummy for urban hukou and local hukou) and city characteristic (i.e., log GDP, GDP growth rate, log population, log fixed investment, log hospital beds per 10,000 people, the number of public libraries). Standard errors are clustered at the city level. The default category is household in all other industries.

Table 4.7 Heterogeneous Impacts by Age Cohorts

	Log Total Income	Log Wage Income	Having Business Income	Having Property Income
	(1)	(2)	(3)	(4)
CONNECT	0.0324	0.377*	-0.153*	-0.156**
	(0.192)	(0.167)	(0.0691)	(0.0565)
CONNECT * Age 35-44	0.171	0.0885	0.112**	0.171***
	(0.102)	(0.115)	(0.0409)	(0.0524)
CONNECT * Age 45-54	0.323***	-0.0469	-0.00438	0.167***
	(0.0763)	(0.141)	(0.0695)	(0.0316)
CONNECT * Age 55-65	0.351***	0.304*	0.000693	0.167***
	(0.0878)	(0.143)	(0.0709)	(0.0415)
Observations	5,006	4,476	5,006	5,006
R-squared	0.524	0.392	0.081	0.560
City FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
Age Cohort by City FE	YES	YES	YES	YES
Age Cohort by Year FE	YES	YES	YES	YES

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. Household data is from CHIP 2002, 2007, 2013. The corresponding dependent variable for each column is log total income, log wage income, the dummy of whether having business income, or property income, respectively. Columns (3) and (4) are estimated using linear probability model. The regression of wage income only uses households with positive wage income. All regressions include a constant, household characteristics (i.e., family size, dummy for female head, head's age, years of education, self-reported health status, dummy for urban hukou and local hukou) and city characteristic (i.e., log GDP, GDP growth rate, log population, log fixed investment, log hospital beds per 10,000 people, the number of public libraries). Standard errors are clustered at the city level. The default category is households whose heads are aged between 25 to 34.

Table 4.8 Heterogeneous Impacts by Skill Levels

	Log Total Income	Log Wage Income	Having Business Income	Having Property Income
	(1)	(2)	(3)	(4)
Panel A: Three-Year Colleges (Associate's Degree)				
CONNECT	0.306**	0.405***	-0.134**	-0.0262
	(0.123)	(0.103)	(0.0473)	(0.0595)
CONNECT * High-Skill	-0.0145	0.0439	0.0545*	0.0675
	(0.0877)	(0.0933)	(0.0269)	(0.0402)
Observations	5,006	4,476	5,006	5,006
R-squared	0.545	0.400	0.081	0.558
Panel B: Four-Year Colleges (Bachelor's Degree)				
CONNECT	0.315*	0.464**	-0.135*	-0.0101
	(0.155)	(0.150)	(0.0611)	(0.0645)
CONNECT * High-Skill	-0.0684	-0.137*	0.0527	0.00177
	(0.0715)	(0.0731)	(0.0446)	(0.0576)
Observations	5,006	4,476	5,006	5,006
R-squared	0.520	0.375	0.073	0.558
City FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES
College by City FE	YES	YES	YES	YES
College by Year FE	YES	YES	YES	YES

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. Household data is from CHIP 2002, 2007, 2013. The corresponding dependent variable for each column is log total income, log wage income, the dummy of whether having business income, or property income, respectively. Columns (3) and (4) are estimated using linear probability model. The regression of wage income only uses households with positive wage income. Panel A defines the high-skilled as people who have a three-year college degree or more, and Panel B defines the high-skilled as people who finished four-year college education. All regressions include a constant, household characteristics (i.e., family size, dummy for female head, head's age, years of education, self-reported health status, dummy for urban hukou and local hukou) and city characteristic (i.e., log GDP, GDP growth rate, log population, log fixed investment, log hospital beds per 10,000 people, the number of public libraries). Standard errors are clustered at the city level. The default category is households whose heads are aged between 25 to 34.

