Review of Finance, 2023, 1119–1154
https://doi.org/10.1093/rof/rfac049
Advance Access Publication Date: 10 August 2022

OXFORD

Expanding Footprints: The Impact of Passenger Transportation on Corporate Locations*

Yatang Lin[1], Yu Qin[2], Johan Sulaeman[3], Jubo Yan[4], and Jialiang Zhang[5]

[1]Department of Economics, Division of Public Policy and Social Science, HKUST, Hong Kong, [2]Department of Real Estate, NUS Business School, National University of Singapore, Singapore, [3]Sustainable and Green Finance Institute (SGFIN) and NUS Business School, National University of Singapore, Singapore, [4]School of Social Sciences, Nanyang Technological University, Singapore and [5]School of Economics, Central University of Finance and Economics, China

Abstract

This article investigates how transportation networks shape firms' geographic footprint by reducing monitoring costs of distant investments. Exploiting the staggered expansions of China's passenger high-speed rail (HSR) network, we document that the amount of intercity investment between a pair of cities increases by 45% with the introduction of an HSR line connecting the cities. We enhance the causal inference by applying high-dimensional fixed effects, and focusing on city pairs that are "accidentally" connected in the network. The HSR effect is the strongest in industries that require on-site monitoring, as well as for controlling stakes in large distant investments.

Keywords: High-speed rail, Home bias, Investment flows, Geographic footprint

JEL classification: G14, G34, H54, R42

Received December 30, 2020; accepted December 8, 2021 by Editor Zhiguo He.

* We thank Sumit Agarwal, Loren Brandt, Hanming Fang, Bo Li, Wenlan Qian, Xiaobo Zhang, Xiaodong Zhu, and the seminar participants at the National University of Singapore, National Chengchi University, Nanyang Technological University, Peking University, the University of Toronto, and the participants at the 2018 SFS Cavalcade Asia-Pacific, the 2018 Econometric Society China Meeting, the 2019 Urban Economic Association Meeting, the 2019 HKUST Conference on International Economics, the 2019 Conference on Transition and Economic Development, the 2020 AREUEA Meeting, and the 2021 Royal Economic Society Annual Conference for helpful discussions and valuable comments. All errors are our own.

1. Introduction

The tendency of firms to focus on geographically close investments remains prevalent over the last several decades (Carvalho and Gabaix, 2013; Bernile, Kumar, and Sulaeman, 2015). This runs counter to two potential benefits of geographic expansions. First, the diversification benefit espoused in basic investment courses. Expanding a firm's geographic footprint would reduce its exposure to local economic shocks by bringing in a new set of potential customers who may not be affected by these localized shocks. Second, a seminal work by Lucas (1990) points out a surprising paradox that investment capital does not seem to flow from developed to developing areas despite the presumably higher return to capital (ROC) in developing areas. Geographic expansions into areas with lower existing investment levels would allow firms to exploit potential investment opportunities that offer higher expected returns.

Despite these potential benefits, corporate expansions of geographic footprints face substantial challenges. As firms develop local operational advantages, corporate managers have the incentive to build on these advantages—making it difficult to justify exploring distant investments associated with higher monitoring cost and more difficult access to information. Recent advances in information technology (e.g., Internet and video conferencing) have facilitated more information flow across physical distances, allowing firms to explore distant investments with reduced need for personal travel. Nevertheless, examining the effect of physical access to distant information on the geographic footprints of corporate investments is not straightforward, as the decision to obtain distant information is endogenous (Van Nieuwerburgh and Veldkamp, 2010). Our article attempts to measure this effect by examining an exogenous shock to the monitoring and information cost of distant investments.

Our analysis is motivated by the rapid improvement in China's transportation connectivity over the last several decades. This improvement is particularly salient in the growth of its passenger high-speed rail (HSR) network. This network has grown by over 25,000 km over just the last decade, accounting for more than half of the world's total length of HSR tracks, with ambitious plans to increase the network coverage by >50% over the next decade.[1]

We exploit this massive rollout of passenger transportation infrastructure on corporate geographic footprint. Whereas railways and highways reduce the cost of moving goods and facilitate interregional trade (Donaldson, 2018), the improvements in passenger transportation connectivity accelerate passenger flows and reduce travel time. By allowing for more frequent physical, face-to-face interactions, the improved connectivity opens up new possibilities of direct communications and interactions across cities—resulting in exogenous shocks to the monitoring and information cost of distant investments. Our analysis is relevant for readers interested in the effect of information on corporate investment patterns, as well as those interested in the effects of large-scale infrastructure projects on the dynamics of the Chinese economy—the fastest growing economy in the world during this period.

We link the expansion of China's passenger HSR network with the registration information of the universe of Chinese firms and analyze how the expansion of the HSR services allows firms to expand their geographic footprint. Before delving into the analysis, it would

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

be useful to describe the main dataset that we use to measure the geography of corporate investments. Our data come from the registration information of the universe of Chinese firms from 2004 to 2015. These data are obtained from the China State Administration for Market Regulation.[2] For any firm registered during this period, we can identify its shareholders and the addresses of these shareholders. Therefore, we can trace the geographic footprint of a firm by linking it to any new firms that it invests in, either as a controlling or non-controlling shareholder. This universe of corporate registrations captures all new firm activities in China, allowing us to evaluate the investment patterns across all regions and sectors in China. Within this broad universe, we observe that firms display a strong propensity to invest locally, particularly within the same city, consistent with the existing literature.

To operationalize the analysis, we aggregate firm-level investment flows (from a shareholder in city A to a new firm in city B) to investment flows between Chinese city pairs. We then implement 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 between the pair, compared with unconnected city pairs. Our specifications include a stringent set of control variables. In particular, the rich information in city-to-city investment flow dynamics allows us to control for both (i) time-invariant pairwise characteristics such as geographical/cultural distances and baseline business networks with city-pair level fixed effects and (ii) time-varying economic shocks at both the origin and destination cities with (origin city $\times$ month)- and (destination city $\times$ month)-fixed effects.

We first document the effect of the introduction of a *direct* HSR connection between a pair of Chinese cities on the investment flows between those cities. The number of a city's investors investing in another city increases by 8% upon the introduction of a direct HSR connection between the two cities. The corresponding increase in the amount of intercity investments is 45%. The large economic magnitude reflects the importance of transportation infrastructure development in improving investment flows across cities and shaping the expansion of the geographic footprints of firms.

We also investigate the specific channel through which HSR can causally affect investment flows: the reduction in intercity travel time. We first define the time-varying travel time between a pair of Chinese cities as the minimum travel time at any given month using three distinct modes of transportation: air, train (including HSR), and road. The staggered introductions of HSR connections change travel times on some routes but not necessarily all of them. Our estimation indicates that the introduction of a direct HSR connection leads to a 35.6% drop in bilateral travel time on average. The instrumented travel time changes allow us to estimate the direct response of intercity investment flows to the reduction in travel time driven by HSR connections—rather than due to improvements in other modes of transportations, for example, introductions of new airline routes.

Our analyses include high-dimensional fixed effects to control for various factors that can affect either infrastructure development, travel time, and/or investment patterns across cities. These time-varying fixed effects also capture the secular growth of various sectors of Chinese economy, including tourism and business travels, real estate and construction, as

2 The business registration information we use is collected by the China State Administration for Industry and Commerce (SAIC) prior to its split in March 2018 into two separate entities: China State Administration for Market Regulation and China National Intellectual Property Administration.

well as the information technology sector. Nevertheless, drawing causal inference in this setting remains hampered by the concern that HSR lines might be endogenously placed between pairs of cities with *growing* economic linkages. We employ two approaches to address these potential identification issues.

First, we explore the time-series evolution of intercity investment flows before and after the announcement of new HSR connections, as well as around the actual opening of the connections. The HSR effect is not observed prior to the announcement of a new HSR connection, inconsistent with endogenous network placements driving the results. There is only a weak effect between the announcement and the opening of the connection, which may be related to the anticipation of the upcoming connection. The vast majority of the HSR effect on inter-city investments occurs after the two cities are actually connected via HSR. We observe an almost immediate effect after the opening of the connection, with the effect becoming stronger with time afterward.

Second, the staggered expansion of an extensive four-vertical-four-horizontal layout of China's HSR network formation allows us to exploit pairwise connections that are likely to be "accidental" outcomes of the more intentional connections in other parts of the network. We define *indirect* connections as connections between non-nodal cities that are formed when one part of the network is connected to a different part of the network. These connections are less likely to be affected by the omitted variable concerns associated with the likely endogenous rationale for establishing direct (point-to-point) transportation links, such as airline routes. As such, this analysis allows for a cleaner inference regarding the causal effect of HSR connections on investment flows across regions. The estimates of indirect connection treatment effect remain large in magnitude—about half the size of direct effect estimates—and statistically significant.

We also use the indirect HSR connection measure to examine the travel-time channel. After documenting a 10.1% drop in bilateral travel time upon the introduction of an indirect HSR connection, we estimate the response of intercity investment flows to travel time changes instrumented by the introduction of indirect HSR connections. We estimate the elasticity between the instrumented reduction in travel time and bilateral investment value to be –1.42.

To shed light on the mechanisms at work, we explore the heterogeneity in the HSR effect across city pairs by industries and ownership structures. First, the HSR effect is stronger for investments in industries that benefit more from face-to-face communication and on-site monitoring. 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. Nevertheless, HSR treatment effects are also statistically and economically significant for non-controlling investment. In particular, the total amount of small stake investments (0–4.9% of the firm) increases by 9.8% with HSR connection, which 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.

In the final part of the article, we provide a set of auxiliary evidence on the welfare and distributional implication of HSR-induced growth in investment flows. We document that the additional investment flows induced by HSR connections act to close the gaps in the rates of ROC across cities and industries. We find that the HSR effects are larger for industries in which the destination city offers a higher industry-level ROC than the corresponding industry-level return in the source city. Further evidence also indicates stronger HSR treatment effects on investment flows from high-income to low-income cities. These consistent pieces of evidence speak directly to the Lucas' (1990) paradox, and reflect HSR's valuable role in meliorating regional inequality issues and promoting a more inclusive economic growth.

