Travel Costs and Urban Specialization Patterns: Evidence from China's High Speed Railway System

Yatang Lin[1]

CEP, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, United Kingdom

Abstract

How does intercity passenger transportation shape urban employment and specialization patterns? To shed light on this question I study China's High Speed Railway (HSR), an unprecedentedly large-scale network that connected 81 cities from 2003 to 2014 with trains running at speeds over 200 km/h. Using a difference-in-differences approach, I find that an HSR connection increases city-wide passenger flows by 10% and employment by 7%. To deal with the issues of endogenous railway placement and simultaneous public investments accompanying HSR connection, I examine the impact of a city's market access changes purely driven by the HSR connection of other cities. The estimates suggest that HSR-induced expansion in market access increases urban employment with an elasticity between 2 and 2.5. Further evidence on sectoral employment suggests that industries with a higher reliance on nonroutine cognitive skills benefit more from HSR-induced market access to other cities.

Keywords: Transportation infrastructure, High Speed Railway, Urban Employment and Specialization

1. Introduction

Transportation costs play an important role in the location, agglomeration and evolution of economic activities. Yet, despite abundant research on the relationship between the cost of goods transportation and trade patterns, relatively little attention has been paid to the costs of passenger travel and their implications for labour markets. Reducing the cost of travel between cities not only removes obstacles to migration (Morten and Oliveira 2013) [39], but also reduces the cost of face-to-face meeting across cities and allows remote sourcing of jobs. As airplanes or speedy trains make frequent day trips more feasible, firms may be more willing to locate their headquarters or R&D centres in centrally-located cities with large pools of talented employees, who can exert effective control over production plants in smaller cities with much lower urban costs. How significant are the benefits of infrastructure projects that dramatically increase the speed of intercity travelling, and how are these benefits distributed across sectors?

Email address: y.lin18@lse.ac.uk (Yatang Lin)

In this paper, I exploit the high speed railway (HSR) project in China, the largest in the world, as a natural experiment to study the benefits of improving passenger-dedicated transportation infrastructure. I examine the impacts of HSR connection through changes in its access to all the other cities, and focus on the one driven by indirect HSR connection to address endogeneity issues such as non-random route placements or simultaneous investments in other areas. The differential impacts of HSR on employment across sectors suggest that industries benefiting more from enhanced market access are either tourism-related or intensive in nonroutine cognitive skills. These findings highlight the role of improved passenger travel infrastructure in promoting the delivery of services across cities, facilitating cross-city labour sourcing and knowledge exchanges, and ultimately shifting the specialization pattern of connected cities towards skilled and communication intensive sectors.

The HSR expansion in China is an appropriate context for such a study. As of 2014, China had the world's longest and busiest HSR network with 15,399 km of track in service, connections between 81 cities, and an annual ridership of 857 million as of 2014.[1] HSR had a marked impact on people's travel patterns: after being connected to the HSR network, a city experiences an 18% increase in the number of passengers travelling by train and a 9.6% increase in the number of passengers travelling by any forms of transportation. The top-down rapid expansion of the HSR network also creates plausibly exogenous variation in each city's connectivity which does not depend on its own its own direct connection to the HSR system, since passengers from unconnected cities use HSR to travel to other cities by transferring at a nearby HSR hub.

In a difference-in-differences specification, I demonstrate that being connected to the HSR network leads to a significant increase in GDP and urban employment. As a first attempt at dealing with the problem of endogenous routes placement, I restrict my study sample to the cities that are either connected by the end of 2014 or will be connected by HSR according to the HSR plan of the Ministry of Railway (MOR) [38]. Therefore, the analysis is identified through the timing of a city's connection to the HSR, which is affected by idiosyncratic factors such as the length of the line and engineering difficulties. I also test whether connected and unconnected cities have experienced differential trends in GDP or employment growth before the actual HSR connection by including the leads of the initial connection dummy. No HSR effects are found on GDP or employment before the actual connection.

To further explore the mechanisms at work, I adopt a "market access" approach similar to that in Donaldson and Hornbeck (2015) [18], which is micro-founded by a model of cross-city labour sourcing. I measure how the expansion of the HSR network affects each city's "market access", and estimate the impacts of enhanced market access on the city's transportation and specialization patterns, as well as aggregate economic outcomes. A city's market access is approximated using an average of other cities' GDP inversely weighted by the bilateral costs of passenger travel. To account for the changes in cities' market access driven by improvements in different modes of transportation, including HSR, I assemble a network database of highways and railways in China from 2000 to 2014 to construct a time-varying travel cost matrix that takes account of the changes in time and fare costs brought by both

[1] Table A8 and Figure A1

highway and HSR expansions in China.

A major empirical advantage of this "market access" method is that it allows a city's market access to be affected by the HSR connections of other cities. Employees in non-HSR cities can travel to a nearby HSR town and transfer there for other destinations[2]. The separation of HSR-induced market access growth from direct HSR placements assists in dealing with the identification challenges of both endogenous infrastructure placement and simultaneous public spending and investments in relevant sectors. In my main specification, I use a measure of market access which deducts the increase in market access driven by a city's own connections. A one percent increase in this "non-connection-induced market access" (NCIMA) leads to an 8% increase in railway ridership and 2% increase in employment. The impacts are largest for tourism-related employment, followed by skilled employment including IT, finance, business services, education and scientific research, and smallest for other types of service and non-service employment.

Another benefit of this approach is that I can evaluate the impacts of alternative market access measures that capture different sources of improvement in intercity connectivity during this period. In particular, I evaluate the separate effects of improvements in HRS expansion, highway expansion, and parts of HSR development that face little competition from air travel. This helps us to get a complete picture on the way various infrastructure improvements change transportation patterns and economic outcomes. HSR expansion does not reduce highway usage but highway expansion affects railway ridership significantly. A larger drop in air travel usage is observed in cities that benefit more from HSR over shorter distance trips. Over longer trips, HSR does not appear to steal passengers from civil aviation or road transportation, but it manages to attract new passengers. Regarding the economic outcomes, service employment is much more responsive to HSR-induced market access changes, particularly those over shorter distance trips, than highway-induced ones, while manufacturing employment and GDP responds more to highway-induced market access changes.

One of the most important findings in this paper is the implications of an improvement in intercity passenger transportation on cross-city specialization patterns. Conceptually, better intercity passenger transportation reduces the cost of face-to-face interactions across space, and should exert larger impacts on industries that are communication intensive. To test it, I begin by estimating the impacts of HSR across sixteen Chinese industries, and compare the estimated coefficients to the task contents of each industry, as per Autor, Levy and Murname (2003) [5]. It is revealed that the benefits of better intercity passenger transportation increase in the industry-specific requirement of nonroutine cognitive tasks, and decrease in their reliance on manual or routine cognitive skills. On the contrary, the estimated impacts of highway expansion do not correlate with the nonroutine contents of industries. These results highlight the distinctive role of HSR in shifting the specialization patterns of cities towards interactive industries, compared to other forms of transportation.

This paper contributes to a growing literature on estimating the economic impacts of transportation infrastructure projects. Recent papers have studied the skill premia in local

[2]Gerald et al. (2014) [25] documents that 40% of the passengers taking the Wuhan-Guangzhou HSR line come from other railway lines or alternative modes of transportation.

labour markets (Michaels, 2008) [36], long-term GDP growth (Banerjee et al., 2012) [6], income volatility (Burgess et al., 2012) [13], gains from trade (Donaldson, 2013) [17], and asymmetric effects on core and peripheral markets (Faber, 2014) [22]. Also, papers in urban economics have explored the effects of urban transportation improvements on urban growth (Duranton and Turner, 2012) [19] and urban form (Baum-Snow et al., 2015) [9], Baum-Snow 2012) [19]. Relative to the existing literature, this article draws attention to a different type of transportation infrastructure, the inter-city passenger transportation and a new mechanism, the sourcing of labour across cities.

This paper also contributes to an extensive literature on urban growth and specialization. Previous research has emphasized the importance of agglomeration effects (Glaeser, Kallal, Scheinkman, and Schleifer, 1992) [27], amenities (Clark et al. 2002) [14], human capital (Glaeser and Saiz, 2004) [28] and intra-city transportation (Duranton and Turner, 2012) [19]. This paper shows that better intercity passenger transportation can also act as a new engine of urban growth. Since I focus on how improved intercity transportation facilitates labour sourcing across cities and reshapes urban employment patterns, the paper is also related to the international trade literature on offshoring, such as Grossman and Rossi-Hansberg (2008) [29], and Ottaviano, Peri and Wright (2013) [41].

Finally, this paper provides a rigorous empirical evaluation of the largest HSR project in the world up to now. Earlier evaluations of HSR projects in Japan and Europe have presented mixed evidence. Sasaki et al. (1997) [45] suggests that the Shinkansen in Japan promoted local development and did not cause regional inequality. Bernard et al. (2014) [10] further examines the response of Japanese firms to a particular Shinkansen line with a focus on supplier relationships. Berhens and Pels (2011) citebp documents a significant change in passenger travel behavior along the Paris-London corridor after the high speed Eurostar was in operation. Ahlfeldt et al. (2010) [2] present evidence that the HSR line connecting Cologne and Frankfurt in Germany substantially increases the GDP of the regions that enjoy an increase in accessibility. While Marti-Hennenberg (2000) [35] discovers that French HSR has neither accelerated industrial concentration nor promoted economic decentralization from Paris. The HSR project in China provides me with an excellent opportunity to evaluate the potential economic benefits of HSR due to its large scale. An earlier paper (Zheng and Kahn, 2013) [48] on Chinese HSR finds that HSR connection boosts housing prices in China, using cross-sectional data. My paper looks at a wider range of economic outcomes with more detailed HSR data and different identification strategies.

2. Background and Discussion of HSR Usage in China

In 2008, the State Council in its revised Mid-to-Long Term Railway Development Plan set the goal of a national high-speed rail grid composed of four north-south corridors and four east-west corridors, with a budget of around 4,000 billion yuan (State Council, 2008) [16]. The construction costs of HSR range from 80-120 million RMB per km (US$13-20 million) excluding stations [12]. The expansion of the HSR network in China from 2003 to 2014, is shown in Figure 1. Detailed information on the construction start date, opening date, distance and speed of all the operating lines is listed in Table A4 and A5. The objective of this HSR grid, as stated in the Plan, is to connect provincial capitals and other major cities with faster means of transportation. The placement of lines, according to the Ministry of

Railway (MOR), should be based on a comprehensive consideration of the economic development, population and resource distribution, national security, environmental concerns and social stability of each region [38]. Finally, HSR lines are expected to complement existing transportation networks as much as possible.

Before exploring the role of HSR in city specialization patterns, we first need to understand how it affects the way people travel. I pay attention to two key questions. First, how does HSR compete with other forms of transportation? Second, what are the socioeconomic characteristics of HSR passengers? I explore these questions using collected passenger surveys and official ridership data.

Table A7 reports results from three pieces of passenger surveys, conducted by Li (2011) [32], Wu et al. (2013) [47] and Gerald et al. (2014) [25], respectively. Interviewees are drawn from four HSR lines, two short lines, and two long-haul ones. I collect answers to four kinds of question: (1) passengers' income; (2) purposes of travel; (3) means of transport to the HSR station and whether or not transfers on and off HSR are made, and (4) their alternative intercity travel choice before the introduction of HSR. From the second column, we learn that the average monthly income of HSR passengers ranges from ¥4300 to ¥6700, which roughly falls into the high-income group in China.[3]. A large proportion (25% to 40% along shorter HSR lines and 40% to 60% along longer HSR lines) of the passengers travel for business purposes. Regarding the substitution between the HSR and other forms of transport, as reported in column 4, none of the HSR passengers on the two short journeys preferred to fly to their destination before the advent of HSR, while 36% (Changchun-Jilin) and 61.5% (Beijing-Tianjin) of them listed conventional railway as their primary choice of transportation at that time, and 50% (Changchun-Jilin) and 32% (Beijing-Tianjin) of them had preferred road transport (including coaches/buses and private cars). Over the long-haul trip (Tianjin-Jinan), a large proportion (77%) mentioned air travel as their main choice before the HSR was in operation, followed by 18% who chose conventional rail travel, and almost none considered long-distance coach journeys. It is clear that HSR mainly competes with air travel for longer trips and with traditional railway/road travel over shorter lines.

I collect ridership[4] data on HSR and other forms of transportation from two sources: the railway yearbook series and Department of Transportation reports. Table A8 reports selected ridership data on most of the HSR lines in service from 2009 to 2012, collected from China Railway Yearbooks. We observe a clear increasing trend in ridership for all the lines listed that have more than one year's ridership data, showing people's gradual acceptance of HSR as a new form of transportation. Among different lines, the most heavily used ones are median-to-short lines connecting two central cities, such as the Beijing-Tianjin, Shanghai-Nanjing, and Guangzhou-Shenzhen lines.

Turning to the total ridership data across different forms of transportation, presented in Figure A1. As noted, more than 85% of the trips were made by cars, buses, and coaches,

[3]According to the Chinese Statistics Bureau, in 2013, the average monthly income of urban residents is ¥2462, residents with average monthly income over ¥4700 are categorized into the high-income group.

[4]Ridership is defined by the National Statistics Bureau of China as the total number of trips made on a particular kind of transportation device. For road ridership, it only includes paid road trips (coaches, etc.), and excludes self-driving trips. For railway, the trips can start and end at any stops, including both destinations.

followed by somewhat more than 10% using the railway. Air and water transport services were relatively less used, accounting for less than 2%. From 2010 to 2014, we observe a steady increase of HSR ridership, from 300 million in 2010 to 830 million in 2014. Over the course of HSR expansion, few changes can be seen in the percentage of passenger trips made by air or water. But we do observe a slight drop in the proportion of passengers carried by conventional railway from 8% to a little less than 7% and by road from 87% to 86%, although the number of railway and road ridership increases steadily. This evidence is consistent with the view that conventional railways face the strongest competition from HSR, most likely because a few of the services on some conventional railway lines are cut when the parallel HSR starts operation [43] and the fact that conventional railway and HSR are closer to each other in the fare/time-cost trade-off spectrum.

To put these numbers into perspective, we can compare the transportation pattern in China to that in the US. In 2014, the total HSR ridership in China was 830 million, more than the combined ridership by air and intercity rail/Amtrak of about 698 million in the US [5]. On the contrary, the total passenger-miles count on US highway (excluding private passenger cars) doubles its Chinese counterpart[6]. In a relative term, HSR appears to be twice as important as air and intercity railway combined in the US.

3. Conceptual framework

The main focus of this paper is to examine the impacts of improved cross-city transportation on specialization patterns across cities. Conceptually, when we think about the differential impacts of HSR connection on different industries, the reliance of these industries on face-to-face contact is essential. By facilitating face-to-face interactions of people, particularly skilled workers, HSR effectively reduces the unit cost of production in communication-intensive sectors, either through productivity boost brought by intense knowledge sharing or through enhanced remote sourcing.

To formalize the intuition on the general equilibrium effects of an improvement in cross-city passenger transportation, I begin with a simple model of labour sourcing based on Grossman and Rossi-Hansberg (2008) [29][7]. In this model, a tradable final good is produced using multiple tasks, and the production of a task can take place in one city while using the technology from another city, subject to communication costs between them. Therefore, a reduction in face-to-face communication cost through HSR connection allows a high productive city to source more tasks from other cities. Some quantitative predictions are generated from the model. The model predicts that the benefits of HSR connection work through enhancing the passenger access to other cities, which leads to employment growth proportional to the change in HSR-induced market access measure, approximated using an

[5] The data on passenger travel in the US are obtained from the National Transportation Statistics by the Bureau of Transportation Statistics. In 2014, the total passenger-miles on air were 607,772 million, and the average length of travel by air is 1,440 miles in 2013, wich translates into 679 million trips made by air. Similarly, we obtain a ridership of 19 million on intercity rail/Amtrak.

[6] In 2014, the total passenger-miles on highways that excludes passenger cars are 1,492,801 miles in the US. In China, total road ridership is 19,082 million, with an average length of 39.14 miles, which adds up to 744,198 million. The information is obtained from the Department of Transportation annual report.

[7] Details of the model are reported in Appendix A.

average of other cities' GDP inversely weighted by the bilateral costs of passenger travel. One could interpret an increase in this "passenger market access" not only as better chances in labour sourcing, but also as better access to other cities knowledge or customer pools.

The model features a single sector, where an improvement in market access translates directly into growth in aggregate city employment. To further evaluate the implications of HSR-induced market access on urban specialization, it could be extended to a multisector one. Think of the simplest case with two different industries, one interactive and one non-interactive; and two different sets of tasks, interactive and non-interactive. The production of the final good in the interactive industry depends on interactive tasks only, and vice versa. A reduction in passenger travel cost reduces the unit cost of interactive tasks production, to an extent that is proportional to the growth in passenger market access. With free final goods and labor mobility across cities and industries, HSR connection leads to a shift in comparative advantage towards the interactive industry, and ultimately to relatively higher growth in this industry in connected cities.[8]

Based on this line of thinking, I derive the following empirically testable hypothesis:

Hypothesis *HSR connection will lead to relatively higher growth in employment of communication-intensive industries in connected cities. And the benefits of HSR connection will appear as improved passenger access to other cities.*

To obtain a sensible measure of industry-specific dependence on face-to-face interactions, I consider the types of tasks required in each industry following Autor, Levy, and Murname (2003) [5]. They divided the required task contents in a particular industry into the following five categories: routine manual, routine cognitive, nonroutine manual, nonroutine analytical and nonroutine interactive; further, they came up with a measure of task intensity of these five types across 140 consistent census industries. According to their analysis, industries high in nonroutine analytical and interactive tasks involve more abstract thinking, problem-solving and complex communication activities. Naturally, a reduction in communication cost will more directly impact the cross-city employment patterns of these industries, other than the ones that focus more on manual or routine cognitive tasks.

The impacts of HSR on city economic outcomes are realized through improvements in the accessibility of connected cities to other cities as a result of a faster and more convenient means of transportation. Therefore, we should expect larger impacts for cities that are connected to a greater number of more prosperous destinations. In my subsequent empirical analysis, I will first check how direct connections to the HSR network lead to employment growth across different industries, before exploring the relationship between HSR-induced market access changes and urban specialization patterns.

[8]It is noted that the aggregate cross-industry relocation of employment will depend on the equilibrium relative price of the two final goods. Therefore, the model may have clear predictions over the absolute effects of HSR on the interactive industry employment in connected cities. However, to generate an expansion in interactive industry as a whole, we can introduce further assumptions on non-homothetic preferences or international markets to regulate the possible relative price changes.

4. Empirical Specification

4.1. Data

Prefecture-level socioeconomic data are drawn from China City Statistical Yearbooks from 2000 to 2013 and China Regional Economic Statistical Yearbooks from 2000 to 2011, since many of the variables are missing for years before 2000. The City Statistical Yearbook series report prefecture-level passenger ridership and volumes of goods transported by different modes of transportation, GDP, population, employment in 18 sectors[9], average wage, local government revenue and expenditure, total number and revenue of industrial firms, total number and sales of retail firms and a variety of city level infrastructure measures. The Regional Economic Statistical Yearbooks report statistics on prefecture-level housing prices. In my analysis, I only focus on prefecture-level cities, excluding prefecture-level autonomous regions. Haikou and Sanya are left out of my sample because they are cities on the Hainan Island and their accessibility to all the other cities cannot be changed by HSR connection easily. Some key socioeconomic data are missing along the time series for a few cities, which leaves me with 278 cities to work with throughout most of my analysis. Table A1 summarizes the source, year range, the total number of observations, number of cities with at least one year of observation and number of cities without missing values along the time series of all the outcome and control variables I use in this paper. Since I control for average city and provincial level GDP and population growth in the past three years throughout my analysis, I am effectively using only outcomes from 2003 to 2013 as my dependent variables.

The yearbooks also report ridership data on railway, road, air and water. Ridership is defined as the number of paid trips made on each form of transportation. Self-drive trips and trips on public transportation are not included in road ridership count. It should be noted here that ridership here is not limited to intercity ridership by definition. Road ridership could include coach trips made across towns or villages within the same prefecture city. But we have reasons to believe that most parts of the ridership come from intercity trips, especially for railway and air travel, as a city typically has only one main railway station and airport.

A prefecture-level city usually has an urban core (Shixiaqu) that consists mainly of urban residents and surrounding counties with a relatively larger proportion of rural population. For each variable, two separate statistics, one aggregated only to urban-ward (Shixiaqu) level and the other covering the whole area of the prefecture, are reported in the yearbooks. Throughout my analysis, I use the statistics counted at urban ward (Shixiaqu) level of prefecture cities since I am interested in the employment and resources flows across urban areas.[10]

High speed railway (HSR) lines are defined as railway lines running at an average speed of 250km/h or more, or passenger-dedicated-intercity-lines running at an average speed of

[9]Detailed descriptions of these sectors, as well as the comparison with NAICS 2-digit industries, are reported in Table A3

[10]Although I stick to statistics counted at the urban core level. This level of aggregation is not available for some variables, such as ridership on variables modes of transport, and I have to use the one counted at the whole prefecture level.