Table 4.9 Impacts of HSR Connection on Household Characteristics

	Local Hukou	Urban Hukou	Family Size	Female Head	Age	Education	Associate	Bachelor	Health
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: OLS Estimation									
CONNECT	-0.0474**	0.00251	-0.0613	0.0847	-3.988***	1.099*	0.131*	0.151***	-0.188
	(0.0152)	(0.0295)	(0.134)	(0.0647)	(0.583)	(0.496)	(0.0698)	(0.0298)	(0.166)
Observations	5,006	5,006	5,006	5,006	5,006	5,006	5,006	5,006	5,006
R-squared	0.025	0.019	0.018	0.039	0.034	0.064	0.029	0.031	0.039
Panel B: IV Estimation									
CONNECT	-0.0766***	-0.0245	-0.264	0.0324	-6.296***	2.182***	0.266***	0.216***	0.189
	(0.0247)	(0.0519)	(0.183)	(0.0746)	(1.368)	(0.845)	(0.100)	(0.0550)	(0.244)
Observations	5,006	5,006	5,006	5,006	5,006	5,006	5,006	5,006	5,006
R-squared	0.025	0.019	0.017	0.039	0.034	0.062	0.028	0.030	0.037
City FE	YES	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES	YES	YES	YES

Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. Household data is from CHIP 2002, 2007, 2013. The dependent variables in each column are the dummy variable indicating local hukou, or urban hukou, family size, dummy for female household head, head's age, years of education, dummy for whether the household head has at least an Associate's degree, or at least a Bachelor's degree, and self-reported health status, respectively. Panel A presents the OLS estimation, as Panel B presents the IV estimation. All regressions include a constant and city characteristic (i.e., log GDP, GDP growth rate, log population, log fixed investment, log hospital beds per 10,000 people, the number of public libraries). Standard errors are clustered at the city level.

Figures

Figure 4.1 The Geography of HSR

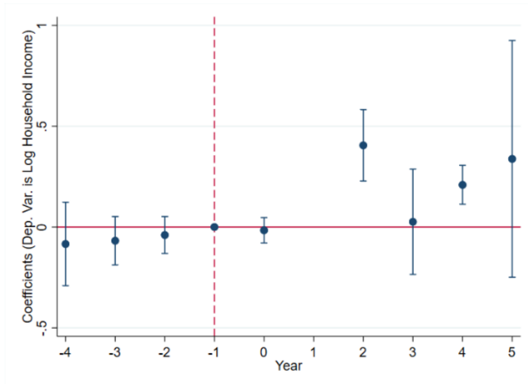
2008

2013

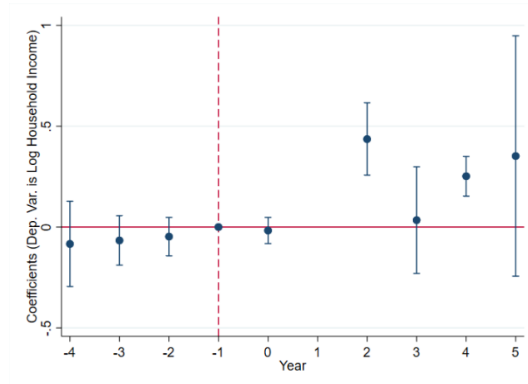
Source: Lawrence, M., Bullock, R. and Liu, Z., 2019. China's High-Speed Rail Development. International Development in Focus. Washington, DC: World Bank.

Figure 4.2 Test on Common Trend Assumption and Dynamic Effect of HSR Connection on Household Income