The current study contributes to our understanding of how firms' geographic footprints are shaped. These footprints have potentially important implications on the propagation of economic shocks across regions and even across countries. In a recent study, Giroud and Mueller (2019) highlighted how *within-firm* reallocation of resources can propagate negative local shocks throughout each firm's existing footprints. Our study highlights how these footprints are shaped, potentially foretelling how future shocks would be propagated across regions.

The current study is also closely related to the literature that examines the effects of *passenger* transportation on economic outcomes.[3] Most studies in this literature examine the potential effects of air travel improvements in various settings (e.g., Giroud, 2013; Bernstein, Giroud, and Townsend, 2016), which largely focus on the impact of travel cost reduction on the allocation of investments from headquarter to existing subsidiary factories. In addition to offering a cleaner causal inference due to the exogenous nature of indirect HSR connections, a major contribution of the current article is that we examine the impact of travel cost on the geographic footprints of firms, which are reflected in investments to new, distant subsidiaries, and branches. In addition, we leverage the comprehensive firm registration data to explore potentially different mechanisms at work for various types of investments—such as controlling versus non-controlling stakes—and industries—for example, manufacturing versus service industries.

A few recent studies focus on the economic impacts of HSR developments, including a study by Bernard, Moxnes, and Saito (2019) on the impact of Shinkansen line on supplier relationship among Japanese firms.[4] 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 (Dong, Zheng and Kahn, 2020). Against this backdrop, the current article identifies a *causal* mechanism at the city-pair level of how infrastructure developments can affect investment flows and resource allocations across cities.

Analyzing the expansions of the 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

3 Our study is also related to the literature that examines roads and railways that reduce trade cost. 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, Duflo, and Qian, 2012), and asymmetric effects on core and peripheral markets (Faber, 2014). Other papers have explored the effects of urban transportation improvements on urban growth (Duranton and Turner, 2012) and urban form (Baum-Snow, 2007; Baum-Snow *et al.*, 2017) . 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 additional support for the consensus in this literature that transportation infrastructure reduces trade costs.

4 Other studies examine the effects of HSR on passenger travel behavior (high-speed Eurostar; includes: Behrens and Pels, 2012) and the economy of regions that are made more accessible (HSR connecting Cologne and Frankfurt; Ahlfeldt and Feddersen, 2018).

and Venables, 2004; Glaeser and Resseger, 2010). 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, that is, "home bias" in both cross-country and within-country setting, across various types of investments (French and Poterba, 1991; Coval and Moskowitz, 1999; Ivković and Weisbenner, 2005; Alcácer and Delgado, 2016). 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 article highlights the effects of exogenous shocks to physical "distance" on the propensity to seize non-local investment opportunities.

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

2. Background

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 started in 2008 when China's State Council came up with the Mid-to-Long Term Railway Development Plan, setting the goal of developing a national HSR grid consisting of four north–south corridors and four east–west corridors (MRC, 2008).

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

According to the Ministry of Railway's (MOR) document (2008),[5] 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 the existing transportation networks to a possible extent.

The construction costs of HSR are estimated between 80 and 120 million *yuan* per kilometer (US$13–20 million), excluding stations (Bullock, Salzberg, and Jin, 2012). The

5 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.

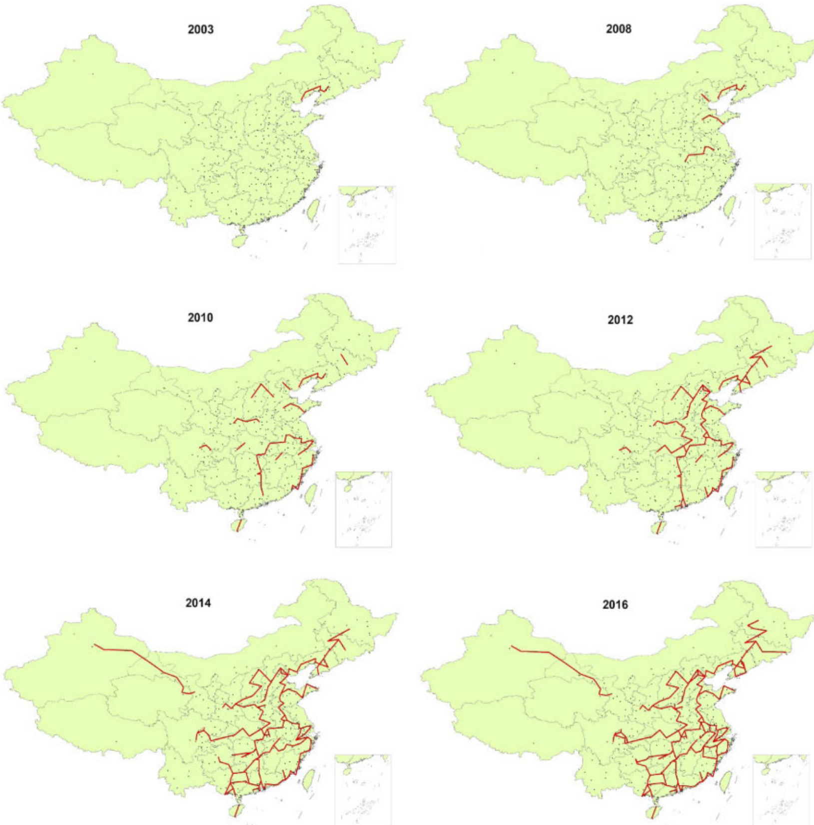

Figure 1. Evolution of HSR network from 2003 to 2016. These graphs display the expansion of HSR network from year 2003 to 2016. The lines in bold are lines in use by the end of that year. Each dot represents a prefecture-level city.

financing of HSR involves very limited private investment. The national government provides about half of the financing through lending by state-owned banks and financial institutions, another 40% by bonds issued by the MOR. Provincial and local governments, mainly through compensation for land use, contribute the remaining.

As the network has grown rapidly over the past decade, so has the ridership of HSR. China's HSR network is the world's most extended and the most extensively used, with 1.713 billion trips taken in 2017, bringing its total cumulative number of trips to 7 billion. Traveling by HSR is particularly attractive for short-to-medium distance business trips owing to its convenience, high frequency, and punctuality, relative to its main alternative for air travel. For example, traveling from Beijing to Shanghai by air typically takes 2.5 h, from taking off to landing; traveling by HSR takes about 4.5–5 h. 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 trip to and from the airports, check-in procedures, and other delays. HSR also costs less than air tickets for most routes.

3. Data

3.1 Firm Investment Dataset

The Firm Registration Database is maintained by the China's State Administration for Market Regulation. The database contains the administrative information of the whole universe of firms in China, covering over 15 million registered firms in 2015. At the date of registration, all firms are required to disclose to the State Administration for Industry and Commerce (SAIC; see footnote 1) 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 that we use in this article 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. Since the database includes investments by non-financial firms in various sectors of the economy, we can examine investment flows that are not captured in prior literature on arms-length investments by financial intermediaries—for example, institutional investors or banks—or within-firm (re)allocation of resources—for example, across existing plant locations in manufacturing firms.

We use the records in the Firm Registration Database to measure the financing activities at the firm-to-firm level, that is, a firm contributing capital to another firm and thereby becoming its shareholder.[6] 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 as well as from natural persons to firms from 2004 to 2015. Our article focuses on the firm-to-firm investment activities, of which there are 1,633,710 observations in the database from 2004 to 2015.

We restrict the sample to investments in which the firm receiving the capital is a new firm, that is, the investment activity occurs within the calendar year in which the receiver firm is firstly registered with the SAIC. With this restriction, our sample excludes the dynamics of ownership structure of incumbent firms.[7] After this trimming, our final sample consists of 1,114,268 firm-to-firm investment observations over the 12-year period, with the number of observations increasing from 45,283 in 2004 to 281,436 in 2015.

3.1.a. City-pair investment flow measure

Our main empirical specification exploits the time-series variation in HSR connectivity between pairs of cities. To match this granularity, we further aggregate the firm-to-firm investment flows into an investment flow measure between city pairs. We perform the aggregation at the monthly frequency. Explicitly, each observation in our empirical analysis

6 If a firm sets up a new plant (that belongs to the same firm) in a different city, the new plant is also considered a separate legal entity, so such investment follows are also captured in our dataset.

7 Such changes are recorded with less precision in the Firm Registration Database. If an existing shareholder contributes additional capital to an incumbent firm, the initial investment record of this shareholder will be replaced by the additional investment record. In other words, we can only observe the second record of investment instead of two separate investment records from the shareholder to the incumbent firm, which create additional noise in determining the precise date and value of capital provided for financing activity. Such circumstances constitute roughly 10% of the total firm-to-firm investment sample. This means that we exclude any observations that are related to acquisitions of incumbent firms to avoid this additional noise. Our main results are virtually identical if investments to incumbent firms are not excluded.

captures the firm-to-firm investments from an origin city i to a destination city j at year-month t, which is the aggregation of all firms located in city i that contribute the capital to newly registered firms located in city j.

We develop two related measures for this aggregation to capture both the frequency as well as the intensity of firm-to-firm investments between the city pair (i, j). The first measure captures the extensive margin of investments, that is, the number of firm-to-firm investments from origin city i to destination city j at year-month t. The second measure reflects the intensive margin, that is, the total value of firm-to-firm investments (in Renminbi, RMB) from city i to city j at year-month t.

3.1.b. Patterns of intercity investment flows

In this section, we document the secular patterns of the investment flows across cities in China. First, we document the time-series trends in the aggregated cross-city and within-city investment flows. In Figure 2, we observe that both cross-city and within-city investment flows have grown dramatically during 2004–15. In Panel A of Figure 2, the aggregated number of cross-city investments increased from 11,138 to 64,245, whereas the amount increased from 143 billion to 1,214 billion *yuan*. In comparison, the aggregated number of within-city investments in Panel B of Figure 2 increased from 34,145 to 217,191, and the amount increased from 249 billion to 1,871 billion *yuan*. The aggregate number of both cross-city and with-city investments has grown around six times in 12 years. Although the aggregate amount of within-city investments is larger than the aggregate amount of cross-city investments, the amount of cross-city investments grows at a slightly faster pace, growing more than eight-fold during our sample period.

