Facilitating Investment Flows:

Evidence from China's High-Speed Passenger Rail Network

Yatang Lin

Department of Economics, Division of Environment and Sustainability, Division of Social Science, HKUST

Yu Qin

Department of Real Estate, National University of Singapore

Johan Sulaeman

Department of Finance, National University of Singapore

Jubo Yan

Division of Economics, School of Social Sciences, Nanyang Technological University

Jialiang Zhang

National School of Development, Peking University

Facilitating Investment Flows:

Evidence from China's High-Speed Passenger Rail Network

Abstract

This paper investigates how transportation infrastructure projects facilitate interregional flows of private investments. Exploiting the staggered expansions of China's passenger high-speed rail (HSR) network as plausibly exogenous shocks to the ease of travel between cities, we document that the introduction of a direct HSR connection between a pair of cities increases the amount of cross-city investment between the pair by 45%, in the presence of high-dimensional fixed effects: city-pair FE to capture static linkages (e.g., geographical distance, cultural proximity) and city-time FE to capture time-series variations in economic dynamics. We enhance the causal inference by exploiting the vertical-horizontal layout of China's HSR network. This allow us to identify the effect of indirect connections that are formed when one part of the network is "accidentally" connected to a different part of the network. The HSR effect is the strongest for investments in industries that benefit more from face-to-face communication and onsite administration, as well as for investors contemplating controlling stakes in large distant investments, consistent with the reduction in monitoring costs associated with travel infrastructure development. The additional investment flows due to (indirect) HSR connections are also associated with potential reductions in the differences of return-to-capital between cities.

Keywords:

High-speed rail, home bias, investment flows, transportation infrastructure development

1. Introduction

The development of transportation infrastructure has improved connection across faraway cities. Railways and highways reduce the cost of moving goods, facilitating interregional trade (Donaldson 2018). Similarly, the improvements in passenger transportation connectivity accelerate passenger flows and open up new possibility of direct communication and interactions across cities. Of particular interest is the recent rapid development of passenger high-speed rail (HSR) network around the world, which provides a fast, convenient, and reliable mode of transportation across a connected network of cities. The most prominent example is China's passenger HSR network; over just the last decade, this network has grown by over 25,000 km, accounting for more than half of the world's total length of HSR tracks, with ambitious plans to increase the network coverage by more than 50 percent over the next decade (Economist, 2017).[1]

We first document the effect of the introduction of a *direct* HSR connection between a pair of cities on the investment flows between those cities. Upon the introduction of a direct HSR connection between a pair of cities, the number of cross-city investors between the city pair increase by 8%, and the amount of investment increases by 45%. The large economic magnitude of the effects of transportation infrastructure reflects its importance in improving investment flows. Identifying this effect complements existing studies on point-to-point air travel connection. However, drawing a causal inference from direct HSR connections is hampered by the concern that HSR lines, similar to air travel routes, might be endogenously placed between pairs of cities with growing economic linkages.

Fortunately, the staggered development and the vertical-horizontal layout of China's HSR network formation allow us to identify the effect of *indirect* connections on cross-city investment flows. These indirect connections are formed when one part

[1]https://www.economist.com/china/2017/01/13/china-has-built-the-worlds-largest-bullet-train-network

of the network is connected to a different part of the network. These connections are likely to be "accidental", and therefore they are less likely to be affected by omitted variable concerns associated with the likely endogenous rationale for establishing direct (point-to-point) transportation links. As such, this analysis allows for a cleaner inference regarding the causal effect of HSR connections on investment flows across regions. We find that the magnitude of the causal effect is about half of the direct effect.

To operationalize our analysis, we employ the dataset of firm registrations in China that includes the information on shareholders of the universe of Chinese firms from 2000 to 2015. This restricted-access firm-level administrative data is uniquely obtained from the China State Administration for Commerce and Industry. We use this dataset to develop inter-city investment flow measures that cover firm activities across all regions and sectors in China. As the dataset covers the whole universe of Chinese firms, we are able to examine and compare various industry sectors, including manufacturing (Giroud, 2013) and services. The dataset also covers investment in private firms, allowing us to extend the analyses to less developed cities that may have been overlooked in analyses that focus on public firms. We can therefore gain insights on how these typically smaller cities are affected by the introduction of transportation network, particularly in relation to larger and more dominant cities in the region.

We adopt a difference-in-differences specification to examine whether the introduction of an HSR connection between a given city pair is associated with an increase in bilateral investments, compared to unconnected city pairs. Our specifications include a relatively stringent set of control variables. In particular, the rich information in city-to-city investment flows allows us to control for the time variation in each city's economic dynamics by including the full set of time varying city fixed effects, i.e., both (origin city × month) and (destination city × month) fixed effects. These allow us to tease out potential omitted variables associated with various conditions, e.g., economic growth, at the city level.

After documenting a positive link between direct HSR connection and cross-city investment flows, we sharpen the causal inference by examining less direct connections.

First, we focus on the pairs of cities that are already on the HSR network but not yet connected to each other with a direct line. Each of these cities have been (previously) selected to be on the HSR network, mitigating the potential bias associated with the (timing of) selection into the network. We observe a similar positive effect on cross-city investment flows when these cities become directly connected to each other.

Second, as mentioned above, we examine pairs of cities that become *indirectly connected* in the HSR network: non-nodal cities in two different HSR lines that become connected as the lines cross each other. In this analysis, we exclude pairs of cities that are ever directly connected by HSR, mitigating any potential bias associated with direct route connection. We find that the introduction of an indirect HSR connection also leads to an increase in cross-city investments: the number of investors between indirectly connected city pairs grow by 4%, whereas the amount of investments grow by 20%, or about half of the effect of direct HSR connection.

The staggered nature of the HSR network formation also allows us to examine the time-series dynamic of the HSR effect. In particular, we analyze the cross-city investment flows from the announcement to the opening of new HSR connections. This allows us to detect potential pre-event effects that could hint at endogenous (pair) selection concerns. The HSR effect is not observed prior to the announcement of the HSR connection, consistent with the network placement decision not directly related to (expected) economic linkages.

There is a weak effect between the announcement and the opening of the connection, which may be related to the anticipation of the upcoming connection. The bulk of the effect, however, occurs after the two cities are actually connected via HSR. While the post-opening effect is almost immediate – we observe statistically significant estimate even in the first three months of the connection, the effect becomes stronger with time afterwards.

We attempt to illustrate the mechanisms underlying the effects of HSR connections by examining the heterogeneity in the HSR effect across city pairs, industries, and ownership structures. We first find that the HSR effects are observed only for city pairs within the same urban cluster, reflecting the superiority of HSR relative to alternative modes of transportation, e.g., air transportation, for these relatively short distances. We then identify core and peripheral cities within each cluster, and examine investment flows across city-pairs within each combination of origin and destination groups (C-to-C, C-to-P, P-to-C, and P-to-P). We observe that within each cluster, the HSR effect is mostly concentrated in investment flows to peripheral cities, both C-to-P and P-to-P, indicating that HSR connection facilitates the decentralization of some industries to smaller satellite cities within urban clusters.

The within-cluster analysis resonates with the renowned Lucas Paradox (Lucas, 1990) that capital does not seem to flow from rich to poor countries/regions. Our analysis provides an indirect answer to the important question of whether the resulting patterns in investment flows reflect capital movements from rich to relatively poor cities that might close regional gaps in development. We perform two more direct tests of whether HSR connections help to alleviate these developmental gaps. First, we find that HSR connections are associated with higher cross-city investment flows to cities in lower income groups. Second, we find affirmative evidence that the additional investment flows induced by HSR connections act to close the gaps in the rates of return to capital across cities and industries. The HSR effect is associated with capital flows from areas with lower rates of return to capital to areas with higher such rates. In sum, these consistent pieces of evidence reflect HSR's valuable role in promoting economic growth and ameliorating regional inequality issues.

We also find evidence consistent with the reduction in travel and monitoring costs associated with the HSR introduction. First, the HSR effect is stronger for investments in industries that benefit more from face-to-face communication and onsite administration. Second, the HSR treatment effects are particularly strong for investors contemplating controlling stakes in large distant investments. The total amount of investments associated with sole (100%) ownership increases by about 35% with HSR connection. This indicates that HSR connection improves the monitoring capabilities of distant investors, allowing them to invest more when taking on larger ownership stakes. Meanwhile, the amount of cross-city investments in small stakes (0-4.9% of the firm) increases by only about 10%, whereas the amount of investments in minority stakes (5-49.9% of the firm) increases by about 14%. Albeit smaller in magnitudes, the existence of the HSR effect for these non-controlling stakes suggests that HSR also helps in reducing information frictions in portfolio investment decisions and helps investors better discover and evaluate investment opportunities in other cities.

Understanding how HSR connections shape investment flows among cities of different sizes, economic developments, and specializations is a crucial step in evaluating the impacts of passenger transportation infrastructure on economic integration. Our analyses are related to two strands of recent literature that examine the impact of transportation infrastructure on various economic outcomes. The first strand examines roads and railways that reduce trade costs. These studies examine the effects of these trade-related infrastructure on interregional trade (Donaldson 2018), local labor markets (Michaels 2008), long-term GDP growth (Banerjee et al. 2012), and asymmetric effects on core and peripheral markets (Faber 2014).[2] To the extent that passenger connectivity also improves communication among agents in different cities that could lead to higher levels of interregional trades, our results provide an additional support for the consensus in this literature that transportation infrastructure reduce trade costs.

Our research is more closely related to the second strand of literature that examines the effects of *passenger* transportation on economic outcomes. Most studies in this literature examine the potential effects of air travel improvements in various settings (e.g., Giroud, 2013). A few recent studies focus on the economic impacts of HSR developments, including a study by Bernard et al. (2019) on the impact of Shinkansen

[2] Other papers have explored the effects of urban transportation improvements on urban growth (Duranton and Turner 2012) and urban form (Baum-Snow et al. 2015, Baum-Snow 2012).

line on supplier relationship among Japanese firms.[3] Several recent studies also examine the impact of China HSR system on housing prices in secondary cities (Zheng and Kahn, 2013), distributional effect in the core and peripheral areas (Qin, 2017), local employment (Lin, 2017), and collaboration in research between Chinese cities (Zheng and Kahn, 2017). Against this backdrop, the current paper identifies a *causal* mechanism at the city-pair level of how infrastructure developments can affect both investment flows and resource allocations across cities.

Analyzing the expansions of HSR network also allows us to isolate the causal effect of information frictions, particularly those associated with spontaneous information and tacit knowledge whose transmission depends critically on face-to-face contacts (Storper and Venables 2004, Glaeser 2011). These frictions are associated with an extensive body of empirical research that documents the strong tendencies of individuals and companies to invest in assets that are geographically close, i.e., “home bias” in both cross-country and within-country setting, across various types of investments (French and Poterba 1991; Coval and Moskowitz 1999; Ivkovic and Weisbenner 2005). The relationship between proximity and investments remains substantial despite the rapid progress in information and communication technologies over the last several decades, highlighting the importance of direct, face-to-face contacts. This paper highlights the effects of exogenous shocks to “distance” on the propensity to seize non-local investment opportunities.

The remainder of the article is organized as follows. Section 2 provides background on the evolution of HSR network in China. We describe our dataset on cross-city investments as well as HSR connection in section 3, and explain our empirical strategy in section 4. The main regressions results at the city-pair level are discussed in section 5. We provide additional discussions in Section 6 and conclude in Section 7.

[3] Other studies examine the effects of HSR on passenger travel behavior (high-speed Eurostar; includes: Berhens and Pels, 2011) and the economy of regions that are made more accessible (HSR connecting Cologne and Frankfurt; Ahlfeldt and Feddersen, 2018)

2. Background

High speed railway (HSR) lines are defined as specially built railway lines running at an average speed of 250 km/h or more, or specially upgraded existing lines running at an average speed of 200 km/h or more (European Union Council Directive 96/48/EC). China's HSR expansion started in 2003, with the first line connecting Qinhuangdao and Shenyang. But the subsequent development of HSR was inhibited by the debate of whether the HSR should be built using conventional tracks or the magnetic levitation (maglev) technology. The rapid development of HSR network began in earnest in 2008, when China's State Council set the goal of forming a national high-speed rail grid consisting of four north-south corridors and four east-west corridors in their Mid-to-Long Term Railway Development Plan using conventional tracks.

The stated aim in 2008 was to develop more than 16,000 kilometers of HSR network before 2020. The network had grown beyond this ambitious goal. By the end of 2017, there were more than 40 HSR lines in operation, with a total mileage of over 22,000 km and 7 billion cumulative number of trips. The expansion of HSR network in China from 2003 to 2016 is displayed in Figure 1.