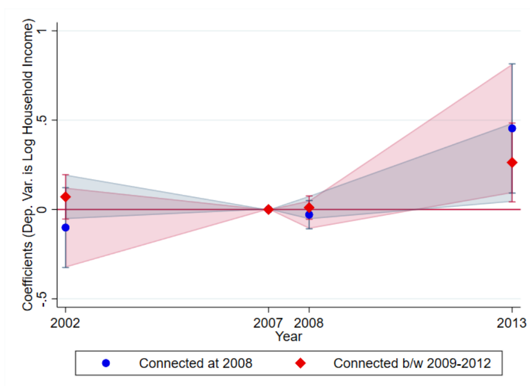
(A) All Households



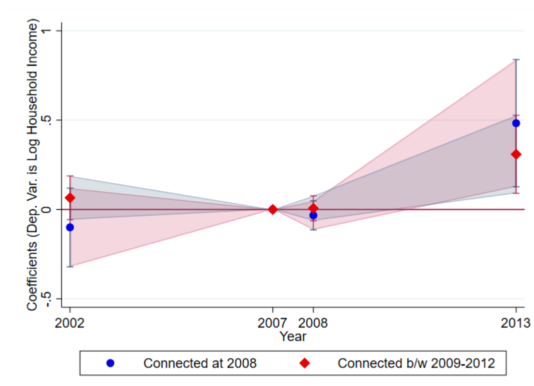
(B) Household with Local Hukou



(C) All Households



(D) Household with Local Hukou



Notes: Fig. (A) and (B) are the dynamic effects of HSR connection on household income using event study model. Fig. (C) and (D) show yearly differences in average household income between the HSR connected cities and the unconnected cities. While Fig. (A) and (C) analyze the whole sample, Fig. (B) and (D) focus on the households with low-skilled (i.e., no college degree) head only.

APPENDICES

APPENDIX A. Additional Tables and Figures for Chapter 2

Table A 1 Robustness Check with Alternative Dependent Variables

Dep. Variable:	(1) Station	(2)	(3)	(4) Route	(5)	(6)
College Score - Tier1 Cutoff	Score	Avg. Score	High. Score	Score	Avg. Score	High. Score
Panel A: OLS						
HSR	2.086***	1.086	0.521	0.811	0.904	0.701
	(0.718)	(0.642)	(0.462)	(0.843)	(0.664)	(0.713)
Observations	22,251	22,251	22,251	21,945	21,945	21,945
R-squared	0.744	0.862	0.838	0.491	0.578	0.639
Panel B: Tobit						
HSR	1.620**	1.094*	0.524	0.326	0.892	0.696
	(0.685)	(0.632)	(0.456)	(0.836)	(0.644)	(0.692)
Observations	22,251	22,251	22,251	21,945	21,945	21,945
Fixed Effects	university, exam package by home province by year			college-city by student-home-province, exam package by home province by year		

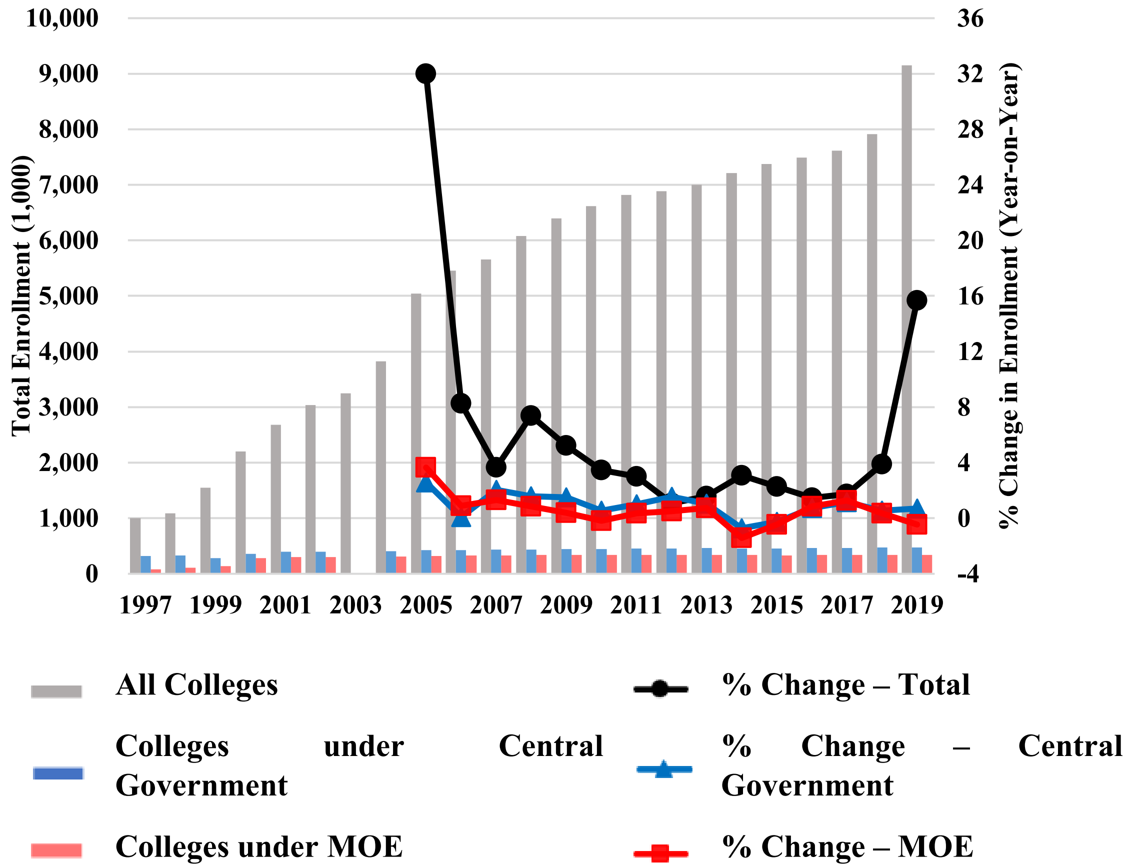
Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. Dependent variables are the differences between the college scores and the tier-1 cutoff scores. All regressions include a constant, college enrollment, and characteristics in the college city -- GDP per capita, population, public green land area, the number of hospital beds. Standard errors are clustered at the college province level for the station effects and at the province pair level for the route equation.

