# 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

* We thank seminar participants at National University of Singapore for helpful discussions and valuable comments. Johan Sulaeman acknowledges research support from NUS Start-Up Research Grant WBS R-315-000-113-133. All errors are our own.

# 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, by exploiting the staggered expansion of passenger high-speed rail (HSR) network as plausibly exogenous shocks to the ease of travel between cities. Drawing on a unique dataset of firm registrations in China, we document that the introduction of a direct HSR connection between a pair of cities increases the amount of cross-city investment by 38%. We control for city-pair fixed effects to capture static linkages between the cities –e.g., geographical distance, cultural proximity– as well as city-time fixed effects to capture variations in economic dynamics across cities. We find similar patterns when examining new connections between cities that are already on the HSR network but not yet connected to each other, and city-pairs that are *indirectly* connected by new routes. Being connected by HSR is particularly important for investors considering controlling stakes in large distant investments, indicating that HSR improves monitoring capabilities of distant investors.

**Keywords:**

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

## 1. Introduction

Public investments in infrastructure projects have featured prominently as stimulus tools during economic downturns. Beyond their direct effects on the economy, these publicly funded projects are likely to affect private investments, e.g., by "crowding out" more efficient private investments in similar projects. On the other hand, these infrastructure projects can facilitate transfers of goods and people, which could increase the feasibility of privately funded investments. Despite the high costs of public infrastructure projects, their effects on private investment flows, are poorly understood.

Over the last decade, China has built a passenger high-speed rail (HSR) network of over 20,000 km, accounting for more than half of the world's total length of high-speed rail tracks in 2016, with plans to increase the network coverage by more than 50 percent over the next decade (Economist, 2017).[1] More countries plan to develop their own high-speed rail systems, including the high-speed rail construction currently underway in California to connect San Francisco and Los Angeles by 2029. These projects are very costly – the current cost estimate for the SF-LA's 770 km connection is $77 billion.[2] Therefore, it is crucial to understand the economic benefits of such transportation infrastructure projects, particularly its impact on private investment dynamics.

Donaldson (2018) investigates how railways facilitate interregional trades, by examining the traditional railroad system built by the British government in colonial India in the 1860-1930 period that facilitated goods transfers across India. In contrast, the HSR network is aimed at accelerating passenger traffic. HSR provides a fast, convenient, and reliable mode of transportation across cities. The increase in travel ease facilitates direct face-to-face contacts between economic agents in different cities, and

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

[2] http://www.latimes.com/local/california/la-me-bullet-train-cost-increase-20180309- story.html. The California project's average cost is $100 million per km of rails. The World Bank estimated that the cost of HSR construction in China to be much lower: around $20 million per km. http://www.worldbank.org/en/news/press-release/2014/07/10/cost-of-high-speed-rail-in-china-one-third-lower-than-in-other-countries

opens up new possibilities of communication and interactions, with potentially transformative effects on private investment dynamics.

This paper examines the effect of HSR introduction on interregional investment flows. Distance matters when it comes to investment decisions, as evidenced by the strong tendency of individuals and companies to hold investments that are geographically close, i.e., "home bias" in both cross-country and within-country setting (French and Poterba 1991; Coval and Moskowitz 1999; Ivkovic and Weisbenner 2005). The relationship between proximity and cross-region investments remains strong and significant despite the rapid progress in information and communication technology (ICT) over the last several decades, highlighting the importance of face-to-face contacts in mitigating information frictions. But exactly how does distance play a role as an obstacle for information flows, especially those relevant for investment decisions, and how can cross-city investments be promoted by reducing the cost of face-to-face contacts? These questions have not been extensively examined in the literature partly because of the strong correlation between geographic proximity and a variety of factors that also matter for cross-regional investment decisions, such as trade cost, language and cultural similarities, and institutional barriers.

This paper revisits these questions by combining the expansion of HSR network with restricted-access firm-level administrative data that is uniquely obtained from the China State Administration for Commerce and Industry. This administrative data of firm registrations in China report the shareholders for the universe of Chinese firms from 2000 to 2015. We use this dataset to develop cross-city investment flow measures that cover firm activities across all regions and sectors in China.

We focus on the role of travel cost in shaping cross-city investment patterns by evaluating the effect of the reduction in travel cost due to passenger transportation infrastructure project on such patterns. The HSR network's rapid expansion provides an ideal setting for this analysis. While highways, traditional railways, and air travel make business trips across cities possible, the advent of HSR makes large-scale cross-city travel faster and much more convenient, allowing an unprecedented amount of people move back and forth across cities, sometimes even within a day. As of 2014, China has the world's longest and most heavily used HSR network with 15,399 km of track in service, connecting over 81 cities, with an annual ridership reaching 857 million (Lin, 2017). This offers a unique opportunity to study how investment flows are affected by 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).

We adopt a difference-in-differences specification to examine whether or not HSR connection between a given city pair leads to an increase in bilateral private investments, compared to unconnected city pairs. The rich information in city-to-city investment flows allows us to control for a full set of origin/destination city × month fixed effects.

We find that direct HSR connection increases the number of investors between the city pair increase by 8%, and amount of investment increases by 38%. To address the concern that HSR lines might be endogenously places between cities with growing economic linkages, we perform two distinct analyses. First, we examine pairs of cities that are already on the HSR network but are not yet connected to each other. 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 when these cities are subsequently connected to each other.

Second, we examine cities that are *indirectly connected* by HSR: non-nodal cities in different HSR lines that become connected as the two lines cross each other. In this analysis, we exclude pairs of cities that are ever directly connected by HSR, mitigating potential bias associated with direct route connection. We again find that indirect HSR connection also leads to an increase in cross-city investments: the number of investors between indirectly connected city pairs grow by 4%, and the amount of investments grow by 18%, about half of the effect of direct connection.

The HSR effect is not observed prior to the announcement of the HSR connection. There is a weak effect between announcement and connection, which seems to be associated with variations in city-level economic growth. However, the bulk of the effect occurs after the two cities are actually connected via HSR. While the effect is almost immediate – we observe statistically significant estimate even in the first three months of the connection, the effect becomes stronger with the length of connection. This indicates the importance of reliable transportation connecting cities in promoting the intercity investment flows.

We then examine the heterogeneity in the benefits of HSR connectedness across city pairs and ownership structures, to tease out the underlying mechanisms. We first find the HSR effects are larger for city pairs in relatively close geographical proximity (e.g., within 50km of each other), reflecting the superiority of HSR relative to air transportation for these small distances. The positive effect of HSR seems to taper off at around 400km (slightly more than an hour on China's highest speed rail route, and about an hour by airplane). The HSR treatment effects seem to be concentrated in the sample of non-finance firms and outside of the State Owned Enterprises (SOE) sample.

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 28% with HSR connection, while the number of such investments increases by about 5.2%. In comparison, the number of small stakes (0-4.9% of the firm) increases by only about 2%, whereas the number of minority stakes (5-49.9% of the firm) increases by about 4%. This pattern indicates that HSR connection improves the monitoring capabilities of distant investors, allowing them to invest in larger stakes.

Finally, we examine the welfare implications of HSR connection by directly investigating the relationship between the number of cities connected via CSR to a focal city, and the number of cities (1) from which the focal city receive investments and (2) in which the focal city's investors invest. We find that being connected to more cities through HSR leads to investment flows to more cities, even conditional on provincial-specific year fixed effects. This indicates that the expansion of HSR in China leads to more diversification in terms of investment destination cities. To the extent that the increased diversification improves welfare, our finding provides a positive evidence for HSR's welfare impact.