According to Ministry of Railway's document (2008)[4], the main objective of this expansion is to connect provincial capitals and other major cities with faster means of transportation. Consistent with this objective, HSR connected 29 of China's 33 provincial-level administrative divisions and 163 of 283 prefectural level cities by 2016. This objective guided the placements of lines, which are based on a comprehensive consideration of each region's economic development, population and resource distribution, national security importance, environmental concerns, and social stability. The HSR lines are also expected to complement existing transportation networks to the extent possible.

[4] On March 10 2013, it was announced that the MOR would be dissolved and its duties taken up by the Ministry of Transportation, National Railway Administration, and China Railway Corporation.

China's HSR expansion is centrally managed, planned, and financed by the government. The initial planning allocated the budget of 4 trillion RMB to build the four north-south corridors and four east-west corridors (State Council, 2004). The construction costs of HSR are estimated between 80 to 120 million RMB per km (US$13–20 million) excluding stations (Bullock et al., 2012). The financing of HSR involves very limited private investment. About half of the financing is provided by the national government through lending by state owned banks and financial institutions, another 40% by bonds issued by the Ministry of Railway (MOR) and the remaining by provincial and local governments, mainly through compensation for land use.

As the network has grown rapidly over the past decade, so has the ridership of HSR. China's HSR network is the world's longest and also the most extensively used, with 1.713 billion trips taken in 2017 bringing its total cumulative number of trips to 7 billion. In the recently revised version of Mid-to-Long Term Railway Development Plan approved by the State Council in 2016, the HSR network will be expanded to eight north-south corridors and eight east-west corridors.

Travelling by HSR is particularly attractive for short-to-medium distance business trips owing to its convenience, high frequency, low price, and punctuality, relative to its main alternative of air travel. This can be illustrated with a simple example. Travelling from Beijing to Shanghai by air typically takes 2.5 hours, from taking off to landing; travelling by HSR takes about 4.5-5 hours. While the HSR takes longer in terms of pure travel time, the total travel time is quite similar as HSR allows passengers to skip the procedure of arriving at the airport at least two hours in advance, the check-in process, and the additional traveling time to/from the airports as HSR stations are often located closer to downtown areas. As HSR is more comfortable than air travel and costs only half, it has become a major means of transportation, particularly for short-to-medium distance business trips.

3. Data

3.1 Firm investment dataset

The Firm Registration Database is maintained by the China's State Administration for Industry and Commerce (SAIC). The database contains the administrative information of the whole universe of enterprises in China, covering over ten million registered firms. At the date of registration, all firms are required to disclose to the SAIC the following information: the firm location, industry code, and ownership type; their legal representatives, shareholders, and executives; the value of registry capital; and the year of establishment. The database we use in this paper is restrictively obtained from the SAIC and it is thus far the most comprehensive data on firm activities across all regions and sectors in China.

We use the records in the Firm Registration Database to measure financing activities at the firm-to-firm level, i.e. a firm contributing capital to another firm and thereby becoming its shareholder. When such activities occur, the firms and natural persons involved are required to report the investment to the SAIC within the same calendar year. As a result, the Firm Registration Database contains records for all such investments between firms or from natural persons to firms during the period of year 2000 to year 2015.

Our paper focuses on the firm-to-firm investment activities. The total number of observations of such activities in the database is 1,814,851. We also restrict the sample to investments in which the receiver is a new firm, i.e., the investment activity occurs within the calendar year in which the receiver firm is firstly registered with the SAIC. With this restriction, our sample essentially excludes changes in shareholdings of incumbent firms, which are recorded with less precision in the Firm Registration Database, and therefore are likely to be plagued by measurement errors. Our analysis sample consists of 1,312,416 firm-to-firm investment observations.

3.2 HSR network and rail travel times

Most of the information on the Chinese HSR system, including construction starting date, opening date, track length, designed speed, and ridership on selected lines, is obtained from the China Railway Yearbook's Major Events, Finished, and Ongoing Projects sections from 1999 to 2012. For a small proportion of lines that are opened in 2013 and 2014 as well as future HSR lines under planning, this information is not available from the most recent (2012) yearbook. We employ official news published on http://news.gaotie.cn as well as other online news sources for this subset. We verify the information on the stops along each existing line using the official railway service website (www.12306.cn). The announcement dates of each HSR line are collected from online official news as well.

In the analysis, we focus only on prefecture-level cities, which exclude prefecture-level autonomous regions, leaving 283 cities in each cross-section. The prefecture level social economic variables are drawn from China City Statistical Year Books from 2007 to 2015, such as GDP, population, average income, average ridership, etc.

4. Empirical strategy

Our empirical investigation is motivated by the literature exemplified by a recent study by Giroud (2013), which documents that the introduction of new airline routes increases plant-level investment and total factor productivity, by making it easier for headquarters to monitor and acquire information about plants with the shorter travel time.

4.1 Baseline specification

We adopt a difference-in-differences specification to examine for a given city pair, whether or not HSR connection between them leads to an increase in bilateral firm investments, compared to the unconnected ones. The rich information available in city-to-city investment flows allows us to control for a full set of origin/destination city interacting with month fixed effects. The specification takes the form of:

$$y_{ijt} = \theta connect_{ijt} + \alpha_{ij} + \beta_{it} + \gamma_{jt} + \delta_t + \epsilon_{ijt} \quad (1)$$

where subscript i denotes the origin city, j denotes the destination city, t denotes time at year-monthly frequency. We aggregate the firm-level portfolio investment records in the Firm Registration Database to city pair and year-monthly level.

In our analysis, we examine both the intensive and extensive margins of investment flows. To examine extensive margin, we define the dependent variables y_{ijt} as the logarithm of the number of unique investment pairs between city i and city j within year-month t. Alternative, we define y_{ijt} as the logarithm of the sum of investment flow from city i to city j within year-month t, to examine the intensive margin. The main coefficient of interest is θ, which measures the effect of the introduction of new HSR connections on cross-city investments. In the dataset for the benchmark setting, each observation is a directed dyad for two different cities. Altogether, the sample consists of 283 prefectural cities and 11,499,062 city pairs for the 2004 to 2015 period, at monthly frequency.

We control for bilateral city pair fixed effects (α_{ij}) throughout the baseline specification to address the following potential endogeneity problem. Some city pairs may have systematically more cross-city investments and are more likely to be connected to HSR than other cities, for these city pairs have a closer social and economic relationship. By inserting city pair fixed effects into our regression framework, we can control for all unobserved and non-time varying heterogeneity at the city pair level. Therefore, in all regressions, the relation between connection to HSR and the outcomes of interest is generated by the expansion of the HSR network over time.

We also include the full set of (origin city × time) and (destination city × time) fixed effects (β_{it}, γ_{jt}) to address the potentially endogenous placements of HSR lines. As mentioned in section 2, the placement of China's high-speed rail lines is centrally managed, planned and financed by the government, taking economic development, population and resource distribution, national security, environmental concerns and social stability into consideration. One major concern is that there might exist some local unobserved heterogeneity, e.g. local growth potential, which could have determined the governmental decision to build the HSR infrastructure, and which might be also correlated with outcome variables, thus biasing our results. Another concern is that the HSR infrastructure itself might have promoted local economic growth, a phenomenon we want to control for in our regressions. There are innumerable local factors could be omitted variables that are driving both the connection of HSR and firm investment, thus rendering any inference regarding the observed relationship between the two problematic. We address this concern using the identification strategy similar to Giroud (2013) and Giroud and Mueller (2015), i.e., by including the full set of origin/destination city × time fixed effects (β_{it} and γ_{jt}), on top of city pair fixed effects (α_{ij}). These time-varying dummy variables should capture the time-varying local heterogeneity and dynamics that affect the attractiveness of destination cities and the investment capacity of investor cities, and therefore remove these local factors from mitigating concerns in our setting.

Lastly, even with all the solutions taken above, if the new HSR connection between two cities is still partially endogenous because of some (shared) pre-existing shock, then we should observe the treatment "effect" even before the plan to build a new HSR line is announced. To investigate this issue, we estimate the dynamic effects of HSR announcement on investment flows in a city pair. We collect news published on official sites to identify the month in which a new HSR line is announced to be constructed in the future. It usually takes three to four years from the announcement to the operation of a new HSR line. The equation for this estimation is specified as the following:

$$y_{ijt} = \sum_{\tau=-12}^{-1} \theta_\tau \, Announcement_{ijt} \times 1\{Month_{ijt} = t + \tau\}$$

$$+\pi \, PostAnnouncement_{ijt} + \sum_{\tau=0}^{25+} \vartheta_\tau \, Connect_{ijt} \times 1\{Month_{ijt} = t + \tau\}$$

$$+\alpha_{ij} + \beta_{it} + \gamma_{jt} + \delta_t + \epsilon_{ijt} \qquad (2)$$

where τ stands for the event year-month of announcement or the event year-month of connection; $Announcement_{ijt}$, $PostAnnouncement_{ijt}$, and $Connect_{ijt}$ are dummy variables denoting the events of announcement, post announcement (the period between announcement and connection), and connection between city i and city j at time t. $1\{Month_{ijt} = t + \tau\}$ is a dummy variable that takes the value 1 for month $t + \tau$ and otherwise 0. Notice that the timings of treatment are not the same for all city pairs. We use 13-24 months before the announcement as the benchmark period. Therefore, the coefficient θ_τ estimates the impact of announcement of HSR connection before the actual announcement, relative to 13-24 months before announcement. If the parallel trend hypothesis holds for the difference-in-differences specification, θ_τ should be indifferent from zero for any time prior announcement. π estimates the announcement effect while ϑ_τ estimates the connection effect. The rest of the specification is the same as Equation (1).

4.2 Measurement of indirect HSR connection

In the baseline identification strategy, we take advantage of the abundant information of city dyads setting to control for large sets of fixed effects, in order to address the time varying local heterogeneity problem. Another threat to identification is that transportation linkages are more likely to be established between two cities with closer and strengthening economic ties. In other words, part of the city pairs experiencing connection to HSR and decreases in travel times is endogenously selected. The strategies we discuss in the previous section may be inadequate depending on the causes of this endogenous selection. Ideally, if we have information indicating criteria of governmental decisions about where HSR extensions are taken, we can restrict our sample to more exogenously constructed city pairs. However, these governmental documents could be confidential and are difficult to be observed by researchers.

To mitigate the endogeneity concern of city pair connection patterns, we use an indirect connection measure as an independent variable. This indirect connection concept exploits the idea that the whole HSR network in China is an extensive network consists of four main horizontal lines and four vertical lines (see Figure 2). When each horizontal line is attached with a vertical line, non-nodal cities along both lines become indirectly connected. These indirect connections are largely unplanned, particularly once we condition on the interaction of origin/destination and time trend. As such, by restricting our treatment group to city pairs that are indirectly connected by HSR, the threat to identification caused by endogenous selection can be more credibly ruled out. In this setting, we drop city pairs that are ever directly connected by HSR. The specification takes the form of:

$$y_{ijt} = \theta indirect_{connect_{ijt}} + \alpha_{ij} + \beta_{it} + \gamma_{jt} + \delta_t + \epsilon_{ijt} \quad (3)$$

We construct the measurement of indirect connection of HSR as dummy variable that equals 1 for a city pair-month triad *ijt* if (1) the origin *i* (destination city *j*) is located along a segment A of a horizontal line mentioned above and the destination city (origin city) is located along a segment B on a vertical line; (2) the line segments A and B are connected at month *t*; (3) the pair is not directly connected.[5] It should be noted that indirect connection is an intention-to-treat measure as the two cities might not be accessible through HSR at month *t* if there are gaps between the two line segments they are at.

5. Empirical Findings

5.1 Descriptive statistics

Table 1 reports the summary statistics for all 11,499,062 city pair- month observations in Panel A. In Panel B, observations are categorized by whether the city pair is ever directly connected by HSR at any point during our sample period. The ever-connected and never-connected samples are very different in terms of the extensive and intensive margin of inter-city investment flows. On average, city pairs that are ever connected to HSR are more likely to have inter-city investment flow on the extensive margin, and

[5] A line segment is defined as a part of the whole vertical/horizontal line that started operating at the same date.

the amount of inter-city investment is 28 times higher than city pairs that are never connected to HSR on the intensive margin. Therefore, as mentioned in the empirical strategy section, we only use the city pairs that have ever connected by high-speed rail as the regression sample to ensure the comparability of the control and treated units.

In Panel C, we segregate the ever-connected city pairs into the periods before and after they are connected by HSR. These city pairs experience three times increase on the extensive margin and four times increase on the intensive margin after being connected to HSR network, compared to before the connection is established.

Table 2 reports the summary statistics for the sample that we use after switching the treatment to indirect HSR connections. The corresponding gaps of cross-city investments between groups are slightly smaller, but still remain of large economic magnitude.