Table A 2 Robustness Check with Alternative HSR Treatment Definitions

	(1) Station	(2)	(3)	(4) Route	(5)	(6)
	Score	Avg. Score	High. Score	Score	Avg. Score	High. Score
Panel A: HSR connection by May						
HSR	2.172**	0.756	0.00301	0.829	1.389***	1.611***
	(0.940)	(0.628)	(0.425)	(0.636)	(0.495)	(0.569)
Observations	19,950	19,950	19,950	31,098	31,098	31,098
R-squared	0.895	0.930	0.892	0.748	0.758	0.748
Panel B: HSR connection by May of last year						
HSR	1.944**	1.933**	1.171*	0.768	1.142	0.577
	(0.828)	(0.762)	(0.595)	(1.069)	(0.897)	(0.943)
Observations	19,982	19,982	19,982	34,359	34,359	34,359
R-squared	0.895	0.930	0.892	0.731	0.740	0.733
Panel C: HSR in the college province capital						
HSR	2.088**	0.873	0.370	1.019	0.673	-0.271
	(0.831)	(0.720)	(0.502)	(0.884)	(0.687)	(0.778)
Observations	19,117	19,117	19,117	18,868	18,868	18,868
R-squared	0.887	0.927	0.889	0.774	0.774	0.756
Fixed Effects	university, exam package by home province by year			college-city by student-home-province, exam package by home province by year		

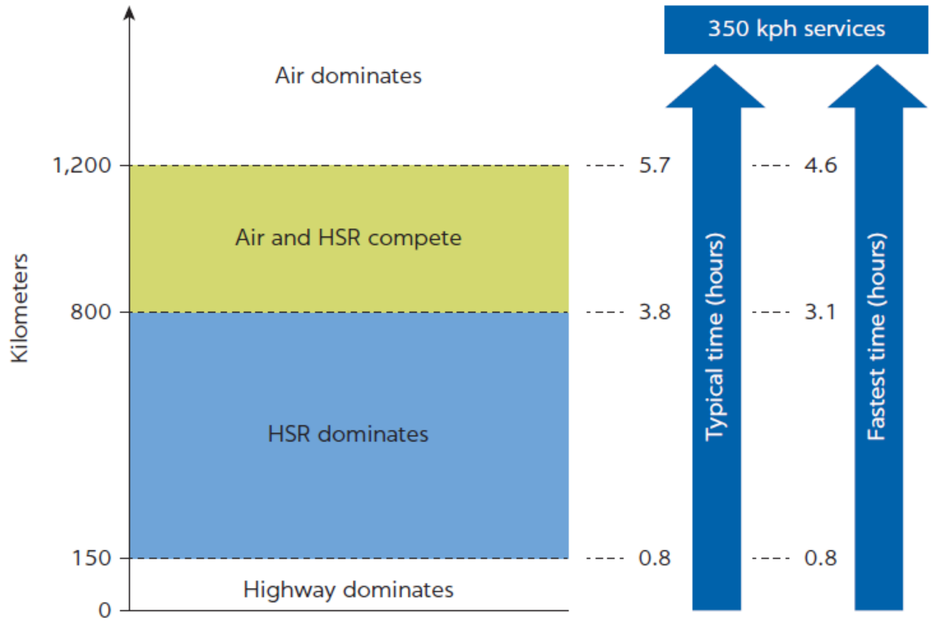
Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. All regressions include a constant, college enrollment, and characteristics in the college city -- GDP per capita, population, public green land area, the number of hospital beds. Standard errors are clustered at the college province level for the station effects and the province pair level for the route equation.