Our paper provides three notable contributions. First, this paper is closely related to two strands of recent literature that examine the impact of transportation infrastructure on various outcomes. The first strand of research examines roads and railways that reduced trade costs, including studies on the gains from interregional trade (Donaldson 2018), local labor markets (Michaels 2008), long-term GDP growth (Banerjee et al. 2012), income volatility (Burgess et al. 2012), and asymmetric effects on core and peripheral markets (Faber 2014).[3] The second strand of research examines the role of passenger transportation on economic integration and other outcomes. While most papers in this literature examine air travel improvements in various settings, a few recent papers specifically investigate the economic impacts of HSR developments, including Bernard et al.'s (2016) study on the impact of Shinkansen line on supplier relationship among Japanese firms.[4,5] Recent studies examine the impact of China HSR system on housing prices in secondary cities (Zheng and Kahn, 2013), local

3 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).

4 Recent studies examine the effects of air travel improvements on research collaboration (Catalini, Fons-Rosen, and Gaule 2016), headquarter-subsidiary location and VC involvement (Giroud 2013, Bernstein, Giroud, and Townsend 2016), trade (Cristea 2011; Poole 2013; Yilmazkuday and Yilmazkuday 2014), and international business and growth (Campante and Yanagizawa-Drott, 2017).

5 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 et al., 2010)

employment (Lin, 2017), and collaboration in research between Chinese cities (Zheng and Kahn, 2017). In this backdrop, the current paper identifies a *causal* mechanism of how infrastructure developments can affect both investment flows and resource allocations across cities.

Second, most of these infrastructure projects are funded either directly or indirectly by governments. Identifying the causal effect of such infrastructure projects on private investments is relevant to policymakers who are considering the role of HSR and other transportation infrastructure projects in promoting economic growth and reducing regional inequality. Within our setting, China has witnessed unprecedented growth since the Opening and Reform, but also stark regional disparities. Coastal provinces' GDP per capita was three times as high as that of the western regions in 2016.[6] In order to reduce these disparities, the government has used numerous strategies, including the most recent strategy of "One Belt, One Road," which involves building roads and railways inland, connecting western Chinese cities to the coast and to cities in Central Asia. The expansion of the HSR network is a core component of the plan. Hence understanding how this expansion facilitates private investments beyond the public funding associated with the infrastructure projects is very relevant from a policy perspective.

Third, the investment flow data allow us to examine the universe of investments beyond the headquarter-subsidiary and VC-startup relationship (Giroud, 2013; Bernstein, Giroud, and Townsend, 2016). We document how the improvement in passenger transport connections reduces information asymmetry for minority investors who do not control firm operations, and even more significantly for majority shareholders who have to monitor the operations. Our results indicate that the availability of fast and reliable transportation could mitigate home bias in investments, although it is important to note that the effect seems to be limited to a 400-km radius

[6] https://www.economist.com/china/2016/10/01/rich-province-poor-province

in our setting.

The remainder of the article is organized as follows. Section 2 provides background on the evolution of HSR network in China. In section 3, we describe our dataset on cross-city investments as well as HSR connection. Our empirical strategy is explained in section 4 and the main regressions results at the city pair level are discussed in section 5. Section 6 concludes.

## 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). The first HSR line in China opened in 2003 connecting Qingdao and Shenyang as a passenger-dedicated line with an operating speed of 200km/h. 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 has grown beyond this ambitious goal. By the end of 2017, there are 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), 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 connects 29 of China's 33 provincial-level administrative divisions and 163 of 283 prefectural level cities by 2016. This objective guide 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.

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 (Freeman, 2010).

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 trip due 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 area. As HSR is more comfortable than air travel and cost 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 uniquely 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 level, i.e. a firm or a natural person contributing capital to another firm and thereby becoming its shareholder. When such activities occur, the firms and natural persons involved is 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 2000 to 2015.

Our main analysis 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 require that the receiver of the investment must be a new firm, i.e., the investment activity occurs within the calendar year in which the receiver firm is first registered with the SAIC. This requirement essentially ignores the change in shareholdings of incumbent firms, as these observations are likely to be plagued by measurement errors due to their relatively noisy recording in the Firm Registration Database. Our final 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, and 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 firms 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 monthly frequency. We aggregate the firm-level portfolio investment records in the Firm Registration Database to city pair and monthly level. The dependent variables $y_{ijt}$ stands for number of unique investment pairs to examine the extensive margin; or the logarithm of the sum of investment flow from city i to city j within month t, to examine to 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 ($\alpha_{ij}$) 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 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 on 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.

Large sets of origin/destination city × time trend effects ($\beta_{it}$, $\gamma_{jt}$) are included to address the endogeneity placement 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. As long as these factors affecting the placement decisions are unrelated to firm investment, this would not be a mitigating concern in our setting.

However, if there are omitted factors that are driving both the connection of HSR and firm investment, then any relationship between the two could be problematic. 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. We address these two concerns using the identification strategy similar to Giroud (2013) and Giroud and Mueller (2015), by introducing large sets of origin/destination city × time fixed effects ($\beta_{it}$ and $\gamma_{jt}$) , on top of city pair fixed effects ($\alpha_{ij}$). These time-varying dummy variables capture the time varying local heterogeneity that affect the attractiveness of destination cities and the investment capacity of investor cities, and help to address the above endogeneity problems.

Lastly, with all the solutions taken above, if the new HSR connection between two cities is still partially endogenous because of some preexisting 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 estimation is specified as the following:

$$y_{ijt} = \sum_{\tau=-12}^{12} \theta_\tau\ Announcement_{ijt} \times yearmonth_\tau + \alpha_{ij} + \beta_{it} + \gamma_{jt} + \delta_t + \epsilon_{ijt} \quad (2),$$

where $\tau$ stands for the event year month of announcement; notice that the timings of treatment are not the same for all city pairs. $Yearmonth_\tau$ is a dummy variable that equals to one if a certain year month belongs to event year $\tau$, otherwise zero. We use three months before the technology transfer, i.e., $\tau = -3$ as the benchmark year month. Therefore, the coefficient $\theta_\tau$estimates the impact of announcement of HSR connection, relative to year month $\tau = -3$. If the parallel trend hypothesis holds for the difference-in-differences specification, $\theta_\tau$should be indifferent from zero for any time prior to the event. 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 a more exogenous constructed city pairs. However, these governmental documents could be confidential and are difficult to 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 to 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.[7] It should be noted that indirect connect 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. Findings

### 5.1 Descriptive statistics

Table 1 reports the summary statistics for all 11,499,062 city pair-year month observations. We divide observations into four groups. In Panel B, observations are grouped by whether the city pair is ever directly connected by HSR at any point during our sample period. The never-connected city pairs and the ever-connected city pairs are very different in terms of the extensive and intensive margin of inter-city investment

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

flows. On average, city pairs that are ever connected to HSR are more likely to have investment flow across on the extensive margin, and the amount of 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. Even within this group, the city pairs experience a three times increase on the extensive margin and four times increase on the intensive margin after getting connected to HSR, 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 results for our baseline specification, equation (1), are reported in Table 3. 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 highly significance, which implies connection of HSR between two cities increases the number of investors by 8%, compared to the control pairs that are connected to HSR later.