200km/h or more[11]. As the end of 2014, there were 43 HSR lines in operation, with a total mileage of 11152 km (Table A4). Most information on the Chinese HSR system, including construction starting date, open date, length, designed speed and ridership on selected lines, is obtained from the Major Events, Finished and Ongoing Projects sections in the China Railway Yearbooks from 1999 to 2012. For a small proportion of lines that are opened in 2013 and 2014 and future HSR lines in plan, this information is not available from the most updated (2012) railway yearbook, so I have to rely on official news published on http://news.gaotie.cn as well as other online news sources[12] I doublechecked the information on the stops along each existing line from the official railway service website (www.12306.cn). Geo-referenced administrative unit data, as well as conventional railway routes, are obtained from the ACASIAN Data Center at Griffith University in Brisbane, Australia. Highway networks data of China in 2000, 2002, 2003, 2005, 2007 and 2010, from Baum-Snow et al. (2015) [9] ,are kindly shared by the authors.

To check other mechanisms at work, I also bring in patent application data in China from the SIPO (State Intellectual Property Office) as a proxy for innovation activities within a city. In China, patents are divided into invention and utility-model patents, with the former category being more stringent and lasts longer.

4.2. Definition of market access variables

As briefly discussed in section 3, the benefits of HSR connection appear as improved passenger access to other cities. To better capture the treatment effects of HSR connection, I introduce measures of market access and examine the impacts of market access growth, induced by HSR or highway, on cross-city transportation and specialization patterns.

In practice, the market access of city k is defined as $MA_k = \sum_{j=1}^{N} \tau_{kj}^{-\theta} X_j$, where τ_{kj} is the travel cost between city k and j, and X_j is the GDP of city j [13]. Intuitively, a city enjoys higher market access if it boasts lower transportation cost to larger cities. τ_{kj} can be reduced through the expansion of transportation infrastructures, such as HSR or highways.

I use four market access measures throughout my analysis:

(1) Market access measure that captures both highway and HSR network expansion (MAall)

(2) Market access measure that captures HSR network expansion only (MAHSR). Given the highway network of each year, the changes in market access purely driven by HSR network expansion are counted.

(3) Market access measure that captures highway network expansion only (MAhighway). Assuming that there is no change in HSR network, I calculate the changes in market access compared to the base year (2000) purely attributable to highway network changes.

(4) Market access measure that captures indirect HSR network expansion only (non-connection-induced market access, NCIMA): For a city i connected in year t, the market

[11] Major technical stipulation on railway, the Ministry of Railway, 2012.

[12] Information on the construction starting date, designed speed and length for lines that started construction before 2012 is also included in the Railway yearbooks major events section. But I need to rely on online news sources for their exact opening date.

[13] This specification of market access can be derived from a labour sourcing model described in Appendix A.

access changes for it using a counterfactual HSR network that bypasses city i but is otherwise the same as the real one. This measure considers only HSR expansion.

(5) Market access measure that captures the impacts of HSR on short distance trips (less than five hours) only (MAHSRless5). Only changes in bilateral travel costs for trips less than five hours are counted in this measure. This measure considers only HSR expansion.

These different measures allow me to examine differential impacts of different sources of improvements in intercity connectivity. We can look at the responses of ridership and economic outcomes to market access induced by highway and HSR connection separately.

The calculation of the market access variable requires the construction of a time-varying transportation cost matrix, τ_{km}, for each city pair. In my definition, I allow τ_{km} to incorporate both time and fare cost in travel. Conventional roads, highways and HSR represent different fare-time cost combinations and passengers face tradeoffs between fare and time cost. The upgrade from conventional roads to highways and to HSR changes τ_{km} by offering passengers alternative options to travel. The information on highway network is obtained from Baum-Snow et al. (2015) [9], and the timing of HSR network expansion is shown in Table A4.

We have to rely on a few assumptions to construct τ_{km}, and the details are reported in Appendix E.

In Figure 2, I plot the distribution of log(MAHSR), the market access measure that considers only HSR expansion, both across cities and through time. The left graph presents the distribution of log(MAHSR) over all the observations. And the right graph shows the distribution of the residuals of log(MAHSR) conditional on city fixed effects. These graphs give us a basic idea on the pooled and within city variation of log market access. Meanwhile, the lower panel of Table 1 shows the mean and standard deviation of five logged market access measures from 2001 to 2014. The growth in HSR-driven market access during this period is roughly 2.4% and the growth in non-connection-induced market access is about 1.9%.

4.3. Difference in differences specification

In this section, I explore the aggregate impacts of HSR connection by regressing my outcome variables on a HSR connection dummy. The baseline estimation strategy is a difference in differences specification of the form:

$$ln(y_{it}) = \alpha_i + \beta_{rt} + \gamma * Connect_{it} + Controls_{it} + \epsilon_{it} \tag{1}$$

where y_{it} is an outcome of interest of city i within region r in year t, α_i is a city fixed effect, β_{rt} is a region[14] by year fixed effect, and $Connect_{it}$ is an indicator of whether city i of region r was connected to HSR in year t. The error term ϵ_{it} is clustered at the city level. Standard errors allow spatial dependence decaying in distance as in Conley (1999)[15].

To test hypothesis 1, I divide the total employment into four groups: tourism-related employment, which includes hotels and catering services and wholesale and retail trade; skilled employment, which includes finance and insurance, real estate, information, business services, scientific research and technical services and education; other service employment,

[14] Cities are divided into 8 regions (Northeast, Northern Costal, Eastern Coastal, Southern Costal, Southwest, Northwest, the middle reaches of Yangtze River, the middle reaches of Yellow River.

and other non-service employment. Apart from employment, I also look at other aggregate outcomes at city level, which include GDP, housing price, total fixed investments, retail sales, and total patents application.

The identifying assumption of difference-in-differences estimations is the parallel trends of outcomes between HSR-connected cities and the other cities should there be no HSR. However, if an HSR placement decision is based on past growth and expected future growth, this assumption may possibly not hold in reality. To mitigate this problem, I restrict my sample to 172 cities that are either connected or planned to be connected by HSR by 2020.

The primary identification challenge is not whether a city is connected to the HSR network, but rather what factors determine the timing of its connection. Several idiosyncratic factors appear to influence the opening time of HSR lines. First, as is evident from Table A4, the construction work on many of the existing lines (12 out of 45) began in 2005, following the passage of the Mid-to-Long Railway Plan in 2004. The timing of when each line opens is determined by the construction progress, which largely depends on engineering difficulties. In Appendix F.1., I exploit variation in HSR connection timing purely driven by engineering difficulties and find similar effects on railway ridership, service and private employment as in the main regressions.

Another obstacle to identification is the possibility of simultaneous investments in other areas. Local governments may take HSR connection as a new engine for local economic growth and invest heavily in related projects. This government spending would create jobs that can be mistakenly interpreted as the direct impacts of HSR connection. At this stage, to deal with these concerns, I control for local government spending, other infrastructure measures such as the length of roads above a certain standard, the total area of new urban roads built, the area of green land, the number of public facilities such as theatres, hospitals, and public libraries. [15] Controlling for region-by-year fixed effects should be able to take care of any region-specific yearly shocks to local economic conditions. Apart from that, I also control for average city and province GDP and population growth for the past three years, as well as interactions of year dummies with distance to regional central cities for fear that the HSR connection decisions of a line connecting multiple nearby cities are correlated with temporary regional shocks at different levels, and that geographical centrality is both correlated with economic growth and connectivity improvement.

To check the parallel trend assumption, I run a variation of equation (1), controlling for the leads and lags of the initial connection dummy.

$$ln(y_{it}) = \alpha_i + \beta_{rt} + \sum_{m=1}^{3} \gamma_m FirstConnect_{i,t-m} + \sum_{n=0}^{4} \gamma_n FirstConnect_{i,t+n} + Controls_{it} + \epsilon_{it} \quad (2)$$

where $FirstConnect_{it}$ is a dummy variable indicating whether a city is first connected to

[15] A potential problem with including these controls on the right-hand side of the regressions is that government spending and investment itself is endogenous—a government expecting better growth prospects invests more with or without HSR. However, if the correlations between the unobservable and government spending and that between HSR connection and government spending are both positive, then the coefficients on HSR connection or market access are underestimated by controlling for local government spending.

the HSR network in year t. It switches to 1 only if the HSR line connecting city i is opened in year t. $FirstConnect_{i,t-m}$ is its m-th lag, and $FirstConnect_{i,t+n}$ is its n-th lead. Controlling for leads allows me to examine the pre-HSR effects of future railways as a placebo test and helps to disentangle anticipatory effects from actual connection effects. Controlling for lags enables me to trace the treatment effects in the years after initial connection. In reality, we expect anticipatory effects to be relevant for some outcome variables but not for others. For example, housing prices can respond positively to the HSR connection before the railway becomes operative because news about a future HSR station is usually known about five years before it opens and people make investment decisions based on this information. However, HSR-induced changes in employment are expected to be observed only after actual HSR lines are in operation since people can only benefit from HSR for travelling and commuting after it is in operation.

4.4. Market Access Approach

As briefly outlined in the conceptual framework, the benefits of HSR connection appear as improved passenger access to other cities. In this section, I use HSR-induced market access measures as the main independent variables to capture finer variations in the treatments of HSR connection.

One major empirical advantage of this market access approach is that it allows me to exploit the variation in a city's access to other cities driven by HSR expansion but has nothing to do with its own HSR placement. In short, an unconnected city can still benefit from HSR connection if a city close to it receives connection, which allows its passengers faster trips to other destinations through transfers. It effectively deals with the identification challenges of both endogenous route placement and simultaneous investments. In my main empirical specification, I examine the response of outcomes of interest to increase in a non-connection-induced market access for both connected[16] and unconnected cities.

Another benefit of this approach is that it enables me to examine differential impacts of different sources of improvements in intercity connectivity. We can look at the responses of ridership and economic outcomes to market access induced by highway and HSR connections separately. Another market access measure worth considering is the one that captures HSR impacts on shorter trips only. As I do not explicitly account for the changes in air travel costs during this period, I could be overestimating the effects of HSR on connectivity if passengers prefer air travel to HSR over longer-distance trips. Focusing on the impacts of HSR-induced market access changes for short distance trips only helps to alleviate the concern, as HSR holds strong advantages over air travel for shorter trips.

Throughout my analysis, my preferred specification is:

$$ln(y_{it}) = \alpha_i + \beta_{rt} + \gamma * NCIMA_{it} + \theta * MAhighway_{it} + Controls_{it} + \epsilon_{it} \tag{3}$$

where y_{it} is an outcome of interest of city i within region r in year t, α_i is a city fixed effect, β_{rt} is a region by year fixed effect, $NCIMA_{it}$ and $MAhighway_{it}$ are non-connection-induced market access and highway access respectively.

[16]For a connected city, its non-connection-induced market access is defined as its access to other cities through a hypothetical HSR network which is the same as the actual one except for bypassing this city.

Two concerns over identification using the non-connection-induced market access measure may arise here. First, apart from basing the HSR placement decision on the economic conditions of each city, the railway authority might design the network expansion plan to maximize the market access of connected cities intentionally and target at prosperous regions first. So "better located" cities that are close to other developed cities might experience larger increase in market access that may or may not depends on its own connection. Thus, throughout my analyses, I control for the past average city and provincial GDP as well as the interaction terms of year fixed effects and distance to provincial capital cities. Another challenge for identification is the common regional shocks that might jointly affect a city's economic outcome and the HSR placement of nearby cities. It is less of an issue after controlling for region-year effects and past city and provincial average GDP. Furthermore, I also adopt a specification similar to equation (2), by including the leads and lags of the increments in non-connection-induced market access (NCIMA) as main independent variables. The formal specification is:

$$ln(y_{it}) = \alpha_i + \beta_{rt} + \sum_{m=1}^{3} \gamma_m \Delta logNCIMA_{i,t-m} + \sum_{n=0}^{4} \gamma_n \Delta logNCIMA_{i,t+n} + Controls_{it} + \epsilon_{it} \quad (4)$$

Similarly, this specification helps us to check if the growth in the non-connection-induced-market-access (NCIMA) measure correlates with the trends of a variety of outcomes prior to the year when the actual increase in market access takes place. I plot the coefficients on the leads and lags of $\Delta logNCIMA$ in Figures 4 and 7.

4.5. Event-study for non-connection-market access

A potential challenge for identification using my non-connection-induced market access (NCIMA) measure is that some cities closer to each other may be able to drag HSR lines to their region, and their incentive and capacity of lobbying is correlated with their growth prospect. Controlling for region-specific year fixed effects may not be able to fully address this issue as the collective bargaining might occur at a smaller geographical scale. So apart from controlling for past provincial GDP, I also rerun my non-connection-induced market access regressions by including the leads and lags of increments in market access. If regions with better growth prospects are more likely to get HSR placement, then we should expect positive coefficients for leads of market access increments in the cities in these regions, no matter whether they are connected or not. The specification is:

The results are shown in Figures A2 and A3. We do not observe large differences in the pretrends across cities that experience higher and lower non-connection-induced market access growth, and there are clear trend breaks around the time when a city actually experiences changes in market access.

5. Estimation Results

5.1. Transportation patterns

I start by examining the effects of HSR connection on passenger transportation patterns before studying its implications on economic outcomes. The upper panel of Table 2 presents

the difference-in-differences estimation results on the usage of different modes of transport. Controlling for city and region-by-year fixed effects, it is observed from Column (1) that HSR connection significantly increases railway ridership by 18%. The coefficient of total ridership on all forms of transportation (Column 4) is a smaller 9.6%) but still positive and significant, suggesting that HSR creates extra demands for transportation, and the poaching of ridership from other forms of transport is small. From Columns (2) and (3), it is clear that HSR connection leads to a small positive effects on road ridership, but a large 12% drop in the number of passengers who travel by air, which indicates that travelling by HSR is a close substitute for air travel, but not road transportation. It is also worth noting that no significant impacts of HSR are observed on the volume of goods transported by railway (Column 5), confirming our intuition that HSR is bringing changes to intercity passenger travel cost, but not goods trade cost.

A back-of-the-envelope calculation suggests that HSR brings in about 901 million extra passengers since its inception until 2013, taking into consideration the substitution and complementarity relation between HSR and other forms of transport [17]. To put this number in perspective, my HSR ridership data on the main existing lines (Table A8) suggests that the aggregate HSR ridership on 11 major lines (out of 28) in 2012 is 320 million, and the aggregate annual HSR ridership data reported by the Department of Transportation is 300 million in 2010, 420 in 2011, 470 in 2012 and 602 in 2013. It is observed that the estimated aggregate increase in total ridership is about half the size of the aggregate HSR ridership during this period. The difference could either be attributed to the substitution between HSR and other means of transport, most notably conventional railway and air travel, or due to the fact that in this simple difference-in-differences design, we assume cities with no HSR station to be untreated while in reality, they could still benefit from HSR connection in other cities by transferring. But without a doubt, such sizable newly generated passenger flows, with a large proportion of business travellers, indicate large improvement in the communication and economic ties across Chinese cities, yielding large impacts on local economic outcomes possible.

The lower panel of Table 2 reports the responses of passenger travelling patterns to a city's non-connection-induced HSR and highway market access changes. The estimates are largely consistent with our intuition about the substitution and complementarity between different modes of transportation. One percent increase in HSR-induced market access[18] translates into an 8 percent increase in railway ridership, 2 percent drop in road transportation, and 7.4 percentage drop in air travel, while a one percent increase in highway-induced market access leads to 5.7% growth in road ridership and 4.2% drop in railway ridership and do not seem to affect air travel negatively. In sum, both HSR and highway expansions increase total passenger ridership but the estimated elasticity is larger for highway market access.

Table 3 reports how ridership on different forms of transportation responds to other

[17] My estimation suggests that cities with HSR connection experience 9.6% increase in total ridership. Using 2008 total ridership of all the cities with HSR connection by 2013 (9387 million) as a benchmark, this translates into 901 million new ridership.

[18] Summary statistics on different measures of market access are reported in Table 1. From 2001 to 2014, the average NCIMA grows by 2.23% as a result of HSR expansion, and the average MAhighway grows by 1.86% following concurrent highway expansion.

market access measures, including the market access measure that incorporates both HSR and highway expansions (MAall), and that limited to short distance trips. For each variant of the market access measures apart from MAhighway, the regressions control for both the connection dummy and highway market access (MAhighway), to show that the main results on the relationship between market access and employment growth are not driven by direct connections *per se* or simultaneous highway expansion. It is worth noting that for the market access measure capturing only HSR impacts over short distance trips (MAHSRless5), the estimated elasticities are in general larger in magnitude, showing that connectivity improvements are more important when they decrease the travel costs between cities that are closer to each other. Interestingly, the estimated elasticity is six times larger in absolute value than the one in baseline for air travel ridership, but only less than doubles for railway ridership. A story consistent with this observation is that the substitution between HSR and air travel is strong over shorter-distance trips: HSR expansion at a closer range leads to a much larger drop in air travel ridership. For longer distance trips, civil aviation faces little competition from the newly available HSR service so the increase in HSR ridership is more likely to be newly-generated passenger flows. However, even for short distance trips where the substitution between HSR and air travel is stronger, the effect of a boost in market access on total ridership is large and positive, with an estimated elasticity of 8.2%. Piecing together the evidence, it is not hard to conclude that the estimated impacts of different aspects of market access changes echo with our intuition on the ways different modes of transport compete with each other and the qualitative evidence shown in section 2. Therefore, we should have more confidence in the validity of the "market access" approach and the estimation results on the impacts of HSR-led market access changes on economic outcomes of interest.

5.2. Specialization patterns

In my conceptual framework, I hypothesize that industries that rely more on communication and interpersonal skills will benefit more from HSR that reduces the cost of face-to-face interaction and communication across cities. In this section, I will directly test the implications of HSR connection on the specialization patterns across cities. I will first roughly divide total employment into skilled service employment, tourism-related service employment, other service and non-service employment and examine the impacts of HSR on them separately. I will then carry on a more detailed analysis on the effects of HSR on 16 industries and rank the estimated effects according to the cognitive content of each one.

The estimated coefficients on connection dummy are presented in Table 4. A city is estimated to experience a statistically significant 7% higher growth in employment after being connected to HSR compared to the unconnected. Columns (2) to (4) report the results on employment of four subcategories[19]. It is certain that the aggregate employment effects are largely driven by growth in tourism-related employment, which grows by 13% following the HSR connection. The effect on skilled employment (4.8%) is moderate, lower than that on non-service employment such as manufacturing, utility, and construction. One possible explanation is that the HSR connection effects on non-service sectors, such as manufacturing

[19] Total employment is divided into tourism-related employment (retail and wholesale, hotel and restaurant), skilled employment (FIRE, IT, business services, research and technical services and education), non-service employment (agriculture, manufacture, construction, utilities) and other service employment.

and construction, are driven by subsequent investments in public infrastructure and fixed assets. Market access results reported in the lower panel of Table 4 supports this interpretation. One percent increase in NCIMA leads to 2.5 and 3.9 percent growth in skilled and tourism-related employment, both higher than the estimated elasticity on aggregate employment, but exerts almost no effects on other non-service employment. It suggests that the benefits of better passenger access to other cities are mostly limited to skilled and tourism-related employment. Therefore, the significant positive impact of direct HSR connections on construction and manufacturing comes partly from subsequent investments accompanying HSR connections. On the contrary, highway-led market access growth leads to growth in non-service employment, most notably manufacturing, construction and utility, which is consistent with the idea that highway expansion has larger effects on trade cost, which benefits the tradable sectors more.

Table 5 reports the estimation results using alternative market access measures, controlling for both the direct HSR connection and highway market access expansion. The evidence presented here is consistent with the main effects: both HSR-driven and total market access growths have positive impacts on total employment, and the coefficients are largest for tourism-related sectors, followed by skilled sectors. Controlling for direct connection dummy does not take away much of the explanatory power of this non-connection-induced market access (NCIMA), compared to the results presented in Table 4, which suggests that this measure captures mostly market access improvements orthogonal to cities' own connections. The estimated elasticities are also larger in magnitude for the market access measure that captures only HSR impacts over short distance trips (MAHSRless5), which is consistent with the transportation patterns results. It shows that the connectivity boost across cities that are relatively close to each other plays a more important role in explaining both variation in railway ridership and employment growth.

Figure 4 shows event study results on employment of different categories, specified by equation (4). In these graphs, I plot the estimated elasticity on the leads and lags of the changes in non-connection-induced market access measure (NCIMA). It is clear that the coefficients on the leads of NCIMA are close to zero for all four categories of employment, which supports the parallel trends assumption: the boost in HSR-induced market access does not correlate with previous trends in employment across different sectors. Meanwhile, all four categories of employment grow gradually following HSR-induced market access boost, and the effects are largest for tourism-related and skilled service sectors. The treatment effects of HSR-induced market access are smallest in magnitudes for non-service employment, which is different from the results obtained using the connection dummy.

Subsequently, I present the estimated HSR effects across sixteen different industries in Table 6. It is clear that most industries respond positively to both non-connection-induced market access (NCIMA) and highway induced market access (MAhighway), but there exists considerable heterogeneity across sectors. For instance, manufacturing employment responds significantly to highway-led market access growth but little to HSR-induced one, while employment in finance, retail, medical services, etc. grow significantly as a result of better passenger access through HSR.

To establish a closer link between an industry's employment response to better intercity passenger transportation and its dependence on human interactions, I compare the industry-specific estimated coefficients to Autor, Levy, and Murname (2003) [5]'s measures of the task

requirement of each industry in nonroutine interactive activities, quantitative reasoning, routine cognitive skills, nonroutine and routine manual activities. In their context, routine tasks are defined as procedural, rule-based activities that are codifiable, and nonroutine tasks as those relying a lot on abstract skills such as problem-solving, intuition, persuasion, and creativity. We believe that industries with a focus on nonroutine tasks should benefit more from HSR connection as it improves face-to-face contacts of people across space.