Second, we examine the geographical distribution of investment destinations. In Panel A of Figure 3, we show the total number of cross-city investments received by each prefectural city in 2005 and 2015, with a darker color reflecting a larger number of investment inflow. In 2005, twenty-five cities (8.83%) in China receive over 100 inter-city firm investments. In contrast, 131 Chinese cities (46.29%) receive >100 inter-city investments in 2015. In addition, we observe a more dispersed geographical distribution of investments over the years. Panel B of Figure 3 depicts the total amount of investment inflow at the prefectural city level in 2005 and 2015. The pattern we observe for this intensive margin is similar to that of the extensive margin.

3.2 HSR Network

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 sections on Major Events, Finished, and Ongoing Projects from 1999 to 2012. We collect official news published on http://news.gaotie.cn as well as other online news sources for a small proportion of lines that are opened in 2013 and 2014 as well as future HSR lines under planning, as this information is not available from the most recent (2012) yearbook. 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 official news online as well.

In the analysis, we only focus on prefecture-level cities and 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 2004 to 2015, such as GDP, population, average income, average ridership, and so on.

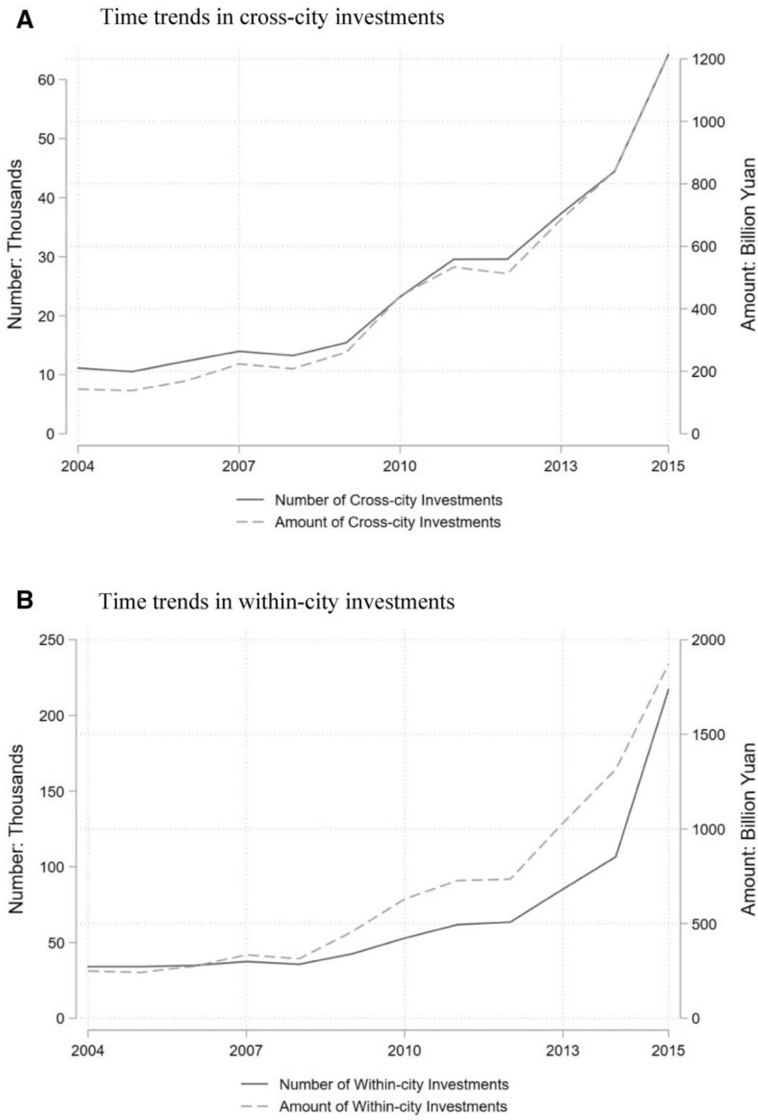

Figure 2. Time trends in cross-city and within-city investments in China. These graphs display the evolution of aggregated cross-city and within-city investments in China from year 2004 to 2015. (A) The aggregated number and amount of cross-city investments and (B) the aggregated number and amount of within-city investments.

3.3 Travel Time on Railway and Competing Modes

We also examine how effective the HSR expansion is in reducing bilateral city-to-city travel time. To do this, we collect comprehensive information on three modes of transportation: air, train (including HSR), and road. For train services, we obtain thirteen snapshots of

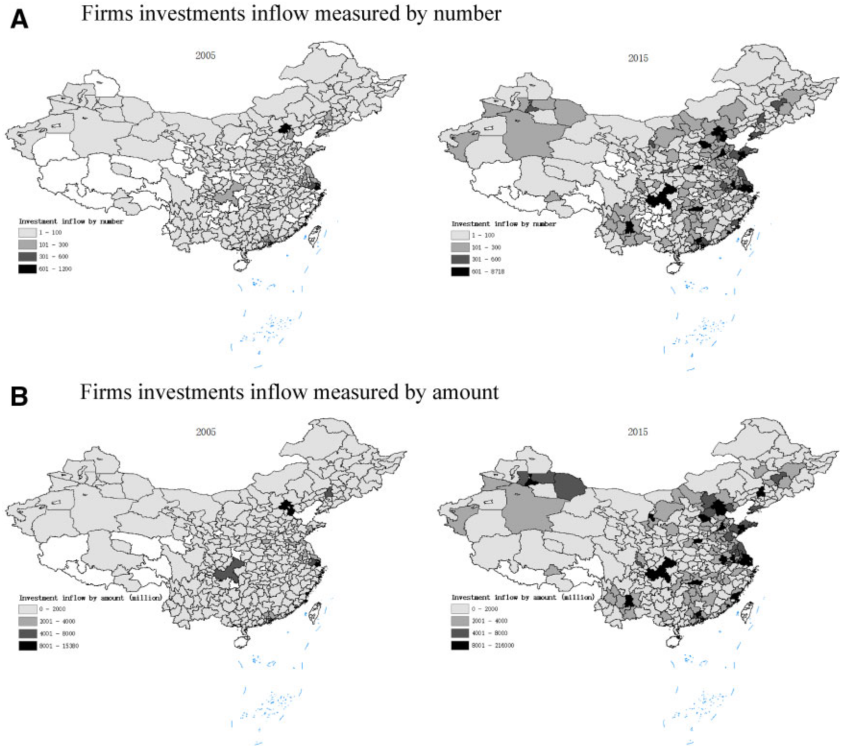

Figure 3. Geographical distribution of firms' investments inflow at city level. These graphs display the geographical distribution of firm investment inflow at the city level, in 2005 and 2015. (A) The aggregated number of investment inflow to cities, ranged from (0, 100), (101, 300), (301, 600), and (601, 8718). (B) The aggregated amount of investment inflow to cities (in million *yuan*), ranged from (0, 2000), (2001, 4000), (4001, 8000), and (8001, 216000).

national railway timetables from 2007 to 2015.[8] These timetables provide information on the duration and fare prices for all the train services, including traditional trains and HSR. For air transport, we obtain monthly data from Official Airline Guide (OAG) that reports the carrier information, schedule, duration, distance, origin and departure airports, and the number of seats on all the flights in 2011. We calculate the distance and duration of travel by car between any two cities' centroids based on OpenStreetMap (OSM) in 2017.[9]

4. Empirical Strategy

4.1 Baseline Specification

We adopt a difference-in-differences specification to examine, for a given city pair, whether HSR connection between them leads to an increase in bilateral firm investments, compared with the not-yet-connected ones. The rich information available in city-to-city investment

8 The months covered are March 2007, January 2008, May 2008, July 2009, May 2010, January 2011, April 2012, August 2013, December 2013, April 2014, October 2014, April 2015, and October 2015.

9 http://wiki.openstreetmap.org/wiki/WikiProject_China

flows allows us to control for a full set of origin/destination city interacting with year-month fixed effects. The specification takes the form of

$$y_{ijt} = \theta \text{connect}_{ijt} + \alpha_{ij} + \beta_{it} + \gamma_{jt} + \delta_t + \epsilon_{ijt}, \tag{1}$$

where subscript i denotes the origin city, j denotes the destination city, and t denotes the time at 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 extensive and intensive margins of investment flows. For the extensive margin, we define the dependent variables y_{ijt} as the logarithm of the number of n firms at city j invested by investors from city i within year-month t. Alternatively, for the intensive margin, we define y_{ijt} as the logarithm of the sum of investment flow in monetary terms from city i to city j within year-month t. The main coefficient of interest is θ, which measures the effect of the introduction of new HSR connections on intercity investments.

In our analyses, each observation is a directed dyad for two different cities. We include only city pairs that have ever been connected by HSR by the end of the sample, thus the identification variation comes from the comparison between the "early treated" pairs and the "later treated" pairs. Altogether, this ever-connected sample consists of 1,188 city pairs at the cross-section over 144 months (12 years), that is, 171,072 monthly city pairs for the 2004–15 period. As our sample includes monthly city-pair observations that experience no cross-city investment records during the month, we transform the dependent y_{ijt} variables using the inverse hyperbolic sine transformation (Burbidge, Magee, and Robb, 1988): $\log(y_{ijt} + \sqrt{y_{ijt}^2 + 1})$, to handle observations with zero y_{ijt} values.

Endogenous placement is always a concern when identifying the causal impact of transportation infrastructure on outcomes of interests. Some city pairs are more likely to be connected by HSR than other pairs, potentially because of historically stronger social and economic ties between the former. Growing cities that are expected to attract more capital and population inflow are also more likely to receive HSR placement. We control for bilateral city-pair fixed effects (α_{ij}) throughout the baseline specification to capture all the time-invariant unobservables at the city-pair level, including pre-existing pairwise economic linkages. Our empirical framework allows us to include the full set of (origin city $\times$ time)- and (destination city $\times$ time)-fixed effects (β_{it}, γ_{jt}). These fixed effects capture dynamic local conditions—even time-varying, unobservable variables at the city level—that affect the time-series variation of both the attractiveness of the destination city and the investment capacity of the investor city. In addition, we apply two-way clustered standard errors at city-pair and monthly levels.