Columns (3) and (4) of Table 3 present the effects 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. According to the strictest set up in Column (4), the coefficient on the treatment dummy is 0.375 with statistically significance, which implies that investment increases by 37.5 percentage points on average. Given that the sample mean of cross city investment is 4.29 million RMB in the pre-treatment period, being connected to HSR implies that cross-city private investment increases by 37.5%, corresponding to an increase in capital expenditures of 1.61 million RMB.[8]

### 5.3 Endogenous selection

One may worry about the endogenous selection of the HSR stations. Even though our identification strategy at the city pair level can handle the origin and destination city time trend and time invariant unobservables at the city pair level, endogeneity may still be an issue if the selection of HSR stations is based on the economic activities at the city pair level. If that is a valid concern, we should observe that HSR's positive effect is mainly driven by newly connected city pairs involved with at least one new HSR station, instead of new connection of two existing HSR stations. As presented in Table 4, the interaction of "NewHSR" (that equals to 1 if at least one new HSR station is involved in this newly connected pair) and Connect is negative in all the specifications and significant for the extensive margin analysis (columns 1 and 2). This indicates that

[8] 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 A1 (for the extensive margin) and Appendix Table A2 (for the intensive margin). These results are consistent with the main results in Table 3.

the positive effect of HSR connection is due to the new connection of existing HSR stations, which also mitigates the concern that the results are driven by the selection of cities hosting new HSR stations. Instead, the positive effect is driven by the additional exposure to the HSR network.

In order to further mitigate the endogeneity concern of high-speed rail placement, we drop the city pairs that are directly connected by high-speed rail, and use the city pairs that are indirectly connected by high-speed rail as treatment. 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. 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 coefficient on the treatment is almost the same with direct connection, which implies indirect connection of HSR between two cities also increases the number of investors 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 look at the intensive margin of cross city investments. The coefficient on the treatment dummy of different indirect connection measurements is 18.5% in column (4), which implies investment flow increases by 18.5%. The economic magnitude is a bit lower compared to direct connection, but still economically significant, and it can be considered as a lower bound of the estimation.

### 5.4 Robustness Checks

In this sub-section, we deal with two concerns regarding our identification strategy. First, the construction of high-speed rail takes time. Investors may respond to the news once they learn about the announcement of high-speed rail connection between two cities. Therefore, we need to incorporate the announcement time of the high-speed rails into the regression equation. Second, the difference-in-differences specification requires the treatment group and control group hold parallel trend in absence of the treatment.

We deal with the above concerns in two ways. First, we study the dynamic effect of high-speed rail connection using event study as shown in Equation (2), including the pre-announcement window, the after-announcement-before-connection window and the post-connection window. The coefficients are plotted in Figure 3 for each $\theta_\tau$ as well as the 95% confidence interval. The benchmark group is 13 to 24 months before announcement of each HSR connection. The first panel presents the estimates using invest dummy as dependent variable, and the second panel presents the estimates using the logarithm of investment amount as dependent variable.

Compared to the benchmark period, there is no significant increase in city-pair investment one year before the announcement of HSR connection, both at the extensive and intensive margin. This verifies the parallel trend assumption of the difference-in-difference design. There is a 2.9% increase in the number of newly registered firms and 15.3% increase in the dollar investment flow in a city pair after the announcement but before the real connection of HSR. These two coefficients are marginally significant at the 10 percent level though. The effects of HSR connection become significant and increases over time after the real connection timestamp, which increase from 6.5% at the extensive margin and 37.9% at the intensive margin in the first four months of the connection, to 21.5% at the extensive margin and 83.4% at the intensive margin even after two years from the connection time.

Second, in addition to the connection dummy, we further include an announcement dummy to control for the announcement effect, which equals to 1 if the time period is after the announcement but before the operation of the HSR, and a pre-announcement dummy to test for the parallel trend, which equals to 1 if the time period is six months or one year before the announcement of HSR. The coefficient on pre-announcement should be insignificant if the parallel trend assumption holds. Table 6 shows the results on such placebo tests. The dummy on HSR connection is still significantly positive and large in magnitude after controlling for the announcement and pre-announcement dummy variables. More importantly, the pre-announcement dummy is not significant in all the specifications, suggesting a parallel trend between the treated and control city pairs before announcement of HSR. In addition, the announcement effect is quite limited, which is only marginally significant in specifications without fully controlling for origin and destination specific time trend.

### 5.5 Discussions on Mechanisms

Admittedly, the results presented in the previous sections do not shed light on the mechanisms that drive the responses of cross-city investments to reductions in travel cost. The branch of literature on investment home bias has suggested theories including improved monitoring capabilities and access to information, i.e. the discovery of new investment opportunities. However, there is limited empirical evidence to identify or disentangle these underlying mechanisms. We explore three kinds of firm heterogeneity to test the monitoring and information channel. Monitoring requires that controlling shareholders to travel to plants. According to Stein (2002, 1891), the same is true for collecting "soft" information, that is, information that "cannot be credibly transmitted" and "cannot be directly verified by anyone other than the agent who produces it".

First, we divide 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. Controlling shareholders have advantage over both monitoring capability and information accessibility, while non-controlling shareholders have information advantage only. 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 they are investing in. For 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.

Table 7 reports the estimation of baseline specification, equation (1), with different cutoff of shareholders. As expected, the size of coefficients for controlling investors with shareholding more than 50% is significantly larger than that obtained using the subsample of non-controlling investors. Still, the estimation of non-controlling investors yield with statistically significant coefficients with smaller magnitude, indicating that the access of information is an important force driving inter-city investment flows. Overall, we can both identify the role of monitoring channel and information channel.

Second, we divide all firm-level investment by whether the investor or receiver side belongs to finance sector. Finance firms might find it easier to overcome the information barriers than non-finance firms by nature. Institutional investors are likely to be more dependent on airlines, and not affected by the introduction of HSR. Therefore, we expect the effects to be more significant in non-finance sectors. As shown in Figure XXX, the estimates exploring sectorial heterogeneity are consistent with our conjecture.

Third, we distinguish types of ownership for all firms through identifying whether the effective controller is state related agents or non-SOE agents in the registry information. The non-SOE firm types include collectively owned, privately owned, and foreign owned enterprises. We expect that the results to be more significant for non-SOE firms than for state-owned enterprises, as SOEs firms may acquire information through a patronage network provided by political leaders, which is less likely to be affected by reduction in travel costs. In addition, non-SOE firms might find it harder to finance and suffer more from information frictions. Figure XXX shows supporting evidence that non-SOE investors are affected by the reduction of travel costs brought by HSR connection with large economic magnitude, while SOE investors are not significantly affected.

Fourth, we divide all city pairs into five groups based on their geographic proximity. To be specific, we identify the centroid of a city and compute the straight-line distance of each city pair, and create five categories of city pairs at the cutoffs of 50 km, 100 km, 200 km and 400 km. Then, we interact these distance dummies with the connect dummy. Table 7 presents the results. Overall, as distance increases, the effect of HSR connection decreases both at the extensive margin and intensive margin. The positive effect disappears when the straight-line distance of a city pair is further than 400 km. This finding is consistent with Bernard et al. (2016) that HSR benefits close connections more than connections further away.