It is evident from the left panel of Figure 5 that there exists a strong positive relationship between the importance of nonroutine interactive cognitive skills and the estimated impacts of better HSR-induced market access across all 16 industries. Industries that require a lot of intuition, creativity and human interactions such as FIRE, IT and business services both have higher measures of nonroutineness and respond more strongly to HSR-induced market access. On the contrary, the correlation between the impacts of highway-induced market access and the industry-specific task interactive task requirement drops to almost zero, as shown in the right panel. It suggests that HSR promotes specialization towards more interactive tasks much more than the highway and shows a clear distinctive role of passenger-dedicated transportation device in shaping cross-city specialization patterns. Similar contrasts can be drawn by comparing the sector-specific impacts of HSR and highway to the manual task requirement of each industry. Contrary to the patterns with interactive task content, we discover in Figure 6 that the estimated impacts of HSR decrease in the reliance of each industry on manual tasks, while the impacts of highway do not differ much across industry-specific manual task intensity. Analogous results on the other two dimensions of task intensity (routine cognitive skills and nonroutine quantitative reasoning skills) are presented in A3 and A4, with similar patterns: the estimated HSR impacts on sector-specific employment increase in nonroutine quantitative skills but decrease in routine cognitive skills.

Finally, I explore deeper into the impacts of HSR on tourism. It is clear that employment in catering and retailing experience huge boosts following the HSR connection. In the left graph of Figure 5, both the coefficients on retailing and catering employment lie above the fitted line, suggesting that there are factors other than the reliance on ”people skills” that contribute to the observed impacts of HSR on these two industries. It would be interesting to account for the contribution of tourism to the observed patterns. In Table A10, I check if the responses of tourism-related economic activities to HSR connection is disproportionally larger for cities with better tourism resources. The China National Tourism Administration (CNTA) has divided major Chinese tourist attractions into five categories based on the code “Categories and Rating Standard of Tourist Attractions”. [20] Here I define the indicator for tourism resources of a city as the number of 5A-class tourist attractions in that city. An additional 5A tourist attraction makes a city's total employment 26% more responsive to growth in market access compared to a city with none, but its retail/wholesale employment 130% more responsive.[21] Similar patterns show up if we compare the results across GDP and

[20]Chris Ryan, Gu Huimin and Fang Meng (2009). “Destination planning in China”. In Chris Ryan and Gu Huimin. Tourism in China: Destination, Cultures, and Communities (1 ed.). pp. 1137. ISBN 9780203886366.

[21]Both ratios are calculated by dividing the coefficients on the interaction terms of tourism resources and log(NCIMA) and the coefficients on log(NCIMA), yielding 0.571/2.187=0.26 and 2.826/2.171=1.3, respectively.

the total sales of wholesale/retail firms. The coefficient on the interaction term of tourism resources and market access is negative for GDP but is substantial and significant for retail sales. An additional 5A site makes the effect of HSR market access 56% stronger compared to the zero 5A site baseline on retail sales.

5.3. Broader Economic Impacts

In the previous sections, I have presented evidence on the impacts of HSR in shifting urban specialization patterns towards more communication-intensive and human-centred sectors. But it still remains a question whether or not HSR connections will exert any aggregate effects at the city level to justify the cost of HSR construction. In this section, I take a look at the impacts of HSR on other aggregate outcomes, such as GDP, investments and housing price.

The results are reported in the first column of Table 7. The direct HSR connection effect on GDP is estimated to be 3%, though not statistically significant, which is very similar to the estimates of 2.7% to 4.7% in Ahlfeldt and Fedderson (2010)[2]. Housing price also grow by 3% following the HSR connection. The patterns are similar with market access measures as the explanatory variables, reported in the lower panel of Table 7. One percent growth in non-connection-induced market access leads to significant 2.1%, 3.9% and 1.7% increases in aggregate employment, GDP and housing price. In the meantime, market access growth created by highway expansion only significantly improves GDP with an elasticity of 2.1%, with little effects on employment. It is consistent with the fact that the reduction in trade costs has larger impacts on goods market as opposed to the labour market, and consequently exerts more substantial effects on GDP other than employment.

To check if the GDP growth is largely driven by simultaneous investments, I observe the effects of HSR on the aggregate fixed investments of all the industrial firms[22] in China in column (4). The coefficient on connection dummy is negative and statistically significant, suggesting a decrease in fixed investments following HSR connections. I believe it is mostly due to the fact that the construction of HSR lines and stations leads to a boost in fixed investments before the actual operating of HSR lines. Considering the next row, the total fixed investments of all industrial firms do not respond to HSR or highway driven market access changes, which indicates that the observed GDP growth is not a result of expansion in fixed assets. On the contrary, HSR connection is associated with a significant 10% growth in total wholesale and retail sales, with an estimated elasticity in response to NCIMA growth being a significant 3.4. It confirms that a boost in retailing accounts for a significant part of the HSR-driven growth in GDP.

Finally, it is conceivable that HSR might generate growth in innovation through knowledge spillovers and enhanced scientific collaboration across cities. In the last column of Table 7, I check whether or not HSR connection leads to growth in patenting activities. The coefficient is positive but not statistically significant. However, in the last figure of Figure 7 and A5, I find a strong effect of direct HSR connections on patent applications after 2 to 3 lags, and a positive and significant coefficient on the third lag of NCIMA, as well. A

[22]For industrial firms, the data include all the state-owned firms and all the private firms with industrial output larger than 5 million RMB per year. For wholesale/retail firms, the data include all the firms with annual sales greater than 5 million RMB

plausible explanation is that innovation and patent application take time, and it is natural for the response in patents to take longer to show up.

6. Robustness Checks

6.1. Robustness to alternative parameters and specifications

Table A12 reports the sensitivity analysis of the estimated market elasticity to alternative definitions of market access.

First of all, in constructing the travelling cost matrix, I assume that on average two trips per month are necessary for a task to be sourced, which is quite arbitrary. So I experiment with other required number of trips ranging from 0.5 time per month to 10 times per month. Given the same expansion in transportation networks, the percentage increase in market access induced by HSR should be mechanically larger for more frequent face-to-face contacts requirement. As shown in rows (1)-(4) of Table A12, most of the coefficients on ridership, employment, GDP and housing price remain positive and statistically significant at 0.01 level.

The calculation of market access also requires an approximate of θ, a parameter on the productivity distribution of cities. Following Donaldson (2014) [17], my baseline measure assumes a value of 3.6. For robustness, I adopt two other values of θ proposed by Eaton and Kortum (2002), 8.28 and 12.86. The signs and significance levels of most of the estimates are maintained with different θ.

6.2. Results on cities without an airport

As briefly mentioned, I do not incorporate air travel into my calculation of market access because air travel is less popular in China compared to railway and highway, and it is hard to get reliable information on the fare cost of planes as it fluctuates a lot across time. A potential problem with this oblivion is that I could be overestimating the impacts of HSR on city connectivity if planes and HSR are very close substitutes. Although I have shown that the substitutability of HSR and air travel is limited over mid-to-long distance trips, the increase in HSR ridership far more than compensate for the drop in civil aviation usage, and my results are robust to an alternative market access measure that captures only changes over short distance trips that face little competition from air travel, it makes sense to double-check my main results with a sample of cities that do not have an airport by the end of year 2013. The information on civil airports is obtained from the official website of CAAC (Civil Aviation Administration of China).

As shown in Table A11, The estimated coefficients on railway ridership and total ridership are larger in this small sample, and those on most economic outcomes are slightly lower than the baseline results. But the sign and significance of all the main results are preserved in the subsample of cities without airports.

7. Conclusion

This paper aims to make two primary contributions. First, I evaluate the impacts of one of the largest transportation infrastructure projects in the world, the high speed railway project in China, on local passenger travel patterns and economic outcomes. Second, I

extend Donaldson and Hornbeck (2015)'s [18] market access approach to evaluate different sources of intercity transportation improvements on cross-city sectoral employment patterns.

Despite the abundance of research on the cost of goods transportation and trade patterns, minimal attention has been paid to the cost of moving people or to the implications of it for the labour market. This paper exploits China's high speed railway (HSR) network expansion as a source of plausibly exogenous variation in passenger travel cost across Chinese cities over time. I find that an increase in HSR-induced market access leads to higher growth in industries with higher requirements in nonroutine cognitive skills rather than manual or routine cognitive skills, which is notably distinctive from the effects of highway development.

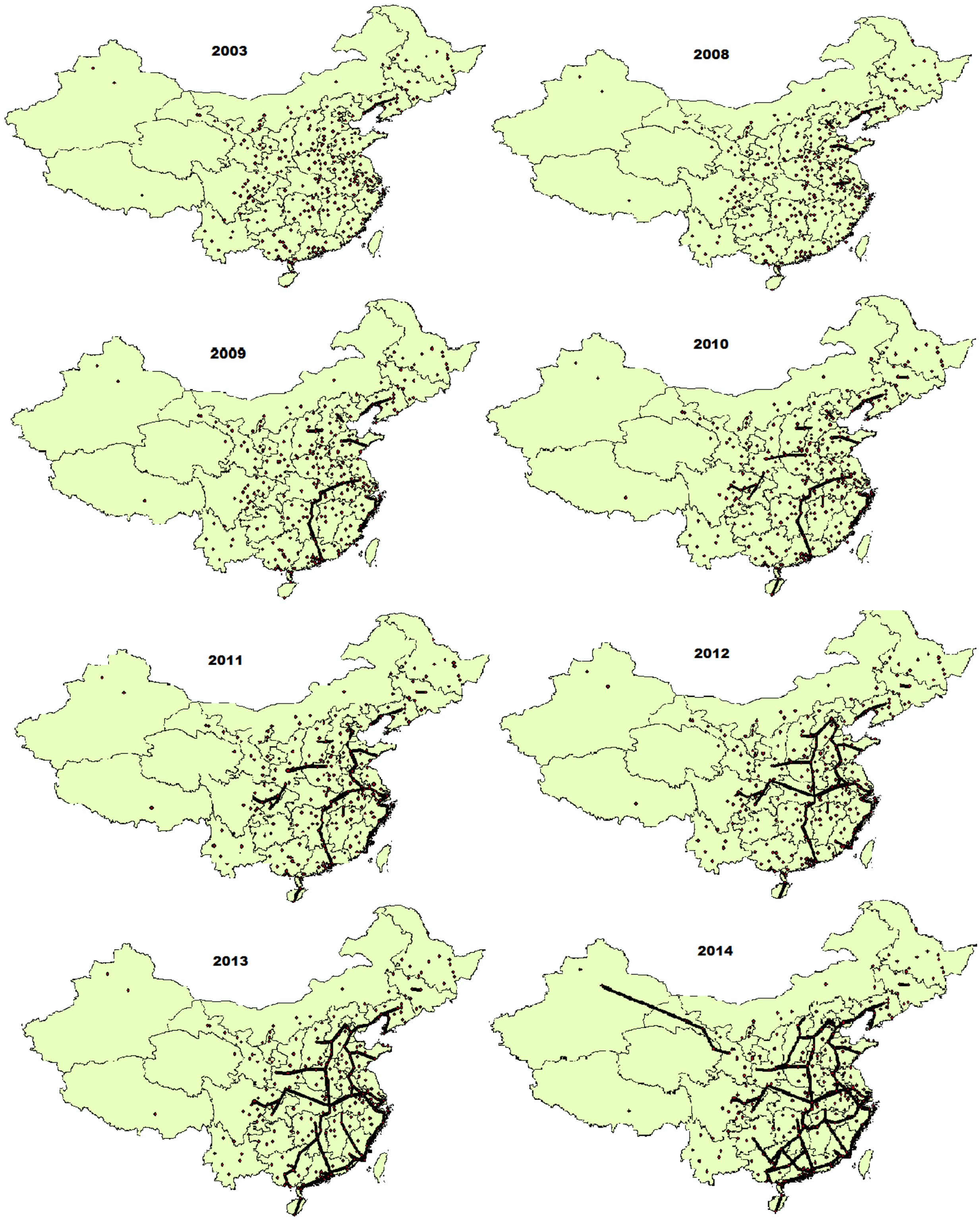


Figure 1: Evolution of HSR expansion from 2003 to 2014

Notes: These figures display the evolution of HSR expansion from year 2003 to 2014. The lines in bold black are lines in use by the end of that year. Each dot represents a prefecture-level city.

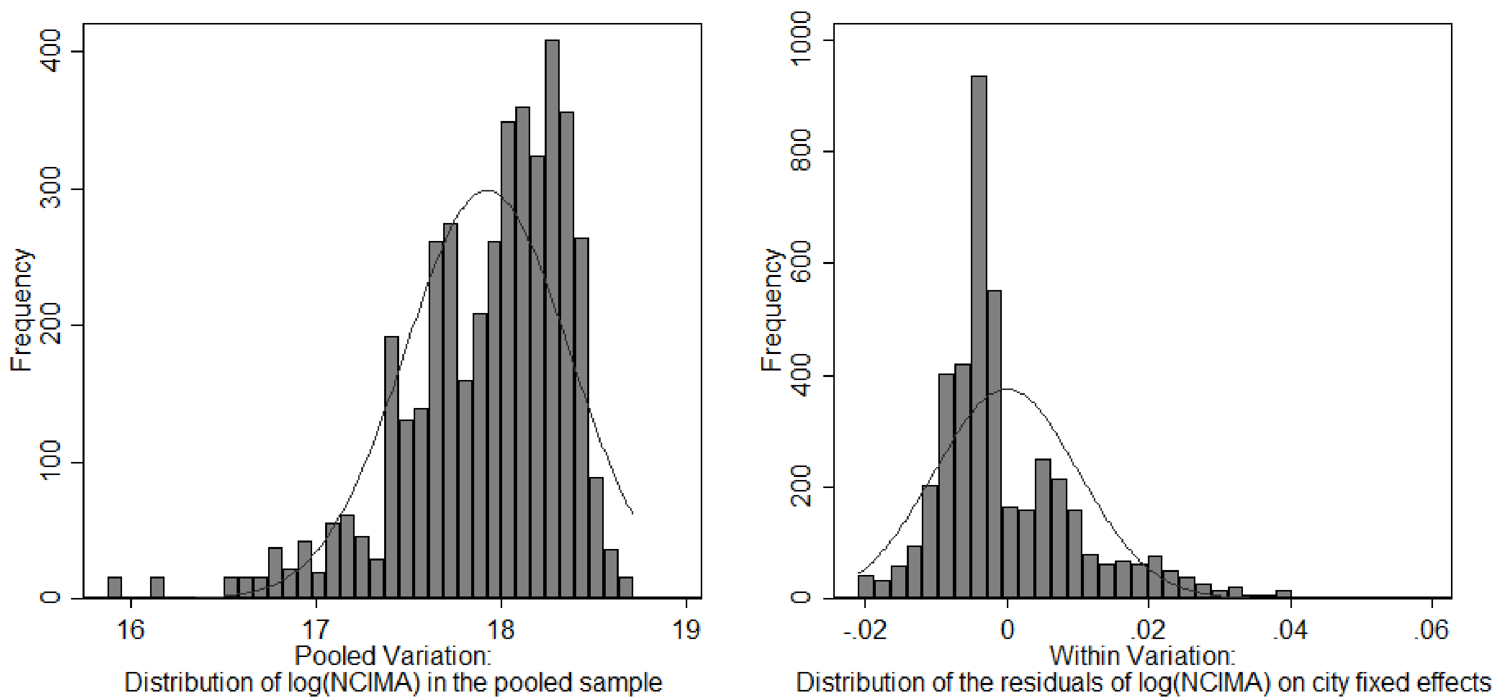
Table 1: Summary Statistics

	Connected		Non-connected	
	2001-2007	2007-2013	2001-2007	2007-2013
Employment growth	0.143	0.539	0.088	0.392
	(0.106)	(0.388)	0.(428	(0.361)
GDP growth	1.718	1.713	1.733	1.79
	(0.945)	(0.563)	(1.64)	(0.688)
Population growth	0.312	0.112	0.223	0.099
	(0.643)	(0.201)	(0.714)	(0.177)
	2001-2006	2006-2011	2001-2006	2006-2011
Housing price growth	0.810	1.399	0.695	1.330
	(0.453)	(0.464)	(0.432)	(0.338)
	2007	2013	2007	2013
Employment	43.71	67.59	14.6	20.04
	(7.16)	(11.79)	(1.41)	(2.05)
GDP	104.36	202.14	28.92	59.59
	(18.13)	(32.66)	(3.71)	(7.45)
Population	196.11	215.87	95.26	103.25
	(23.78)	(26.16)	(6.27)	(6.59)
	2006	2011	2006	2011
Housing price	2633.68	5577.08	1811.75	3712.23
	(1555.28)	(3421.74)	(700.96)	(1435.61)

	Growth in market access measures				
Year	2001	2003	2007	2010	2014
log(MAall)	17.9249	17.9264	17.9338	17.9422	17.9656
	(0.4431)	(0.4453)	(0.4439)	(0.4456)	(0.4420)
log(MAHSR)	17.9187	17.9193	17.9193	17.9274	17.9415
	(0.4452)	(0.4454)	(0.4454)	(0.4475)	(0.4433)
log(MAHSRless5)	17.9187	17.9187	17.9187	17.9239	17.9266
	(0.4453)	(0.4453)	(0.4453)	(0.4464)	(0.4464)
log(NCIMA)	17.9186	17.9191	17.9191	17.9261	17.9374
	(0.4453)	(0.4454)	((0.4454)	(0.4472)	(0.4428)
log(MAhighway)	17.9248	17.9257	17.9332	17.9335	17.9433
	(0.4443)	(0.4450)	(0.4438)	(0.4434)	(0.4439)

Notes: Unit: GDP: 1 billion RMB; Population and employment: 10000 people; Wage: RMB; Housing price: RMB/m^2; Market access: 1 billion RMB. A city is defined to be connected if it is connected by HSR by the end of year 2014. As housing price data only exist until 2011, I look at it over a slightly different time frame.

Figure 2: Pooled and within city variation of log(MAHSR)



Notes: Left: the distribution of log(MAHSR) over all the observations (city*year level). Right: the distribution of the residuals of log(MAHSR) netting out city fixed effects (within city variation)

Table 2: Impacts of HSR connection on transportation patterns

VARIABLES	log railway ridership	log road ridership	log air ridership	log total ridership	log railway goods
connect	0.186***	0.0283	-0.127**	0.0957***	0.106
	(0.0450)	(0.0344)	(0.0541)	(0.0261)	(0.0815)
Observations	1,637	1,746	731	1,773	1,642
R-squared	0.033	0.036	0.044	0.057	0.024
log(NCIMA)	8.096***	-2.095	-7.437**	1.244*	0.696
	(2.447)	(1.629)	(3.559)	(0.737)	(3.713)
log(MAhighway)	-4.213**	5.722**	1.958	3.312*	3.631
	(1.871)	(2.419)	(3.837)	(1.972)	(3.376)
Observations	2,480	2,791	1,175	2,808	2,544
R-squared	0.020	0.020	0.039	0.025	0.015

Notes: Data from the upper table are a panel of 172 Chinese prefecture cities annually from 2003 to 2013 that are connected or planned to be connected by HSR lines by the end of 2014. Connect is a dummy which is zero unless a city is connected to HSR before the end of that year, in which case it takes the value one. Data from the lower table are a panel of 278 Chinese prefecture cities annually from 2003 to 2013. NCIMA is the non-connection-induced-market-access measure. MAhighway accounts for market access changes by highway network expansion only. The dependent variables as listed are logs of railway passenger ridership (not restricted to high speed railway ridership), road ridership, air travel ridership, total ridership and volume of goods transported by railway (in tons). All outcome variables are counted at urban wards (shixiaqu) of prefecture cities. Cities are divided into 8 regions (Northeast, Northern Costal, Eastern Coastal, Southern Costal, Southwest, Northwest, the middle reaches of Yangtze River, the middle reaches of Yellow River. All regressions include city fixed effects and region-by-year fixed effects. Controls include government spending, other infrastructure measures, past city and provincial GDP, and interactions of year dummies with geographical centrality measures. Standard errors, reported in parentheses, are heteroskedasticity robust and clustered at the city level, allowing spatial dependence decaying in distance as in Conley (1999).* significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3: Impacts of HSR on transportation patterns: other market access variables

VARIABLES	log railway ridership	log road ridership	log air ridership	log total ridership	log railway goods
log(MAHSR)	5.610**	0.588	-5.300	1.883*	-1.370
	(2.408)	(1.767)	(3.862)	(1.059)	(3.264)
Observations	2,480	2,791	1,175	2,808	2,544
R-squared	0.024	0.016	0.020	0.017	0.014
log(NCIMA)	7.433***	-2.140	-6.435*	1.212	-0.280
	(2.604)	(1.746)	(3.535)	(0.803)	(3.853)
Observations	2,480	2,791	1,175	2,808	2,544
R-squared	0.026	0.016	0.021	0.017	0.013
log(MAall)	5.895**	0.847	-5.483	2.052*	-1.131
	(2.444)	(1.803)	(4.915)	(1.080)	(3.336)
Observations	2,480	2,791	1,175	2,808	2,544
R-squared	0.024	0.016	0.021	0.017	0.013
log(MAHSRless5)	12.80**	5.075	-41.42***	8.261**	3.072
	(6.346)	(4.787)	(15.96)	(3.832)	(11.28)
Observations	2,480	2,791	1,175	2,808	2,544
R-squared	0.022	0.016	0.029	0.018	0.013
log(MAhighway)	-2.653	4.781**	1.803	3.421**	1.664
	(1.949)	(2.537)	(4.191)	(1.976)	(3.440)
Observations	2,480	2,791	1,175	2,808	2,544
R-squared	0.020	0.020	0.039	0.025	0.015

Notes: Data are a panel of 278 Chinese prefecture cities annually from 2003 to 2013. Each cell corresponds to a separate regression. MAHSR is the market access measure accounting for HSR network changes only. NCIMA is the non-connection-induced-market-access measure. MAall is the market access measure accounting for both changes in HSR and highway networks. MAHSRless5 only accounts for market access changes by HSR when the bilateral travel time is less than 5 hours. MAhighway accounts for market access changes by highway network expansion only. For the regressions with log(MAall), log(MAHSR), log(NCIMA) and log(MAHSRless5) as independent variables, both the direct HSR connection dummy and log(MAhighway) are controlled for. The dependent variables as listed are logs of railway passenger ridership (not restricted to high speed railway ridership), road ridership, air travel ridership, total ridership and volume of goods transported by railway (in tons). All outcome variables are counted at urban wards (shixiaqu) of prefecture cities. All regressions include city fixed effects and region-by-year fixed effects. Controls include government spending, other infrastructure measures, past city and provincial GDP, and interactions of year dummies with geographical centrality measures. Standard errors, reported in parentheses, are heteroskedasticity robust and clustered at the city level, allowing spatial dependence decaying in distance as in Conley (1999).* significant at 10%; ** significant at 5%; *** significant at 1%.