5.2 Main specification

The parameter estimates from our baseline specification in equation (1) are reported in Table 3. In Panel A, we include all city-pair-month observations in our sample. Columns (1) and (2) show the effect of the direct connection to HSR on cross city investment on the extensive margin. The dependent variable $Lnumber_{ijt}$ is the logarithm of unique firm investment pairs from city i to city j within month t. *Connect* is a dummy variable indicating whether a city pair (i,j) is connected by HSR at month t. Column (1) includes city-pair fixed effects and year-month fixed effects. Column (2) is the more complete specification which controls for origin city × year-month and destination city × year-month fixed effects that allows for a very flexible functional form of origin and destination city time trend. As reported in Column (2), the connection dummy is 0.08 with statistically high significance, which implies that the introduction of an HSR connection between two cities increases the number of inter-city investor-receiver pairs by 8%, compared to the control city pairs that are connected to HSR at a later date.

Columns (3) and (4) of Table 3 present the effect of HSR connection on the intensive margin of cross city investments. The dependent variable $Linvestment_{ijt}$ is the logarithm of the sum of investment flow from city *i* to city *j* within month *t*. In the more stringent specification that include city-month fixed effects in Column (4), the coefficient on the treatment dummy is 0.375, which implies that the amount of inter-city investment increases by 45.5% on average with the introduction of HSR connection.[6] Given that the sample mean of cross-city investment flow is 4.29 million RMB in the pre-treatment period, the increase associated with being connected to HSR corresponds to an increase in capital expenditures of 1.95 million RMB per city-pair-month in the post-treatment period.[7]

Panel B reports estimates using inter-city investments originated and received by Privately Owned Enterprises (POEs). This allows us to alleviate the concerns owing to the large flow of investments made by State Owned Enterprises (SOEs) in China. This type of investments is problematic in our context as these SOEs might be directed by the government to make investments in certain cities, which could be done in concert with the introduction of HSR connections. The estimates obtained using the sample that excludes investments in which either the investor or the receiver is an SOE are almost identical to those in Panel A.[8]

Another potential concern with the data is that some firms being registered in the SAIC database are not intended for legitimate business purposes but rather for other purposes such as receiving subsidies from the government. To address this concern, we replicate our main regression using a subsample of the dataset that is restricted to

[6] We calculate the implied growth rate using exp(coefficient)-1 for all the significant coefficients larger than 0.1.

[7] Readers may worry about the quality of those newly registered firms. Ideally, we would like to investigate the performance of these firms. Unfortunately, we do not have such data. However, we note that the China State Administration for Industry and Commerce conducts audits on certain registered firms to record the survivorship of those firms, which we use to further refine the outcome variables by only counting the number and the corresponding investment flows of firms that survived by the end of 2015, or have survived for at least 3, 4, and 5 years for each city pair-month observation. The results are reported in Appendix Table A2 (for the extensive margin) and Appendix Table A3 (for the intensive margin). These results are consistent with the main results in Table 3.

[8] Regression results on other groups: SOE to SOE, SOE to POE, and POE to SOE are reported in Table A1 in the appendix. The results show that HSR has no significant effect on investment if the investors are SOEs.

investments before the end of year 2013, during which the SAIC imposed more stringent rules for new firm registration.[9] The results are reported in Table A4 in the appendix. The coefficient estimates obtained using this subsample are of very similar magnitudes to the estimates using the whole sample reported in Table 3. This indicates that the incidence of potentially bogus firm registrations is unlikely to be correlated with the observed effects of HSR connections.

5.3 Endogenous HSR placement

We are acutely aware of the potentially confounding effects of the endogenous selection of HSR stations. Our baseline identification strategy includes the origin and destination city time fixed effects in addition to the time invariant city-pair fixed effects that should capture unobservable variations at the city pair level. This allows us to rule out endogeneity concerns that are related to the time-series dynamics at the city level.

However, this approach is not immune to a potential endogeneity concern owing to the possibility that the selection of HSR stations is based on the (shared) economic activities at the city pair level. To evaluate whether this is a valid concern, we perform two related tests. First, we examine whether the HSR's positive effect is mainly driven by newly connected city pairs that includes at least one new HSR station (reflecting the endogenous selection of that HSR station), instead of a new connection between two existing HSR stations. As presented in Table 4, the interaction of "NewHSR" (an indicator variable that takes the value of 1 if at least one new HSR station is involved in this newly connected pair) and Connect is negative in all specifications, and statistically significant for the extensive margin analysis (columns 1 and 2). This indicates that the positive effect of HSR connection is mostly due to the new connections of *existing* HSR stations, mitigating the concern that the results are driven by the endogenous selection of cities hosting new HSR stations. Instead, the

[9] The *Regulation of the People's Republic of China on the Administration of Company Registration* together with several other regulations were amended in early 2014 to ease the procedures of registering a firm. One of the major amendments is that SAIC no longer requires the registered capital to be fully credited to the newly registered firm. Before this amendment, SAIC requires the investors to credit the registered capital to the account of the newly established firm before the new firm can be registered.

positive effect is driven by the additional exposure to the HSR network experienced by existing host cities on the network.

Second, in order to further mitigate the endogeneity concern of high-speed rail placement, we examine whether the HSR's positive effects also exist for cities that are indirectly connected to the network. In particular, we drop city pair observations that are directly connected by high-speed rail, and identify the city pairs that are indirectly connected by high-speed rail as the treated sample. As illustrated in section 4, by restricting our treatment group to city pairs that are indirectly connected by HSR, which are much less likely to be planned in advance, the threat to identification caused by endogenous selection can be ruled out more explicitly. Table 5 reports the results of estimating equation (3), using indirect connection as the treatment variable. In the first two columns of Table 5, we investigate the relationship between HSR indirect connection and city pair investment on the extensive margin. The magnitude of the coefficient estimates on this indirect treatment is almost the same with direct connection. In particular, we observe that the indirect connection of HSR between two cities increases the number of investments between the city pair within the same month by approximately 3.8%, taking the coefficient in column (2), the most complete specification. In the last two columns of Table 5, we examine the intensive margin of cross city investments. The coefficient estimate on the treatment dummy of indirect connection is 18.5% in column (4), which implies that investment flow increases by 20.3%. The magnitude is lower relative to direct connection, but still economically significant, and can be considered as a lower bound of the estimated effect.[10]

5.4 Robustness checks

In this sub-section, we address two potential concerns regarding our identification strategy. First, the construction of high-speed rail tracks takes time. Investors may

[10] Similar to the results on direct connection, we also replicate the analysis with indirect connection using the subsample before the end of 2013. The results are reported in Table A5 in the appendix. With the subsample, the estimated coefficients are also statistically significant with magnitudes very similar to the full sample., w

respond to the news once they learn about the announcement of high-speed rail connection between two cities. Second, our difference-in-differences approach requires the treatment group and the control group to maintain parallel trends in the absence of the treatment.

We address the above concerns in two ways. First, we examine the dynamic effect of HSR connection using the event study specification shown in Equation (2), which incorporates the announcement time of the high-speed rails into the regression equation. This specification includes the pre-announcement window, the after-announcement-before-connection window and the post-connection window. This allows us to examine whether the effect of HSR connection starts immediately after the announcement and before the actual HSR connection is established. The coefficients are plotted in Figure 3 for each θ_τ as well as the 95% confidence intervals[11]. The benchmark group is 13 to 24 months before announcement of each HSR connection. The first panel presents the estimates for the extensive margin, i.e., using the indicator variable of the incidence of cross-city investment as the dependent variable, whereas the second panel presents the estimates for the intensive margin, i.e., using the logarithm of investment amount as dependent variable.

Compared to the benchmark period, there is no significant increase in city-pair investment in the one year before the announcement of HSR connection, both at the extensive and intensive margins. This helps us validate the parallel trend assumption of the difference-in-difference design. After the announcement of the HSR route connecting a pair of cities, there is a 3% increase in the number of newly registered firms and a 16.6% increase in the dollar investment flow across city pair after the announcement but before the real connection of HSR. However, these two coefficients are only marginally significant at the 10 percent level. The effects of HSR connection become significant and increase over time after the connection is initiated. The HSR effects increase from 6.5% at the extensive margin in the first four months of the

[11] Please refer to Appendix Table A6 for the regression coefficients.

connection, to 24% after two years of the connection. The corresponding increase in the intensive margin is 46.2% in the first four months to 130.7% after two years of the HSR connection.

Second, instead of using event study, we include an announcement dummy to control for the announcement effect, and a pre-announcement dummy to test for the validity of the parallel trend assumption. The announcement dummy equals to 1 if the city-pair-month observation is during the period after the announcement but before the operation of the HSR connection, whereas the pre-announcement dummy equals to 1 if the city-pair-month observation is during the six-month or one-year period before the announcement of the HSR connection.

Table 6 reports the estimates of this regression model. The parameter estimate on the HSR connection indicator variable is still significantly positive and large in magnitude after controlling for the announcement and pre-announcement dummy variables. Furthermore, the pre-announcement dummy variable is not significant in all the specifications, indicating a parallel trend between the treated and control city pairs before the announcement of the HSR connection. In addition, the announcement effect is quite weak; it is only marginally significant in the specifications that do not include origin and destination specific time trend.

6. Discussions

6.1 Discussion on potential mechanisms

The results presented in the previous sections do not shed light on the potential mechanisms that drive the sensitivity of cross-city investments to reductions in travel cost associated with HSR connections. Some theories in the literature on investment home bias argue that proximity is associated with improvements in monitoring capabilities. In particular, direct monitoring requires shareholders to travel to plants so that they can gather "soft" information, that is, information that "cannot be credibly transmitted" and "cannot be directly verified by anyone other than the agent who produces it" (Stein, 2002, p. 1891). Proximity can also facilitate access to information, i.e. the discovery of new investment opportunities. However, there is limited empirical evidence to identify and disentangle these underlying mechanisms. In this section, we explore several dimensions of firm heterogeneity to evaluate the monitoring and information channels.

First, we categorize all firm-level investments into four groups according to the shares held by the investors, namely investors who hold 100%, [50%, 100%), [5%, 50%), and (0, 5%) of registered capital of the new firm, separately. HSR connection can benefit controlling shareholders in terms of both monitoring capability and information accessibility, while non-controlling shareholders are more likely to benefit only through the information advantage channel. In particular, we hypothesize that if the access to information channel is important, we should observe positive impacts of HSR on investments even from investors who do not have real control over the firms, e.g., because they only invest in 5% of the firm. For these small shareholders, it could be easier to identify new investment opportunities, and build connections with local government or businesses to gain private information through frequent travels via HSR.

Figure 4 displays the estimated coefficients for each different group with different cutoff of shareholders, along with the 95% confidence intervals. We observe that the magnitude of the coefficient estimates of the HSR connection variable obtained using the subsample of controlling investors (e.g., those with shareholding more than 50%) is about twice those obtained using the subsample of non-controlling investors. However, the estimation using non-controlling investors also yields statistically significant (10% level) coefficients, indicating that the better access to information after the HSR introduction constitutes a relevant channel driving inter-city investment flows. Nevertheless, we can identify the role of both monitoring channel and information channel.[12]

Second, to provide evidence on sectoral heterogeneity, we divide receiver firms

[12] We report the full estimation results in Appendix Table A8.

into 20 different groups based on their specific industries. Figure 5 shows the estimated coefficients as well as their 95% confidence intervals using our main specification.[13] Panel A presents the extensive margin and Panel B presents the intensive margin. The results show heterogeneous effects of HSR connections across different industries in terms of both the extensive and intensive margins. In terms of extensive margins, the estimated coefficients are statistically significant (at 5% level) if the receiver firms fall into the following four industries: “leasing and business service”, “wholesale and retail trade”, “real estate”, and “scientific research and technology”. In terms of intensive margins, the coefficients are also significant for firms in the “information technology” and “construction” industries, in addition to the four industries above. Compared to other industries, the industries for which the HSR connection seems to have the strongest effect tend to require more face-to-face communications and more on-site monitoring. In other words, the ease of people movement owing to more efficient transportation infrastructures is more important for the operation of businesses in these industries. As such, the observed patterns in Figure 5 is consistent with our conjecture that HSR reduces travel cost, and facilitates the movement of people and the flow of information.

Figure 5 also provides information on each industry’s share of skilled workers, defined as the share of employees with college education in the industry. The industries in both panels are ranked according to their share of skilled works from the largest to the smallest. Industries employing a high proportion of skilled workers are likely to require more direct interactions to maximize their marginal value. The higher value of direct interactions with high skilled workers can be in the form of synergy – as they are more likely to be involved in non-standardized job tasks (such as idea generation) instead of standardized job tasks (such as running machines) common to low skilled workers – or in the form of monitoring – as preventing abuses by high skilled workers

[13] We included all the 20 industries when estimating the coefficients and corresponding confidence intervals but only showing those industries with received investment of more than 1% of the total investment in the graph.

are more crucial due to their high compensation and access to company's sources of value (e.g., financial information or trade secrets). As such, we expect industries with higher fractions of skilled workers to benefit more from the introduction of new HSR connections. The pattern in Figure 5 is again consistent with this hypothesis.