### 5.6 Discussions on Welfare Implications

In this sub-section, we would like to discuss about the welfare implications of high-speed rail connections based on our empirical findings. A full-fledged structural model to evaluate the welfare consequences of HSR connection is beyond the scope of the current paper. Instead, we aim to provide some discussions regarding the welfare consequences of HSR in reduced form models. Following the discussion in Bhamra and Uppal (2018) regarding the negative welfare implications of portfolio underdiversification, we employ city level diversification of investment flows as a measure of welfare.

In Table 8, we conduct city level (instead of city pair level) regression analysis to examine the effect of HSR on investment diversification. In particular, we look at the HSR connections from both the receiver's and the investor's perspectives. In columns (1) to (3) the dependent variable is the number of cities that a particular city i invests in (i.e. number of receiver cities); in columns (4) to (6) the dependent variable is the number of cities that city i receives investment from (i.e. number of investor cities). The first dependent variable measures outbound investments from city i and reflect the diversity of city i's portfolios, whereas the second dependent variable reflects the diversity in the sources of inbound investments into city i.

The main independent variable in all regressions is the number of connections from city i or to city i. In all regressions, we control for time fixed effects and city fixed effects; we further add city specific cubic year trend in columns (2) and (4). The coefficient on the number of connections is positive and significant at the 1 percent level in column 2, which suggests that additional HSR connections encourage investors in city i to diversify their investment destinations. While city i's number of HSR connections is also positively associated with the number of cities that city i receives investments from, the parameter estimate in column (6) is about one third of the magnitude observed in column (3), and it is not statistically significant ($t<1.4$). The patterns observed in this table indicate that the potential welfare gains associated with increased diversification are predominantly coming from the increased portfolio diversification of a city's investors, rather than the increased diversification in each city's sources of inbound investments. Admittedly, even with all the fixed effects and time trends that are included, we cannot fully address the endogeneity concerns and argue for a causal relation. Hence, these results provide only suggestive evidence of welfare gains from the HSR connections.

We have shown that high-speed rail connections facilitate investment flows across cities. Do such changes in investment flows increase or decrease regional inequality? It depends on the direction of capital flows. Lucas (1990) illustrated the paradox of why capital does not flow from rich to poor countries and regions. One major explanation is that higher information frictions for investors from rich areas to invest in poorer areas than the other way around. We ask the question does HSR have larger impacts on rich-to-poor flows than poor-to-rich ones? To answer it, we divide cities into three groups: rich, middle and poor according to their GDP per capita at base year 2004, and interact connect dummies with dummies of rich, middle and poor. For example, dummy variable R(rich) to M(iddle) equals one if investor city belongs to rich group and destination city belongs to middle group. The omitted or baseline group is poor to poor dummy variable.

As shown in Table 9, when controlling for the most flexible origin and destination city specific time trend (column (2) and column (4)), the coefficients imply that HSR connection increases the investment flow from rich cities to rich cities of the largest magnitude, which is true at both the extensive and intensive margins. Even more, the entire extensive margin is driven by the increase of investment flows from rich cities to rich cities. At the intensive margin, the investment flow from rich to rich still shows the largest effect. The coefficient on the interaction of middle to middle cities dummy and the connection dummy is marginally significant too, with smaller magnitude compared to the coefficient on rich to rich. Overall, these results suggest that the additional inter-city investment flows induced by HSR connections may further aggravate regional inequalities in China.

## 6. 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. This project focuses how the reduction in passenger travel cost facilitates information and capital flows cross-city. We make two contributions to the impacts of large transportation infrastructure projects, in the context of a rapid and enormous expansion of high-speed rails in China. First, we document "home bias" patterns in firm-to-firm investments in China using the whole universe of firm investment records from 2007 to 2015. The database we use for analyses is uniquely obtained from the China State Administration for Industry and Commerce and it is thus far the most comprehensive data on firm activities cross all regions and sectors in China. Second, by exploiting plausibly exogenous shocks to travel cost from the expansion of HSR system in China, we evaluate how reduction in travel cost facilitate cross-city investments. We find that direct HSR connection increases the number of investors between the city pair increase by 8%, and amount of investment increases by 38%. The results are robust when we control for city pair heterogeneity and time-varying local shocks that could potentially drive the introduction of new HSR routes, and when we consider only city pairs that are indirectly connected to HSR. Moreover, connection to HSR increases inter-city investment for both controlling and non-controlling shareholders, which indicates that improved monitoring capabilities and access to information are both important underlying mechanisms.

Our paper is notably distinct from the studies on the effects of goods-shipping transportation infrastructure as highways or traditional railroads development and the ones on the partial effect of HSR. This paper abstracts from some important aspects of the real world that could affect the impacts of HSR connection. Reducing the cost of passenger travel between cities makes face-to-face contacts much more convenient. 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 the first step in evaluating its general economic impacts.

## References

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

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? The Quarterly Journal of Economics, 122 (2): 775–805.

Bernard, Andrew B., Andreas Moxnes, Yukiko U. Saito (2016): "Production networks, geography and firm performance," CEP Discussion Paper No.1435.

Bernstein, Shai, Xavier Giroud, and Rick Townsend (2015), "The Impact of Venture Capital Monitoring," Journal of Finance, forthcoming.

Charnoz, Pauline, Claire Lelarge and Corentin Trevien. (2017). "Communication Costs and the Internal Organization of Multi-Plant Businesses: Evidence from the Impact of the French High-Speed Rail," Banque de France Working Paper 635.

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.

Donaldson, Dave. (2014). "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,” The Quarterly Journal of Economics, 128(2): 861–915.

Huberman, Gur. (2001). Familiarity Breeds Investment. The 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. The Journal of Finance, 11 (1): 267-306.

Lin, Yatang. (2017). “Travel costs and urban specialization patterns: Evidence from China high speed railway system,” Journal of Urban Economics, 98: 98–123.