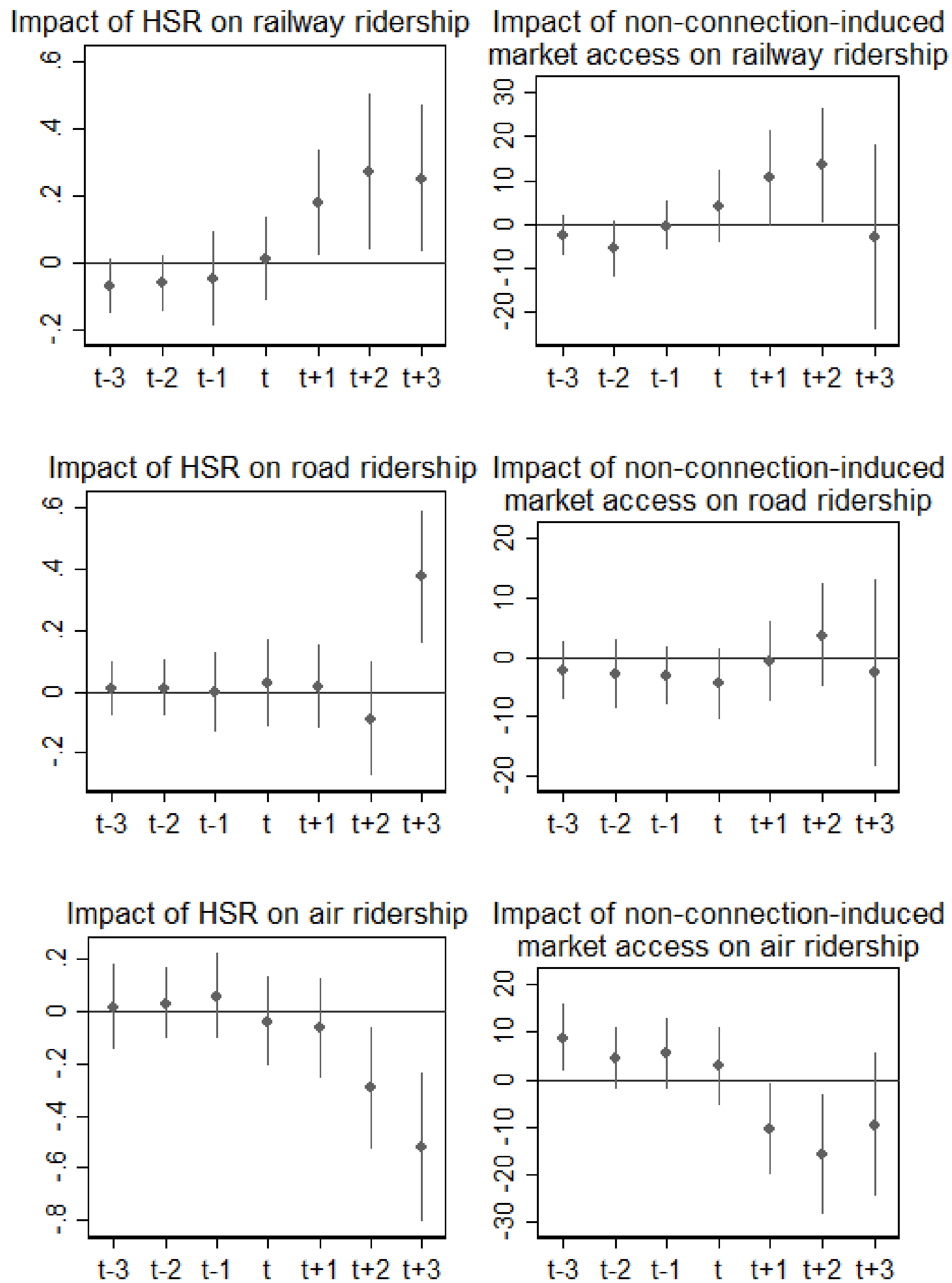


Figure 3: Event study: Transportation patterns

Notes: The figures display the estimated coefficients and 95% confidence intervals in regressions where dependent variables (from up to down) are the logs of railway ridership, road ridership and air travel ridership. The independent variables on the left column are the leads and lags of the initial connection dummy. The independent variables on the right column are the leads and lags of the increments in non-connection-induced market access (NCIMA). For both sets of regressions, the sample is a balanced panel from 2003-2011, as HSR connection information is available only until 2014, and the third lead is a missing value for observations after 2011. For connection dummy, the panel includes 172 Chinese prefecture cities that are connected or planned to be connected by HSR lines by the end of 2014. For NCIMA, the panel includes 278 Chinese prefecture cities.

Table 4: Impacts of HSR on specialization patterns

VARIABLES	log employment	log skilled employment	log tourism employment	log other service employment	log other non-service employment
connect	0.0736***	0.0482***	0.131***	0.0225**	0.0650**
	(0.0123)	(0.0131)	(0.0247)	(0.0123)	(0.0283)
Observations	1,801	1,774	1,795	1,680	1,751
R-squared	0.035	0.030	0.031	0.011	0.029
log(NCIMA)	2.156***	2.466***	3.917***	1.032	0.551
	(0.677)	(0.515)	(1.165)	(0.718)	(1.937)
log(MAhighway)	-0.718	1.686***	-0.401	-0.346	6.393*
	(0.602)	(0.584)	(1.290)	(0.717)	(3.683)
Observations	2,877	2,813	2,847	2,604	2,788
R-squared	0.026	0.022	0.026	0.011	0.047

Notes:Data from the upper table are a panel of 172 Chinese prefecture cities annually from 2003 to 2013 that are connected or planned to be connected by HSR lines by the end of 2014. Connect is a dummy which is zero unless a city is connected to HSR before the end of that year, in which case it takes the value one. Data from the lower table are a panel of 278 Chinese prefecture cities annually from 2003 to 2013. NCIMA is the non-connection-induced-market-access measure. MAhighway accounts for market access changes by highway network expansion only. The dependent variables as listed are logs of total employment, skilled employment (includes IT, FIRE, education, business service and scientific research), tourism-related employment (includes wholesale and retail trade, hotels and catering service), other service and non-service employment. All outcome variables are counted at urban wards (shixiaqu) of prefecture cities. All regressions include city fixed effects and region-by-year fixed effects. Controls include government spending, other infrastructure measures, past city and provincial GDP, and interactions of year dummies with geographical centrality measures. Standard errors, reported in parentheses, are heteroskedasticity robust and clustered at the city level, allowing spatial dependence decaying in distance as in Conley (1999).* significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: Impacts of HSR on specialization patterns: other market access variables

VARIABLES	log employment	log skilled employment	log tourism employment	log other service employment	log other non-service employment
log(MAall)	1.365**	2.328***	3.708***	1.150**	-0.182
	(0.640)	(0.521)	(1.112)	(0.560)	(1.127)
Observations	2,877	2,813	2,847	2,604	2,788
R-squared	0.028	0.027	0.029	0.012	0.015
log(MAHSR)	1.343**	2.281***	3.571***	1.120**	-0.217
	(0.628)	(0.511)	(1.096)	(0.547)	(1.108)
Observations	2,877	2,813	2,847	2,604	2,788
R-squared	0.028	0.027	0.029	0.012	0.015
log(NCIMA)	1.557**	1.969***	2.806**	1.434**	0.795
	(0.653)	(0.508)	(1.119)	(0.567)	(1.094)
Observations	2,877	2,813	2,847	2,604	2,788
R-squared	0.029	0.025	0.028	0.013	0.015
log(MAHSRless5)	9.022***	7.322***	13.46***	6.136	-0.144
	(2.012)	(1.919)	(3.543)	(3.770)	(3.345)
Observations	2,877	2,813	2,847	2,604	2,788
R-squared	0.033	0.026	0.029	0.014	0.015
log(MAhighway)	-0.500	1.937***	-0.0142	-0.258	6.451*
	(0.579)	(0.589)	(1.267)	(0.711)	(3.579)
Observations	2,877	2,813	2,847	2,604	2,788
R-squared	0.021	0.015	0.022	0.011	0.047

Notes: Data are a panel of 278 Chinese prefecture cities annually from 2003 to 2013. MAHSR is the market access measure accounting for HSR network changes only. NCIMA is the non-connection-induced-market-access measure. MAall is the market access measure accounting for both changes in HSR and highway networks. MAHSRless5 only accounts for market access changes by HSR when the bilateral travel time is less than 5 hours. MAhighway accounts for market access changes by highway network expansion only. For the regressions with log(MAall), log(MAHSR), log(NCIMA) and log(MAHSRless5) as independent variables, both the direct HSR connection dummy and log(MAhighway) are controlled for. The dependent variables as listed are logs of total employment, skilled employment (includes IT, FIRE, education, business service and scientific research), tourism-related employment (includes wholesale and retail trade, hotels and catering service), other service and non-service employment. All outcome variables are counted at urban wards (shixiaqu) of prefecture cities. All regressions include city fixed effects and region-by-year fixed effects. Controls include government spending, other infrastructure measures, past city and provincial GDP, and interactions of year dummies with geographical centrality measures. Standard errors, reported in parentheses, are heteroskedasticity robust and clustered at the city level, allowing spatial dependence decaying in distance as in Conley (1999).* significant at 10%; ** significant at 5%; *** significant at 1%.

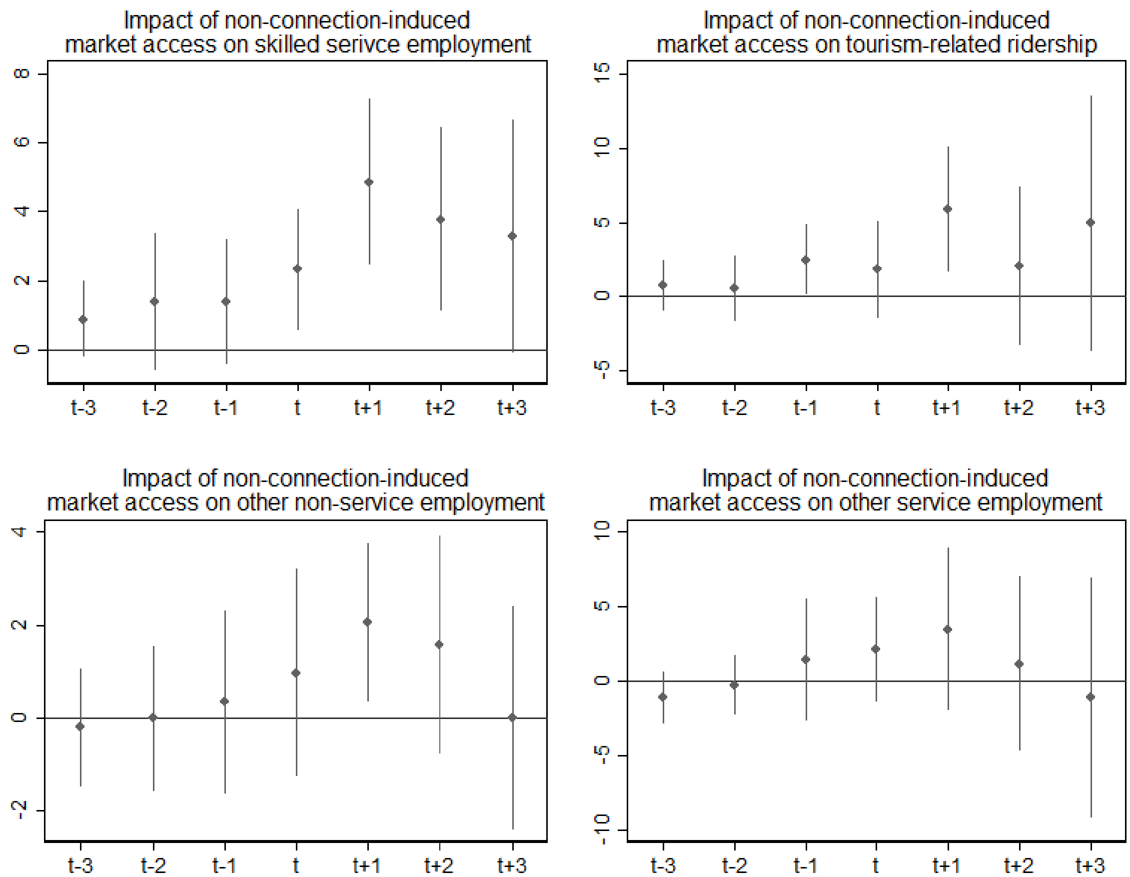


Figure 4: Event study: Specialization patterns

Notes:The figures display the estimated coefficients and 95% confidence intervals in regressions where the dependent variables are log skilled service employment, tourism-related service employment, other service and non-service employment. The independent variables are the leads and lags of the increments in non-connection-induced market access (NCIMA). For all the regressions, the sample is a balanced panel of 278 Chinese prefecture cities from 2003-2011, as HSR connection information is available only until 2014, and the third lead is a missing value for observations after 2011.

Table 6: Impacts of market access across industries

VARIABLES	log manufacture employment	log utility employment	log construct employment	log retail employment	log hotel/restaurant employment	log transport employment	log finance employment	log IT employment
log(NCIMA)	0.599	-0.250	-1.358	3.516**	1.991	-1.556	4.929***	1.794
	(3.541)	(1.515)	(1.771)	(1.425)	(1.536)	(1.240)	(1.077)	(1.927)
log(MAhighway)	12.13**	-0.925	-5.519***	-2.784*	2.027	-1.608	-1.552	-1.068
	(5.892)	(1.446)	(1.839)	(1.528)	(1.425)	(1.164)	(0.962)	(1.623)
Observations	2,796	2,350	2,354	2,621	2,363	2,618	2,362	2,338
R-squared	0.043	0.014	0.009	0.019	0.009	0.009	0.025	0.007

VARIABLES	log real estate employment	log research employment	log public employment	log medical employment	log education employment	log business service employment	log facility employment	log entertain employment
log(NCIMA)	2.327	2.138*	0.399	2.817***	1.628***	1.602	2.678*	2.645***
	(1.717)	(1.160)	(0.739)	(0.553)	(0.487)	(1.827)	(1.466)	(0.839)
log(MAhighway)	2.640	2.962**	0.231	0.477	2.831***	-0.769	1.676	-1.178
	(1.773)	(1.305)	(0.844)	(0.699)	(0.829)	(1.854)	(1.354)	(1.220)
Observations	2,847	2,864	2,269	2,363	2,363	2,557	2,358	2,354
R-squared	0.014	0.010	0.016	0.019	0.027	0.028	0.017	0.008

Notes The dependent variables are logs of sectoral employment across sixteen industries. Detailed definitions of these sectors are reported in Table A3. NCIMA is the non-connection-induced-market-access measure. MAhighway accounts for market access changes by highway network expansion only. All regressions include city fixed effects and region-by-year fixed effects. Controls include government spending, other infrastructure measures, past city and provincial GDP, and interactions of year dummies with geographical centrality measures. Standard errors, reported in parentheses, are heteroskedasticity robust and clustered at the city level, allowing spatial dependence decaying in distance as in Conley (1999).* significant at 10%; ** significant at 5%; *** significant at 1%.

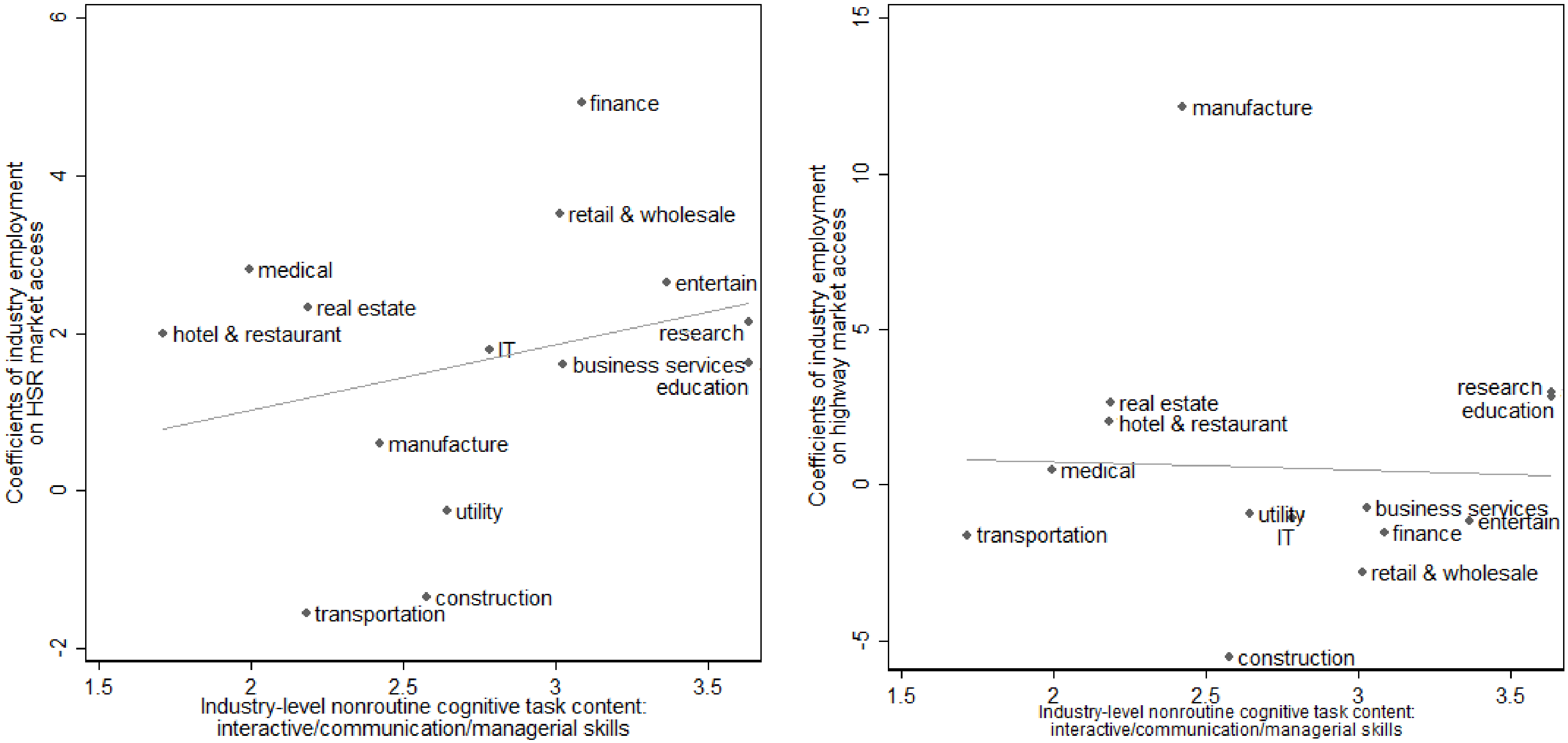


Figure 5: Sector-specific HSR and highway impacts and nonroutine interactive task intensity

Notes:The figures plot estimated coefficients on the impacts of NCIMA and highway-induced market access (MAhighway) on sectoral employment reported in Table 6 against sector-specific nonroutine interactive task intensity (interactive, communiction and managerial skills) reported in Autor, Levy and Murname (2003) [5]. The matching between Chinese industries and the US industries is reported in Table A3.