4.2 Anatomy of HSR Connection

The strict baseline regression framework with high-dimensional fixed effects absorbs unconditional economic linkages between each pair of cities, as well as the time-varying unobservables at each city. Nevertheless, it leaves one remaining challenge to our identification strategy: time-varying unobservables at the *city-pair* level that correlate with pairwise HSR connection. In particular, the new HSR connection between two cities could be a consequence of *improving* economic linkages between the two cities. If indeed this channel drives the HSR effect, we should observe the trending up of bilateral investments even before the actual completion of the HSR linkage, or even before the announcement of the construction plan.

To investigate in this issue, we estimate the dynamic effects of HSR announcement and connection on intercity investment flows in an event-study framework. 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 3–4 years from the announcement to the initial operation of a new HSR line. The equation for this estimation is specified as follows:

$$\begin{aligned} y_{ijt} = &\sum_{\tau=-12}^{-1} \theta_\tau \text{ Pre-Announcement}_{ijt} \times 1\{\text{Month}_{ijt} = t+\tau\} \\ &+ \pi \text{ Post-Announcement}_{ijt} + \sum_{\tau=0}^{25+} \vartheta_\tau \text{ Connect}_{ijt} \times 1\{\text{Month}_{ijt} = t+\tau\} \\ &+ \alpha_{ij} + \beta_{it} + \gamma_{jt} + \epsilon_{ijt}, \end{aligned} \tag{2}$$

where τ stands for the event year-month of announcement or the event year-month of connection; Pre-Announcement$_{ijt}$, Post-Announcement$_{ijt}$, and Connect$_{ijt}$ are dummy variables denoting 1 year before announcement, post-announcement (the period between announcement and connection), and connection between city i and city j at time t. $1\{\text{Month}_{ijt} = t+\tau\}$ is a dummy variable that takes the value 1 for month $t+\tau$ and otherwise 0. We pick the 13–24 months before the announcement as the benchmark period. Therefore, the coefficient θ_τ estimates the "impact" of future HSR announcements within a year, relative to the base period of 13–24 months prior to announcement. If the parallel trend hypothesis holds for the difference-in-differences specification, θ_τ should be significantly different from zero. π estimates the announcement effect while ϑ_τ estimates the connection effect. The rest of the specification is the same as Equation (1).

4.3 Measurement of Indirect HSR Connection

In the baseline identification strategy (described in Section 4.1), we take advantage of the abundant information of city dyads setting to control for large sets of fixed effects. In addition, we also test the parallel trend assumption of the DID setting to falsify the possibility of pairwise unobservables in the trend of investment opportunities growth that might correlate with bilateral HSR connection. The empirical evidence supports the common trend assumption. Our estimates are unbiased unless the time-varying unobservables at the city-pair level display a structural break at exactly the same time of HSR connection. Although this scenario is unusual, it is not impossible. As described in Section 4.2, we attack this problem by examining the time-series dynamic of investment flows before, around, and after the announcement and completion of each new HSR linkage between pairs of cities.

We go further to mitigate the endogeneity concern by exploiting the network structure of China's HSR system. The system consists of four main horizontal lines and four vertical lines that are constructed in a staggered fashion (Figure 4). When a horizontal line is attached with a vertical line, non-nodal cities along both lines become indirectly connected. These indirect connections are largely unplanned, particularly after conditioning on the interaction of origin/destination city and time trend, that is, (origin city $\times$ time)- and (destination city $\times$ time)-fixed effects (β_{it}, γ_{jt}) in our baseline analysis.

In order to strip away threats to identification caused by endogenous selection, we focus on city pairs that are "indirectly connected" as a result of the network structure of China's HSR expansion. Restricting our treatment group to city pairs that receive HSR train

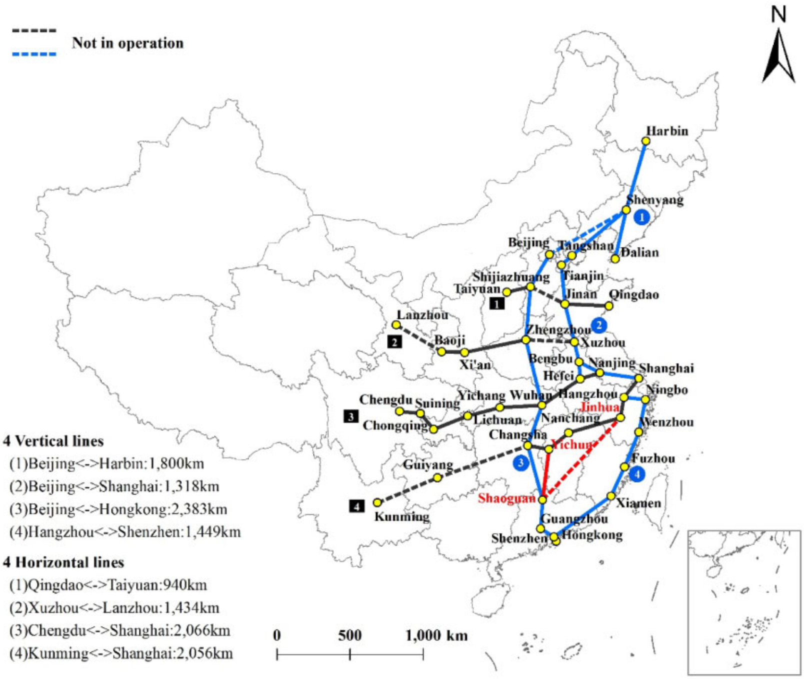

Figure 4. Construction of indirect connect measures of HSR network. This figure illustrates the concept of an indirect connection by HSR. Yichun and Shaoguan are considered to be indirectly connected after both Changsha–Nanchang and Changsha–Guangzhou lines are in operation.

services as a result of connections in other parts of the network helps to rule out endogenous selection in a more credible manner. We construct the measurement of indirect connection of HSR, Indirect_{ijt}, as a dummy variable that equals 1 for a city-pair-month triad *ijt* if (i) the origin city *i* (destination city *j*) is located along a segment A of a horizontal line mentioned above and the destination city *j* (origin city *i*) is located along a segment B on a vertical line; (ii) the line segments A and B are connected at month *t*; and (iii) the pair is not directly connected.[10]

In this analysis, we include all city-pair observations that are never directly connected by HSR, which include 11,320,992 city-pair-month observations. The specification takes the form of

$$y_{ijt} = \theta \text{Indirect}_{ijt} + \alpha_{ij} + \beta_{it} + \gamma_{jt} + \epsilon_{ijt}. \quad (3)$$

A large fraction of this sample is populated by monthly city-pair observations that experience no cross-city investment records (i.e., zero y_{ijt} values) during the month. Similar to the analysis detailed in Sections 4.1 and 4.2, we transform the dependent variables y_{ijt} using the inverse hyperbolic sine transformation (Burbidge, Magee, and Robb, 1988): $\log(y_{ijt} + \sqrt{y_{ijt}^2 + 1})$.

10 A line segment is defined as a part of the whole vertical/horizontal line that started operation at the same date.

Table I. Summary statistics of city pairs

Information on firm investment is collected from Firm Registration Database maintained by China State Administration for Market Regulation. The first row in each panel reports the average monthly number of intercity investments between each city pair, while the second row reports the average monthly intercity total investment amount (in 10,000 *yuan*). Information on opening dates of HSR lines is from the China Railway Yearbooks. The sample period covers 2004–15 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 4.3.

Panel A. All city pairs (283 cities; 11,499,064 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) Mean	Never directly connected pairs (11,320,992 observations) 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) Mean	After direct connection (48,870 observations) Mean
Number of investments	0.302	1.026
Total investment amount	429.428	1797.944

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

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

5. Empirical Findings

5.1 Summary Statistics

Table I reports the summary statistics for all the 11,499,064 city-pair-month observations (=283 cities in our sample $\times$ 282 other cities to choose from $\times$ 144 month) in Panel A. The average incidence of monthly inter-city investment is 0.0253, corresponding to an average of 7.13 inter-city investments per month from each source city (=0.0253 × 282 destination cities), or >85 investments per year. Aggregating the magnitude across all cities in China, we obtain an average number of investments of 0.0253 × 283 cities in our sample × 282 other cities to choose from = 2,019 per month, or 24,229 per year. These contribute to around a quarter of all corporate investments in the Firm Registration Database sample.

In Panel B, observations are categorized by whether the city pair is ever directly connected by HSR at any point during our sample period. There are 171,072 city-pair-month observations that are ever directly connected by HSR (i.e., the estimation sample for Equations (1) and (2)), and 11,320,992 city-pair-month observations that are never directly connected by HSR (the estimation sample for Equation (3)). 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 directly connected to HSR are substantially more likely to have intercity investment flow on the extensive margin (twenty-eight times higher), and the amount of intercity investment is twenty-five times higher than city pairs that are never connected to HSR on the intensive margin. Therefore, as described in Section 4, we segregate city pairs that have ever versus never been connected by HSR as the regression sample to ensure the comparability of the control and treated units for each analysis.

In Panel C, we segregate the ever-connected city pairs further 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 the HSR network, compared with before the connection is established. In Panel D, we focus on the never-connected sample, and segregate these city pairs further into those that are ever indirectly connected (although never directly connected) versus those that are never directly nor indirectly connected. The corresponding gaps of intercity investments between groups are slightly smaller, but still remain of large economic magnitude. It is important to note that these cross-group variations are swept away in our regression analysis by the (origin city × time)- and (destination city × time)-fixed effects (β_{it}, γ_{jt}).

5.2 Main Specification

The parameter estimates from the baseline direct connection specification in Equation (1) are reported in Table II. Columns (1) and (2) report the effect of the direct connection to HSR on cross city investment on the extensive margin, in which the dependent variable y_{ijt} is the number of unique investments from city i to city j within month t.[11] 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, whereas Column (2) adds origin city × year-month and destination city × year-month fixed effects, allowing for a highly flexible functional form of origin and destination city time trend. The coefficient estimate on the connection dummy in Column (2) is 0.080 with high statistical significance, which implies that the introduction of a direct HSR connection between two cities increases the number of intercity investor–receiver pairs by 8%, compared with the control city pairs that are connected to HSR at a later date.