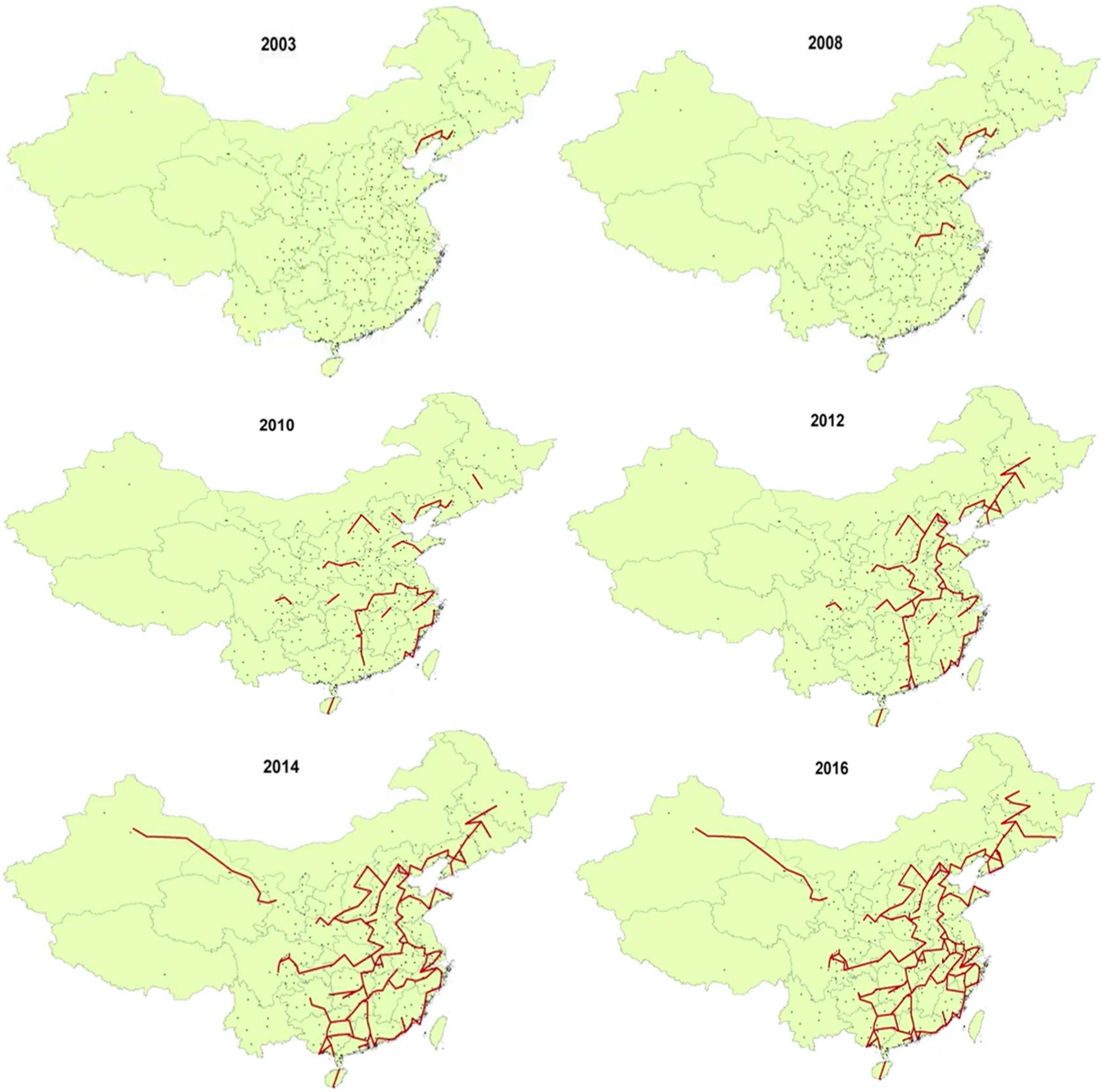
MRC (2008): “Mid-to-Long Term Railway Network Plan,” Report of the Ministry of Railways of China.

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.

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.

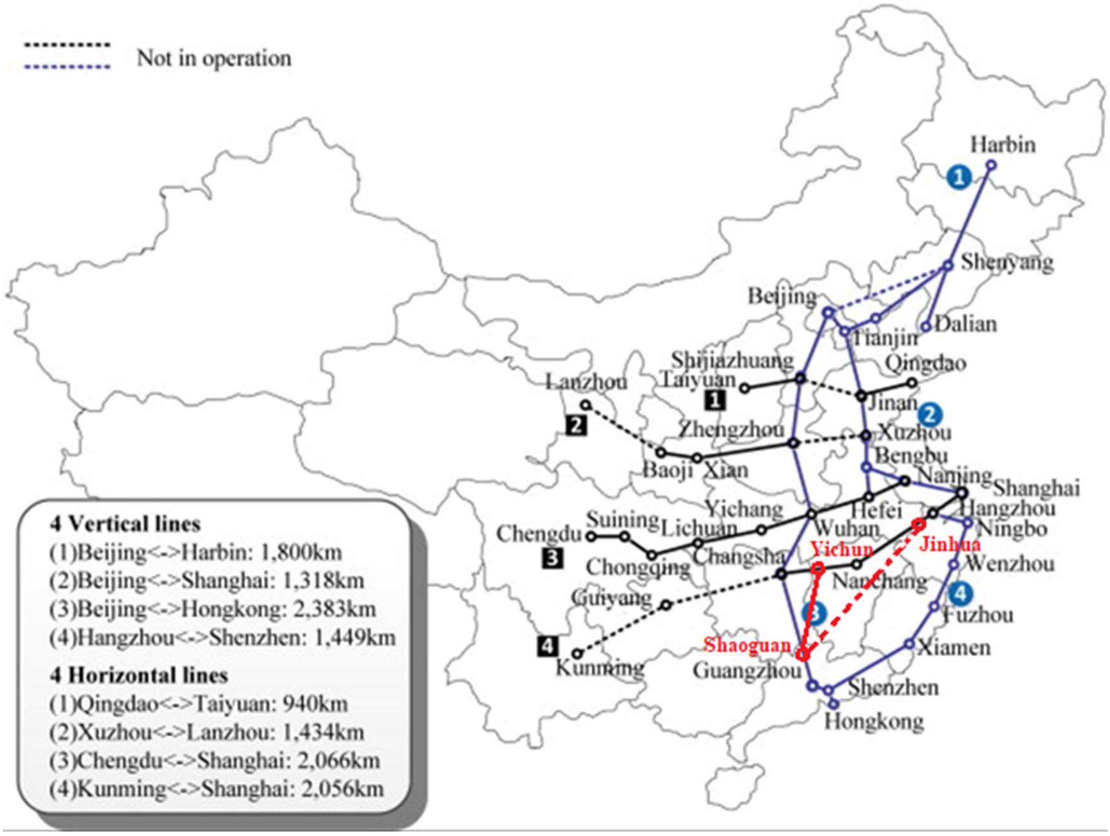
Seasholes, Mark S. and Ning Zhu. (2010). Individual Investors and Local Bias. The Journal of Finance, 105 (5): 1987-2010.