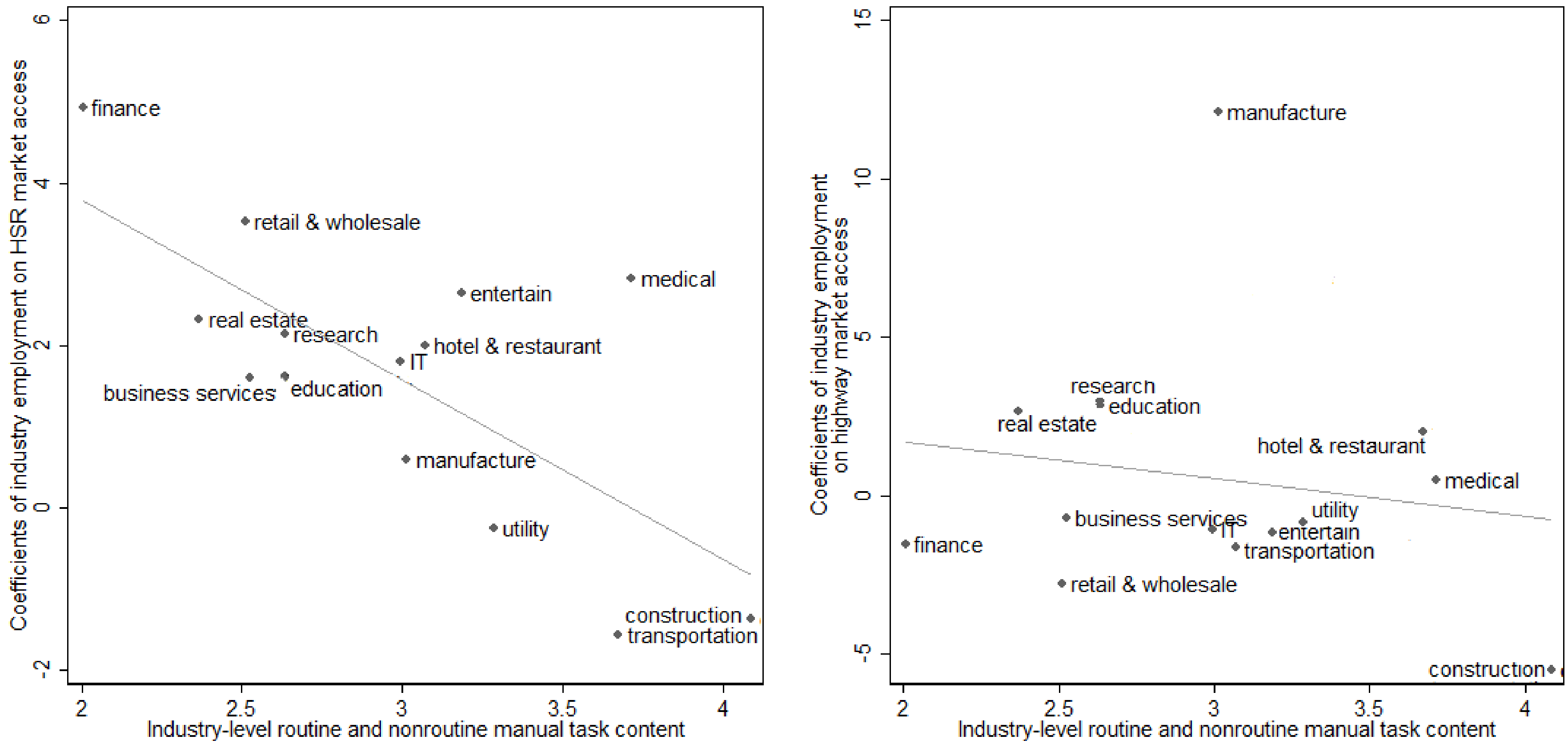


Figure 6: Sector-specific HSR and highway impacts and manual task intensity

Notes:The figures plot estimated coefficients on the impacts of NCIMA (left) and highway-induced market access (MAhighway) (right) on sectoral employment reported in Table 6 against sector-specific manual task intensity reported in Autor, Levy and Murname (2003) [5]. The matching between Chinese industries and the US industries is reported in Table A3.

Table 7: Impacts of HSR on aggregate economic outcomes

VARIABLES	log employment	logGDP	log housing price	log industrial fixed investment	log retail sales	logpatent
connect	0.0736***	0.0354	0.0303**	-0.0846***	0.101**	0.0352
	(0.0123)	(0.0250)	(0.0152)	(0.0278)	(0.0423)	(0.0232)
Observations	1,801	1,801	1,468	1,315	1,803	1,563
R-squared	0.035	0.023	0.023	0.050	0.028	0.132
log(NCIMA)	2.156***	3.932***	1.669**	-1.348	3.431**	1.616
	(0.677)	(1.483)	(0.763)	(1.537)	(1.378)	(1.263)
log(MAhighway)	-0.718	2.141**	0.612	-0.0346	1.301	0.0414
	(0.602)	(0.913)	(0.801)	(1.280)	(2.149)	(0.598)
Observations	2,877	2,871	2,332	2,098	2,879	2,397
R-squared	0.026	0.051	0.074	0.053	0.022	0.079

Notes:Data from the upper table are a panel of 172 Chinese prefecture cities annually from 2003 to 2013 that are connected or planned to be connected by HSR lines by the end of 2014. Connect is a dummy which is zero unless a city is connected to HSR before the end of that year, in which case it takes the value one. Data from the lower table are a panel of 278 Chinese prefecture cities annually from 2003 to 2013. NCIMA is the non-connection-induced-market-access measure. MAhighway accounts for market access changes by highway network expansion only. The dependent variables as listed are logs of total employment, GDP, housing price, total fixed investment of industrial firms, sales of retail firms and total patents. All outcome variables are counted at urban wards (shixiaqu) of prefecture cities. All regressions include city fixed effects and region-by-year fixed effects. Controls include government spending, other infrastructure measures, past city and provincial GDP, and interactions of year dummies with geographical centrality measures. Standard errors, reported in parentheses, are heteroskedasticity robust and clustered at the city level, allowing spatial dependence decaying in distance as in Conley (1999).* significant at 10%; ** significant at 5%; *** significant at 1%.

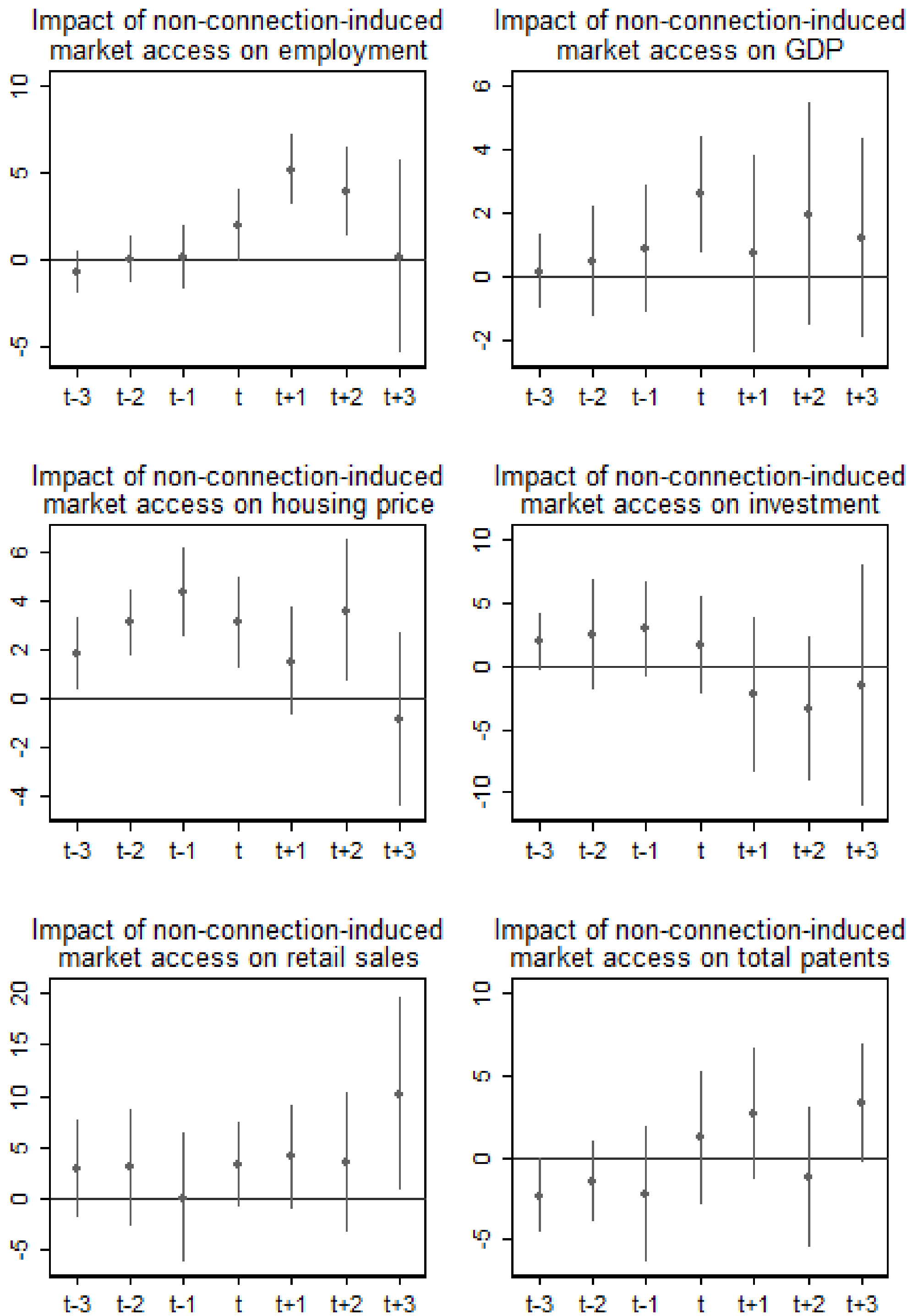


Figure 7: Event study: Aggregate economic outcomes

Notes:The figures display the estimated coefficients and 95% confidence intervals in regressions where the dependent variables are log total employment, GDP, housing private, fixed investments, total retail sales and patents. The independent variables are the leads and lags of the increments in non-connection-induced market access (NCIMA). For all the regressions, the sample is a balanced panel of 278 Chinese prefecture cities from 2003-2011, as HSR connection information is available only until 2014, and the third lead is a missing value for observations after 2011.

Appendix A. A Model of labour Sourcing

Appendix A.1. Preferences and Endowments

The economy consists of N cities and a single final good sector. Each city j is endowed with an inelastic supply of land ($\bar{H}_j$). The economy as a whole is endowed with an inelastic supply of workers who are perfectly mobile across cities. The representative consumer's preferences are defined over a final consumption good (C_j), housing (H_j) , and local-amenities (δ_j) taking the Cobb-Douglas form:

$$U_j = \delta_j C_j^{\alpha} H_j^{1-\alpha}, 0 < \alpha < 1 \tag{A.1}$$

Consumption good price is normalized to be 1. Housing rental rate is represented as q_j. The indirect utility function is:

$$V_j = b\delta_j w_j (q_j)^{\alpha-1} \tag{A.2}$$

where $b = \alpha^{\alpha}(1-\alpha)^{(1-\alpha)}$ is a constant.

For simplicity we assume that the final good is freely traded across cities. Cobb-Douglas utility function indicates that housing expenditures should take up $1-\alpha$ of total income. We assume that housing expenses are redistributed as a lump-sum to consumers, as in Helpman (1998) [30]. Therefore for city j, its total income can be represented as:

$$v_j L_j = w_j L_j + (1-\alpha) v_j L_j = w_j L_j / \alpha \tag{A.3}$$

Combined with housing market clearing condition, we obtain equilibrium housing rental rate:

$$q_j = \frac{1-\alpha}{\alpha} \frac{w_j L_j}{\bar{H}_j} \tag{A.4}$$

where $\bar{H}_j$ represents city j's inelastic housing supply.

Appendix A.2. Production

The final good y in city j is produced from a continuum of tasks $i \in [0, 1]$ as in Grossman and Rossi-Hansberg (2008) [29] under CES technology. Perfect competition and constant returns to scale apply.

$$y_j = [\int_0^1 x(i)^{(\sigma-1)/\sigma} dj]^{\sigma/(\sigma-1)} \tag{A.5}$$

Tasks are produced using labour according to a linear technology:

$$l_j(i) = \frac{x_j(i)}{z_j(i)} \tag{A.6}$$

A city j can remote source tasks from another city k. The productivity of producing task i at city k for city j is $z_{jk}(i)$. The origin city j's productivity on task i can be summarized by a vector $\mathbf{z_j(i)} = (z_{j1}(i), ...z_{jI}(i))$. I assume that the productivity vector for $i \in [0, 1]$ and $k = 1, 2, ..., I$, is a random variable drawn independently across tasks and destination cities from a multivariate Frechet distribution with zero correlation

across draws, $F_j(\mathbf{z_j}(\mathbf{i}) = exp(-\sum_l T_j(z_{jl}(i))^{-\theta})$. The marginal distribution of $z_{jk}(i)$ is then $F_j(z_{jl})) = exp(-T_j(z_{jl}(i))^{-\theta})$.

Here I use the multivariate Frechet distribution instead of the single-variable version because I would like to have the ex-ante probabilities of a city j sourcing out its tasks to any cities to be positive.

By remote sourcing a task i from city k to city j, firms take advantage of city j's higher productivity $z_{jk}(i)$ and city k's lower labour cost w_k, subject to iceberg transportation costs, $\tau_{jk} > 1$. The iceberg cost can be interpreted as the loss of efficiency in management over longer distances.

When a task i is sourced from city k to city j, the labour requirement $l_{jk}(i)$ and total cost of production $g_{jk}(i)$ are:

$$l_{jk}(i) = \frac{\tau_{jk} x_j(i)}{z_{jk}(i)} \tag{A.7}$$

$$g_{jk}(i) = \frac{\tau_{jk} w_k(i)}{z_{jk}(i)} \tag{A.8}$$

Firms in city j that sources in task i looks for a lowest cost source of supply for that task. Following Eaton and Kortum (2002) [21], the cost of sourcing task i from city k to city j has the distribution:

$$G_{jk}(g) = Pr[g_{jk}(i) \leq g] = Pr(\frac{w_k(i)}{z_{jk}(i)} \tau_{jk} \leq g) = 1 - exp((-T_j(w_k \tau_{jk})^{-\theta}) g^{\theta}) \tag{A.9}$$

The cost of production in city j then has the distribution:

$$G_j(g) = Pr[g_{jk}(i) \leq g, k = 1, 2...N] = 1 - \prod_{k=1}^{N} [1 - G_{jk}(g)] = 1 - exp(-\Phi_j g^{\theta}) \tag{A.10}$$

where the parameter Φ_j of city j's cost distribution is:

$$\Phi_j = \sum_{k=1}^{N} T_j (w_k \tau_{jk})^{-\theta} \tag{A.11}$$

The actual unit cost of production of final good for the CES production function, assuming $\sigma < 1 + \theta$, is

$$c_j = \gamma \Phi_j^{-1/\theta} \tag{A.12}$$

Appendix A.3. Flows in Tasks Sourcing

Perfect competition implies that all firms receive zero profits. Prices of the final good are equalized across cities because of free trade. Therefore, unit cost of production of final goods should also be equalized across cities, that is:

$$c_j = \gamma \Phi_j^{-1/\theta} = c, \forall j \tag{A.13}$$

A direct implication is that $\Phi_j = \sum_{i=1}^{N} T_j (w_i \tau_{ji})^{-\theta}$ is also equalized across cities. In

equilibrium, this cost equalization condition indicates that higher productivity in a central city usually goes with higher wages not only in the city itself, but also in surrounding cities. Additionally, a decrease in communication costs τ_{kj} between two cities is likely to drive up wages in these two cities relative to those in other cities.

Another important result derived following Eaton and Kortum (2002) [21] describes X_{jk}, the total labour cost in city k of producing for city j, as:

$$X_{jk} = T_j(w_k\tau_{kj})^{-\theta}(\sum_{k=1}^{N} T_j(w_k\tau_{kj})^{-\theta})^{-1}X_j \tag{A.14}$$

where X_j is city j's total cost of labour production (domestically produced or sourced to other cities). It is obvious from the above equation that city j sources more tasks to city k if city k has a lower wage, or is closer to j.

Note here $\sum_{k=1}^{N} T_j(w_k d_k)^{-\theta} = \Phi$ is the same across all cities due to the unit labour cost equalization condition. And we have:

$$Y_k = \sum_{j=1}^{N} X_{jk} = \Phi^{-1}(w_k^{-\theta})\sum_{j=1}^{N} \tau_{kj}^{-\theta} X_j T_j \tag{A.15}$$

Here Y_k is the total labour income of city k, and can be represented as $l_k w_k$. To guide the empirical analysis, I define $MA_k = \sum_{j=1}^{N} \tau_{kj}^{-\theta} X_j T_j$ as the "market access" of k. Intuitively, the "market access" in city k is a weighted average of all other cities' "market size" scaled up by their productivity and scaled down by distances. It can be roughly thought to represent the "perceived" demand of k's labour from the whole country.

Appendix A.4. Spatial Equilibrium

Free labour mobility implies that indirect utility is equalized across all cities in equilibrium.

$$V_j = b\delta_j W_j(q_j)^{\alpha-1} = V, \forall i \tag{A.16}$$

Combined with the "market access" equation (18) and the market clearing condition (6), we are now able to characterize a similar spatial equilibrium with endogenous variables N_k, w_k, q_k, and exogenous variables $T_k, \delta_k, \bar{H}_k, MA_k$, as in Glaeser and Gottlieb (2009). [26].

Given $\{T_k, \delta_k, \bar{H}_k, MA_k\}$, a spatial equilibrium is $\{N_k, W_k, q_k\}$ such that

1. $q_j = \frac{1-\alpha}{\alpha}\frac{w_j l_j}{\bar{H}_j}$ (Housing Market Clearing)
2. $V_j = b\delta_j w_j(q_j)^{\alpha-1} = V, \forall i$ (Workers Mobility)
3. $l_k w_k = \Phi^{-1}(w_k^{-\theta}) * MA_k$ (labour Income and Market Access)
4. $\sum_k l_k = N$ (Aggregate labour market clearing)
5. $\sum_k C_k = \sum_k y_k$ (Aggregate goods market clearing)

To summarize, my model has the following implications: First, a city's labour income $w_j l_j$ is increasing in its market access, defined as the access to all the other cities' technology. Second, housing prices increase following wages because of free labour mobility. Whether

the increase in market access translates into higher employment or higher wage depends on the housing supply restrictions of the city.

Appendix B. Data Appendix

The City Statistical Yearbook series reports production approach prefecture-city level GDP broken up into primary, secondary, and tertiary gross value added. The data are collected from local establishments under the supervision of the provincial governments.

The annual population and employment data from 2000 to 2013 are also obtained from City Statistical Yearbook series. The population data report residential population prepared by the Department of Population and Employment Statistics of the National Bureau of Statistics of China (NBS). Data for the years 2000 and 2010 are from the censuses, and the rest of the data covered are estimates from the annual national sample survey on population changes, or the annual Sample Survey on Labour Force. For some employment variables (retail, catering, business service employment), because of a change of employment categorization in 2003, observation from earlier years are dropped.

Data on total wage bills are collected and compiled through The Reporting Form System on Labour Wage Statistics by the National Bureau of Statistics of China (NBS). The Reporting Form System on Labour Wage Statistics covers corporate units in all urban areas.

The real estate price is calculated as the total sales of newly built commodity housing units divided by the total areas of these units. Data on total sales and areas of newly built commodity housing units are obtained from the Regional Economic Statistical Yearbook series by the NBS.

Data on annual city level passenger ridership on different modes of transportation are obtained from the China City Statistical Yearbook series, as well. Ridership is defined as the total number of trips made on a particular kind of transportation device. For road ridership, it only includes paid road trips (coaches, etc.), and excludes self-driving trips. For railway, the trips can start and end at any stops, including both destinations. I double check this source with the aggregate ridership data from China Railway Yearbook series and find them to be consistent with each other.

Table A1 summarizes the source, year range, the total number of observations, number of cities with at least one year of observation, number of cities with missing values along the time series of all the variables that I use in this paper. It is noted that the number of the observations are not the same for most of the variables. There are several possible explanations. First, the time span is different for data obtained from the Regional Economic Statistical Yearbooks and those obtained from the City Statistical Yearbooks, with the former covering two years less than the latter. Second, the geographical definition of urban core is not consistent for a few cities over the period I am studying. Usually, new suburbs are included into the urban cores as they get sufficiently urbanized. Third, around 2002 there was a change in employment categorization, which makes it impossible to get time-consistent employment data for some sectors before and after 2003, such as IT, business service, catering, retail, and education. Fourth, apart from key socioeconomic indicator variables such as GDP, population, employment and wage, the collection and compilation of some variables are not strictly overseen by the local statistical authority. Instead, the data

are usually obtained directly from government branches that collect and work with these data for other purposes. This practice leads to missing values from time to time.

Yearbook data are sometimes criticized for inaccuracy because of inadequate quality controls. To examine the reliability of yearbook data, I compare the 2009 population and employment data from the City Statistical Yearbook with those from the 2010 census. [23] The discrepancy between these two data sources should come from either measurement errors or actual changes in one year. Since population and employment are likely to be highly autocorrelated across years, low correlation between census and yearbook data is a sign for the poor quality of yearbook data. However, as shown in Table A2, the correlations between these two sources are quite high (more than 0.90 for most of the variables).

Similarly, Zheng and Kahn (2013) [48] has compared the housing price data obtained from the Regional Economic Statistical Yearbook series with their quality-controlled hedonic price index compiled by the Institute of Real Estate Studies at Tsinghua University and the correlation coefficient for these 35 cities for which both price sequences are available is 0.90. Hence we have confidence in the quality of the housing prices data from the yearbooks.

Appendix C. Services and fares of HSR

Most of the HSR lines in China operate intensively, but there is a large variety in the number of services across lines: the busiest Nanjing-Shanghai line carries more than 160 trains one way per day while the number of services for the Western Chengdu-Dazhou line is only 7. Detailed information on the number of services for each line is listed in Table A5. A typical HSR train normally carries eight cars, with a total capacity of about 600.

There are two main categories of high speed trains in China. One set of trains runs at a designed speed of 350km/h, and their train numbers start with G. The other runs at a designed speed of 250km/h with their train numbers led by D, compared to a top speed of 120km/h for pre-HSR trains. There are national fare scales for the two speeds but in practice, HSR fares vary slightly from line to line. The price, travelling time and cost per kilometer for passengers are listed in Table A4. For the lines served by trains of different speeds I report the price and travelling time of the category starting with G (350km/h).