6.2 Discussion on capital distribution

Our results so far are consistent with the hypothesis that high-speed rail connections facilitate investment flows across cities. A potential concern for policymakers is that the effects of the HSR connections, which are typically portrayed as linking mostly more developed cities, on investment flows reflect only capital moving between more developed cities instead of redistribution of capital to relatively less developed cities. More importantly, one may be concerned that HSR connections may even reduce investments to less developed cities. The answer to such questions depends on the source and direction of capital flows. In a seminal work, Lucas (1990) points out the surprising paradox that capital does not always flow from developed to developing areas despite the presumably higher return to capital in developing areas. One potential explanation for this paradox is that investors from more developed areas face information frictions in investing in less developed areas. In our context, we expect that the infrastructure development reducing intercity transportation cost would bring about substantial effects on flows from developed to less-developed areas.

The development of HSR network has been cited as instrumental in China’s grand plan of developing 19 urban clusters.[14] The plan is to foster the integration of giant urban clusters, anchored around central hubs surrounded by smaller cities. The plan calls for 19 clusters in all, which would account for nine-tenths of economic activity. HSR could improve the integration of cities within an urban cluster by fostering the outsourcing of central hub functions to nearby smaller satellite cities following reductions in travel cost.

[14] https://www.economist.com/china/2018/06/23/china-is-trying-to-turn-itself-into-a-country-of-19-super-regions

In Table 7, we divide all cities in our sample into city clusters based on the official definition.[15] To answer the question that whether the impact of HSR connections on investment happens within or cross different clusters, we add an extra interaction term of HSR connection status and an indicator of whether or not both cities belong to the same city cluster. As shown in column (1) and (3) of the table, the impact of HSR connection only pertains to the within cluster investments in terms of both the extensive and intensive margins. These effects are also much larger when compared with the estimated main effects in Table 3.

In column (2) and (4) of Table 7, we further explore the direction of investment flows within city clusters by focusing on origin-destination pairs within clusters and divide them into four subgroups: core-to-core, core-to-peripheral, peripheral-to-core and peripheral to peripheral. The results suggest that HSR's within cluster effects are mostly concentrated in core-peripheral and peripheral-peripheral flows, which indicates that HSR connection facilitates the decentralization of some industries within urban clusters from the core city to smaller satellite cities.[16]

The assessment of the overall impact of HSR on the aggregate national investment also depends fundamentally on whether HSR connection creates new economic activities or simply reallocates investments from unconnected to connected cities. Therefore, we examine the potential spillover effects to the neighboring cities of HSR-connected cities. Conceptually, the sign of these spillover effects is ambiguous. On the one hand, being close to the treated cities could be associated with a stronger displacement effect: investments intended for some cities might flow to their neighbors that are selected for HSR connection. So a city's inward investments might drop after its neighbor is connected to the HSR network. On the other hand, the proximity to a nearby HSR station might also lead to a boost in accessibility for untreated neighbors

15 Table A9 lists official city cluster definitions in details.

16 We conducted similar analysis by dividing cities into poor, median, and rich groups. We obtain consistent inference from the results reported in Table A10.

as potential investors can also travel to other destinations via HSR after a transfer. We investigate the possible direction and magnitude of spillover effects in Tables A7 by examining investment inflows into neighboring unconnected cities. We find that the positive accessibility effect dominates: cities within 100km to an HSR city experienced on average 7% additional growth in inward investments following the introduction of the neighbor's HSR connection relative to cities further away (100-200km) in the control group. This result suggests that the positive spillover effect dominates the potential displacement effect.

Lastly, we conduct analyses to examine whether the HSR connections help capital to flow to cities with higher return rates. Such flows may help to reduce the gap in Lucas paradox. Using firm inspection data[17], we calculate the rate of return to capital (ROC) as the ratio of profit to asset for all inspected firms. We then aggregate this measure to the city-industry level, and calculate the gap between each pair of cities by industry. This ROC gap variable is positive if the destination city has a higher return-to-capital rate than the source city for a particular industry, and negative otherwise. We interact this gap variable with the HSR connection dummy and report the parameter estimates in Table 8.[18] The results indicate that the HSR connections help capital to flow to cities with higher return-to-capital rates. As shown in Panel A, for every percentage point increase in ROC gap, the direct HSR connections induce an additional 2.2% newly established firms (extensive margin) and a corresponding increase of 8.5%

[17] The SAIC's inspection database includes annual firm-level information on assets, sales and profits from 2009 to 2012, with coverage expanding over time. The information is self-reported by each firm randomly inspected by SAIC. We link the registration and inspection database by firm's ID.

[18] When calculating the difference in ROC within each city pair, we always compare the same industry. That is, a particular industry in the destination city is only compared with the same industry in the source city. If the firm inspection data measuring the return to capital for an industry does not exist in either the destination or the source city, the industry is then dropped. The specification of the regression models in Table 8 are the same as equation (1) and (3) except that the dependent variable is investment (number or amount) from source city i to industry k in destination city j within month t. Only the ever-connected city pairs are included in the analysis.

in investment amounts (intensive margin). The results using the indirection connection measure are similar, as shown in Panel B.

7. Conclusions

Transportation plays an important role in the location, agglomeration, and evolution of economic activities. Yet, relatively little attention has been paid to the cost of moving people and their implications for economic integration and development. The current study focuses on how the reduction in passenger travel cost facilitates information and capital flows cross-city. By exploiting shocks to travel cost from the expansion of HSR system in China, we evaluate how the reduction in travel costs facilitates cross-city investments. To better deal with the possible endogenous city-pair level HSR network formation, we exploit the indirect high-speed rail connections of non-nodal cities on vertical and horizontal railway lines, which are unlikely connected on purpose.

We find that direct HSR connection increases the number of investments between the city pair by 8%, and amount of investment increases by 45%. The results are robust when we control for city-pair heterogeneity and time-varying local shocks that could potentially drive the selection of new HSR routes, and when we consider only city pairs that are indirectly connected to HSR. Moreover, the introduction of HSR connection increases inter-city investment for both controlling and non-controlling shareholders, indicating that both improved monitoring capabilities and access to information serve as important underlying channels of the HSR effect. Industry-wise, the effect is the largest among sectors that require more face-to-face communications, such as leasing, real estate, wholesale and retail trade, scientific research and so on, manifesting the role of HSR on reducing the costs of moving people around. From a regional perspective, the effect of HSR on cross-city investment is largely within a regional cluster (rather than cross regional clusters), and mostly concentrated in core-to-peripheral and peripheral-to-peripheral flows, suggesting that HSR connection facilitates the decentralization of some industries within urban clusters from the core city to smaller satellite cities.

It is worth noting that the quantification of HSR's welfare effect is not within scope of the current paper. However, we find that the displacement effect of HSR is limited for our outcome. Moreover, we show that the incremental cross-city investments associated with HSR connections help to close the return-to-capital gap between the origin and destination cities, which suggests welfare improvement from the capital misallocation perspective.

With our focus on the passenger rail network, our paper is notably distinct from prior studies on the effects of goods-shipping transportation infrastructure such as highways or traditional railroads. HSR is a convenient mode of passenger transportation between cities, facilitating face-to-face contacts among economic agents. This opens up new possibilities of communication and interactions, with potentially transformative effects on economic integration and regional development. Understanding how the reduction in passenger travel cost facilitates information and capital flows is therefore an important step in evaluating the general economic impacts of HSR and other large travel infrastructure projects.

References

Ahlfeldt, Gabriel M and Arne Feddersen. (2018). From Periphery to Core: Measuring Agglomeration Effects using High-speed Rail. Journal of Economic Geography, 18(2): 355–390.

Baik, Bok, Jun-Koo Kang and Jin-Mo Kim. (2010). Local Institutional Investors, Information Asymmetries, and Equity Returns. Journal of Financial Economics, 97: 81–106

Behrens, Christiaan, and Eric Pels. (2012). Intermodal Competition in the London-Paris Passenger Market: High-speed Rail and Air Transport. Journal of Urban Economics, 71(3): 278-288.

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

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

Bernard, Andrew B., Andreas Moxnes, Yukiko U. Saito (2019). Production Networks, Geography and Firm Performance. Journal of Political Economy, 127(2): 639-688.

Bernstein, Shai, Xavier Giroud, and Rick Townsend (2016). The Impact of Venture Capital Monitoring. Journal of Finance, 71 (4): 1591-1622.

Bullock, Richard, Andrew Salzberg, and Ying Jin. (2012). High-Speed Rail – The First Three Years: Taking the Pulse of China's Emerging Program. China Transport Topics No. 04, World Bank Office, Beijing.

Charnoz, Pauline, Claire Lelarge and Corentin Trevien. (2018). Communication Costs and the Internal Organization of Multi-Plant Businesses: Evidence from the Impact of the French High-Speed Rail. The Economic Journal. 128 (610): 949-994.

Coval, Joshua D., and Tobias J. Moskowitz. (1999). Home Bias at Home: Local Equity Preference in Domestic Portfolios. The Journal of Finance, 6 (12): 2045-2072.

Coval, Joshua D., and Tobias J. Moskowitz. (2001). The Geography of Investment: Informed Trading and Asset Prices. Journal of Political Economy. 109 (4): 811-841.

Dong, Xiaofang, Siqi Zheng, Matthew E. Kahn. (2018). The Role of Transportation Speed in Facilitating High Skilled Teamwork. NBER Working Paper No. 24539.

Donaldson, Dave. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. American Economic Review, 108 (4-5): 899-934.

Donaldson, David and Richard Hornbeck. (2016). Railroads and American Economic Growth: a "Market Access" Approach. Quarterly Journal of Economics, 131: 799-853.

Duranton, Gilles, Peter M. Morrow, Matthew A. Turner. (2014). Roads and Trade: Evidence from the US. Review of Economic Studies, 81(2): 681–724.

Faber, Benjamin. (2014). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. Review of Economic Studies, 81: 1046-1070.

French, K. R., and J. M. Poterba. (1991). Investor Diversification and International Equity Markets. American Economic Review, 81:222–26.

Giroud, Xavier. (2013). Proximity and Investment: Evidence from Plant-level Data. Quarterly Journal of Economics, 128(2): 861–915.

Giroud, Xavier and Holger M. Mueller. (2015). Capital and Labor Reallocation within Firms. Journal of Finance, 70(4): 1767-1804.

Edward L. Glaeser and Matthew G. Resseger. (2010). The Complementarity Between Cities and Skills. Journal of Regional Science, 50 (1): 221-244.Huberman, Gur. (2001). Familiarity Breeds Investment. Review of Financial Studies, 14 (3): 659-680.

Ivkovic, Zoran, and Scott Weisbenner. (2005). Local Does as Local Is: Information Content of the Geography of Individual Investors' Common Stock Investments. Journal of Finance, 11 (1): 267-306.

Krugman, Paul. (1991). Increasing Returns and Economic Geography. Journal of Political Economy, 99: 483–499

Lin, Yatang. (2017). Travel Costs and Urban Specialization Patterns: Evidence from China High Speed Railway System. Journal of Urban Economics, 98: 98–123.

Lucas, Robert. (1990). Why Doesn't Capital Flow from Rich to Poor Countries? American Economic Review, 80 (2): 92-96.

Michaels, Guy. (2008). The Effect of Trade on the Demand for Skill - Evidence from the Interstate Highway System. Review of Economics and Statistics, 90(4): 683-701.

MRC. (2008). Mid-to-Long Term Railway Network Plan. Report of the Ministry of Railways of China.

Ollivier, Gerald, Richard Bullock, Ying Jin and Nanyan Zhou. (2014). High-Speed Railways in China: Traffic Analysis. The World Bank Report: China Transport Topics No.11.

Petersen, Mitchell A., and Raghuram G. Rajan, (2002) Does Distance Still Matter? The Information Revolution in Small Business Lending. Journal of Finance, 57(6): 2533–2570.

Qin, Yu. (2017). No County Left Behind? The Distributional Impact of High-speed Rail Upgrades in China. Journal of Economic Geography, 17(3): 489–520.