Figure A 1 Enrollment in Higher Education Institutions by Providers



Data Source: Tables of "Number of Students for Regular and Adult Programs by Providers in HEIs," Educational Statistics from Ministry of Education of the People's Republic of China.

Figure A 2 Competitiveness of High-Speed Rail



Note: The competitive ranges of the three modes are indicative. The air and high-speed rail (HSR) competitiveness was studied with a sample of 300–350-kilometer-per-hour (kph) lines. With different price and speed assumption of 200–250 kph lines, the dominance range will be slightly different.

Source: Lawrence, M., Bullock, R. and Liu, Z., 2019. China's High-Speed Rail Development. International Development in Focus. Washington, DC: World Bank.

APPENDIX B. Additional Tables and Figures for Chapter 4

Table B 1 Cities in Sample by HSR Station Opening Time

Opening Time	City	Province
2008/04/18	Nanjing	Jiangsu
2008/04/18	Hefei	Anhui
2009/04/01	Wuhan	Hubei
2009/12/26	Guangzhou	Guangdong
2010/02/06	Zhengzhou	Henan
2010/05/10	Chengdu	Sichuan
2010/07/01	Wuxi	Jiangsu
2011/06/30	Bengbu	Anhui
2013/12/28	Chongqing	Chongqing
2014/12/20	Leshan	Sichuan

Notes: Data is obtained from the Chinese High-Speed Rail and Airline Database (CRAD) of the Chinese Research Data Services (CNRDS) Platform.

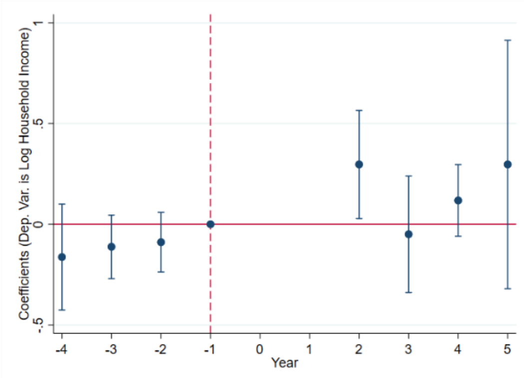
Table B 2 Impacts of HSR Connection by Income Sources (IV Estimation)

	Log Total Income	Log Wage Income	Having Business Income	Having Property Income
	(1)	(2)	(3)	(4)
Panel A: All Households				
CONNECT	0.417**	0.481***	-0.174**	-0.00578
	(0.169)	(0.148)	(0.0753)	(0.0921)
Observations	5,006	4,476	5,006	5,006
R-squared	0.514	0.378	0.072	0.555
Panel B: Households with Local Hukou				
CONNECT	0.474***	0.534***	-0.174**	0.00899
	(0.168)	(0.139)	(0.0775)	(0.0877)
Observations	4,871	4,364	4,871	4,871
R-squared	0.517	0.377	0.065	0.561
City FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