11 As described in Sections 4.1 and 4.2, the dependent y_{ijt} variables are transformed using the inverse hyperbolic sine transformation (Burbidge, Magee, and Robb, 1988): $\log(y_{ijt} + \sqrt{y_{ijt}{}^2 + 1})$.

Table II. The impact of HSR connection on intercity investments

The table reports difference-in-differences estimation results from Equation (1). The main dependent variables are: Number$_{i,j,t}$, which is the number of unique investments from city *i* to city *j* within month *t*; and Investment$_{i,j,t}$, which is the total investment flow from city *i* to city *j* within month *t*. Each of these y_{ijt} dependent variables is transformed using the inverse hyperbolic sine transformation (Burbidge, Magee, and Robb, 1988): $\log(y_{ijt} + \sqrt{{y_{ijt}}^2 + 1})$, to handle observations with zero y_{ijt} values. 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 (Table I, Panel C). As Columns (2) and (4) control for origin city × year-month and destination city × year-month fixed effects, 2,880 singleton observations are dropped from the regressions. Robust standard errors are two-way clustered at city-pair and month levels. The corresponding *t*-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

	(1)	(2)	(3)	(4)
Variables	Number		Investment	
Connect	0.028***	0.080***	0.116***	0.375***
	(2.884)	(3.760)	(2.787)	(3.908)
Observations	171,072	168,192	171,072	168,192
R^2	0.582	0.746	0.478	0.653
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Origin (*i*) × year-month FE		✓		✓
Destination (*j*) × year-month FE		✓		✓

Columns (3) and (4) of Table II present the effect of HSR connection on the intensive margin of intercity investments, in which the dependent variable y_{ijt} is the sum of investment flow from city *i* to city *j* within month *t*. In the more stringent specification that includes origin/destination city-month fixed effects in Column (4), the coefficient on the treatment dummy is 0.375, which implies that the amount of intercity investment increases by 45.5% on average with the introduction of HSR connection.[12] Given that the sample mean of monthly intercity investment flow is 4.29 million *yuan* in the pre-treatment period, the increase associated with being connected via the HSR corresponds to an increase in capital expenditures of 1.95 million *yuan* per city-pair for each month in the post-treatment period.

5.3 Investment Dynamics around HSR Connection

The endogeneity of HSR connection remains a valid concern even after controlling for a full set of fixed effects. We conduct two additional tests to further mitigate such concern. 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 HSRs into the regression equation in case that investors may respond to the news once they learn about the

12 We calculate the implied growth rate using exp(coefficient)−1 for coefficients that have an absolute magnitude >0.1. Such approximation is still valid for the inverse hyperbolic sine transformation if the dependent variable is not extremely small.

announcement of HSR connection between two cities. Second, we explore the indirect HSR connection as described in Section 5.4.

The dynamic specification includes the pre-announcement window, the after-announcement-before-connection window, and the post-connection window, which allows us to examine whether the effect of HSR connection starts immediately after the announcement and before the actual HSR connection is established. More importantly, if cities with stronger economic ties (and thus stronger intercity investment demand) are connected by HSR, we may observe that the positive increase in investment flows between two connected cities surged up before the completion of the HSR connection (or even before its announcement).

The coefficients and the 95% confidence intervals for each coefficient in the event study are plotted in Figure 5.[13] The benchmark period is 13–24 months before announcement of each HSR connection. The first panel presents the estimates for the extensive margin, whereas the second panel presents the estimates for the intensive margin. Compared with the benchmark period, there is no significant increase in city-pair investment in the 1 year before the announcement of HSR connection (i.e., the first three estimates in each panel), both at the extensive and intensive margins. This helps us to validate the parallel trend assumption of the difference-in-differences design and mitigates the concern on reverse causality. 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 investment flow in monetary terms across city pair after the announcement but before the real connection of HSR. However, these two coefficients are only marginally significant at the 10% level.

The effects of HSR connection become significant and increase over time after the connection is initiated. The HSR effects on the extensive margin increase from 6.5% in the first 4 months of the connection to 24% after 2 years of the connection. The corresponding increase in the intensive margin is from 46.2% in the first 4 months to 130.7% after 2 years of the HSR connection.

Table III presents an alternative specification, in which we include a pre-announcement dummy to test for the validity of the parallel trend assumption, and an announcement dummy to control for the announcement effect. The pre-announcement dummy equals to 1 if the city-pair-month observation is during the 1-year (Panel A) or 6-month (Panel B) period before the announcement of the HSR connection. 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.

The announcement dummy is set to 1 if the city-pair-month observation is during the period after the announcement but before the completion and operation of the HSR line. The results consistently suggest little (if any) HSR announcement effect. The announcement dummy variable is weakly significant in models that do not control for the full set of fixed effects, but are generally small in economic magnitude. They also become insignificant once the high-dimensional fixed effects are controlled for. In contrast, the parameter estimates on the HSR connection indicator variable are still significantly positive and large in magnitude, after controlling for the announcement and pre-announcement dummy variables.

13 Refer to Online Appendix Table A1 for the regression coefficients.

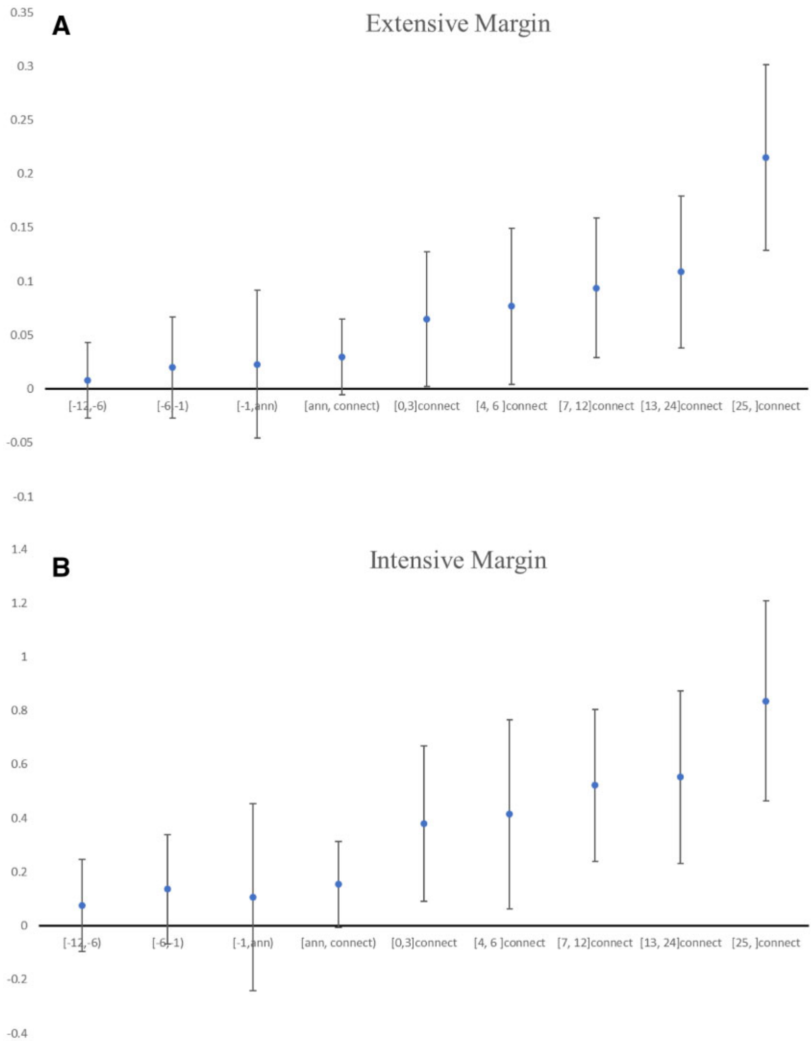

Figure 5. Dynamic effect of HSR announcement and connection. This figure visualizes the coefficients θ in Equation (2). The top panel reports the extensive margin (number of investments) of intercity investment flows around the introduction of direct HSR connection, whereas the bottom panel reports the intensive margin (the total amount of investments). The regression estimates are available in Online Appendix Table A1 (Columns (2) and (4)). The coefficients are presented in dots, with their 95% confidence intervals.

5.4 Indirect HSR Connection

To further mitigate the concern on the endogenous placement of HSR stations, we examine whether the HSR effects also exist for cities that are indirectly connected via the network. In particular, we drop all city-pair observations that are ever directly connected by HSR (i.e., the sample used in Sections 5.2 and 5.3), and instead use only city-pair observations that are never directly connected by HSR. In the latter sample, we identify city pairs that are *indirectly* connected by HSR as the treated sample. By restricting the treatment group

Table III. Parallel trends and announcement effects

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 (1 year in Panel A; 6 months in Panel B). The main dependent variables are: $\text{Number}_{i,j,t}$, which is the number of unique investments from city *i* to city *j* within month *t* and $\text{Investment}_{i,j,t}$, which is the total investment flow from city *i* to city *j* within month *t*. Each of these y_{ijt} dependent variables is transformed using the inverse hyperbolic sine transformation (Burbidge, Magee, and Robb, 1988): $\log(y_{ijt} + \sqrt{y_{ijt}^2 + 1})$, to handle observations with zero y_{ijt} values. $\text{Connect}_{i,j,t}$ 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 (Table I, Panel C). Robust standard errors are clustered in two dimensions: city-pair and monthly levels. The corresponding *t*-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

Panel A. Using 1 year before announcement as benchmark

	(1)	(2)	(3)	(4)
Variables	Number		Investment	
Pre-announcement (1 year)	0.007	0.005	0.034	0.076
	(0.856)	(0.279)	(1.010)	(1.049)
Announcement	0.019*	0.007	0.082*	0.091
	(1.909)	(0.388)	(1.953)	(1.073)
Connect	0.047***	0.089***	0.202***	0.478***
	(3.758)	(2.928)	(3.672)	(3.620)
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		✓		✓

Panel B. Using 6 months before announcement as benchmark

	(1)	(2)	(3)	(4)
Variables	Number		Investment	
Pre-announcement (6 months)	0.005	0.010	0.015	0.091
	(0.546)	(0.466)	(0.424)	(0.953)
Announcement	0.017*	0.008	0.072*	0.083
	(1.949)	(0.429)	(1.915)	(1.050)
Connect	0.045***	0.089***	0.190***	0.470***
	(3.837)	(3.047)	(3.675)	(3.711)
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		✓		✓

Table IV. The impact of indirect HSR connection on intercity investments

The table reports difference-in-differences estimation results from Equation (3). The main dependent variables are: $\text{Number}_{i,j,t}$, which is the number of unique investments from city i to city j within month t and $\text{Investment}_{i,j,t}$, which is the total investment flow from city i to city j within month t. Each of these y_{ijt} dependent variables is transformed using the inverse hyperbolic sine transformation (Burbidge, Magee, and Robb, 1988): $\log(y_{ijt} + \sqrt{y_{ijt}^2 + 1})$, to handle observations with zero y_{ijt} values. Indirect $\text{Connect}_{i,j,t}$ 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 (Table I, Panel D). Robust standard errors are clustered in two dimensions: city-pair and monthly levels. The corresponding t-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

	(1)	(2)	(3)	(4)
Variables	Number		Investment	
Indirect Connect	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		✓		✓

to these city pairs, which are much less likely to be planned in advance, the threat to identification due to potential endogenous selection can be ruled out more explicitly.