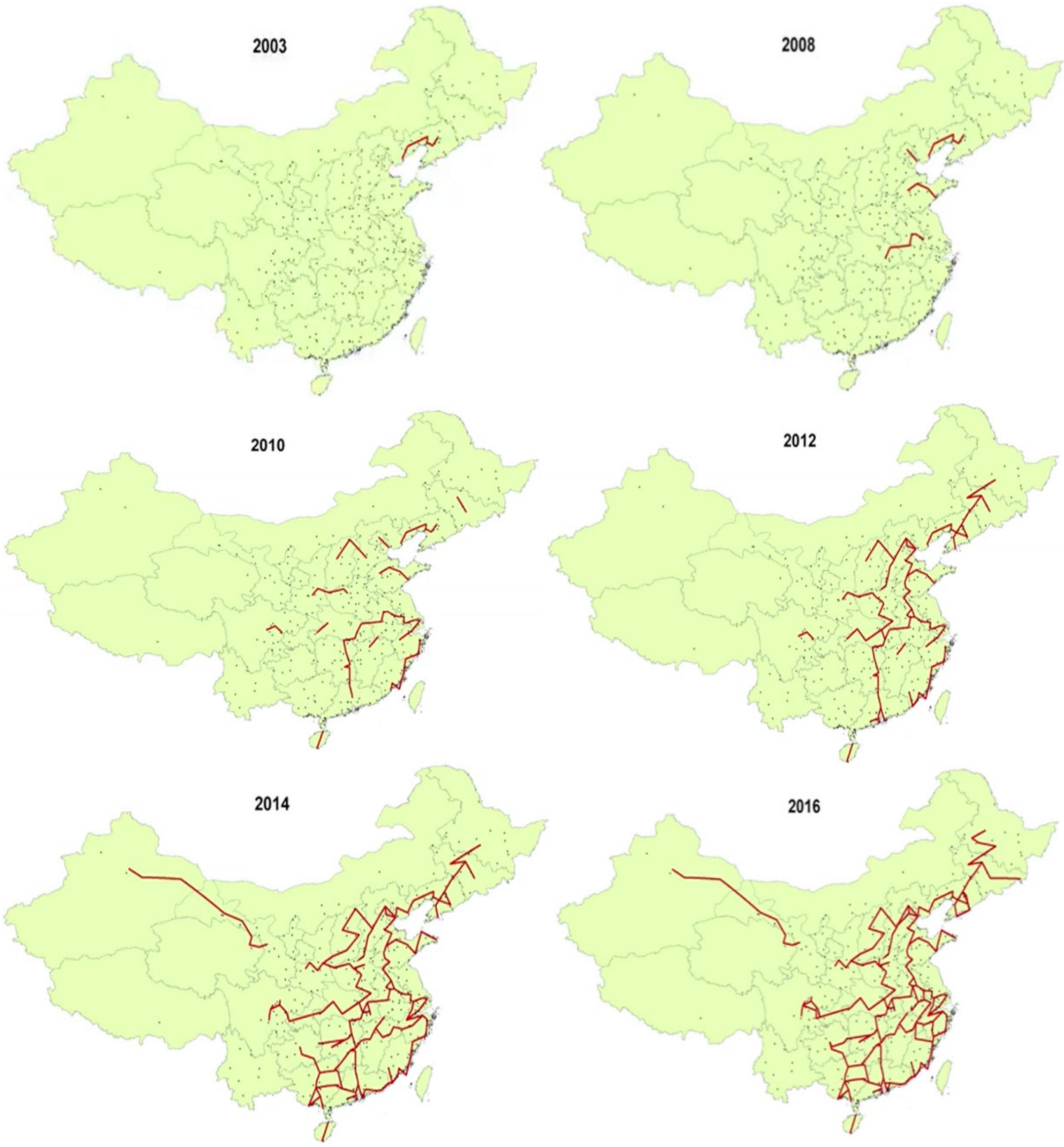
Seasholes, Mark S. and Ning Zhu. (2010). Individual Investors and Local Bias. The Journal of Finance, 105 (5): 1987-2010.

Stein, Jeremy C. (2002). Information Production and Capital Allocation: Decentralized versus Hierarchical Firms. Journal of Finance, 57(5): 1891-1921.

Storper, Michael and Anthony J. Venables. (2004). Buzz: Face-to-face Contact and the Urban Economy. Journal of Economic Geography, 4(4): 351–370.

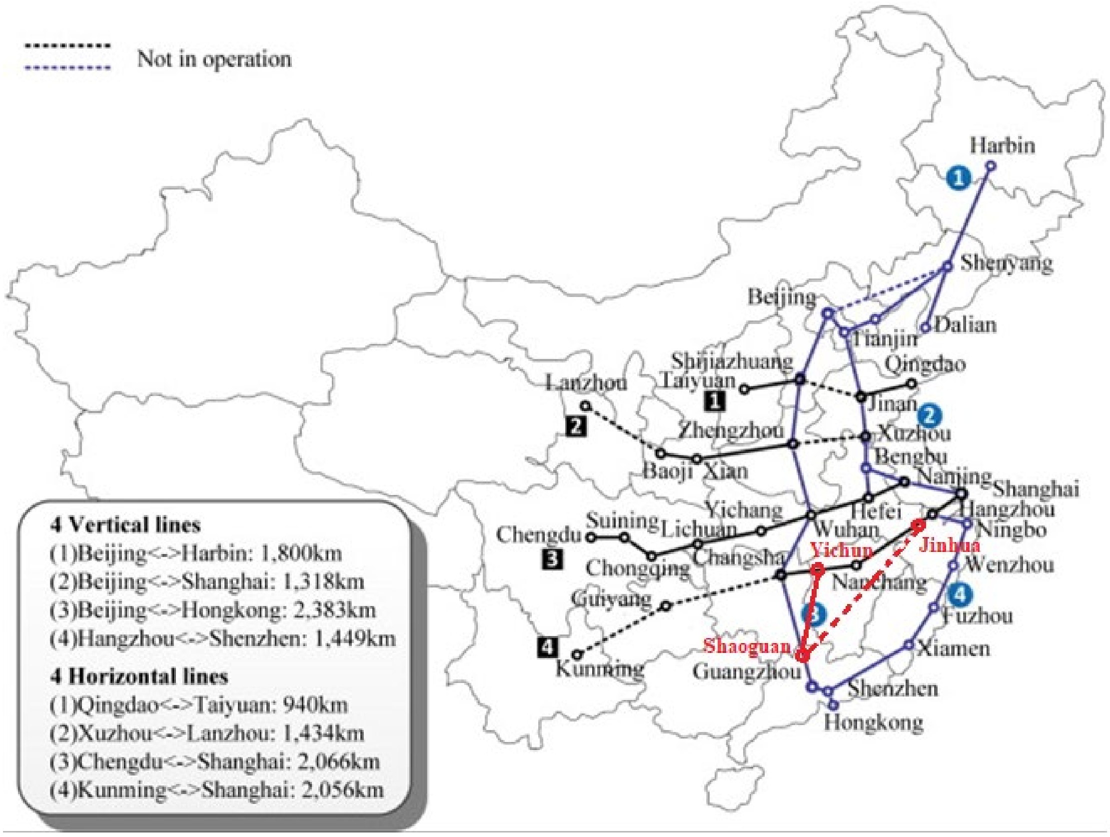
Zheng, Siqi and Matthew E. Kahn. (2013). China's Bullet Trains Facilitate Market Integration and Mitigate the Cost of Megacity Growth, PNAS, 110 (14): 1248-1253.

Figure 1. Evolution of HSR Network from 2003 to 2016



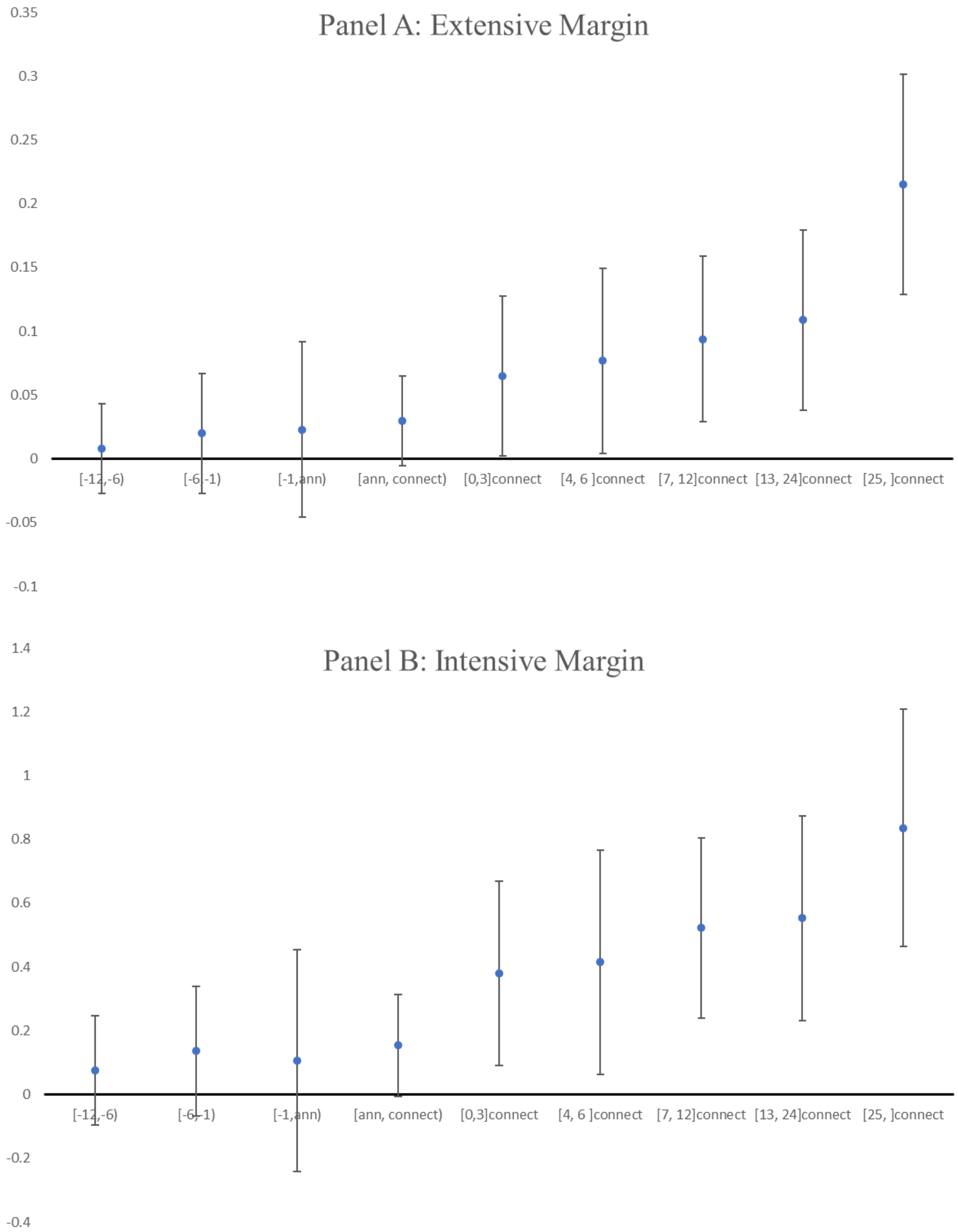
Notes: These figures display the evolution of HSR expansion from year 2003 to 2016. The lines in bold red are lines in use by the end of that year. Each dot represents a prefecture-level city.

Figure 2. Construction of Indirect Connect Measures of HSR Network



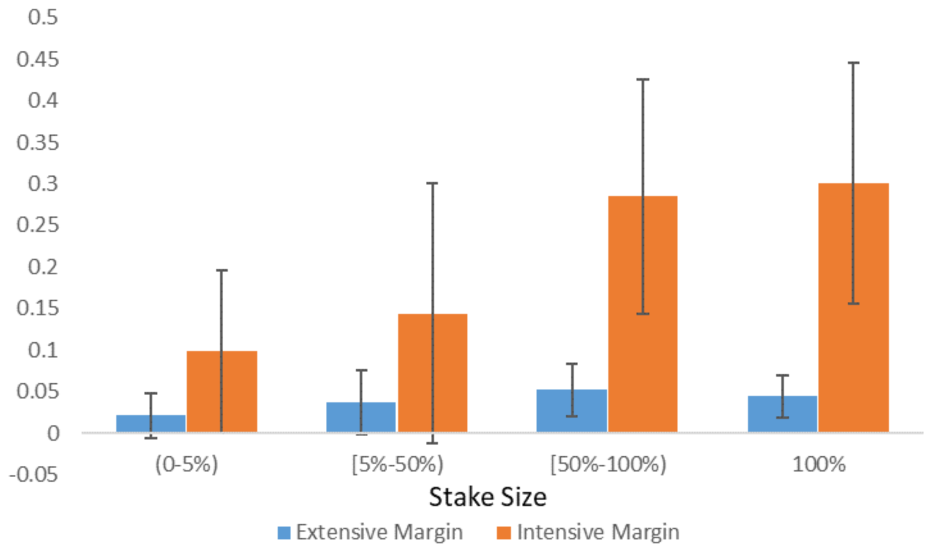
Notes: This figure illustrates the concept of an indirect connection by high-speed rail. Yichun and Shaoguan is considered to be indirectly connected after both Changsha-Nanchang and Changsha-Guangzhou lines are in operation.