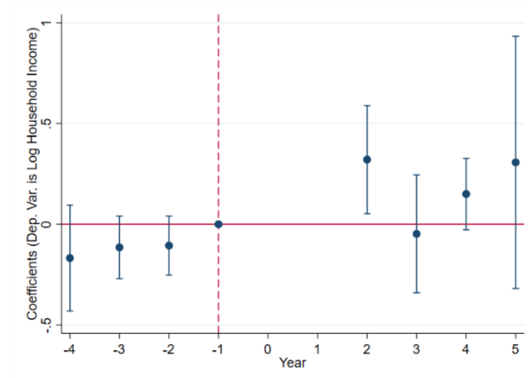
Notes: *, **, *** indicate statistical significance at the 10%, 5%, 1% level. Household data is from CHIP 2002, 2007, 2013. The corresponding dependent variable for each column is log total income, log wage income, the dummy of whether having business income, or property income, respectively. IVs are the historical railroad information, which is the same as the baseline regressions. Columns (3) and (4) are estimated using linear probability model. The regression of wage income only uses households with positive wage income. Panel A reports the results for all households, and Panel B reports the results for households with local hukou. All regressions include a constant, household characteristics (i.e., family size, dummy for female head, head's age, years of education, self-reported health status, dummy for urban hukou and local hukou) and city characteristic (i.e., log GDP, GDP growth rate, log population, log fixed investment, log hospital beds per 10,000 people, the number of public libraries). Standard errors are clustered at the city level.

Figure B 1 Test on Common Trend Assumption and Dynamic Effect of HSR Connection on Household Income (for CHIP 2002, 2007, 2013 only)

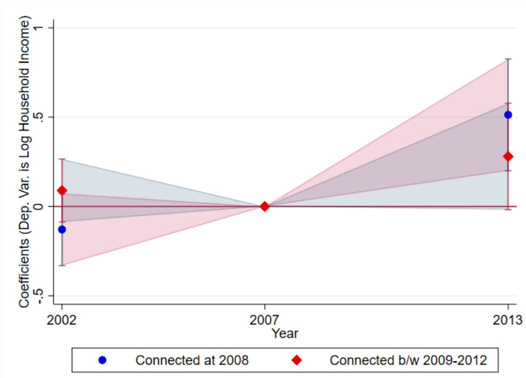
(A) All Households



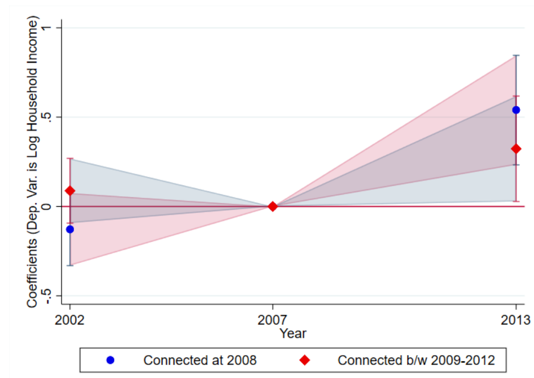
(B) Household with Local Hukou



(C) All Households



(D) Household with Local Hukou



Notes: Fig. (A) and (B) are the dynamic effects of HSR connection on household income using event study model. Fig. (C) and (D) show yearly differences in average household income between the HSR connected cities and the unconnected cities. While Fig. (A) and (C) analyze the whole sample, Fig. (B) and (D) focus on the households with low-skilled (i.e., no college degree) head only.

BIBLIOGRAPHY

Agrawal, A., Galasso, A. and Oettl, A., 2017. Roads and innovation. Review of Economics and Statistics, 99(3), pp.417-434.

Au, C.C. and Henderson, J.V., 2006. Are Chinese cities too small?. The Review of Economic Studies, 73(3), pp.549-576.

Banerjee, A., Duflo, E. and Qian, N., 2012. On the road: Access to transportation infrastructure and economic growth in China. Technical report, NBER Working Paper no. 17897.

Baum-Snow, N., 2007. Did highways cause suburbanization?. The quarterly journal of economics, 122(2), pp.775-805.

Baum-Snow, N., 2019. Urban transport expansions and changes in the spatial structure of us cities: Implications for productivity and welfare. Review of Economics and Statistics, pp.1-45.