Figure 1. Evolution of HSR Network from 2003 to 2016



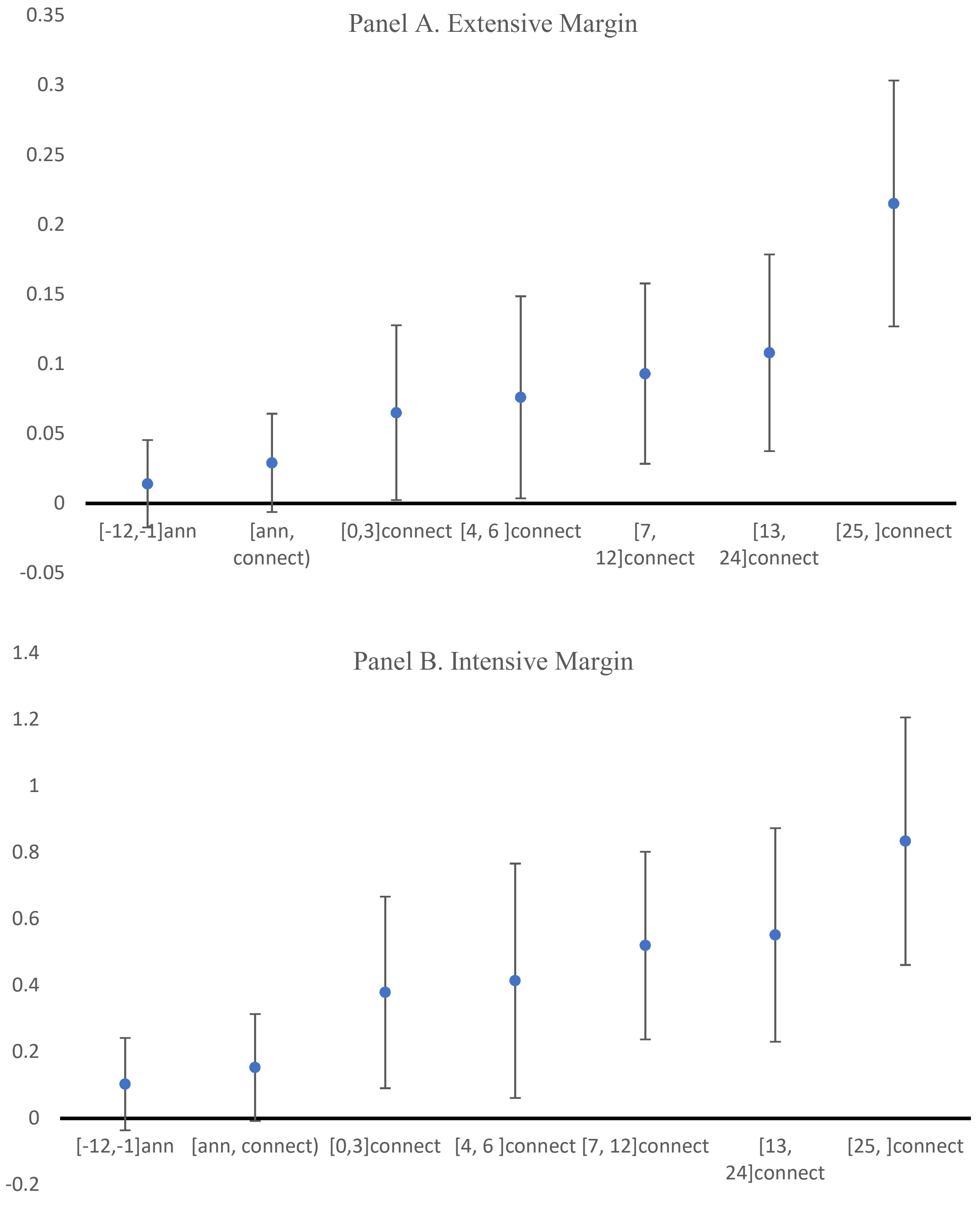

Notes: These figures display the evolution of HSR expansion from year 2003 to 2016. The lines in bold red are lines in use by the end of that year. Each dot represents a prefecture-level city.

Figure 2. Construction of Indirect Connect Measures of HSR Network



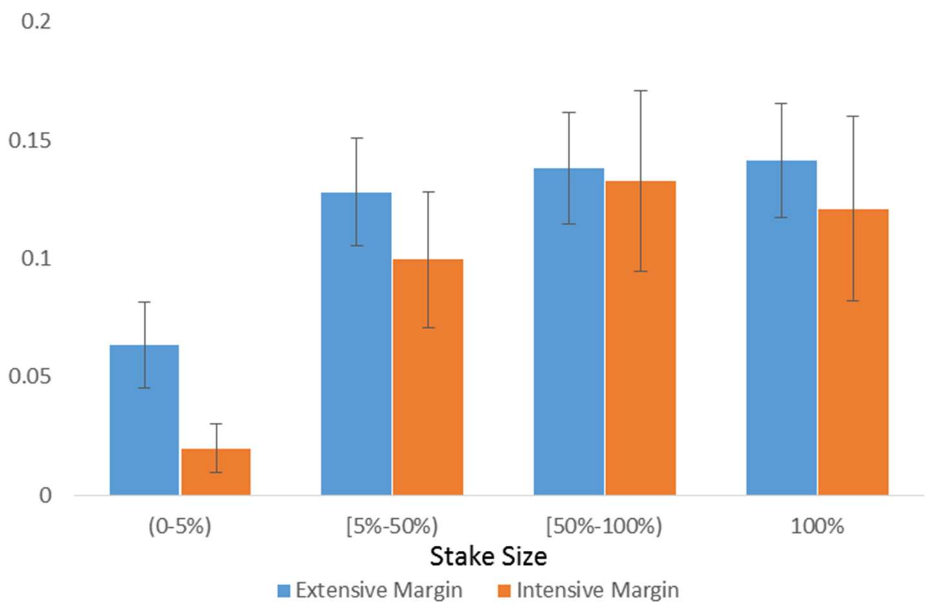

Notes: Yichun and Shaoguan is considered to be indirectly connected after both Changsha-Nanchang and Changsha-Guangzhou lines are in operation.

Figure 3. Dynamic Effect of HSR Announcement and Connection



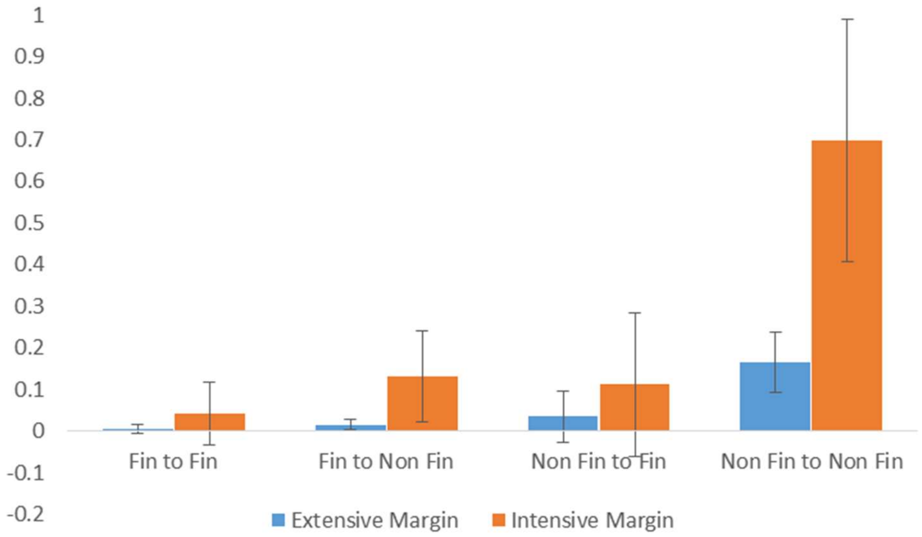

Notes: This figure visualizes the coefficients $\theta$ in Equation (2). The top panel reports the extensive margin (# of investments) of cross-city investment flows around the introduction of direct HSR connection, whereas the bottom panel reports the intensive margin (total amount of investments). The regression estimates are available in Appendix Table A3 (columns 2 and 4). The coefficients are presented in dots, with their 95% confidence intervals.