Table IV reports the results of estimating Equation (3), using indirect connection as the treatment variable. As shown in Column (2), the indirect connection of HSR between two cities increases the number of investments between the city pair by approximately 3.8%. On the intensive margin, the coefficient estimate on the treatment dummy of indirect connection is 0.185 in Column (4), which implies that the investment flow increases by 20.3% after indirectly connected to HSR network. The magnitude is lower relative to direct connection, but still economically significant, and can be considered as a lower bound of the estimated effect.[14]

14 We conduct two additional analyses to further mitigate potential endogeneity concerns over the "indirect" approach. First, we replicate all regression analyses in Table IV but drop all observations that include Beijing, Shanghai, Guangzhou, or Shenzhen in the city pair, either as the origin or the destination city. The results are reported in Online Appendix Table A2. The coefficient estimates are largely consistent with those reported in Table IV. Second, we conduct a regression analysis that includes an interaction term for city pairs that involve at least one newly constructed HSR station. As presented in Online Appendix Table A3, 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 indirect HSR connection is mostly confined to the additional exposure to the HSR network experienced by *existing* host cities on the network, further mitigating the concern that the results are driven by the endogenous selection of cities hosting new HSR stations.

5.5 Robustness Checks

In this section, we perform various robustness checks to provide further evidence that the observed HSR impacts indeed come from firms' own investment decisions. First, a large flow of investments is made by State-Owned Enterprises (SOEs) in China and this type of investments might be 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. To alleviate such concern, we divide the firms in our sample into SOEs and non-SOEs using the guideline provided by the National Bureau of Statistics.[15]

Table V reports the parameter estimates obtained using the baseline specification used in Table II, but estimated for different types of investment flows. Panel A shows the effect of HSR on investment flows originating from non-SOEs to non-SOEs, aggregated from 937,099 firm-to-firm investment records, making 84.1% of the full firm-level sample. Panel B reports the result from SOEs to SOEs, with 51,256 firm-to-firm investment observations or 4.6% of the full sample. Panel C reports the results from SOEs to non-SOEs, with 75,770 firm-to-firm records or 6.8% of total sample, and Panel D reports results from non-SOEs to SOEs, with 50,143 firm-to-firm records or 4.5% of the full sample.[16]

Examining the four different panels helps us to further pin down the impact of HSR on investment decisions. The estimates obtained in Panel A (i.e., non-SOEs as both investors and receivers) are almost identical in magnitude and statistically significant to those in Table II, whereas the coefficients in Panels B and C (i.e., SOEs as investors) are quite small in magnitude and most of them are not statistically significant. These patterns indicate that our main results are not driven by investments made by SOEs, further alleviating potential concerns regarding the possibly joint decisions of regional investments and HSR connections made by the (central) government. Panel D displays investment flows from non-SOEs to SOEs—the estimates are mostly statistically significant, but with much smaller magnitudes relative to Panel A.

Aside from the concerns over SOEs' investments, some may also have concerns regarding the firm registration data. In particular, 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 investments before the end of year 2013, during which the SAIC imposed more stringent rules for new firm registration.[17] The results are reported in Table A4 in the Online Appendix. The coefficient estimates

15 We use the three-digit code provided by the National Bureau of Statistics to identify the SOEs in our dataset. SOE is one type of firm under the category of domestic firms. The other types of firms under the same category include collective enterprises, joint venture, associated enterprises, limited liability company, joint stock limited liability company, and private enterprise. There are also other categories including Hong Kong, Macau, Taiwan invested firms, and foreign investment firms. http://www.stats.gov.cn/tjsj/tjbz/200610/t20061018_8657.html

16 Similar to the benchmark specifications, we construct a fully balanced panel, which include all observations including those that have no cross-city investment records, and transform the dependent variables using $\log(y_{ijt} + \sqrt{y_{ijt}^2 + 1})$ form to mitigate the zero flow situations.

17 The *Regulation of the People's Republic of China on the Administration of Company Registration* together with several other regulations were approved by State Council in October 2013, and launched in March 2014 to ease the procedures of registering a firm. From March 2014 to August 2017, the SAIC implemented a series of reforms which aim to reduce the institutional costs of entrepreneurial activities. 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

Table V. Heterogeneity on ownership: SOE versus non-SOE firms

Notes: The table reports difference-in-differences estimation results by SOE and non-SOE firms. The main dependent variables are: $\text{Number}_{i,j,t}$, which is the number of unique investments from city *i* to city *j* within month *t* and $\text{Investment}_{i,j,t}$, which is the total investment flow from city *i* to city *j* within month *t*. Each of these y_{ijt} dependent variables is transformed using the inverse hyperbolic sine transformation (Burbidge, Magee, and Robb, 1988): $\log(y_{ijt} + \sqrt{y_{ijt}^2 + 1})$, to handle observations with zero y_{ijt} values. $\text{Connect}_{i,j,t}$ is an indicator variable, taking the value of 1 if a city pair (*i*, *j*) is directly connected by HSR at month *t*. Robust standard errors are clustered in two dimensions: city-pair and monthly levels. The corresponding *t*-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

	(1)	(2)	(3)	(4)
Panel A: Non-SOEs to non-SOEs				
Variables	Number		Investment	
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
Panel B: SOE to SOE				
Variables	Number		Investment	
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 C: SOE to non-SOE				
	Number		Investment	
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 D: Non-SOE to SOE				
Variables	Number		Investment	
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 FE	✓		✓	
City-pair FE	✓	✓	✓	✓
Origin (*i*) × year-month FE		✓		✓
Destination (*j*) × year-month FE		✓		✓

obtained using this subsample are of very similar magnitudes to the estimates using the whole sample reported in Table II. This indicates that the incidence of potentially bogus firm registrations is unlikely to be correlated with the observed effects of HSR connections.

Similarly, readers may further worry about the quality of those newly registered firms after the new HSR connections. Ideally, we would investigate the overall performance of these newly registered firms but data are not available for such analysis. Nonetheless, we note that the SAIC 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 year 2015, or have survived for at least 3, 4, and 5 years for each city-pair-month observation. The results are reported in Online Appendix Table A5 (for the extensive margin) and Table A6 (for the intensive margin). These results are largely consistent with the baseline results in Table II.[18]

5.6 Travel Time Elasticity

The empirical patterns we have documented so far are consistent with robust and likely causal effects of HSR on intercity investments. Next, we explore a specific channel through which transportation infrastructure expands the geographical scope of investment opportunities: reductions in travel time. This analysis would allow us to estimate the elasticity of bilateral investment flows on travel time reductions, particularly those driven by the introduction of HSR connections.

Conceptually, the benefit of HSR on bilateral investment flows should work through reductions in travel cost and more frequent physical, face-to-face contact between agents from different cities. As such, we should expect the HSR connection effect to be an intention-to-treat estimate: the impact of HSR connection on travel time largely hinges on the availability of competing modes of transportation. For instance, the HSR connection between Beijing and Guangzhou (~2,000 km apart, 3 h by air and 8 h by HSR) will play a much less important role in facilitating pairwise business travel than the HSR connection between Beijing and Tianjin (120 km apart, 0.5 h by HSR and 1.5 h by car).

We start by constructing a time-varying matrix of bilateral travel time between each pair of Chinese cities in our sample. We define pairwise travel time as the shortest time across the three modes of transportation, assuming that travelers always choose the least time-consuming way of travel. To construct a time-varying bilateral travel time matrix, we collect comprehensive trip duration and price information from train timetables and OAG air schedules database.[19] To capture the additional time spent on the way to airports and going through security checks, we add 2 h to air travel.

For air and highway travel time, we fill in the full panel using the 2011 and 2017 cross-sectional observations, respectively. By building our time-varying travel time measure using time-invariant air and road travel times, we constrain the time-series variation in our travel

requires the investors to credit the registered capital to the account of the newly established firm before the new firm can be registered. http://www.gov.cn/xinwen/2017-09/12/content_5224235.htm

18 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 A7 in the Online Appendix. With the subsample, the estimated coefficients are also statistically significant with magnitudes very similar to the full sample.