Figure 3. Dynamic Effect of HSR Announcement and Connection



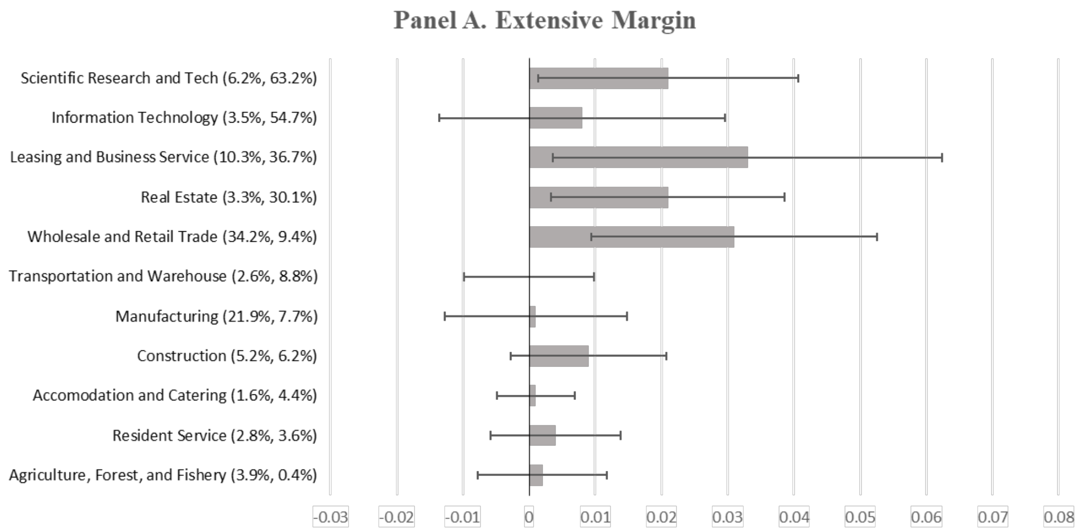
Notes: This figure visualizes the coefficients θ in Equation (2). The top panel reports the extensive margin (# of investments) of cross-city investment flows around the introduction of direct HSR connection, whereas the bottom panel reports the intensive margin (total amount of investments). The regression estimates are available in Appendix Table A6 (columns 2 and 4). The coefficients are presented in dots, with their 95% confidence intervals.

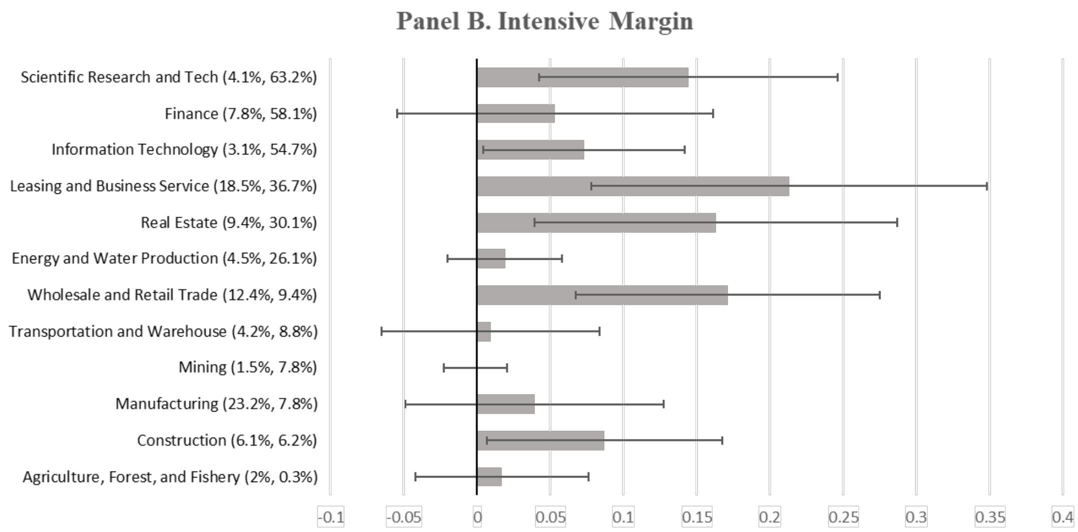
Figure 4. Coefficients by Different Ownership Stakes



Notes: The sample is split into four categories based on the investor's stake size. The four categories are: (1) the investor holding 0% to 5% (not inclusive) of the invested firm's share; (2) the investor holding 5% (inclusive) to 50% (not inclusive) of the invested firms' share; (3) the investor holding 50% (inclusive) to 100% (not inclusive) of the invested firms' share; and (4) the investor holding all (100%) of the shares of the invested firm. The heights of the bars represent the magnitudes of the coefficients while the lines denote the 95% confidence intervals.

Figure 5. HSR Effects, by Industries





Notes: The sample is split into 20 categories based on the receiver firm's industry. The first number in the parentheses is the proportion of investment flow into each industry and the second number is the college share of the industry. For example, firms in the Leasing and Business Service industry represent 10.3% of the total number of newly established firms across all industries over the whole sample period, and 18.5% of the total investment capital across all industries. The college share of this industry is 36.7%. The length of each bar represents the magnitude of the Connect coefficient for the industry, while the lines denote the 95% confidence intervals. Industries are sorted by the magnitudes of their coefficients. We included all 20 industries when estimating the coefficients and confidence intervals but only display those industries with proportion of more than 1% in this graph.

Table 1. Summary Statistics of City Pairs

Panel A. All city pairs (283 cities; 11,499,062 observations)

	Mean	Std. Dev	Min	Max
Number of investments	0.0253	0.503	0	354
Total investment amount	44.486	1292.301	0	1,109,168

Panel B. All city pairs, sorted by eventual direct connection

	Ever directly connected pairs (171,072 observations)	Never directly connected pairs (11,320,992 observations)
	Mean	Mean
Number of investments	0.509	0.0179
Total investment amount	820.371	32.762

Panel C. Ever directly connected city pairs; before and after connection

	Before direct connection (122,202 observations)	After direct connection (48,870 observations)
	Mean	Mean
Number of investments	0.302	1.026
Total investment amount	429.428	1797.944

Notes: Information on firm investment is collected from Firm Registration Database conducted by China State Administration for Industry and Commerce. The first row in each panel reports the number of cross-city investments between each city pair during the month, while the second row reports the total investment amount (in 10 thousands of RMB). Information on opening dates of HSR lines is from the China Railway Yearbooks. The sample period covers 2004 to 2015 with monthly frequency. Each cross-section includes 283 prefectural level cities with 79,806 city dyads.

Table 2. Summary Statistics of Never Directly Connected City Pairs

Panel A. Never directly connected city pairs (11,320,992 observations)

	Mean	Std. Dev	Min	Max
Number of investments	0.0179	0.359	0	181
Total investment amount	32.762	990.963	0	832,080

Panel B. Never directly connected pairs, sorted by eventual indirect connection

	Ever indirectly connected pairs (96,480 observations)	Never indirectly connected pairs (11,224,512 observations)
	Mean	Mean
Number of investments	0.140	0.0169
Total investment amount	240.135	30.980

Panel C. Ever indirectly connected pairs; before and after indirect connection

	Before indirect connection (68,166 observations)	After indirect connection (28,314 observations)
	Mean	Mean
Number of investments	0.0784	0.288
Total investment amount	123.104	521.890

Notes: Information on firm investment is collected from Firm Registration Database conducted by China State Administration for Industry and Commerce. The first row in each panel reports the number of cross-city investments between each city pair during the month, while the second row reports the total investment amount (in 10 thousands of RMB). Information on opening dates of HSR lines is from the China Railway Yearbooks. The sample period covers 2004 to 2015 with monthly frequency. Each cross-section includes 283 prefectural level cities with 79,806 city dyads. The definition of indirect connection is described in Section 4.2.

Table 3. The Impact of HSR Connection on Cross-City Investments

Panel A: All Firms

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
Connect	0.028***	0.080***	0.116***	0.375***
	(2.88)	(3.76)	(2.79)	(3.91)
Observations	171,072	168,192	171,072	168,192
R^2	0.58	0.75	0.48	0.65
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Origin (i) * year-month FE		✓		✓
Destination (j) * year-month FE		✓		✓

Panel B: POEs to POEs

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
Connect	0.029***	0.078***	0.126***	0.353***
	(3.03)	(3.75)	(3.05)	(3.81)
Observations	171,072	168,192	171,072	168,192
R^2	0.57	0.74	0.46	0.64
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Origin (i) * year-month FE		✓		✓
Destination (j) * year-month FE		✓		✓

Notes: The table reports difference-in-differences estimation results from Equation 1. The main dependent variables are: $Lnumber_{i,j,t}$, which is the logarithm of the number of unique investments from city i to city j within month t; and $Linvestment_{i,j,t}$, which is the logarithm of the total investment flow from city i to city j within month t. $Connect_{i,j,t}$ is an indicator variable, taking the value of 1 if a city pair (*i,j*) is directly connected by HSR at month *t*. The sample includes only city pairs that are ever connected (see Table 1 Panel C). Panel A includes all investment while Panel B only considers investment from privately owned enterprises (POE) to privately owned enterprises. Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses. The coefficient estimates are statistically significant at 1% level, as indicated by the asterisks ***.

Table 4. New HSR Stations versus Existing HSR Stations

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
Connect	0.029***	0.080***	0.118***	0.375***
	(2.924)	(3.76)	(2.763)	(3.908)
Connect * New HSR Station	-0.039**	-0.126***	-0.045	-0.257
	(-2.264)	(-2.982)	(-0.480)	(-1.108)
Observations	171,072	168,192	171,072	168,192
R-squared	0.582	0.746	0.478	0.653
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Origin (i) * year-month FE		✓		✓
Destination (j) * year-month FE		✓		✓

Notes. The table reports difference-in-differences estimation results from Equation 1, augmented with an indicator variable for new HSR stations. The main dependent variables are: $Lnumber_{i,j,t}$, which is the logarithm of the number of unique investments from city i to city j within month t; and $Linvestment_{i,j,t}$, which is the logarithm of the total investment flow from city i to city j within month t. $Connect_{i,j,t}$ is an indicator variable, taking the value of 1 if a city pair (*i,j*) is directly connected by HSR at month t. Connected city pairs are divided into two groups: (1) connections of two existing HSR stations and (2) connections that involve at least one new HSR station. The *New HSR Station* dummy is set to 1 if at least one of the HSR stations involved in this connection is new, and zero otherwise. The standalone variable is not included in the regression as it is subsumed by either the Origin*year-month FE or the Destination*year-month FE. The sample includes only city pairs that are ever connected (see Table 1 Panel C). Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

Table 5. The Impact of Indirect HSR Connection on Cross-City Investments

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
IndirectConnect	0.062***	0.038***	0.324***	0.185***
	(8.396)	(6.394)	(10.192)	(7.548)
Observations	11,320,992	11,320,992	11,320,992	11,320,992
R-squared	0.369	0.406	0.269	0.299
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Origin (i) * year-month FE		✓		✓
Destination (j) * year-month FE		✓		✓

Notes: The table reports difference-in-differences estimation results from Equation 3. The main dependent variables are: $Lnumber_{i,j,t}$, which is the logarithm of the number of unique investments from city i to city j within month t; and $Linvestment_{i,j,t}$, which is the logarithm of the total investment flow from city i to city j within month t. $IndirectConnect_{i,j,t}$ is an indicator variable, taking the value of 1 if a city pair (*i,j*) is indirectly connected by HSR at month *t*. The sample includes only city pairs that are never directly connected (see Table 2). Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses. The coefficient estimates are statistically significant at 1% level, as indicated by the asterisks ***.