Undoubtedly, the introduction of HSR drastically decreases the travel time across cities. The travel time between Beijing and Shanghai has been shortened from 13 hours to 5 after the introduction of the Beijing-Shanghai HSR line in 2011, for a second-class price of 553 RMB (90 USD). By comparison, a flight from Beijing to Shanghai usually takes a bit more than two hours for a full price of 1290 RMB (210 USD, including 160 RMB fees and taxes). Even if discounts are usually available for flights, HSR still proves to be a good value of money if we consider the extra time to get to the airport and occasional delays by air.

Appendix D. International Comparison

In this section, I will compare the Chinese HSR system to two of its most successful international predecessors, the Japanese Shinkansen and the French TGV, both of which

[23] The 2010 yearbook data are automatically adjusted according to census data

heavily used and profitable. The comparison is reported in Table A6, which helps me to put the current performance of Chinese HSR in context.

By the year 2011, the 2388 km Shinkansen HSR network in Japan generates an annual ridership of 345 millions[24]. In the meantime, the French TGV network, consisting of 2036 km of railway in service by 2011, reports an annual ridership of 114 million, including that of Eurostar[25] and Thalys[26]. Compared to these successful examples, China's 7021 km HSR network in services by 2011 boasts an annual ridership of 440 millions[27], overtaking its French and Japanese counterparts. From these numbers, we could say that although the HSR in China is not as heavily used as that in Japan, where the population density and the level of urbanization are much higher, it is quite comparable to its French counterpart.

On the profitability side, both the Japanese and French systems are in black after years of operating. JR Central, the main railway corporation in Japan operating several Shinkansen lines, reports a net income of USD 1.61 billion in year 2011.[28] In 2007, SNCF generated profits of 1.1 billion Euros (approximately USD 1.75 billion) driven largely by higher margins on the TGV network.[29]

For concreteness, here I make a comparison of the most representative lines from these three HSR systems, the Tokaido/Sanyo Shinkansen of Japan that runs through its eastern coastal corridor, the LGV SudEst of France that connects Paris and Lyon, and the Beijing-Shanghai line of China. Table A6 lists information on the distance, speed, price, ridership, as well as the size of main cities connected by these three lines. It is quite clear that the success of the Tokaido-Sanyo Shinkansen is largely attributable to the giant cities along the line. Given the size of Chinese cities along the Beijing-Shanghai corridor and their growth prospects, as well as the general growth trend of ridership on all the HSR lines in China, I would expect the ridership of Beijing-Shanghai line to grow rapidly in the future since there seems to be a lot of potentials to be realized. In fact, the annual ridership of Beijing-Shanghai line (excluding the Beijing-Tianjin and Shanghai-Nanjing segments) has increased from 48 million in 2011 to about 92 million in 2014. In the meantime, an insider from the China Railway Corporation revealed that the Beijing-Shanghai line could start earning profits by the end of 2014, three years after its opening. [30]

However, it is worth noting that although the Japanese and French HSR systems managed to make balances even, the Chinese HSR network is a much more extensive one, covering some lines that are not potentially profitable shortly. Therefore, it is still quite early to predict the profitability of Chinese HSR as a whole.

Appendix E. Market Access: Construction of cost parameters

As briefly mentioned in section 4.2, the calculation of the market access variables requires the construction of a time-varying transportation cost matrix for each city pair. We have to

[24] Data Book 2011, Central Japan Railway Company

[25] International HSR between London and Paris

[26] International HSR between Brussels and Paris

[27] http://english.peopledaily.com.cn/90778/7754100.html

[28] Databook 2011, Central Japan Railway Company

[29] David Gow (9 July 2008). "Europe's rail renaissance on track". guardian.co.uk (London).

[30] http://finance.ifeng.com/a/20140607/12494925_0.shtml

rely on a few assumptions to construct τ_{km}.

First, I allow τ_{km} to incorporate both time and fare cost in travel. I generalize passenger choices among different modes of transportation into three distinct combinations of time-fare tradeoffs. (1) Travel by conventional roads (not highways) or slower railway at a speed of 60 km/h and monetary cost of 0.1 km/h (2) Travel by highways or faster conventional rail (K or T initials) at a speed of 100 km/h, with a fare/monetary cost of ¥0.23/km [31] (3) Travel by HSR at a speed of 220 km/h, with a fare cost of ¥0.43[32] per kilometer. I assume that all cities are connected by conventional roads had they not been connected by highway or HSR. The information on highway network is obtained from Baum-Snow et al. (2015) [9]. A city is defined to be connected by highway if the distance between the centroid of the city and the nearest highway is less than 30 km. [33]

Second, wage plays an essential role in passengers' fare-time cost tradeoff: people who earn more have a higher value of time and are more likely to take high speed trains at a higher cost. For a particular city pair, I assume that the value of time is determined by the average wage of these two cities. If the value of hours saved by taking HSR does not cover the extra cost in fares, I assume that the HSR connection between these two cities does not reduce bilateral travel cost. To avoid additional endogeneity, I use the wage data in 2007, the year before most HSR lines were opened, in all the calculations, since there are a few wage data missing from 2000.

Third, I restrict the maximum number of transfers to be two to better capture people's travelling behavior. It is therefore assumed that if passengers need to make more than two transfers travelling by HSR, they will stick to alternative forms of transportation and HSR does not change their travel cost.

Fourth, to capture the nature of travel cost facilitate intercity communication, we need an approximation of the required frequency of face-to-face contacts from one city to another for an intercity link to be successfully built. Without sensible estimates on labour sourcing cost, I assume that on average, two trips per month are necessary for all kinds of tasks. In Section 8, I experiment with multiple values of the number of trips from 0.5 to 10 per month as robustness checks.

Finally, an estimation of market access as $MA_k = \sum_{m!=k}^{N} \tau_{km}^{-\theta} Y_m$, requires an estimation of the decay parameter θ. In my model, the parameter θ measures the dispersion of productivity. I follow Eaton and Kortum (2002) [21] and Donaldson (2014) and try several

[31]The fare of traditional railway is set according to a formula, adding up three parts: 0.06/km for seaters, 0.016/km extra for air-conditioning, 0.024/km extra for pre-HSR top speed, and usually scaled up by 50% for new air-conditioned trains, which are prevalent nowadays. The fare is also allowed to be scaled up by local railway authorities according to the operational costs; here I use an arbitrary scaling up rate of 50%. As for highways, the monetary cost usually consists of toll fees and fuel cost. The former is around ¥0.5/km for a small vehicle and the latter around 0.8/km for average fuel efficiency. Assuming on average four passengers occupy a small vehicle, the cost per person is 0.33/km. Taking long-distance coach would cost slightly less. For convenience I assume the fare and time cost are the same for conventional trains and road travel.

[32]Imputed from Table A3. For long-haul trains with a travel time above 10 hours, I scale up the average fare rate by a half to account for the extra cost of sleepers.

[33]Baum-Snow et al. (2015) [9] only reports changes in highway network for the year 2000, 2002, 2003, 2005, 2007 and 2010. According to them, data on the year 2000, 2005 and 2010 are more accurate. For years without data, I assume the highway network follows the previous year.

different measures over a wide range (3.6, 8.28 and 12.86) for robustness, but stick to 3.6 from Donaldson (2014) in my main analysis. I find the results to be robust to the selection of θ, as reported in Table A12. To avoid endogeneity caused by geographically correlated local shocks, I use base year (2000) GDP in calculations of all the market access variables.

Appendix F. Additional robustness checks

Appendix F.1. Exploiting variation in HSR opening time due to variation engineering difficulties only

As mentioned in Section 3.2, one way to get around the concerns that there are economic and political reasons to push some lines to be finished earlier that are correlated with the expected growth prospects of the connected cities is to focus on a subset of connected cities where the construction of HSR starts at the same year and to exploit the variation in project completion solely driven by engineering difficulties. Specifically, I look at the cities where HSR construction work commences in 2005, following the passage of "Mid-to-long run HSR plan immediately. I collect data on the construction duration, length, the number of stations, bridge and tunnel length of the eleven HSR lines that started construction in 2005. From an engineering point of view, an HSR line takes longer to be built if it is longer, has more stations and has a higher bridge and tunnel ratio. To implement the idea, I regress the duration of construction on HSR line length, the number of stations and bridge/tunnel ratio and get a predicted duration of construction of these lines as the fitted value of this regression. I then set the engineering-difficulty-determined opening date of each earlier HSR lines to be the construction starting date plus predicted duration of construction. The new engineering-difficulty-determined opening year changes for four lines (Wuhan-Guangzhou, Zhengzhou-Xi'an, Fujian-Xiamen, and Jiangmen-Xinhui). Although this operation leaves me with a very small sample (27 cities), I still find similar effects on railway ridership, service employment and private employment as in the main regressions (Table A13). However, the effects on manufacturing employment are negative in this case, which might be attributable to especially large construction effects as they are the earlier connectors, and the construction work might be more costly and requires more input with fewer experiences. It means that we should observe larger manufacture and construction employment increase before the actual connection of HSR, leaving the coefficient of connection dummy to be negative.

Appendix F.2. HSR impacts on peripheral areas

Table A14 presents the results on the impacts of HSR connection or HSR-induced market access on the peripheries of prefecture cities. I run separate regressions and replace the dependent variables as the outcomes of the whole prefecture city excluding its central urban area. I find small but insignificant negative impacts of direct HSR connection on GDP and employment in the peripheral areas of prefecture cities. It suggests that a small proportion of the positive impacts of HSR connection on urban-core employment is a result of an accelerating urbanization process. Direct HSR connection or HSR induced market access boost promote the attractiveness of central cities and draws people and economics activities from rural areas to urban centers.

Table A1: Data Appendix

Variables	Source	Year Range	Obs.	No. of cities	No. of cities with no missing values
GDP	City Statistical Yearbook	2000-2013	3888	282	261
Population	City Statistical Yearbook	2000-2013	3887	282	261
Employment	City Statistical Yearbook	2000-2013	3887	282	261
manufacturing employment	City Statistical Yearbook	2000-2013	3799	282	261
Service Employment	City Statistical Yearbook	2000-2013	3887	282	261
Wage	City Statistical Yearbook	2000-2013	3794	282	261
Housing Price	Regional Economic Statistical Yearbooks	2000-2011	3311	281	261
Railway Ridership	City Statistical Yearbook	2000-2013	3441	255	230
Road Ridership	City Statistical Yearbook	2000-2013	3940	284	268
Air Ridership	City Statistical Yearbook	2000-2013	1488	134	98
Total Ridership	City Statistical Yearbook	2000-2013	3960	284	268
Goods Transported by Railway	City Statistical Yearbook	2000-2013	3465	252	231
SOE employment	Regional Economic Statistical Yearbooks	2000-2011	3355	282	261
Utility employment	City Statistical Yearbook	2000-2013	3870	282	261
Construction employment	City Statistical Yearbook	2000-2013	3880	281	261
FIRE employment	City Statistical Yearbook	2000-2013	3705	282	261
IT employment	City Statistical Yearbook	2003-2013	3069	282	274
Business service employment	City Statistical Yearbook	2003-2013	3033	282	271
Retail employment	City Statistical Yearbook	2002-2013	3331	282	270
Catering employment	City Statistical Yearbook	2003-2013	3092	282	277
Educaion employment	City Statistical Yearbook	2003-2013	3101	282	277
Mining employment	City Statistical Yearbook	2000-2013	2887	206	198
Government Spending	City Statistical Yearbook	2000-2013	3641	282	261
Area of urban paved roads	City Statistical Yearbook	2002-2013	3423	282	277
Urban green land area	City Statistical Yearbook	2003-2013	3105	284	281
Number of theatres	City Statistical Yearbook	2000-2013	3856	282	262

Table A2: Correlation between the 2009 Yearbook and 2010 Census data

Variable	Correlation
Total employment	0.915
Total population	0.950
Industrial employment	0.886
Service employment	0.93
Manufacturing employment	0.878
Utility production employment	0.919
Construction employment	0.759
Hotel and restaurants employment	0.871
Finance employment	0.965
Real estate employment	0.874
Business services employment	0.855
Scientific research employment	0.982
Public services employment	0.971
Education employment	0.986
Medical employment	0.987

Notes:The table reports the correlation of data on population and employment for 282 cities in the 2009 yearbook and 2010 census. The classification of sectors within service industry in yearbooks is slightly different from the census. Hence I only calculate the correlations between two sequences when the classifications are exactly the same.

Table A3: Match between Chinese industries and US two-digit industries

Code	Chinese industries	US industries	NAICS
A	Agriculture, forestry, animal production and hunting, fishing	Forestry, Fishing, Hunting and Agriculture Support	11
B	Mining and quarrying	Mining	21
C	Manufacturing	Manufacturing	31
D	Production and distribution of electricity heating power, gas and water	Utilities	22
E	Construction	Construction	23
F	Wholesale and retail trade	Wholesale and retail trade	42,43
G	Transportation, warehousing and post	Transportation and Warehousing	48
H	Hotels and catering services	Accommodation and Food Services	72
I	Information transmission, computer services and software	Information	51
J	Finance and Insurance	Finance and Insurance	52
K	Real estate	Real Estate and Rental and Leasing	53
L	Leasing and business sevices	Professional, Scientific, and Technical Services	54
M	Scientific research, technical services	Professional, Scientific, and Technical Services	54
N	Management of water conservancy, environment and public facilities		
O	Household services, repair and other services		
P	Education	Educational Services	61
Q	Health,social work	Health Care and Social Assistance	62
R	Culture, sports and entertainment	Arts, Entertainment, and Recreation	71

Table A4: HSR lines in use before the end of 2014

HSR lines	Construction Time	Openning Time	Length(km)
Qinhuangdao-Shenyang	01/01/1999	01/07/2003	405
Hefei-Nanjing	11/06/2005	19/04/2008	154
Beijing-Tianjin	07/04/2005	01/08/2008	120
Qingdao-Jinan	28/01/2007	20/12/2008	393
Shijiazhuang-Taiyuan	11/06/2005	01/04/2009	190
Hefei-Wuhan	01/08/2005	01/04/2009	351
Dazhou-Chengdu	01/05/2005	07/07/2009	148
Ningbo-Taizhou-Wenzhou	27/10/2005	28/09/2009	268
Wenzhou-Fuzhou	08/01/2005	28/09/2009	298
Wuhan-Guangzhou	23/06/2005	26/12/2009	968
Zhengzhou-Xian	01/09/2005	06/01/2010	455
Fuzhou-Xiamen	01/10/2005	26/04/2010	275
Chengdu-Dujiangyan	04/11/2008	12/05/2010	65
Shanghai-Nanjing	01/07/2008	01/07/2010	301
Nanchang-Jiujiang	28/06/2007	20/09/2010	131
Shanghai-Hangzhou	01/04/2009	26/11/2010	150
Yichang-Wanzhou	01/12/2003	22/12/2010	377
Wuhan-Yichang	17/09/2008	23/12/2010	293
Haikou-Sanya (Eastern Coastal line)	29/09/2007	30/12/2010	308
Changchun-Jilin	01/04/2008	30/12/2010	111
Jiangmen-Xinhui	18/12/2005	07/01/2011	27
Beijing-Shanghai	18/04/2008	30/06/2011	1433
Guangzhou-Shenzhen	20/08/2008	26/12/2011	116
Longyan-Xiamen	25/12/2006	01/07/2012	171
Zhengzhou-Wuhan	15/10/2008	28/09/2012	536
Hefei-Bengbu	20/05/2009	16/10/2012	132
Haerbin-Dalian	23/08/2007	01/12/2012	921
Beijing-Zhengzhou	26/12/2007	26/12/2012	693
Nanjing-Hangzhou	01/04/2009	01/07/2013	249
Panjin-Yingkou	31/05/2009	12/09/2012	89
Tianjin-Qinhuangdao	08/11/2008	01/12/2013	261
Xiamen-Shenzhen	23/11/2007	28/12/2013	502
Xi'an-Baoji	18/12/2009	28/12/2013	148
Guangxi Coastal (Nanning-Qinzhou-Beihai)	11/12/2008	28/12/2013	261
Liuzhou-Nanning	27/12/2008	28/12/2013	223
Wuhan-Xianning	26/03/2009	28/12/2013	90
Taiyuan-Xi'an	03/12/2009	01/07/2014	678
Nanchang-Changsha	26/02/2009	16/09/2014	344
Hangzhou-Nanchang	18/04/2010	10/12/2014	582
Lanzhou-Wulumuqi	01/01/2010	26/12/2014	1776

Guangzhou-Nanning	11/09/2008	18/04/2014	577
Huanggang-Wuhan -Huangshi	02/10/2009	18/06/2014	97

Notes: Data are obtained from the Major Events and the Finished and Ongoing Projects sections in the China Railway Yearbooks from 1999 to 2012.

Table A5: Service and fare information of HSR lines

HSR lines	Trains/day	Speed (km/h)	Price	Duration	Cost/km
Qinhuangdao-Shenyang	20	250	125	02:30	0.31
Hefei-Nanjing	57	250	60.5	00:44	0.36
Beijing-Tianjin	87	350	54.5	00:33	0.47
Qingdao-Jinan	35	250	116.8	02:28	0.32
Shijiangzhuang-Taiyuan	16	250	68	01:35	0.36
Dazhou-Chengdu	7	200	110	02:45	0.74
Ningbo-Taizhou	44	250	49	00:59	0.32
Taizhou-Wenzhou	48	250	35.5	00:45	0.31
Wenzhou-Fuzhou	37	250	88.5	02:05	0.3
Wuhan-Guangzhou	58	350	463.5	04:10	0.48
Zhengzhou-Xian	28	350	229	02:27	0.5
Hefei-Wuhan	41	250	105	02:02	0.3
Fujian-Xiamen	65	250	71.5	01:39	0.26
Chengdu-Dujiangyan	6	220	15	00:33	0.23
Shanghai-Nanjing	162	350	139.5	01:30	0.46
Nanchang-Jiujiang	14	250	42	01:05	0.32
Shanghai-Hangzhou	81	350	73	00:50	0.49
Yichang-Wanzhou	16	200	162	04:47	0.43
Haikou-Sanya (Eastern Coastal line)	26	250	83.5	01:50	0.27
Changchun-Jilin	42	250	31.5	00:50	0.28
Jiangmen-Xinhui	19	200	10	00:07	0.37
Beijing-Shanghai	37	350	553	05:05	0.31
Guangzhou-Shenzhen	127	350	74.5	00:33	0.64
Longyan-Xiamen	15	200	48.5	01:10	0.31
Zhengzhou-Wuhan	54	350	244	02:10	0.46
Hefei-Bengbu	29	350	70	00:57	0.53
Haerbin-Dalian	18	350	403.5	04:14	0.44
Beijing-Zhengzhou	55	350	309	03:22	0.45
Nanjing-Hangzhou	73	350	117.5	01:45	0.47
Panjin-Yingkou	5	350	27.5	00:26	0.31
Xiamen-Shenzhen	39	350	150.5	03:50	0.30
Xi'an-Baoji	18	350	51.5	00:59	0.35
Tianjin-Qinhuangdao	16	350	120	01:25	0.46
Guangxi Coastal (Nanning-Qinzhou-Beihai)	23	250	61	01:38	0.30
Liuzhou-Nanning	31	250	65.5	01:30	0.30
Wuhan-Xianning	48	250	24.5	00:24	0.27
Hangzhou-Nanchang	42	350	263.5	02:57	0.46
Nanchang-Changsha	45	350	157	01:42	0.46
Lanzhou-Wulumuqi	3	250	548.5	11:25	0.31
Taiyuan-Xi'an	15	250	178.5	03:47	0.28

Guangzhou-Nanning	34	250	169	00:39	0.30
Huanggang-Wuhan -Huangshi	127	250	74.5	00:33	0.64

Table A6: Travel cost comparison across different modes of transportation

Transportation modes	Short-distance trip (Beijing-Tianjin, 150 km)			Mid-distance trip (Beijing-Zhengzhou, 600km)			Long-distance trip (Beijing-Shanghai, 1400km)		
	Time	Fare cost	Frequency	Time	Fare cost	Frequency	Time	Fare cost	Frequency
HSR	0.5 h	¥74.5	37/day	3.3 h	¥309	55/day	6 h	¥553	44/day
Air				1.2 h	¥582 (half price)	26/day	2 h	¥568 (half price)	120/day
Conventional trains (Tekuai)	2 h	¥65	2/day	7.5 h	¥174	32/day	15 h	¥325	2/day
Highway	2.5 h	¥40	37/day	9 h	¥158	8/day	16 h	¥341	1/day

Notes:1. Fare, frequency and travel time information on HSR and conventional trains is obtained from the China Railway Corporation official ticket sale website. Fare, frequency and travel time information on coaches is obtained from Keyunzhan, a website for coach information lookup.
2. Fare, frequency and travel time information on planes is obtained from Qunar, a Chinese flight fare comparison website. Since fare on flights fluctuates a lot, I use the half price fare from Air China as a benchmark for comparison.