19 Detailed sources of data are described in Section 3.3.

Table VI. The role of travel time on the HSR–intercity investment relationship

The table reports difference-in-differences estimation results from Equation (4). The main dependent variables are: $\text{Number}_{i,j,t}$, which is the number of unique investments from city *i* to city *j* within month *t* and $\text{Investment}_{i,j,t}$, which is the total investment flow from city *i* to city *j* within month *t*. Each of these y_{ijt} dependent variables is transformed using the inverse hyperbolic sine transformation (Burbidge, Magee, and Robb, 1988): $\log(y_{ijt} + \sqrt{y_{ijt}^2 + 1})$, to handle observations with zero y_{ijt} values. L min Travel time is the minimum travel time (in log form) within a city pair (*i*, *j*) at month *t* among air, HSR, and road. The sample includes only city pairs that are ever directly connected (Table II) in Panel A and city pairs that are never directly connected in Panel B. Panel A reports the results from IV regression using the HSR connection indicator variable $\text{Connect}_{i,j,t}$ as the instrument and Panel B reports results using the indirect indicator variable Indirect $\text{Connect}_{i,j,t}$ as the instrument. Robust standard errors are clustered in two dimensions: city-pair and monthly levels. The corresponding *t*-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

Panel A. HSR connection as instrumental variable

Variables	(1) L min Travel time	(2) Number	(3) Investment
Connect	–0.440***		
	(–41.521)		
L min Travel time		–0.098**	–0.582***
		(–2.149)	–2.824)
First-stage *F*-statistics	1723.973		
Observations	126,144	126,144	126,144
R-squared	0.993	0.001	0.000
City-pair FE	✓	✓	✓
Origin (*i*) × year-month FE	✓	✓	✓
Destination (*j*) × year-month FE	✓	✓	✓

Panel B: HSR indirect connection as instrumental variable regression

Variables	(1) L min Travel time	(2) Number	(3) Investment
Indirect connect	–0.106***		
	(–31.554)		
L min Travel time		–0.291***	–1.415***
		(–5.706)	(–6.539)
First-stage *F*–statistics	995.145		
Observations	8,490,744	8,490,744	8,490,744
R-squared	0.995	–0.003	–0.001
City-pair FE	✓	✓	✓
Origin (*i*) × year-month FE	✓	✓	✓
Destination (*j*) × year-month FE	✓	✓	✓

time measure to capture only the expansion of the HSR network, but not the expansions of airline or highway networks during the same period, which are subject to greater endogeneity bias. We instrument travel times using the (staggered) introductions of both direct and indirect HSR connections to estimate the effect of travel time reduction. The sample includes 126,144 city pairs when using direct connection as the instrumental variable while there are 8,490,744 city pairs that are never directly connected to HSR with indirect connection being the instrumental variable, for the 2007–15 period, measured at monthly frequency.

The instrumental variable analyses take the form of the following regression specifications:

$$y_{ijt} = \theta\big(\text{L min Travel time}_{ijt} = \text{Connect}_{ijt}\big) + \alpha_{ij} + \beta_{it} + \gamma_{jt} + \epsilon_{ijt}, \tag{4}$$

$$y_{ijt} = \theta\big(\text{L min Travel time}_{ijt} = \text{Indirect Connect}_{ijt}\big) + \alpha_{ij} + \beta_{it} + \gamma_{jt} + \epsilon_{ijt}, \tag{5}$$

where subscript i denotes the origin city, j denotes the destination city, and t denotes the time at monthly frequency. We define the dependent variables y_{ijt} as the number of unique investment pairs from city i to city j within month t at the extensive margin, and as the sum of investment flow in monetary terms at the intensive margin. Similar to benchmark the analysis, we transform the dependent y_{ijt} variables using the inverse hyperbolic sine transformation: $\log(y_{ijt} + \sqrt{y_{ijt}^2 + 1})$. L min Travel time is constructed as the logarithm of the minimum travel time within a city pair (i, j) at month t among air, HSR, and road. In Equation (4), we employ Connect as the instrumental variable in the first-stage regression. The empirical model to test Equation (4) includes only city pairs that have ever been connected by direct HSR links by the end of the sample. We employ Indirect Connect as the instrumental variable in Equation (5); the corresponding empirical model includes all city pairs that are never directly connected by HSR. Consistent with our baseline specification in Equation (1), we control for bilateral city-pair fixed effects (α_{ij}) to capture all the time-invariant unobservables at the city-pair level. We further include the full set of (origin city × month)- and (destination city × month)-fixed effects (β_{it}, γ_{jt}) to control for the time-varying local heterogeneity and dynamics at both the origin and destination city levels.

Table VI reports the regression results showing responses of investment flows to changes in travel time with direct connection being the instrumental variable in Panel A while indirect connection being the instrumental variable in Panel B. Column (1) in both panels shows the first-stage regression result. Columns (2) and (3) report the IV regression results for the number and magnitude of investments, respectively. According to Panel A of the table, a direct connection can induce 35.6% decrease in travel time. With this first-stage estimate, we can further estimate that a 1% decrease in pairwise travel time leads to 0.098% and 0.582% growth in intercity investment at the extensive and intensive margins, respectively.

When we use indirect connection as the instrumental variable (Panel B of Table VI), first-stage result (Column (1)) indicates that city pairs indirectly connected by HSR experience a 10.1% reduction in travel time, on average. When travel time is instrumented in Columns (2) and (3) of Panel B, the elasticity of bilateral investment on travel time becomes much larger at –0.291 and –1.415 (relative to the elasticity of –0.098 and –0.582 in Columns (2) and (3) in Panel A) for the number of unique investor–investee pair and total investment value, respectively.

6. Discussions

6.1 Discussion on Potential Mechanisms

The results presented in the previous section have yet to shed light on the potential mechanisms that drive the sensitivity of intercity 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 and lead to the discovery of new investment opportunities. However, there is limited empirical evidence to identify and disentangle these different underlying mechanisms. In this section, we explore several dimensions of firm heterogeneity to evaluate the monitoring and information channels.

First, we provide evidence on sectoral heterogeneity by dividing receiver firms into twenty different groups based on their specific industries. Figure 6 shows the estimated coefficients for each group, as well as their 95% confidence intervals using our main specification.[20] 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 with other industries, the industries for which the HSR connection has the strongest effects tend to require more face-to-face communications and more on-site, physical monitoring. In other words, the ease of people movement owing to more efficient transportation infrastructure is more important for the operation of businesses in these industries. As such, the observed patterns in Figure 6 are consistent with our conjecture that HSR reduces travel cost, and facilitates the movement of people and the flow of information.

Neither extensive nor intensive margin is significant in our estimation using the manufacturing industry. This (non-)result seems to be in contrast to the results documented in Giroud (2013) for US manufacturing companies and plants. There are two potential explanations for this difference. First, our research design allows us to include origin city × year-month FE and destination city × year-month FE in the regression specifications to control for the time-series variation in each city's unobservable characteristics. Second, we include both controlling and non-controlling stakes in our analysis. If we focus on the cases of controlling investments (i.e., stake size of at least 50%) and use a similar specification to best emulate the analysis in Giroud (2013), we observe a significant HSR effect in the manufacturing industry. The results of this manufacturing-specific analysis are reported in Online Appendix Table A8.

20 We included all the twenty industries when estimating the coefficients and corresponding confidence intervals but only show those industries with received investment of >1% of the total investment in the graph.

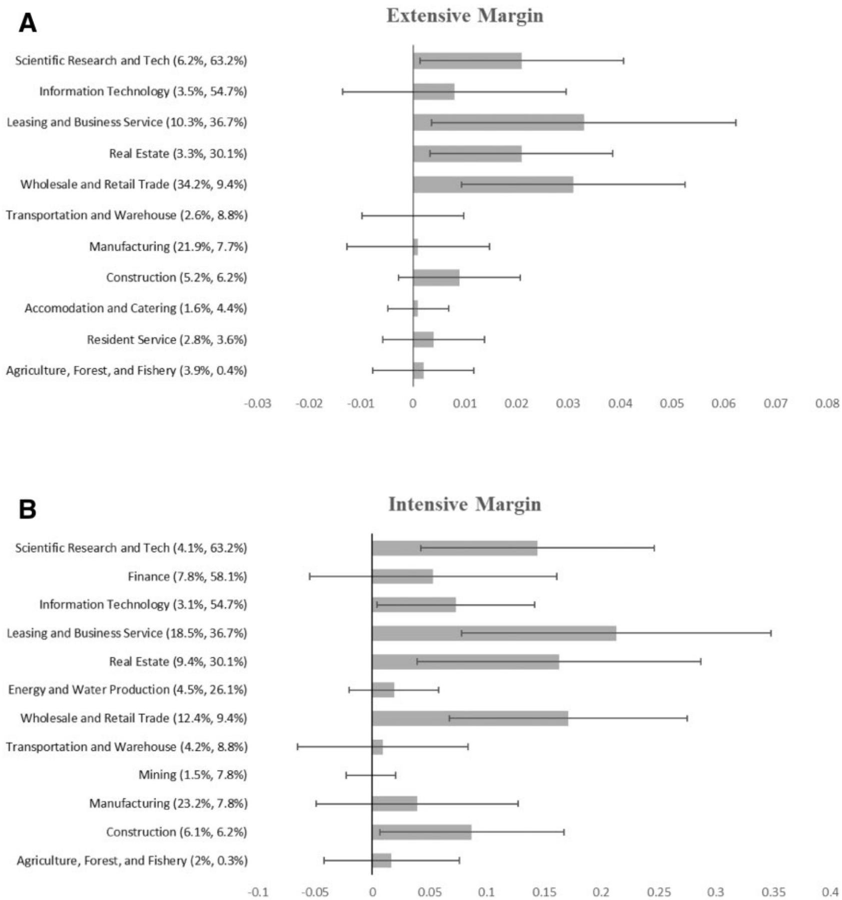

Figure 6. HSR effects, by industries. The sample is split into twenty categories based on the receiver firm's industry. Panel A reports the extensive margin (number of investments) of intercity investment flows, and Panel B reports the intensive margin (total amount of investments). The first number 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 their college shares. We included all twenty industries when estimating the coefficients and confidence intervals but only display those industries with proportion of >1% in this figure.