Table 6. Test of Parallel Trend and the Announcement Effect

Panel A. Using One Year Before Announcement as Benchmark

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
Pre-Announcement (1 Year)	0.01	0.00	0.03	0.08
	(0.86)	(0.28)	(1.01)	(1.05)
Announcement	0.02*	0.01	0.08*	0.09
	(1.91)	(0.39)	(1.95)	(1.07)
Connect	0.05***	0.09***	0.20***	0.48***
	(3.76)	(2.93)	(3.67)	(3.62)
Observations	171,072	168,192	171,072	168,192
R-squared	0.58	0.75	0.48	0.65
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Origin (i) * year-month FE		✓		✓
Destination (j) * year-month FE		✓		✓

Panel B. Using Six Months Before Announcement as Benchmark

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
Pre-Announcement (6 Months)	0.00	0.01	0.02	0.09
	(0.55)	(0.47)	(0.42)	(0.95)
Announcement	0.02*	0.01	0.07*	0.08
	(1.95)	(0.43)	(1.92)	(1.05)
Connect	0.05***	0.09***	0.19***	0.47***
	(3.84)	(3.05)	(3.68)	(3.71)
Observations	171,072	168,192	171,072	168,192
R-squared	0.58	0.75	0.48	0.65
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Origin (i) * year-month FE		✓		✓
Destination (j) * year-month FE		✓		✓

Notes: The table reports difference-in-differences estimation results from Equation 1, augmented with two additional variables: (1) *Announcement*, a dummy variable for the period between the announcement of the HSR lines and the actual introduction of the connection, and (2) *Pre-Announcement*, a dummy variable for the pre-announcement period (one year in Panel A; six months in Panel B). The main dependent variables are: $Lnumber_{i,j,t}$, which is the logarithm of the number of unique investments from city i to city j within month t; and $Linvestment_{i,j,t}$, which is the logarithm of the total investment flow from city i to city j within month t. $Connect_{i,j,t}$ is an indicator variable, taking the value of 1 if a city pair (*i,j*) is directly connected by HSR at month *t*. The sample includes only city pairs that are ever connected (see Table 1 Panel C). Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

Table 7. The Impacts of HSR Connections on Cross- and Within-Cluster Investments

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
Connect	-0.019		-0.035	
	(-0.861)		(0.354)	
ToSameCluster_Connect	0.191***		0.651***	
	(8.652)		(7.417)	
CtoC*Connect		0.054		-0.206
		(0.507)		(-0.573)
CtoP*Connect		0.178***		0.809**
		(2.373)		(2.031)
PtoC*Connect		0.049		0.429
		(0.648)		(1.167)
PtoP*Connect		0.260***		0.891***
		(3.216)		(2.722)
Observations	168,192	60,480	168,192	60,480
R-squared	0.748	0.834	0.654	0.763
City-pair FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	✓
Origin (i) * year-month FE	✓	✓	✓	✓
Destination (j) * year-month FE	✓	✓	✓	✓

Notes: The table reports difference-in-differences estimation results from Equation 1. The main dependent variables are: $Lnumber_{i,j,t}$, which is the logarithm of the number of unique investments from city i to city j within month t; and $Linvestment_{i,j,t}$, which is the logarithm of the total investment flow from city i to city j within month t. $Connect_{i,j,t}$ is an indicator variable, taking the value of 1 if a city pair (*i,j*) is directly connected by HSR at month *t*. The sample includes only city pairs that are ever connected (see Table 1 Panel C). All cities are divided into three categories: core city in a cluster, peripheral city in a cluster, and cities not belonging to a cluster. In column (2) and (4) only cities that belong to a cluster have been included in the regression analyses. Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

Table 8. The Impacts of HSR Connections on Capital Return Gaps

Panel A: Direct Connection (using ever connected city pairs)

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
Connect*ROC	0.012*	0.022***	0.054*	0.085***
	(1.715)	(2.647)	(1.849)	(2.484)
Observations	24,813	24,772	24,813	24,772
R^2	0.46	0.48	0.41	0.43
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Origin city*year-month FE		✓		✓
Destination city*year-month FE		✓		✓

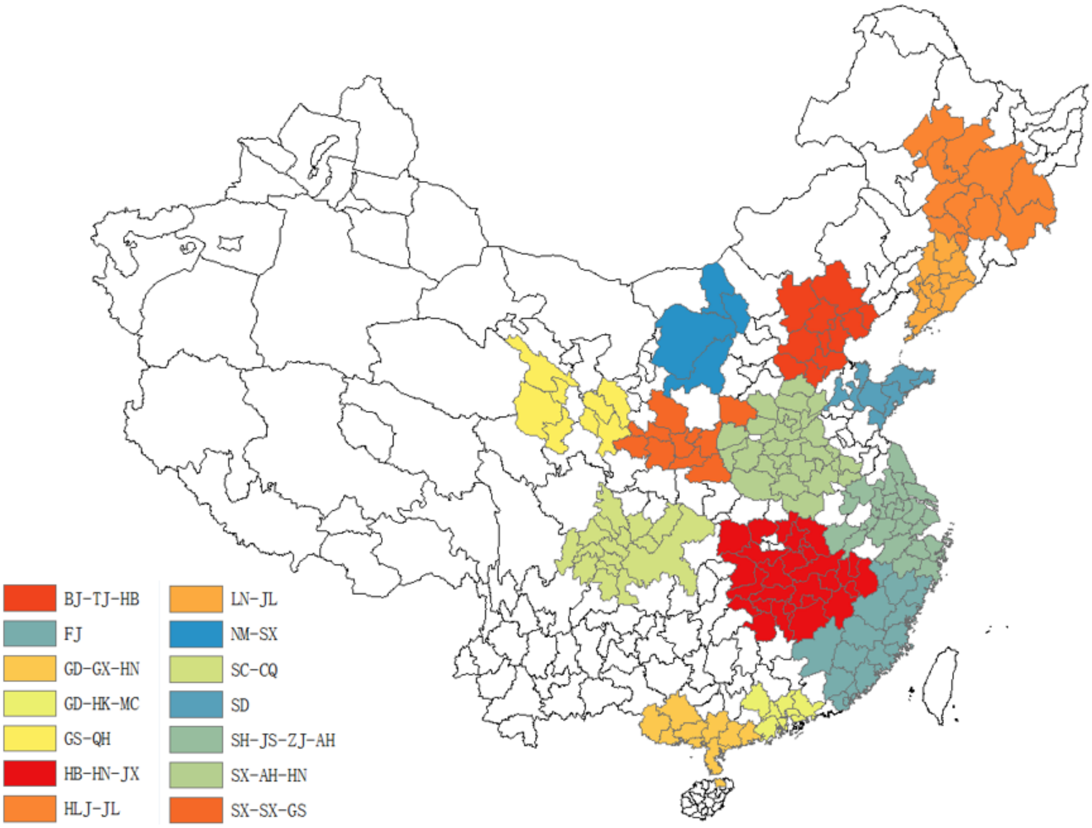
Panel B: Indirect Connection (using never directly connected city pairs)

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
IndirectConnect*ROC	0.007	0.018***	0.020	0.074**
	(1.024)	(2.565)	(0.554)	(2.012)
Observations	1,207,698	1,207,698	1,207,698	1,207,698
R^2	0.36	0.37	0.30	0.30
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Origin city*year-month FE		✓		✓
Destination city*year-month FE		✓		✓

Notes: The table reports difference-in-differences estimation results from Equation 1. The main dependent variables are: $Lnumber_{i,j,k,t}$, which is the logarithm of the number of unique investments from city i to city j in industry k within month t; and $Linvestment_{i,j,k,t}$, which is the logarithm of the total investment flow from city i to city j in industry k within month t. $Connect_{i,j,t}$ is an indicator variable, taking the value of 1 if a city pair (*i,j*) is directly connected by HSR at month *t*. We show the coefficients of the interaction term between *Connect* and *ROC*. *ROC* is the ratio of profit over asset that measures the city-industry level capital return rate by aggregating all the sampled firms within the same city and industry. When conducting the aggregation, *ROC* is weighted by asset so to take the difference in firm size into account. Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

Online Appendix

Figure A1. City Cluster Regions



Notes: see Table A9 for the detailed definitions and explanation of code.

Table A1. Heterogeneity on Ownership: SOE vs Non-SOE Firms

Panel A: SOE to SOE

	(1)	(2)	(3)	(4)
Variables	*Lnumber*	*Lnumber*	*Linvestment*	*Linvestment*
connect	0.000	0.006	0.000	0.044
	(-0.222)	(1.007)	(0.007)	(1.506)
Observations	171,072	168,192	171,072	168,192
R-squared	0.090	0.389	0.091	0.362

Panel B: SOE to POE

	Lnumber	*Lnumber*	*Linvestment*	*Linvestment*
connect	0.003	0.009	0.028*	0.087*
	(1.265)	(1.127)	(1.784)	(1.729)
Observations	171,072	168,192	171,072	168,192
R-squared	0.212	0.450	0.186	0.430

Panel C: POE to SOE

Variables	*Lnumber*	*Lnumber*	*Linvestment*	*Linvestment*
connect	0.003	0.013*	0.020**	0.084**
	(1.498)	(1.779)	(1.981)	(2.484)
Observations	171,072	168,192	171,072	168,192
R-squared	0.423	0.631	0.337	0.558
Year-month dummy	✓		✓	
city-pair FE	✓	✓	✓	✓
Origin city * year-month FE		✓		✓
Destination city * year-month FE		✓		✓

Notes: The table reports difference-in-differences estimation results by SOE and non-SOE firms. *Lnumber* is the logarithm of unique investment pairs from city i to city j within month t. *Linvestment* is the logarithm of the sum of investment flow from city i to city j within month t. *Indirectconnect* is a dummy indicating whether a city pair ij is indirectly connected by HSR at year month t. Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses.

Table A2. Survival Adjusted Extensive Margin

	Survived in 2015		Survived at least 3 yrs		Survived at least 4 yrs		Survived at least 5 yrs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	*Lnumber*	*Lnumber*	*Lnumber*	*Lnumber*	*Lnumber*	*Lnumber*	*Lnumber*	*Lnumber*
connect	0.037***	0.098***	0.030***	0.087***	0.032***	0.089***	0.034***	0.091***
	(3.541)	(4.300)	(3.212)	(4.238)	(3.349)	(4.263)	(3.452)	(4.263)
Observations	171,072	168,192	171,072	168,192	171,072	168,192	171,072	168,192
R-squared	0.533	0.729	0.561	0.738	0.553	0.735	0.546	0.733
Year Dummy	✓		✓		✓		✓	
City Pair FE	✓	✓	✓	✓	✓	✓	✓	✓
Origin city *year FE		✓		✓		✓		✓
Destination city *year FE		✓		✓		✓		✓

Notes: The table reports difference-in-differences estimation results from Equation 1. *Lnumber* is the logarithm of unique investment pairs from city i to city j within month t and survived for certain number of years (by the end of year 2015, at least three years; at least four years; and at least five years); *connect* is a dummy indicating whether a city pair ij is connected by HSR at year month t. Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses.

Table A3. Survival Adjusted Intensive Margin

	Survived in 2015		Survived at least 3 yrs		Survived at least 4 yrs		Survived at least 5 yrs	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Variables	*Linvestment*	*Linvestment*	*Linvestment*	*Linvestment*	*Linvestment*	*Linvestment*	*Linvestment*	*Linvestment*
connect	0177***	0.458***	0.153***	0.432***	0.158***	0.430***	0.168***	0.443***
	(3.537)	(3.962)	(3.301)	(4.022)	(3.379)	(3.973)	(3.494)	(3.979)
Observations	171,072	168,192	171,072	168,192	171,072	168,192	171,072	168,192
R-squared	0.435	0.636	0.452	0.641	0.447	0.64	0.443	0.638
Year Dummy	✓		✓		✓		✓	
City Pair FE	✓	✓	✓	✓	✓	✓	✓	✓
Origin city *year FE		✓		✓		✓		✓
Destination city *year FE		✓		✓		✓		✓

Notes: The table reports difference-in-differences estimation results from Equation 1. *Linvestment* is the logarithm of the sum of investment flow from city i to city j within month t and survived for certain number of years (by the end of year 2015, at least three years; at least four years; and at least five years). *connect* is a dummy indicating whether a city pair ij is connected by HSR at year month t. Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses.

Table A4. The Impact of HSR Connection on Cross-City Investments (Subsample before end of 2013)

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
Connect	0.04***	0.098***	0.157***	0.376***
	(3.540)	(3.789)	(3.238)	(3.353)
Observations	142,560	168,192	142,560	140,160
R^2	0.569	0.724	0.466	0.638
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Origin (i) * year-month FE		✓		✓
Destination (j) * year-month FE		✓		✓

Notes: The table reports difference-in-differences estimation results from Equation 1 using sample before the end of 2013. The main dependent variables are: $Lnumber_{i,j,t}$, which is the logarithm of the number of unique investments from city i to city j within month t; and $Linvestment_{i,j,t}$, which is the logarithm of the total investment flow from city i to city j within month t. $Connect_{i,j,t}$ is an indicator variable, taking the value of 1 if a city pair (*i*,*j*) is directly connected by HSR at month *t*. The sample includes only city pairs that are ever connected (see Table 1 Panel C). Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses. The coefficient estimates are statistically significant at 1% level, as indicated by the asterisks ***.

Table A5. The Impact of Indirect HSR Connection on Cross-City Investments (Subsample before end of 2013)

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
IndirectConnect	0.054***	0.034***	0.297***	0.174***
	(7.354)	(5.784)	(8.565)	(6.357)
Observations	9,434,160	9,434,160	9,434,160	9,434,160
R^2	0.357	0.379	0.260	0.280
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Origin (i) * year-month FE		✓		✓
Destination (j) * year-month FE		✓		✓

Notes: The table reports difference-in-differences estimation results from Equation 1 using sample before the end of 2013. The main dependent variables are: $Lnumber_{i,j,t}$, which is the logarithm of the number of unique investments from city i to city j within month t; and $Linvestment_{i,j,t}$, which is the logarithm of the total investment flow from city i to city j within month t. $IndirectConnect_{i,j,t}$ is an indicator variable, taking the value of 1 if a city pair (*i*,*j*) is indirectly connected by HSR at month *t*. The sample includes only city pairs that are never directly connected (see Table 2). Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses. The coefficient estimates are statistically significant at 1% level, as indicated by the asterisks ***.