Figure 4. Coefficients by Different Shareholding Amounts



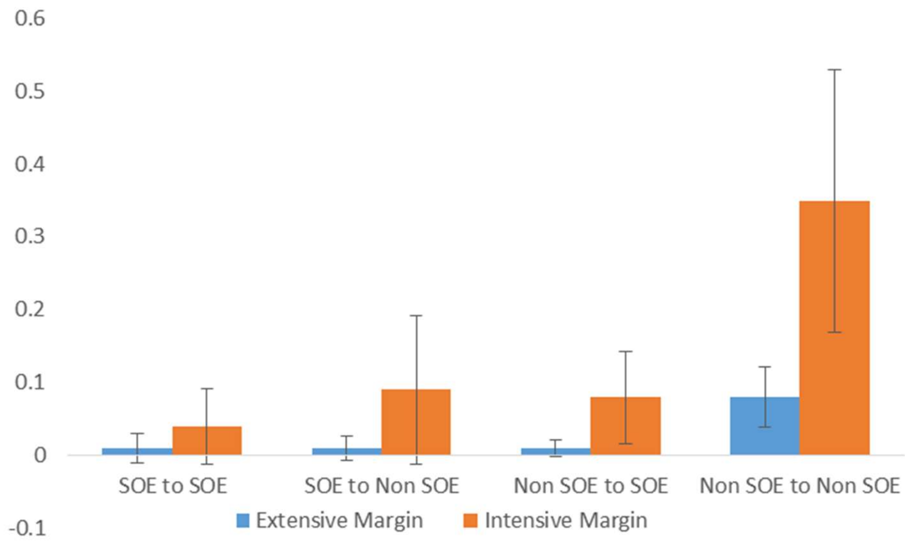

Notes: The sample is split into four categories based on the investor's stake size. The four categories are: (1) the investor holding 0% to 5% (not inclusive) of the invested firm's share; (2) the investor holding 5% (inclusive) to 50% (not inclusive) of the invested firms' share; (3) the investor holding 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. Coefficients by Financial and Non-Financial Enterprises




Notes: The sample is split into four categories based on whether the investor is a firm in the financial sector, and whether the receiving firm is a financial firm. The heights of the bars represent the magnitudes of the coefficients while the lines denote the 95% confidence intervals.

Figure 6. Coefficients by State and Non-State-Owned Enterprises




Notes: The sample is split into four categories based on whether the investor is an SOE firm, and whether the receiving firm is an SOE firm. The four categories are (1) SOE as investor and SOE as receiver; (2) SOE as investor and non-SOE as receiver; (3) non-SOE as investor and SOE as receiver; and (4) non-SOE as investor and non-SOE as receiver. The heights of the bars represent the magnitudes of the coefficients while the lines denote the 95% confidence intervals.

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

| | (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 | | ✓ | | ✓ |

*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). 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

| Variables | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | *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 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. $IndirectConnect_{i,j,t}$ 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. Heterogeneity by Distance

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Variables | *Lnumber* | | *Linvestment* | |
| Connect | -0.006 | -0.043* | -0.001 | -0.044 |
| | (-0.355) | (-1.854) | (-0.016) | (-0.442) |
| Connect*(0,50km] | 0.278** | 0.352*** | 0.800 | 1.326*** |
| | (1.994) | (6.772) | (1.635) | (4.449) |
| Connect*(50,100km] | 0.155*** | 0.268*** | 0.608*** | 0.973*** |
| | (3.9) | (9.553) | (4.557) | (8.937) |
| Connect*(100,200km] | 0.068** | 0.203*** | 0.244** | 0.693*** |
| | (2.418) | (9.062) | (2.313) | (7.623) |
| Connect*(200,400km] | -0.024 | 0.098*** | -0.126 | 0.295*** |
| | (-0.949) | (4.969) | (-1.334) | (4.174) |
| | | | | |
| Observations | 170,496 | 167,616 | 170,496 | 167,616 |
| R-squared | 0.585 | 0.749 | 0.48 | 0.654 |
| 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). These city pairs are divided into five groups (0,50km], (50,100km], (100,200km], (200,400km], and above 400km based on the geographic straight-line distance between the centroid of the two cities. These time-invariant dummies are interacted with time-varying *Connect* to estimate the heterogeneous effect of high-speed rail connection of city pairs with different geographic proximity. The omitted group is the city pairs that are further than 400km. 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. Welfare Implication: HSR Connection and Investment Diversification

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | *Log(NumberReceiver)* | | *Log(NumberInvestor)* | |
| Log(*NumberConnect*) | 0.046*** | 0.046*** | 0.016 | 0.016 |
| | (4.381) | (4.381) | (1.370) | (1.370) |
| Observations | 40,464 | 40,464 | 40,464 | 40,464 |
| R-squared | 0.804 | 0.804 | 0.698 | 0.698 |
| Yearmonth Dummy | ✓ | ✓ | ✓ | ✓ |
| City FE | ✓ | ✓ | ✓ | ✓ |
| City*Year Trend | | ✓ | | ✓ |

*Notes.* This table reports the estimation results from analyses of the HSR connection and the source or destination of investments. In the first two columns, the outcome variable is the logarithm of $NumberReceiver_i$, which is the number of cities that investors from city *i* invests to. In the last two columns, the outcome variable is the logarithm of $NumberInvestor_i$, the number of cities that city *i* receives investments from. The main explanatory variable is the number of cities that city *i* is directly connected to via HSR. City-specific cubic year trend is controlled for 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 9. The Impact of HSR Connection on Regional Inequality

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | *Lnumber* | | *Linvestment* | |
| RtoR*Connect | 0.429*** | 0.287*** | 1.354*** | 0.840*** |
| | (10.133) | (7.307) | (10.161) | (5.388) |
| RtoM*Connect | 0.065** | -0.010 | 0.408*** | -0.004 |
| | (2.548) | (-0.197) | (3.534) | (-0.018) |
| RtoP*Connect | 0.069** | -0.076 | 0.456*** | 0.060 |
| | (2.287) | (-1.481) | (3.146) | (0.240) |
| MtoR*Connect | -0.062*** | -0.034 | -0.091 | 0.278 |
| | (-3.643) | (-0.548) | (-1.098) | (1.010) |
| MtoM*Connect | -0.079*** | 0.066 | -0.258*** | 0.372* |
| | (-4.958) | (1.490) | (-3.650) | (1.753) |
| MtoP *Connect | -0.037* | 0.043 | -0.105 | 0.482 |
| | (-1.656) | (0.689) | (-1.006) | (1.646) |
| PtoR*Connect | -0.098*** | -0.028 | -0.345*** | 0.065 |
| | (-6.869) | (-0.694) | (-4.614) | (0.359) |
| PtoM*Connect | -0.122*** | 0.064 | -0.462*** | 0.210 |
| | (-9.682) | (1.474) | (-7.255) | (1.080) |
| PtoP*Connect | -0.146*** | 0.025 | -0.607*** | 0.219 |
| | (-13.765) | (0.669) | (-12.102) | (1.138) |
| Observations | 170,496 | 167,616 | 170,496 | 167,616 |
| R-squared | 0.599 | 0.749 | 0.487 | 0.654 |
| City-pair FE | ✓ | ✓ | ✓ | ✓ |
| Year-month FE | ✓ | | ✓ | |
| Origin (i) * year-month FE | | ✓ | | ✓ |
| Destination (j) * year-month FE | | ✓ | | ✓ |

*Notes.* Cities are divided into three groups according to GDP per capita at base year 2004: (R)ich, (M)idle, and (P)oor. City-pairs are categorized into 3x3 groups using the categories of the source (*i*) and destination (*j*) cities. 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.