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

Baum-Snow, N. and Turner, M.A., 2017. Transport Infrastructure and the Decentralization of Cities in the People's Republic of China. Asian development review, 34(2), pp.25-50.

Bogart, D., 2009. Turnpike trusts and property income: new evidence on the effects of transport improvements and legislation in eighteenth-century England 1. The Economic History Review, 62(1), pp.128-152.

Chan, J., Wang, Z. and Zhao, S., 2019. The Institution of Merit: A Study of Chinese College Admissions. Available at SSRN 3423896.

Chandra, A. and Thompson, E., 2000. Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. Regional science and urban economics, 30(4), pp.457-490.

De La Roca, J. and Puga, D., 2017. Learning by working in big cities. The Review of Economic Studies, 84(1), pp.106-142.

Dong, L., Du, R., Kahn, M., Ratti, C. and Zheng, S., 2021. “Ghost Cities” versus Boom Towns: Do China’s HSR New Towns Thrive?. Regional Science and Urban Economics, p.103682.

Dong, X., Zheng, S. and Kahn, M.E., 2020. The role of transportation speed in facilitating high skilled teamwork across cities. Journal of Urban Economics, 115, p.103212.

Duranton, G., Morrow, P.M. and Turner, M.A., 2014. Roads and Trade: Evidence from the US. Review of Economic Studies, 81(2), pp.681-724.

Duranton, G. and Turner, M.A., 2012. Urban growth and transportation. Review of Economic Studies, 79(4), pp.1407-1440.

Donaldson, D. and Hornbeck, R., 2016. Railroads and American economic growth: A "market access" approach. The Quarterly Journal of Economics, 131(2), pp.799-858.

Faber, B., 2014. Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System. Review of Economic Studies, 81(3), pp.1046-1070.

Gao, Y., Song, S., Sun, J. and Zang, L., 2020. Does high-speed rail connection really promote local economy? Evidence from China’s Yangtze River Delta. Review of Development Economics, 24(1), pp.316-338.

Gao, Y. and Zheng, J., 2020. The impact of high-speed rail on innovation: An empirical test of the companion innovation hypothesis of transportation improvement with China’s manufacturing firms. World Development, 127,

Grogger, J. and Hanson, G.H., 2011. Income maximization and the selection and sorting of international migrants. Journal of Development Economics, 95(1), pp.42-57.

Grogger, J. and Hanson, G.H., 2015. Attracting talent: Location choices of foreign-born PhDs in the United States. Journal of Labor Economics, 33(S1), pp.S5-S38.

Gibbons, S. and Machin, S., 2005. Valuing rail access using transport innovations. Journal of urban Economics, 57(1), pp.148-169.

Haapanen, M. and Tervo, H., 2012. Migration of the highly educated: Evidence from residence spells of university graduates. Journal of Regional Science, 52(4), pp.587-605.

Kato, T. and Sparber, C., 2013. Quotas and quality: The effect of H-1B visa restrictions on the pool of prospective undergraduate students from abroad. Review of Economics and Statis-tics, 95(1), pp.109-126.

Ke, X., Chen, H., Hong, Y. and Hsiao, C., 2017. Do China's high-speed-rail projects promote local economy?—New evidence from a panel data approach. China Economic Review, 44, pp.203-226.

Kerr, S.P., Kerr, W., Özden, Ç. and Parsons, C., 2017. High-skilled migration and agglomeration. Annual Review of Economics, 9, pp.201-234.

Kleven, H.J., Landais, C. and Saez, E., 2013. Taxation and international migration of superstars: Evidence from the European football market. American economic review, 103(5), pp.1892-1924.

Kong, D., Liu, L. and Yang, Z., 2021. High-speed rails and rural-urban migrants’ wages. Economic Modelling, 94, pp.1030-1042.

Lawrence, M., Bullock, R. and Liu, Z., 2019. China's High-Speed Rail Development. International Development in Focus. Washington, DC: World Bank. doi:10.1596/978-1-4648-1425-9.

Li, H., Meng, L., Shi, X. and Wu, B., 2012. Does attending elite colleges pay in China?. Journal of Comparative Economics, 40(1), pp.78-88.