Table A7: HSR passenger survey results

<table>
<tr><th>Line</th><th>Sample Size</th><th>Average monthly Income</th><th>Purpose of travel[5]</th><th>Connecting or transfer modes[4]</th><th colspan="3">Alternative modes [7]</th></tr>
<tr><td rowspan="3">Changchun-Jilin (110km)[1]</td><td rowspan="3">1001</td><td rowspan="3">4300</td><td>Non-commute business: 26%</td><td>Car/Taxi: 48%</td><td colspan="3">Bus/Coach: 50%</td></tr>
<tr><td>Leisure: 46%</td><td rowspan="2">Public transport: 43%</td><td colspan="3">Ordinary train: 36%</td></tr>
<tr><td>Commuting: 19%</td><td colspan="3">Air: 0%</td></tr>
<tr><td rowspan="3">Beijing-Tianjin[3] (120km)</td><td rowspan="3">1108</td><td rowspan="3"></td><td>Non-commuting Business: 39%</td><td>Car/Taxi: 34.7%</td><td colspan="3">Bus: 20.5%</td></tr>
<tr><td>Leisure: 33%</td><td rowspan="2">Public transport: 60.1%</td><td colspan="3">Other train: 61.5%</td></tr>
<tr><td>Commuting: around 15%</td><td colspan="3">Car: 11.60%</td></tr>
<tr><td rowspan="3">Wuhan-Guangzhou[2] 968km)</td><td rowspan="3">556</td><td rowspan="3">4500</td><td>Non-commute Business: 39%</td><td>No transfer: 49%</td><td rowspan="3" colspan="3"></td></tr>
<tr><td>Leisure: 33%</td><td>Transfer to road: 40%</td></tr>
<tr><td>Commuting: around 15%</td><td>Transfer to air:1.5%</td></tr>
<tr><td rowspan="4">Tianjin-Jinan[1] (Part of Beijing-Shanghai 1318km)</td><td rowspan="4">1001</td><td rowspan="4">6700</td><td>Non-commute Business: 62%</td><td>Transfer to road 67%</td><td></td><td>Short trip [6]</td><td>Long trip</td></tr>
<tr><td>Leisure: 28%</td><td rowspan="3">Other: 29%</td><td>Bus</td><td>32%</td><td>1%</td></tr>
<tr><td rowspan="2">Commuting: 0%</td><td>Other Train</td><td>40%</td><td>18%</td></tr>
<tr><td>Air</td><td>7%</td><td>77%</td></tr>
</table>

Notes: 1. Passenger survey results from Gerald et al. (2014) [25]
2. Passenger survey results Li (2011) [32]
3. Passenger survey results Wu et al. (2013) [47]
4. The exact questions differ across surveys. For the survey on Wuhan-Guangzhou line, the question is "What are your transfer choices off HSR to your final destination, if any?". For the other surveys, the question is "What are your transportation choices to the HSR station?"
5. The percentages of business travel, leisure and commuting do not add up to 100% and the remaining respondents choose other purposes.
6. Short trips are defined as trips shorter than 300 km, and long trip longer than 300 km.
7. Passengers are asked "What are your preferred ways to travel before the introduction of HSR for similar trips?" in the surveys.

Table A8: Annual ridership for major HSR lines 1, (in millions passengers)

HSR lines	2009	2010	2011	2012
Qingdao-Jinan	21.22	23.95	26.74	29.63
Beijing-Tianjin	16.41	20.22	21.04	21.50
Changchun-Jilin			8.84	9.12
Shijiazhuang-Taiyuan 2	4.64 (6.19)	7.46	8.48	8.83
Hefei-Wuhan 2	1.62 (2.16)	2.97	3.49	4.57
Shanghai-Shenzhen 3	2.17	20.89	35.06	
Zhengzhou-Xian	X5	3.74	3.82	6.37
Guangzhou-Shenzhen 4	33.49	36.95	39.05	35.78
Guangzhou-Shenzhen -Hong Kong				10.56
Guangzhou-Zhuhai			14.06	16.31
Beijing-Guangzhou 3		20.52	30.87	38.57
Beijing-Shanghai 2			20.59 (35.29)	54.81
Shanghai-Nanjing			52.46	56.96
Shanghai-Hangzhou			16.29	18.52
Nanchang-Jiujiang			10.80	11.53
Dalian-Shenyang				1.01
Haikou-Sanya (Eastern Coastal line)			9.82	10.48
Hefei-Nanjing			1.37	2.83

Notes:1. The ridership data on Guangzhou-Shenzhen line are obtained from the annual report of the GUANGSHEN Railway Co.,Ltd. The other ridership information is obtained from the China Railway Yearbooks from 1999 to 2012.
2. The Hefei-Wuhan and the Shijiazhuang-Taiyuan lines were opened on April 1st, 2009. The Beijing-Shanghai line was opened on June 30th, 2011. Therefore the ridership is counted for only 9 months and 7 months, respectively. Projected ridership numbers of the whole year for these three lines in 2009 and 2011 are reported in the parentheses.
3. The Shanghai-Shenzhen line contains several segments: Ningbo-Taizhou-Wenzhou (open on 28/09/2009), Wenzhou-Fuzhou (open on 28/09/2009), Shanghai-Hangzhou (open on 26/10/2010), Fujian-Xiamen (26/04/2010), Hangzhou-Ningbo (in construction), Xiamen-Shenzhen (open on 26/12/2013). Similarly, the Beijing-Guangzhou line consists of three line segments: Wuhan-Guangzhou line (opened on 26/12/2009), Wuhan-Shijiazhuang line (opened on 28/09/2012) and Shijiazhuang-Beijing line(opened on 26/12/2012). Therefore the change in ridership reflects both the change for existing lines and the newly-generated ridership on new segments.
4. This line is an upgraded line (D-initial) opened in 2007 with a top speed of 200km/h, different from the newly constructed high speed line (Shenzhen-Guangzhou-Hong Kong) opened in 2011 with a top speed of 350km/h)
5. X denotes missing data. The blank indicates that the line is not opened in that year.

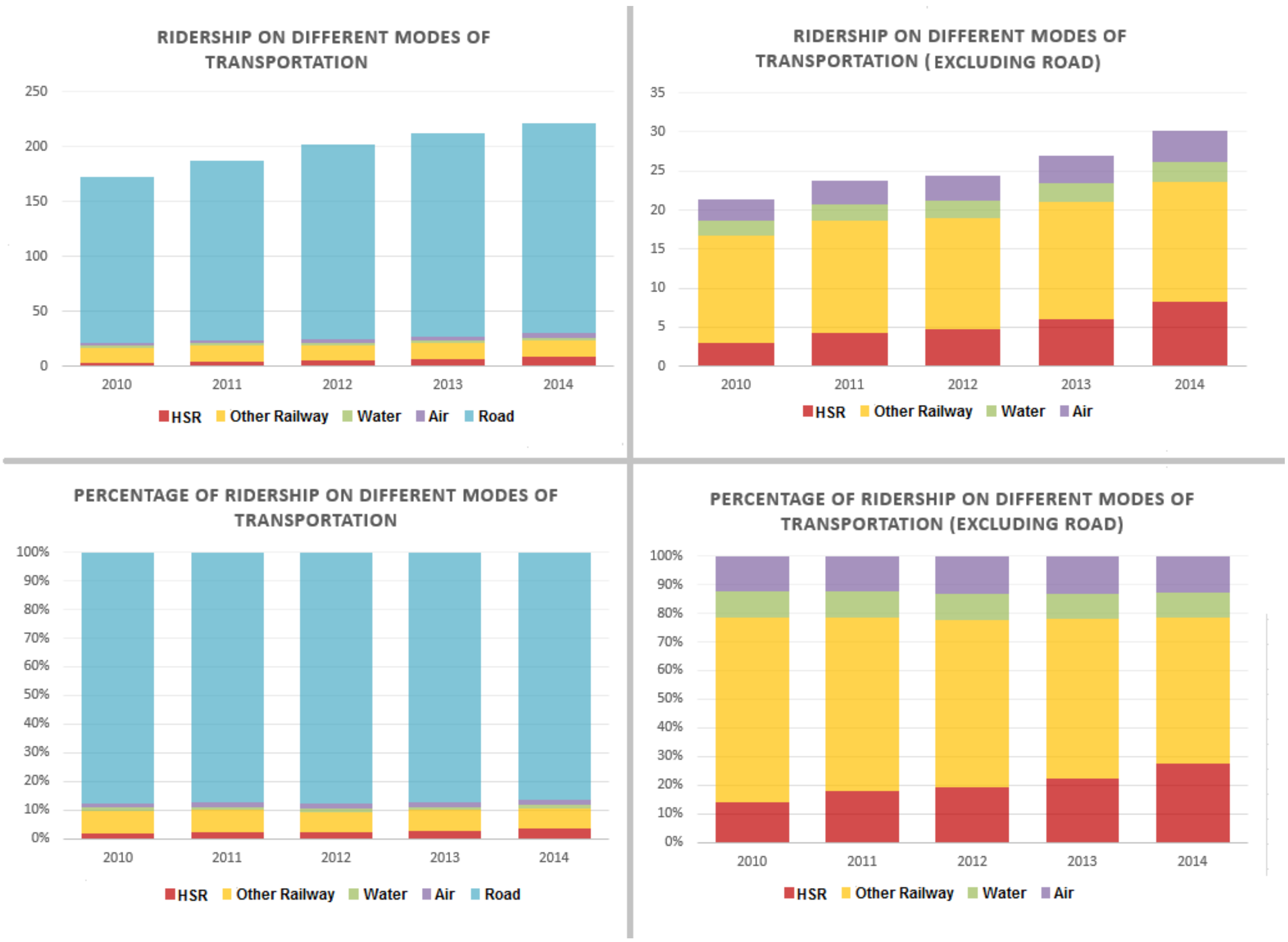


Figure A1: Aggregate ridership on different modes of transportation: 2010 to 2014

Notes:The figures display the aggregate number and percentage of ridership on high speed railway, conventional railway, road, air and water travel from 2010 to 2014. The unit is 100 million.

Table A9: International Comparison

Lines	Tokaido-Sanyo Shinkansen	LGV SudEst	HSR Beijing-Shanghai
Inaguration	1964(Tokaido) 1975(Sanyo)	1981	2008(Beijing-Tianjin) 2010(Shanghai-Nanjing) 2011(Beijing-Shanghai)
Section	Tokyo-Osaka -Hakata	Paris-Lyon	Beijing-Tianjin -Nanjing-Shanghai
Distance	1069 km	431 km	1433 km
Trains/Day	323(Tokaido) 277(Sanyo)	211	162(Shanghai-Nanjing) 87(Beijing-Tianjin) 44 (Beijing-Shanghai)
Price(USD, full section)	125	69-126	90
Max speed(km/h)	300(Tokaido), 270(Sanyo)	300	350
Ridership(2011) Millions	149 (Tokaido), 62 (Sanyo)	About 30	21 (Beijing-Tianjin) 65 (Nanjing-Shanghai) 48 (Beijing-Shanghai)
Main cities GDP (USD Billions, 2011)	Tokyo 1,520 Osaka 654.8 Nagoya 367.0	Paris 669.2 Lyon 90.4	Beijing 427.2 Shanghai 516.5 Tianjin 308.7
Main cities Population (Millions, 2011)	Tokyo 13.19 Osaka 2.67 Nagoya 2.27	Paris 10.52 Lyon 1.57	Beijing 12.07 Shanghai 13.50 Tianjin 5.51

1. Data on city GDP are obtained from "Global city GDP 2011-2012". Brookings Institution. Data on population are obtained from website of Japanese Statistical Bureau, Eurostat and Chinese City Yearbook, respectively.
2. The information on railway inauguration, distance, train frequency and maximum speed for Shikansen and LGV, and the ridership data of Shikansen are obtained from the 2011 Databook by the Central Japan Railway Company. The price information is obtained from the operators' websites.
3. The ridership data on LGV SudEst is obtained from "Brunhouse, Jay. travelling the Eurail Express. Pelican Publishing, 2004."
4. The data reflects the most recent information as of October, 2014, unless otherwise specified.

Table A10: Tourism

VARIABLES	log employment	log retail employment	log cater employment	log skilled employment	log other service employment	log other non-service employment	logGDP	log retail sales
log(NCIMA)	2.187***	2.171	-0.958	1.467***	-0.342	1.228	3.666**	2.426
	(0.648)	(1.655)	(1.728)	(0.533)	(0.767)	(1.791)	(1.538)	(1.515)
tourism resources* log(NCIMA)	0.571**	2.826***	5.021***	1.336***	1.656***	-0.350	-0.683**	1.363*
	(0.263)	(0.761)	(0.560)	(0.326)	(0.370)	(1.132)	(0.322)	(0.735)
Observations	2,805	2,550	2,547	2,747	2,546	2,717	2,800	2,807
R-squared	0.048	0.030	0.032	0.040	0.020	0.060	0.131	0.033

Notes: Data are a panel of 278 Chinese prefecture cities annually from 2003 to 2013. NCIMA is the non-connection-induced-market-access measure. log(NCIMA) is interacted with the number of 5A tourist attractions of each city. The dependent variables as listed are logs of total employment, retail and wholesale trade, hotel and restaurants employment, skilled employment (includes IT, FIRE, education, business service and scientific research), other service and non-service employment, total GDP and total retail sales. All outcome variables are counted at urban wards (shixiaqu) of prefecture cities. All regressions include city fixed effects and region-by-year fixed effects. Controls include government spending, other infrastructure measures, past city and provincial GDP, and interactions of year dummies with geographical centrality measures. Standard errors, reported in parentheses, are heteroskedasticity robust and clustered at the city level, allowing spatial dependence decaying in distance as in Conley (1999).* significant at 10%; ** significant at 5%; *** significant at 1%.

Table A11: Robustness: Limited to cities without an airport

VARIABLES	log railway ridership	log road ridership	log total ridership	log railway goods
connect	0.144**	-0.0450	0.142***	0.200*
	(0.0577)	(0.122)	(0.0438)	(0.102)
Observations	829	735	840	806
R-squared	0.056	0.012	0.069	0.031
log(NCIMA)	10.98***	-0.787	3.265*	-4.562
	(3.894)	(2.254)	(1.808)	(6.670)
log(MAhighway)	-2.159	5.784*	5.110**	-1.395
	(2.054)	(3.459)	(2.559)	(4.191)
Observations	1,207	1,329	1,341	1,229
R-squared	0.051	0.045	0.050	0.037

VARIABLES	log employment	logGDP	log housing price	log skilled employment	log tourism employment	log other service employment	log other non-service employment	logpatent
connect	0.0960***	0.0493**	0.0389*	0.0583***	0.118***	0.0575**	0.0709	0.00541
	(0.0192)	(0.0199)	(0.0198)	(0.0149)	(0.0315)	(0.0260)	(0.0560)	(0.0391)
Observations	840	841	685	835	839	796	817	754
R-squared	0.050	0.034	0.042	0.042	0.044	0.016	0.040	0.172
log(NCIMA)	1.393	3.949***	1.530	2.752***	0.5327	2.193**	1.065	-2.631*
	(1.041)	(1.269)	(0.961)	(0.612)	(2.1036)	(1.079)	(3.062)	(1.578)
log(MAhighway)	-0.136	3.490***	-1.304	1.937***	1.937***	1.632*	4.065	-0.513
	(0.862)	(0.903)	(1.028)	(0.715)	(1.7022)	(0.902)	(2.691)	(0.694)
Observations	1,341	1,339	1,087	1,326	1,320	1,226	1,303	1,141
R-squared	0.024	0.122	0.041	0.033	0.033	0.022	0.029	0.121

Notes:Data are a panel of 130 Chinese prefecture cities without any airports by the end of 2013 annually from 2003 to 2013. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A12: Robustness to Alternative Parameters and Specifications

VARIABLES	log railway ridership	log employment	log manufacture employment	log service employment	loggdp	log housing price
Set required trips to be 0.5 per month	19.31***	8.093***	8.530	9.363***	11.51**	4.801**
	(7.134)	(1.878)	(10.85)	(1.565)	(4.654)	(2.400)
Set required trips to be 1 per month	10.25***	4.386***	4.637	5.070***	6.149**	2.580**
	(3.800)	(1.015)	(5.833)	(0.839)	(2.514)	(1.304)
Set required trips to be 5 per month	3.013***	1.418***	1.470	1.612***	1.777**	0.773*
	(1.115)	(0.316)	(1.785)	(0.252)	(0.762)	(0.418)
Set required trips to be 10 per month	2.101***	1.041***	1.041	1.164***	1.185**	0.532*
	(0.768)	(0.224)	(1.254)	(0.175)	(0.521)	(0.302)
Set parameter θ to 8.28 in equation (19)	5.652**	2.442***	7.559***	2.382***	1.614	1.210
	(2.877)	(0.867)	(2.903)	(0.701)	(1.149)	(1.178)
Set parameter θ to 12.86 in equation (19)	5.667***	2.112***	3.767	2.838***	2.771***	1.589*
	(2.097)	(0.560)	(2.783)	(0.431)	(0.993)	(0.924)

Notes:Each row reports a set of estimates from the indicated specification, as discussed in the text (section 6.1). The relevant market access measure is the one that only captures the HSR-induced market access changes (MAHSR). Robust standard errors clustered at city level are reported in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

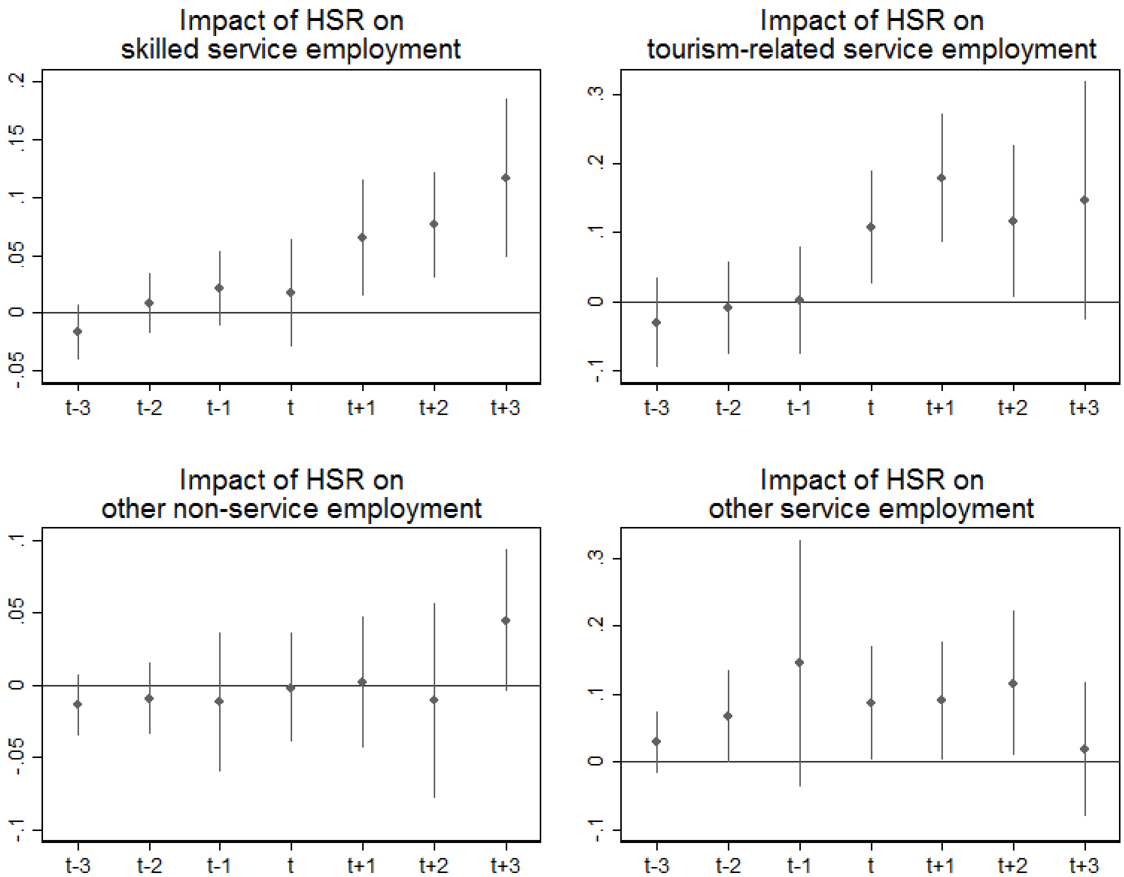


Figure A2: Event study with connection dummy: Specialization patterns

Notes:The figures display the estimated coefficients and 95% confidence intervals in regressions where the dependent variables are log skilled service employment, tourism-related service employment, other service and non-service employment. The independent variables are the leads and lags of the initial connection dummy. For all the regressions, the sample is a balanced panel of 172 Chinese prefecture cities from 2003-2011, as HSR connection information is available only until 2014, and the third lead is a missing value for observations after 2011.