**Online Appendix**

Table A1. 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 | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| City Pair FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Origin city *year FE | No | Yes | No | Yes | No | Yes | No | Yes |
| Destination city *year FE | No | Yes | No | Yes | No | Yes | No | Yes |

*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 A2. 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 | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| City Pair FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Origin city *year FE | No | Yes | No | Yes | No | Yes | No | Yes |
| Destination city *year FE | No | Yes | No | Yes | No | Yes | No | Yes |

*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 A3. Dynamic Effect of HSR Announcement and Connection

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Variables | lnumber | lnumber | linvestment | linvestment |
| preannounce_oneyear | 0.015* | 0.014 | 0.064* | 0.103 |
| | (1.918) | (0.879) | (1.932) | (1.439) |
| announce | 0.034*** | 0.029* | 0.141*** | 0.153* |
| | (3.506) | (1.656) | (3.372) | (1.876) |
| connect [0,3m] | 0.048*** | 0.065** | 0.217*** | 0.379** |
| | (3.714) | (2.019) | (3.606) | (2.579) |
| connect[4,6m] | 0.044*** | 0.076** | 0.182*** | 0.414** |
| | (3.100) | (2.061) | (2.796) | (2.304) |
| connect[7,12m] | 0.045*** | 0.093*** | 0.193*** | 0.520*** |
| | (3.391) | (2.830) | (3.175) | (3.602) |
| connect[13,24m] | 0.058*** | 0.108*** | 0.237*** | 0.552*** |
| | (3.861) | (2.998) | (3.667) | (3.372) |
| connect[25m, ] | 0.129*** | 0.215*** | 0.510*** | 0.834*** |
| | (5.593) | (4.820) | (5.753) | (4.387) |
| Observations | 171,072 | 168,192 | 171,072 | 168,192 |
| R-squared | 0.583 | 0.746 | 0.479 | 0.653 |
| Yearmonth FE | Yes | Yes | Yes | Yes |
| City Pair FE | Yes | Yes | Yes | Yes |
| Origin city*yearmonth FE | No | Yes | No | Yes |
| Destination city *yearmonth FE | No | Yes | No | Yes |

*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_oneyear is a dummy that turns on if year month t is within one year before the announcement of HSR connection; announce is a dummy that turns on if year month t is after the announcement period and before the connection of HSR; connect[i,j] are dummies 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 A4. Heterogeneity on Control: Controlling Investors versus Non-Controlling Investors

| | | Panel A: Non-controlling Investors | | | |
|---|---|---|---|---|---|
| | Share | (1) | (3) | (4) | (6) |
| Variables | (0,5%) | lnumber | lnumber | linvesetment | linvesetment |
| connect | | 0.012*** | 0.021 | 0.048*** | 0.098* |
| | | (3.788) | (1.471) | (3.674) | (1.947) |
| | | 0.003 | 0.014 | 0.013 | 0.050 |
| Observations | | 171,072 | 168,192 | 171,072 | 168,192 |
| R-squared | | 0.193 | 0.469 | 0.156 | 0.435 |
| | [5%,50%) | lnumber | lnumber | linvesetment | linvesetment |
| connect | | 0.017*** | 0.037** | 0.085*** | 0.144* |
| | | (3.19) | (2.17) | (3.11) | (1.87) |
| | | 0.01 | 0.02 | 0.03 | 0.08 |
| 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 | linvesetment | linvesetment |
| connect | | 0.019*** | 0.052*** | 0.095*** | 0.285*** |
| | | (3.259) | (3.232) | (3.411) | (3.953) |
| | | 0.006 | 0.016 | 0.028 | 0.072 |
| Observations | | 171,072 | 168,192 | 171,072 | 168,192 |
| R-squared | | 0.448 | 0.645 | 0.362 | 0.570 |
| Variables | 100% | lnumber | lnumber | linvesetment | linvesetment |
| connect | | 0.013** | 0.044*** | 0.094*** | 0.301*** |
| | | (2.065) | (3.256) | (2.926) | (4.049) |
| | | 0.006 | 0.013 | 0.032 | 0.074 |
| Observations | | 171,072 | 168,192 | 171,072 | 168,192 |
| R-squared | | 0.414 | 0.629 | 0.333 | 0.555 |
| Year-month dummy | | Yes | Yes | Yes | Yes |
| city-pair FE | | Yes | Yes | Yes | Yes |
| Origin city * year-month FE | | No | Yes | No | Yes |
| Destination city * year-month FE | | No | Yes | No | Yes |

*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.

Table A5. Heterogeneity on Industry: Finance and Non-Finance Firms

| | Panel A: Finance to Finance | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Variables | lnumber | lnumber | linvesetment | linvesetment |
| connect | 0.0019** | 0.0057 | 0.0155** | 0.0424 |
| | (2.1713) | (1.0380) | (2.2690) | (1.0801) |
| Observations | 171,072 | 168,192 | 171,072 | 168,192 |
| R-squared | 0.06 | 0.29 | 0.05 | 0.28 |
| | Panel B: Finance to Non-Finance | | | |
| | lnumber | lnumber | linvesetment | linvesetment |
| connect | 0.0070*** | 0.0156** | 0.0405*** | 0.1327** |
| | (3.0695) | (2.5344) | (2.8204) | (2.3833) |
| Observations | 171,072 | 168,192 | 171,072 | 168,192 |
| R-squared | 0.18 | 0.44 | 0.14 | 0.41 |
| | Panel C: Non-Finance to Finance | | | |
| Variables | lnumber | lnumber | linvesetment | linvesetment |
| connect | 0.0079 | 0.0351 | 0.0291 | 0.1121 |
| | (1.5395) | (1.1187) | (1.4062) | (1.2785) |
| Observations | 171,072 | 168,192 | 171,072 | 168,192 |
| R-squared | 0.21 | 0.47 | 0.15 | 0.43 |
| | Panel D: Non-Finance to Non-Finance | | | |
| Variables | lnumber | lnumber | linvesetment | linvesetment |
| connect | 0.0538*** | 0.1647*** | 0.2311*** | 0.6996*** |
| | (3.7864) | (4.4243) | (3.7649) | (4.6928) |
| Observations | 171,072 | 168,192 | 171,072 | 168,192 |
| R-squared | 0.46 | 0.71 | 0.39 | 0.64 |
| Year-month dummy | Yes | Yes | Yes | Yes |
| city-pair FE | Yes | Yes | Yes | Yes |
| Origin city * year-month FE | No | Yes | No | Yes |
| Destination city * year-month FE | No | Yes | No | Yes |

*Notes*: The table reports difference-in-differences estimation results by finance and non-finance 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. Robust standard errors clustered at city pair level and the corresponding t-statistics are reported in parentheses.

Table A6. Heterogeneity on Ownership: SOE vs Non-SOE Firms

| | Panel A: SOE to SOE | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Variables | lnumber | lnumber | linvesetment | linvesetment |
| 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 | linvesetment | linvesetment |
| 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 | linvesetment | linvesetment |
| 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 |
| | Panel D: POE to POE | | | |
| Variables | lnumber | lnumber | linvesetment | linvesetment |
| connect | 0.029*** | 0.078*** | 0.126*** | 0.353*** |
| | (3.025) | (3.745) | (3.047) | (3.807) |
| Observations | 171,072 | 168,192 | 171,072 | 168,192 |
| R-squared | 0.567 | 0.737 | 0.460 | 0.641 |
| Year-month dummy | Yes | Yes | Yes | Yes |
| city-pair FE | Yes | Yes | Yes | Yes |
| Origin city * year-month FE | No | Yes | No | Yes |
| Destination city * year-month FE | No | Yes | No | Yes |

*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.