Table A6. Dynamic Effect of HSR Announcement and Connection

	(1)	(2)	(3)	(4)
Variables	*Lnumber*		*Linvestment*	
preannounce [-12, -6m)	0.014*	0.008	0.069*	0.076
	(1.894)	(0.457)	(1.955)	(0.873)
preannounce [-6, -1m)	0.012	0.020	0.047	0.136
	(1.324)	(0.849)	(1.152)	(1.314)
preannounce [-1, 0m)	0.028*	0.023	0.111	0.106
	(1.954)	(0.657)	(1.544)	(1.600)
announce	0.034***	0.030*	0.140***	0.154*
	(3.497)	(1.667)	(3.355)	(1.892)
connect [0,4m)	0.048***	0.065**	0.216***	0.380**
	(3.716)	(2.028)	(3.599)	(2.595)
connect [4, 7m)	0.044***	0.077**	0.182***	0.415**
	(3.101)	(2.068)	(2.791)	(2.315)
connect [7,13m)	0.045***	0.094***	0.193***	0.522***
	(3.394)	(2.837)	(3.169)	(3.612)
connect [13,25m)	0.058***	0.109***	0.237***	0.553***
	(3.867)	(3.008)	(3.662)	(3.384)
connect [25m,]	0.129***	0.215***	0.510***	0.836***
	(5.603)	(4.844)	(5.751)	(4.406)
Observations	171,072	168,192	171,072	168,192
R-squared	0.583	0.746	0.479	0.653
Year-month FE	✓		✓	
City Pair FE	✓	✓	✓	✓
Origin city*year-month FE		✓		✓
Destination city *year-month FE		✓		✓

Notes: The table reports the event study results from Equation 2. *Lnumber* is the logarithm of unique investment pairs from city i to city j within month t; *Linvestment* is the logarithm of the sum of investment flow from city i to city j within month t. *preannounce[i,j]* are dummy variables that turn on if year month t is no earlier than I month before the announcement and no later than j month before the announcement of HSR connection; *announce* is a dummy variable that turns on if year month t is after the announcement period and before the connection of HSR; *connect[i,j]* are dummy variables that turn on if year month t is no earlier than i month after connection and no later than j month after connection. The benchmark group is 13-24 months before the announcement month. Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses.

Table A7. Spillover: Investment Inflow at City Level

Panel A: within 200km to HSR

	(1)	(2)	(3)	(4)
Variables	*Lnumber*	*Lnumber*	*Linvestment*	*Linvestment*
HSR*Buffer100km	0.067*	0.067*	0.120	0.120
	(1.815)	(1.815)	(1.127)	(1.127)
Observations	12,887	12,887	12,887	12,887
R-squared	0.467	0.467	0.446	0.446

Panel B: within 250km to HSR

Variables	*Lnumber*	*Lnumber*	*Linvestment*	*Linvestment*
HSR*Buffer100km	0.076**	0.076**	0.176	0.176
	(2.077)	(2.077)	(1.642)	(1.642)
Observations	14,429	14,429	14,429	14,429
R-squared	0.453	0.453	0.428	0.428

Panel C: within 300km to HSR

Variables	*Lnumber*	*Lnumber*	*Linvestment*	*Linvestment*
HSR*Buffer100km	0.087**	0.087**	0.217**	0.217**
	(2.357)	(2.357)	(2.030)	(2.030)
Observations	15,329	15,329	15,329	15,329
R-squared	0.455	0.455	0.426	0.426
Year-month dummy	✓	✓	✓	✓
City FE	✓	✓	✓	✓
City * Time Trend		✓		✓

Notes: The table reports difference-in-differences estimation results from analyses of the HSR connection and the investment inflow to nearby cities that are not connected to HSR. The main dependent variables are: *Lnumber*, which is the logarithm of the number of unique investments inflow to city i within month t; and *Linvestment*, which is the logarithm of the total investment inflow to city i within month t. The regression sample contain cities that are never connected to HSR, but are located within 200 km, 250km, 300km to a nearby HSR city, which include 101, 114, 124 cities, respectively. *Buffer100km* is an indicator variable, taking the value of 1 if a city i is not connected to HSR, but locates within 100 km of a hub city j that is ever connected to HSR. 47 cities belong to the treatment group of Buffer100km. *HSR* is a time-varying dummy indicating whether that nearby HSR city j is connected by HSR at year month *t*. City-specific cubic year-month trend is controlled for in in columns (2) and (4). Robust standard errors clustered at city level and the corresponding t-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

Table A8. Heterogeneity on Control: Controlling vs. Non-Controlling Investors

Panel A: Non-controlling Investors

	Share	(1)	(2)	(3)	(4)
Variables	(0,5%)	*Lnumber*	*Lnumber*	*Linvestment*	*Linvestment*
connect		0.012***	0.021	0.048***	0.098*
		(3.788)	(1.471)	(3.674)	(1.947)
Observations		171,072	168,192	171,072	168,192
R-squared		0.193	0.469	0.156	0.435
	[5%,50%)	*Lnumber*	*Lnumber*	*Linvestment*	*Linvestment*
connect		0.017***	0.037**	0.085***	0.144*
		(3.19)	(2.17)	(3.11)	(1.87)
Observations		171,072	168,192	171,072	168,192
R-squared		0.42	0.63	0.34	0.56

Panel B: Controlling Investors

Variables	[50%,100%)	*Lnumber*	*Lnumber*	*Linvestment*	*Linvestment*
connect		0.019***	0.052***	0.095***	0.285***
		(3.259)	(3.232)	(3.411)	(3.953)
Observations		171,072	168,192	171,072	168,192
R-squared		0.448	0.645	0.362	0.570
Variables	100%	*Lnumber*	*Lnumber*	*Linvestment*	*Linvestment*
connect		0.013**	0.044***	0.094***	0.301***
		(2.065)	(3.256)	(2.926)	(4.049)
Observations		171,072	168,192	171,072	168,192
R-squared		0.414	0.629	0.333	0.555
Year-month dummy		✓		✓	
city-pair FE		✓	✓	✓	✓
Origin city * year-month FE			✓		✓
Destination city * year-month FE			✓		✓

Notes: The table reports difference-in-differences estimation results from Equation 1 with different cutoff of shareholders. *Lnumber* is the logarithm of unique investment pairs from city *i* to city *j* within month *t*; *Linvestment* is the logarithm of the sum of investment flow from city *i* to city *j* within month t. Connect is a dummy indicating whether a city pair *ij* is connected by HSR at year month *t*. Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

Table A9. Definition of City Clusters

Code	Cluster Name	Core Cities	Peripheral Cities
BJ-TJ-HB		Beijing, Tianjin	Baoding, Tangshan, Langfang, Shijiazhuang, Qinhuangdao, Zhangjiakou, Chengde, Cangzhou, Hengshui
LN-JL#	Southern of Liaoning	Shenyang, Dalian	Anshan, Fushun, Benxi, Dandong, Liaoyang, Yingkou, Panjin, Tieling
FJ*	West of the Strait	Fuzhou	Quanzhou, Xiamen, Wenzhou, Shantou, Putian, Zhangzhou, Sanming, Nanping, Ningde, Longyan, Lishui, Quzhou, Shangrao, Yingtan, Fuzhou, Ganzhou, Chaozhou, Jieyang, Meizhou
NM-SX			Huhehaote, Baotao, Erduosi, Yulin
GD-GX-HN	Beibu Gulf	Nanning	Beihai, Qinzhou, Fangchenggang, Yulin, Chongzuo, Zhanjiang, Maoming, Yangjiang, Haikou, Chanzhou, Dongfang, Chengmai, Lingao, Changjiang
SC-CQ		Chongqing, Chengdu	Zigong, Luzhou, Deyang, Mianyang, Suining, Neijiang, Nanchong, Leshan, Meishan, Yibin, Guangan, Dazhou, Yaan, Ziyang
GD-HK-MC	Peral River Delta	Guangzhou, Shenzhen, Hong Kong Macau	Foshan, Dongguan, Zhongshan, Zhuhai, Jiangmen, Zhaoqing, Huizhou
SD+	Shandong Peninsular	Jinan, Qingdao	Zibo, Yantai, Weifang, Dongying, Weihai, Rizhao
GS-QH	West of Lanzhou	Xining, Lanzhou	Haidong, Baiyin, Dingxi, Linxia, Haibei, Hainan, Huangnan
SH-JS-ZJ-AH	Yangzi River Delta	Shanghai	Nanjing, Wuxi, Suzhou, Changzhou, Nantong, Yancheng, Yangzhou, Zhenjiang, Taizhou, Hangzhou, Ningbo, Huzhou,

			Jiaxing, Shaoxing, Jinhua, Zhoushan, Taizhou, Hefei, Wuhu, Maanshan, Tongling, Anqing, Chuzhou, Chizhou, Xuancheng
HB-HN-JX	Yangzi River	Wuhan	Huangshi, Ezhou, Xiaogan, Huanggang, Xianning, Xiantao, Qianjiang, Tianmen, Xiangyang, Yichang, Jingmen, Jingzhou, Changsha, Zhuzhou, Xiangtan, Yueyang, Yiyang, Changde, Hengyang, Loudi, Nanchang, Jiujiang, Jingdezhen, Yingtan, Xinyu, Yichun, Pingxiang, Shangrao, Fuzhou, Ji'an
SX-AH-HN	Zhongyuan	Zhengzhou	Kaifeng, Luoyang, Anyang, Nanyang, Shangqiu, Xinxiang, Pingdingshan, Xuchang, Jiaozuo, Zhoukou, Xinyang, Zhumadian, Luohe, Puyang, Hebi, Sanmenxia, Jiyuan, Changzhi, Jincheng, Yuncheng, Xingtai, Handan, Liaocheng, Heze, Huaibei, Bengbu, Suzhou, Puyang, Haozhou
HLJ-JL		Harbin, Changchun	Daqing, Qiqihaer, Suihua, Mudanjiang, Jilin, Siping, Liaoyuan, Songyuan, Yanbian
SX-SX-GS		Xi'an	Baoji, Xianyang, Tongchuan, Weinan, Shangluo, Yuncheng, Linfen, Tianshui, Pingyang, Qingyang

Notes: #plan by Liaoning provincial government; *plan by Ministry of Housing and Urban-Rural Development; +plan by Shandong provincial government. All other clusters have been approved by the State Council.

Table A10. The Impact of HSR Connection on Regional Inequality

	(1)	(2)	(3)	(4)
	Lnumber	*Linvestment*	*Lnumber*	*Linvestment*
RtoR*Connect	0.133***	0.480***		
	(4.766)	(4.045)		
RtoM*Connect	-0.098*	-0.081		
	(-1.769)	(-0.286)		
RtoP*Connect	0.066	0.562*		
	(1.135)	(1.927)		
MtoR*Connect	0.080*	0.420*		
	(1.767)	(1.797)		
MtoM*Connect	-0.021	0.030		
	(-0.395)	(0.110)		
MtoP *Connect	0.188***	0.881**		
	(3.160)	(2.579)		
PtoR*Connect	-0.023	-0.009		
	(-0.576)	(-0.050)		
PtoM*Connect	-0.108*	-0.289		
	(-1.814)	(-0.951)		
PtoP*Connect	0.030	0.178		
	(0.602)	(0.714)		
ToSameGroup*			0.105***	0.403***
Connect			(4.553)	(3.964)
ToHigherGroup*			0.033	0.265**
Connect			(1.245)	(2.257)
ToLowerGroup*			0.056*	0.403***
Connect			(1.880)	(2.993)
Observations	167,616	167,616	167,616	167,616
R-squared	0.749	0.654	0.746	0.653
City-pair FE	✓	✓	✓	✓
Origin city*year-month FE	✓	✓	✓	✓
Destination city*year-month FE	✓	✓	✓	✓

Notes: Cities are divided into three groups according to GDP per capita at base year 2004: (R)ich, (M)iddle, and (P)oor. City-pairs are categorized into 3x3 groups using the categories of the source (*i*) and destination (*j*) cities. *ToSameGroup* means the investment is from (R)ich to (R)ich, (M)iddle to (M)iddle, or (P)oor to (P)oor. T*oHigherGroup* means the investment is from (P)oor to M(iddle), (P)oor to (R)ich, or (M)iddle to (R)ich. *ToLowerGroup* means the investment is from (R)ich to (M)iddle, (R)ich to (P)oor, or (M)iddle to (P)oor. The time-invariant city-pair category indicators are interacted with time-varying $Connect_{i,j,t}$ variable. Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.