Li, S., Li, Z. and Yuan, J., 2020. Examining the impact of high-speed railways on land value and government revenue: Evidence from China. China Economic Review, 63, p.101502.

Li, T. and Zhang, J., 2010. What determines employment opportunity for college graduates in China after higher education reform?. China Economic Review, 21(1), pp.38-50.

Li, X., Huang, B., Li, R. and Zhang, Y., 2016. Exploring the impact of high speed railways on the spatial redistribution of economic activities-Yangtze River Delta urban agglomeration as a case study. Journal of Transport Geography, 57, pp.194-206.

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

Ma, Liping. and Pan, K., 2014. Stay or Migrate? An Empirical Study of the Relationship between Place of Work, Place of Study and Birthplace. Chinese Education & Society, 47(6), pp.80-95.

Ma, L., Niu, D. and Sun, W., 2021. Transportation infrastructure and entrepreneurship: Evidence from high-speed railway in China. China Economic Review, 65, p.101577.

McHugh, R. and Morgan, J.N., 1984. The determinants of interstate student migration: A place-to-place analysis. Economics of Education Review, 3(4), pp.269-278.

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

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

Redding, S.J. and Turner, M.A., 2015. Transportation costs and the spatial organization of eco-nomic activity. In Handbook of regional and urban economics (Vol. 5, pp. 1339-1398). Elsevier.

Rhode, P.W. and Strumpf, K.S., 2003. Assessing the importance of Tiebout sorting: Local heterogeneity from 1850 to 1990. American Economic Review, 93(5), pp.1648-1677.

Rosenzweig, M.R., Irwin, D.A. and Williamson, J.G., 2006, January. Global wage differences and international student flows. Brookings trade forum (pp. 57-96). Brookings Institution Press.

Sheard, N., 2014. Airports and urban sectoral employment. Journal of Urban Economics, 80, pp.133-152.

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

Winters, J.V., 2011. Why are smart cities growing? Who moves and who stays. Journal of regional science, 51(2), pp.253-270.

Zhao, S. and Ma, D., 2017. Ghost city phenomenon along China's high-speed railway grid. International Journal of Sustainable Society, 9(3), pp.210-225.

Zheng, S. and Kahn, M.E., 2013. China's bullet trains facilitate market integration and mitigate the cost of megacity growth. Proceedings of the National Academy of Sciences, 110(14), pp.E1248-E1253.

VITA

Shiyu Cheng

Education

University of Kentucky, Lexington, KY, US
M.S. in Economics, 12/2017
University of California - Santa Barbara, Santa Barbara, CA, US
Visiting student, course work in Math and Economics, 03/2015 - 03/2016
Southwestern University of Finance and Economics (SWUFE), Chengdu, Sichuan, China
Master course work in Public Economics, Fall 2014 & Spring 2016
B.A. in Economics, 06/2014

Employment

Research Assistant, Center for Poverty Research, University of Kentucky, 2017 – 2019
Teaching Assistant, Department of Economics, University of Kentucky, 2016 – 2017, 2019 – 2021

Awards and Fellowship

Luckett Graduate Fellowship, University of Kentucky, 2019 – 2021
Gatton Doctoral Fellowship, University of Kentucky, 2019 – 2020
Poverty Center Understanding SNAP Fellowship, University of Kentucky, 2017 – 2018
Gatton Award for Academic Achievement in First-Year Ph.D. Coursework, 2017
Max Steckler First-Year Graduate Fellowship, University of Kentucky, 2016 – 2017
National Scholarship, China, 2014

Professional and Invited Presentations

The 46th Annual Conference of the Association for Education Finance and Policy (AEFP), 03/2021
The 7th International Conference on the Chinese Economy: Past, Present, and Future, 12/2020
The 12th International Symposium on Human Capital and Labor Markets, 12/2020
The 11th Camphor Economist Circle (CEC) Hangzhou Virtual Seminar, 11/2020
The 67th Annual North American Meetings of the Regional Science Association International, 11/2020
2019 Chinese Economists Society North America Conference, 04/2019
2019 Kentucky Economic Association Conference, 10/2019