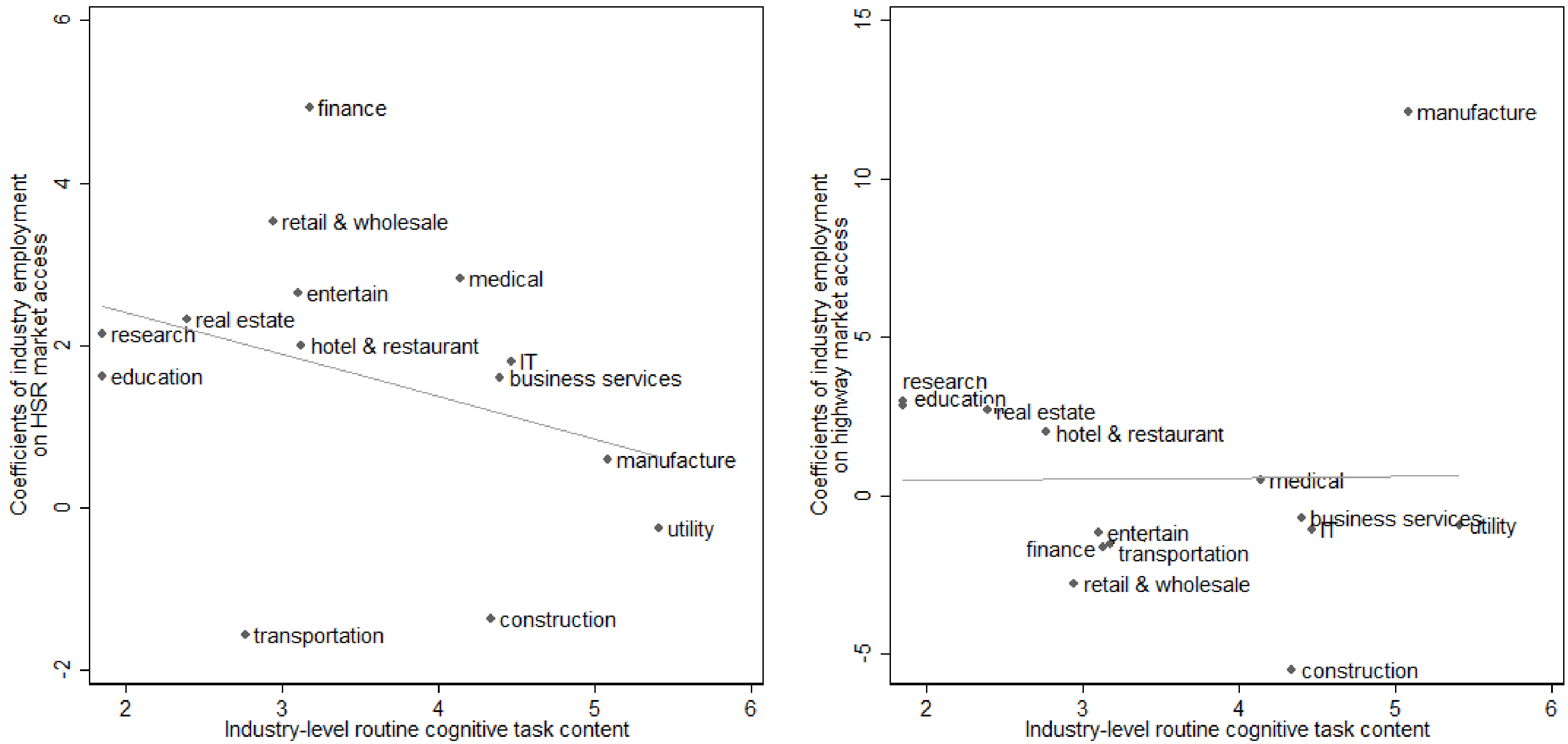


Figure A3: Sector-specific HSR and highway impacts and routine cognitive task intensity

Notes:The figures plot estimated coefficients on the impacts of NCIMA highway-induced market access (MAhighway) on sectoral employment reported in Table 6 against sector-specific routine task intensity reported in Autor, Levy and Murname (2003) [5]. The matching between Chinese industries and the US industries is reported in Table A3.

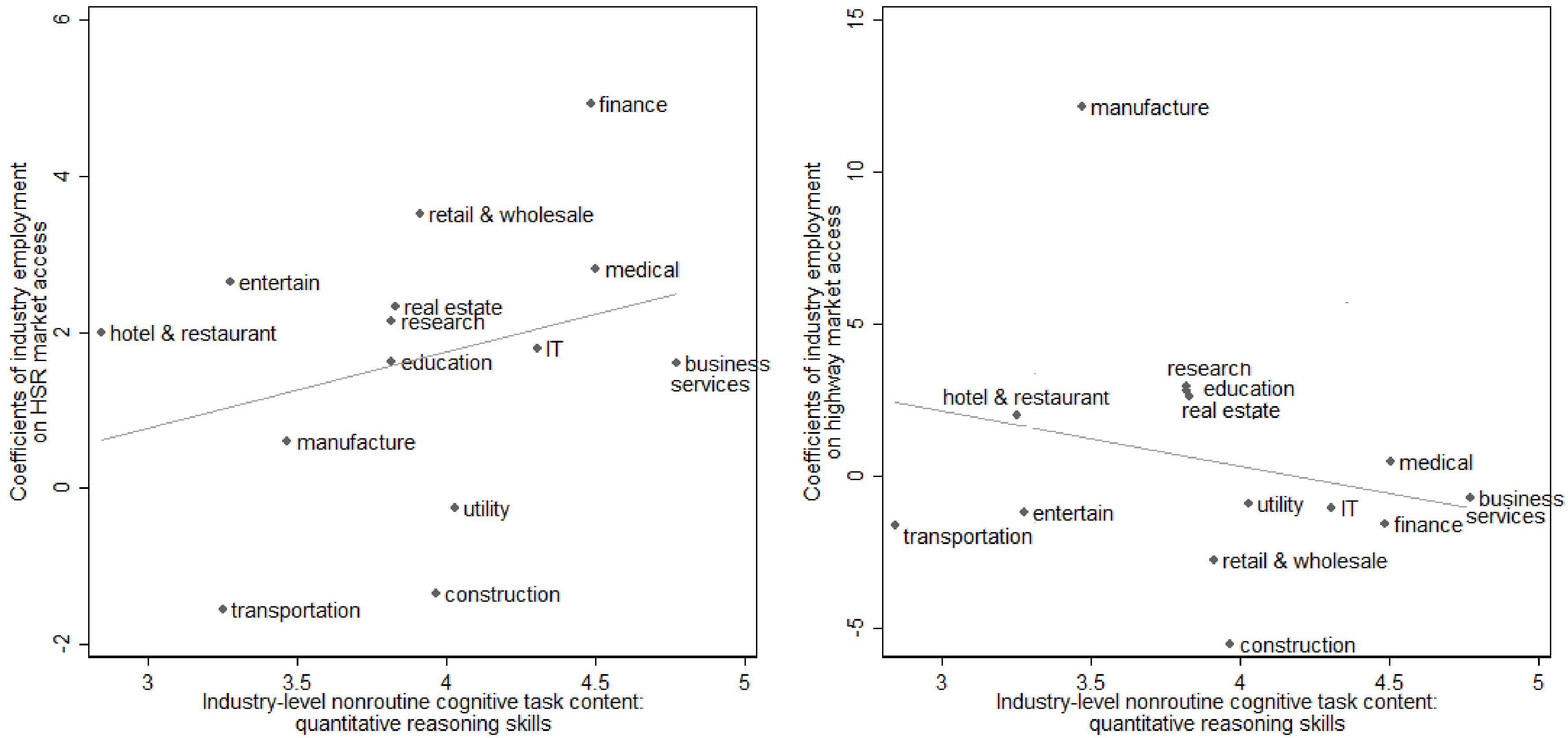
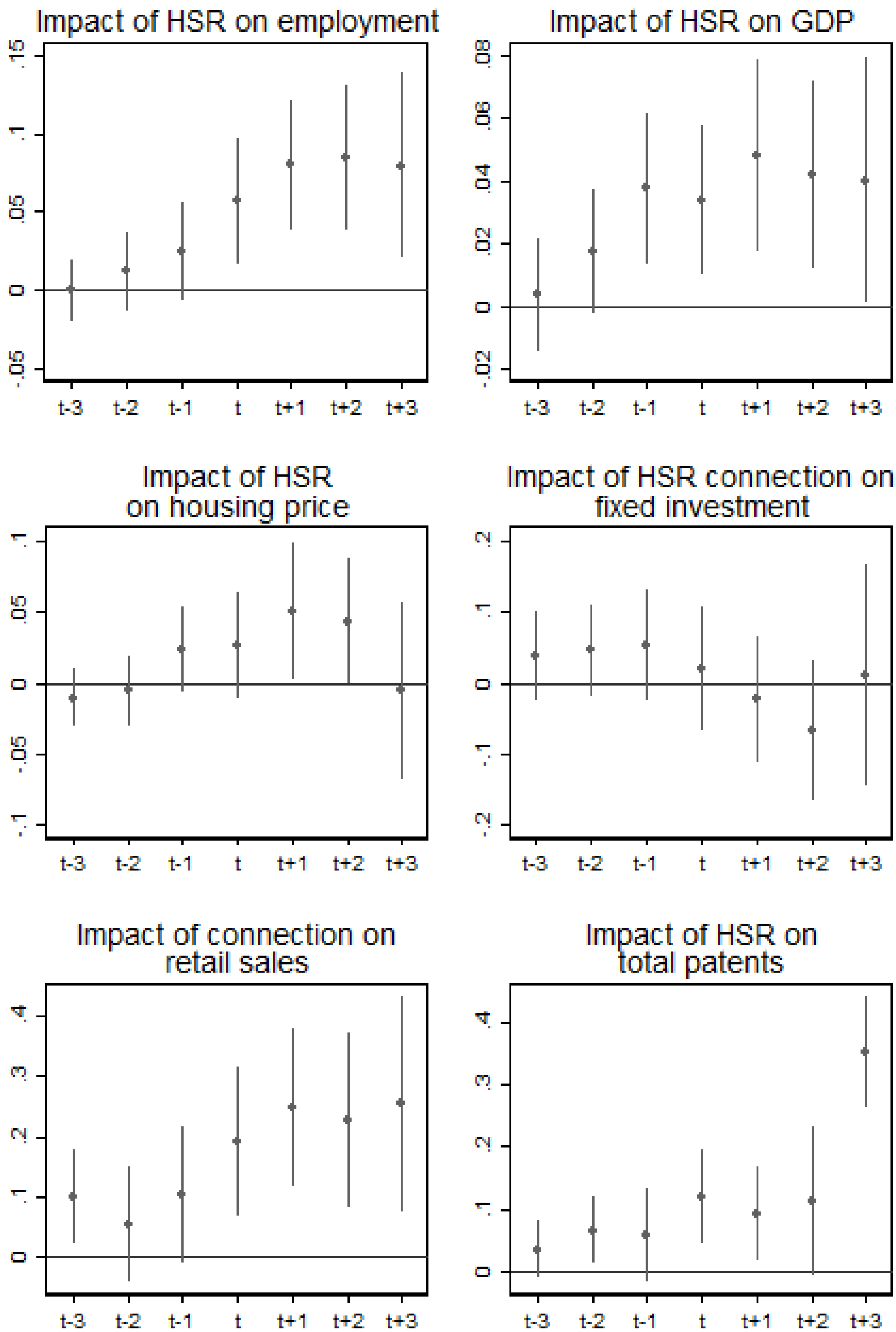


Figure A4: Sector-specific HSR and highway impacts and nonroutine analytical task intensity

Notes:The figures plot estimated coefficients on the impacts of NCIMA highway-induced market access (MAhighway) on sectoral employment reported in Table 6 against sector-specific nonroutine analytical task intensity reported in Autor, Levy and Murname (2003) [5]. The matching between Chinese industries and the US industries is reported in Table A3.

Figure A5: Event study with connection dummy: Aggregate economic outcomes



Notes:The figures display the estimated coefficients and 95% confidence intervals in regressions where the dependent variables are log total employment, GDP, housing private, fixed investments, total retail sales and patents. The independent variables are the leads and lags of the initial connection dummy. For all the regressions, the sample is a balanced panel of 172 Chinese prefecture cities from 2003-2011, as HSR connection information is available only until 2014, and the third lead is a missing value for observations after 2011.

Table A13: Exploiting variation in HSR connection timing purely due to engineering difficulty differences

VARIABLES	lograilpas	logemployment	logmanuemp	logserviceemp	logwage	loggdp	loghouseprice
connectalt	0.159**	-0.00342	-0.0701*	0.0476*	0.0518**	-0.0605**	0.00317
	(0.0680)	(0.0297)	(0.0392)	(0.0242)	(0.0251)	(0.0302)	(0.0417)
Observations	213	226	227	227	225	227	226
No. of cities	27	27	27	27	27	27	27
R-squared	0.323	0.297	0.299	0.220	0.253	0.545	0.121
VARIABLES	log SOE employment	log other service employment	log catering employment	log industrial firm numbers	log industrial revenue	log retail firm numbers	log retail sales
connectalt	0.0199*	0.0656**	-0.0772	0.00412	-0.00751	-0.0118	0.207***
	(0.0108)	(0.0256)	(0.0705)	(0.0424)	(0.127)	(0.0598)	(0.0667)
Observations	224	227	227	227	227	224	227
No. of cities	27	27	27	27	27	27	27
R-squared	0.147	0.265	0.289	0.288	0.215	0.200	0.213

Notes:Data are a panel of 27 Chinese prefecture cities annually from 2003 to 2013, whose HSR construction started in 2005. Connectalt is a dummy of an alternative connection measure. It is the projected year of HSR opening based purely on engineering difficulty driven construction duration predicted by the length of HSR lines and total bridge/tunnel percentage. Standard errors, reported in parentheses, are heteroskedasticity robust and clustered at the city level, allowing spatial dependence decaying in distance as in Conley (1999).* significant at 10%; ** significant at 5%; *** significant at 1%.

Table A14: Impacts of HSR on peripheral areas

VARIABLES	log suburban employment	log suburban population	log suburban GDP	log suburban employment	log suburban population	log suburban GDP
connect	-0.0325	0.00464	-0.0644			
	(0.0302)	(0.00677)	(0.0459)			
log(NCIMA)				-0.724	-0.428	-5.710*
				(1.254)	(0.375)	(3.439)
Observations	1,768	1,771	1,771	2,801	2,806	2,801
No. of cities	172	172	172	279	279	279
R-squared	0.043	0.022	0.036	0.032	0.020	0.034

Notes:Data are a panel of 278 Chinese prefecture cities annually from 2003 to 2013. The outcome variables are employment, population, and GDP at the peripheries (areas out of urban wards) of prefecture cities. Standard errors, reported in parentheses, are heteroskedasticity robust and clustered at the city level, allowing spatial dependence decaying in distance as in Conley (1999).* significant at 10%; ** significant at 5%; *** significant at 1%.

References

[1] Adler, N., Pels, E., Nash, C., 2010. High-speed rail and air transport competition: Game engineering as tool for cost-benefit analysis. Transportation Research Part B: Methodological 44, 812-833.

[2] Ahlfeldt, Gabriel M., and Arne Feddersen. *From periphery to core: economic adjustments to high speed rail.* (2010).

[3] Agrawal, Ajay, Alberto Galasso, and Alexander Oettl. "Roads and Innovation." Rotman School of Management Working Paper 2478752 (2014).

[4] Au, Chun-Chung, and J. Vernon Henderson. *How migration restrictions limit agglomeration and productivity in China.* Journal of Development Economics 80.2 (2006): 350-388.

[5] David, H., Frank Levy, and Richard J. Murnane. "The Skill Content of Recent Technological Change: An Empirical Exploration." Quarterly Journal of Economics 118.4 (2003).

[6] Banerjee, Abhijit, Esther Duflo, and Nancy Qian. *On the road: Access to transportation infrastructure and economic growth in China.* No. w17897. National Bureau of Economic Research, 2012.

[7] Baum-Snow, Nathaniel. *Did highways cause suburbanization?.* Quarterly Journal of Economics 122.2 (2007): 775-805.

[8] Baum-Snow, Nathaniel. "Urban Transport Expansions, Employment Decentralization, and the Spatial Scope of Agglomeration Economies." (2014).

[9] Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., Zhang, Q. (2015). Roads, Railroads and Decentralization of Chinese Cities.

[10] Bernard, A B, A Moxnes, and Y U Saito (2014a) Geography and Firm Performance in the Japanese Production Network, RIETI Discussion Paper 14-E-034.

[11] Christiaan Behrens and Eric Pels, "Intermodal competition in the London-Paris passenger market: High-speed rail and air transport," Journal of Urban Economics, online 29 December 2011

[12] Bullock, Richard Harvey, Andrew Salzberg, and Ying Jin. *High-speed Rail–the First Three Years: Taking the Pulse of China's Emerging* Program. No. China Transport Topics No. 04. World Bank., 2012.

[13] Burgess, Robin, and Dave Donaldson. *Can Openness Mitigate the Effects of Weather Fluctuations? Evidence from India's Famine Era.* (2010).

[14] Clark, Terry Nichols, et al. *Amenities drive urban growth.* Journal of Urban Affairs 24.5 (2002): 493-515.

[15] Conley, T. G. (1999). Gmm estimation with cross sectional dependence. Journal of Econometrics, 92 (1), 1-45.

[16] State Council (2004): *zhong chang qi tie lu wang gui hua*

[17] Donaldson, Dave. *Railroads of the Raj: Estimating the impact of transportation infrastructure.* forthcoming, American Economic Review, 2014.

[18] Donaldson, Dave, and Richard Hornbeck. *Railroads and American economic growth: A market access approach.* No. w19213. National Bureau of Economic Research, 2013.

[19] Duranton, Gilles, and Matthew A. Turner. *Urban growth and transportation.* Review of Economic Studies 79.4 (2012): 1407-1440.

[20] Duranton, Gilles, and Diego Puga. *Micro-foundations of urban agglomeration economies.* Handbook of Regional and Urban Economics (2004): 2063-2117.

[21] Eaton, Jonathan, and Samuel Kortum. *Technology, geography, and trade.* Econometrica 70.5 (2002): 1741-1779.

[22] Faber, Benjamin. *Trade Integration, Market Size and Industrialization: Evidence from China's National Trunk Highway System.* forthcoming, Review of Economic Studies, 2014.

[23] Feng, Shuaizhang, Yingyao Hu, and Robert Moffitt. Long Run Trends in Unemployment and labour Force Participation in China. No. w21460. National Bureau of Economic Research, 2015.

[24] Fogel, Robert William. Railroads and American economic growth. Books on Demand, 1994.

[25] Ollivier, Gerald; Bullock, Richard; Jin, Ying; Zhou, Nanyan. 2014. High-speed railways in China : a look at traffic. China transport topics ; no. 11. Washington, DC: World Bank Group.

[26] Glaeser, Edward L., and Joshua D. Gottlieb. *The Wealth of Cities: Agglomeration Economies and Spatial Equilibrium in the United States.* Journal of Economic Literature 47.4 (2009): 983-1028.

[27] Glaeser, Edward Ludwig, et al. *Growth in Cities.* Journal of Political Economy 100.6 (1992): 1126-52.

[28] Glaeser, Edward L., and Albert Saiz. *The rise of the skilled city.* Brookings-Wharton Papers on Urban Affairs 2004.1 (2004): 47-105.

[29] Grossman, Gene M., and Esteban Rossi-Hansberg. *"Trading tasks: a simple theory of offshoring.* American Economic Review 98.5 (2008): 1978.

[30] Helpman, Elhanan. *The size of regions.* Topics in public economics (1998): 33-54.

[31] Kolko, Jed. "Urbanization, agglomeration, and coagglomeration of service industries." Agglomeration economics. University of Chicago Press, 2010. 151-180.

[32] L. Jianbin, Survey and Analysis on Passenger Travel Characteristics and Distribution Features of Wuhan Guangzhou High Speed Railway, Railway Standard Design, No. 11, 2011, pp. 1-4.

[33] Liu, Z.(2003): tie lu shi shi kua yue shi fa zhan zhan lue.

[34] A. Marshall, Principles of Economics, Macmillan, London, 1890.

[35] Mart-Henneberg, Jordi. 2000. Un Balance del Tren de Alta Velocidad en Francia. Enseanzas para el Caso Espaol [An Evaluation of the High-Speed Train in France: Lessons for the Spanish Case]. Era 52: 131-43.

[36] Michaels, Guy. *The effect of trade on the demand for skill: evidence from the Interstate Highway System.* Review of Economics and Statistics 90.4 (2008): 683-701.

[37] Michaels, Guy, Ferdinand Rauch, and Stephen J. Redding. *Task specialization in US cities from 1880-2000.* No. w18715. National Bureau of Economic Research, 2013.

[38] Mid-to-Long Railway Plan. The Ministry of Railway of China, 2008

[39] Morten, Melanie, and Jaqueline Oliveira. *Migration, roads and labour market integration: Evidence from a planned capital city.* (2014).

[40] Moretti, Enrico. *Workers' education, spillovers, and productivity: evidence from plant-level production functions.* American Economic Review (2004): 656-690.

[41] Ottaviano, Gianmarco IP, Giovanni Peri, and Gregory Wright. *"Immigration, offshoring, and American jobs.* American Economic Review 103.5 (2013): 1925-1959.

[42] Perlman, Elisabeth Ruth. "Dense Enough To Be Brilliant: Patents, Urbanization, and Transportation in Nineteenth Century America Population." (2014).

[43] Qin, Yu. *No County Left Behind? The Distributional Impact of High-Speed Rail Upgrade in China.*, 2015.

[44] Redding, Stephen J., and Matthew A. Turner. *Transportation costs and the spatial organization of economic activity."* Chapter in progress for the Handbook of Regional and Urban Economics (2014).

[45] Sasaki, Komei, Tadahiro Ohashi, and Asao Ando. *High-speed rail transit impact on regional systems: does the Shinkansen contribute to dispersion?.* The Annals of Regional Science 31.1 (1997): 77-98.

[46] Shiue, Carol H., and Wolfgang Keller. *Markets in China and Europe on the eve of the industrial revolution.* American Economic Review 97.4 (2007): 1189-1216.

[47] Wu K, Fang C L, Zhao M X, Chen C, 2013, The Intercity Space of Flow Influenced by High Speed Rail: A Case Study for the Rail Transit Passenger Behavior between Beijing and Tianjin. Working paper 7, Beijing City Lab

[48] Zheng, Siqi, and Matthew E. Kahn. *China's bullet trains facilitate market integration and mitigate the cost of megacity growth.* Proceedings of the National Academy of Sciences 110.14 (2013): E1248-E1253.