Figure 6 also provides information on each industry's share of skilled workers, defined as the share of employees with college education in the industry. This skilled worker measure is reported as the second number in the parentheses next to the industry name. Industries in both panels are sorted on 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

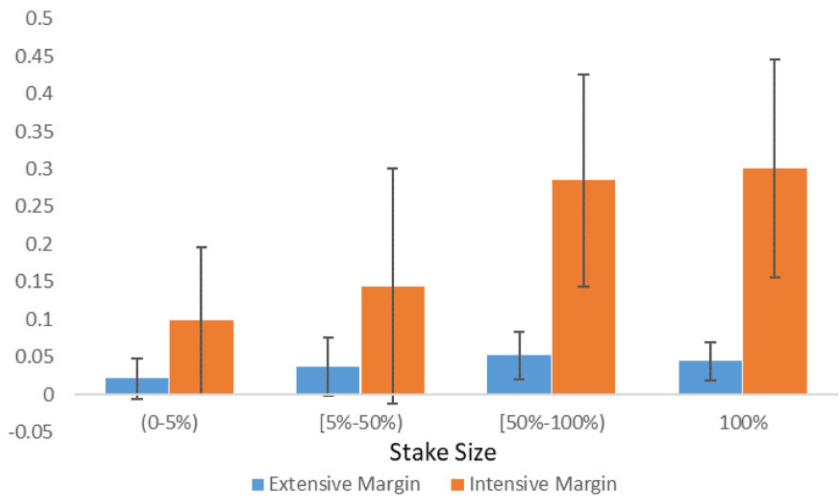

Figure 7. Coefficients by different ownership stakes. The sample is split into four categories based on the investor's stake size. The four categories are: (1) the investor holding 0–5% (not inclusive) of the invested firm's share; (2) the investor holding 5% (inclusive) to 50% (not inclusive) of the 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. The regression estimates are available in Online Appendix Table A9 (Columns (2) and (4)).

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 are more crucial due to their high compensation and access to firm'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 6 is consistent with this hypothesis.

Second, we categorize all firm-level investments into four groups according to the shares held by the investors, namely investors who hold (0, 5%), [5%, 50%), [50%, 100%), and 100%, of registered capital of the new firm, separately. The latter two groups correspond to controlling shareholders, for whom HSR connection can benefit in terms of both monitoring capability and information accessibility. The first two groups correspond to non-controlling shareholders, who are more likely to benefit only through the information advantage channel. As such, we hypothesize that if the access to information channel is important, we should observe positive impacts of HSR on investments even for the non-controlling investors who do not have real control over the firms, for example, because they only invest in 5% of the firm. For these small shareholders, frequent travels on HSR could facilitate connections with local government or businesses to gain private information, helping to identify new investment opportunities.

Figure 7 displays the estimated coefficients for each group with different cutoffs of shareholders, along with the 95% confidence intervals. We observe that the magnitude of the coefficient estimates of the HSR connection variables obtained using the two subsamples of controlling investors with shareholdings of 50% or more (i.e., the two groups on the right-hand side of Figure 7) is about twice as those obtained using the two subsamples of non-controlling investors. However, the estimations using the non-controlling subsamples also yield statistically significant (10% level) coefficients in some specifications, indicating that better access to information after the HSR introduction constitutes a relevant channel driving intercity investment flows.[21]

We conduct three additional heterogeneity analyses. First, we divide the sample period into an early subperiod (2004–10) when the HSR network is relatively scarce and a later subperiod (2011–15) when the HSR network has a wide coverage across the country. We interact the dummies indicating the two subperiods with the DID term (Connect) to examine the differential effects of HSR connection in these two subperiods. As reported in Panel A of Table VII, the impact of HSR connection on cross-city investments is mainly observed in the later subperiod, during which the HSR network is more comprehensive.

Second, the development of HSR network has been cited as instrumental in China's grand plan of developing nineteen urban clusters.[22] The plan is to foster the integration of large urban clusters, anchored around central hubs surrounded by smaller cities. The plan calls for nineteen clusters, which in combination would account for nine-tenths of the economic activity in China. By reducing travel impediments, HSR could improve the integration of cities within each urban cluster by fostering the outsourcing of central hub functions to nearby smaller satellite cities.

In Panel B of Table VII, we divide all cities in our sample into city clusters based on the official definition.[23] To answer the question of whether the impact of HSR connections on investment transpires within or across different clusters, we add an interaction term between the HSR connection status and an indicator of whether both cities belong to the same city cluster. As shown in Columns (1) and (3) of the table, the impact of HSR connection is observed only for within-cluster investments, in terms of both the extensive and intensive margins. These effects are larger when compared with the estimated main effects in Table III.

We further explore the direction of investment flows within each city cluster by focusing on origin–destination pairs within each cluster and divide them into four subgroups: core-to-core, core-to-peripheral, peripheral-to-core, and peripheral-to-peripheral. The results reported in Columns (2) and (4) of Panel B of Table VII indicate that the HSR's within-cluster effects are concentrated in core-to-peripheral and peripheral-to-peripheral flows, which indicates that HSR connection facilitates the decentralization of some industries within each urban cluster from the core city to smaller satellite cities.

Third, we perform an additional analysis on tourist cities. HSR could promote cross-city real-estate investments and tourism, leading to a higher familiarity of other areas, particularly those with popular tourist attractions. To explore the tourism channel, we start by identifying destination cities that are associated with multiple popular tourist attractions.[24] We interact this tourism city indicator (for the destination city) with our measures of direct and indirect HSR connection in regression models that also control for the secular growth

21 We report the full estimation results in Online Appendix Table A9.

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

23 Online Appendix Table A13 lists the official city cluster definitions in details.

24 We define tourist cities as those with at least two 5A tourist attraction sites; for example, Beijing has eight such sites (http://www.xinhuanet.com/english/2021-01/03/c_139638199.htm).

Table VII. The impact of HSR connection by sub-period and different types of cities

The table reports difference-in-differences estimation results from Equation (1). The main dependent variables are: $\text{Number}_{i,j,t}$, which is the logarithm of the number of unique investments from city *i* to city *j* within month *t* and $\text{Investment}_{i,j,t}$, which is the logarithm of the total investment flow from city *i* to city *j* within month *t*. Each of these y_{ijt} dependent variables is transformed using the inverse hyperbolic sine transformation (Burbidge, Magee, and Robb, 1988): $\log(y_{ijt} + \sqrt{y_{ijt}^2 + 1})$, to handle observations with zero y_{ijt} values. $\text{Connect}_{i,j,t}$ is an indicator variable, taking the value of 1 if a city pair (*i*, *j*) is directly connected by HSR at month *t*. In Panel B, the sample includes only city pairs that are ever connected (Table I, 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 Columns (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.

Panel A: Analysis by sub-period

	(1)	(2)	(3)	(4)
Variables	Number		Investment	
Connect × Year (2004, 2010)	−0.007	0.011	−0.018	−0.017
	(−0.695)	(0.442)	(−0.334)	(−0.139)
Connect × Year (2011, 2015)	0.116***	0.165***	0.480***	0.807***
	(4.689)	(2.957)	(4.603)	(3.138)
Observations	170,496	167,616	170,496	167,616
R-squared	0.583	0.746	0.479	0.653
Year-month dummy	✓	✓	✓	✓
City pair FE	✓	✓	✓	✓
Origin city × year-month FE		✓		✓
Destination city × year-month FE		✓		✓

Panel B: Analysis by city type

Variables	Number		Investment	
Connect	−0.019		−0.035	
	(−0.861)		(0.354)	
To Same Cluster × 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
Year-month dummy	✓	✓	✓	✓
City pair FE	✓	✓	✓	✓
Origin city×year-month FE	✓	✓	✓	✓
Destination city×year-month FE	✓	✓	✓	✓

of city-level tourism (using city-year FE). The results are reported in Online Appendix Table A10. The estimates for the interaction variable (capturing the marginal HSR effect for tourist cities) indicate that the HSR effect is stronger for destination cities with multiple 5A tourism attractions, with the incremental effect particularly salient (and statistically significant) for indirect HSR connections. Nevertheless, we continue to observe positive HSR connection effects for the remaining cities with only one or no 5A tourism attraction site. As some of the effects remain for non-tourism destination cities, this pattern indicates that the HSR effects we documented can only be partially attributed to the growth in domestic tourism.

6.2 Discussion on Capital Distribution and Welfare

Our results so far are consistent with the hypothesis that HSR 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 and may even reduce investments to less developed cities. The answer to such questions depends on the source and direction of capital flows.

In theory, capital should flow to cities with higher rates of return, but this theoretical prediction is not always consistent with capital flow patterns observed in reality. In his seminal work, Lucas (1990) points out the surprising paradox that capital does not always flow from developed to developing areas despite the presumably higher ROC 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. These frictions could lead to agglomerations of investment in areas that are already heavily invested (Alcácer and Delgado, 2016). An advantage of our setting is that we are able to directly measure rates of returns of capital across regions and industries, which allows us to test whether capital flows to cities with higher rates of return.

In this section, we explore the direction of HSR-induced capital flow and its welfare implications with two empirical exercises. First, to evaluate the argument of Lucas Paradox, we divide cities into P(oor), M(iddle), and R(ich) categories. We examine the intercity flows within and across categories in Online Appendix Table A11. In general, capital is more likely to flow from richer cities to poorer cities. In particular, the effect on the extensive margin is largely driven by capital flow among the cities within the same group (i.e., R to R, M to M, and P to P) or capital flow to poorer groups as shown in Column (1). Moreover, the estimated coefficient for within-group flows has larger magnitude and higher statistical significance level than for the cross-group estimates. In terms of the intensive margin, Column (2) shows that capital flows to the similar or poorer cities. Results from Columns (3) and (4) further confirm such observations.

Second, we directly test whether capital flows to cities with higher potential rates of return with the help of HSR connections. Using the sample of firms that are included in SAIC inspection database,[25] we calculate each inspected firm's ROC as the ratio of its profit to the total registered capital. We aggregate this measure for all firms in the same industry in each city. For

25 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.

each city pair, we then calculate the return gap for each industry as the difference in the industry ROC measure between the two cities. The return gap variable is positive for a particular industry if the industry firms in the destination city have a higher average ROC than the corresponding industry firms in the source city, and negative otherwise. A positive return gap variable indicates investment in destination city with a higher ROC.

We interact this gap variable with the HSR connection dummy, and report the parameter estimates in Table VIII.[26] The results indicate that HSR's effect on investment is larger if the return gap for a city pair-industry is more positive. In other words, HSR connections facilitate capital flow to destination cities with higher industry return-to-capital rates. As shown in Panel A, for a 1% increase in the return gap, the direct HSR connections induce an additional 0.65% newly established firms (extensive margin) and a corresponding increase of 2.84% in investment amounts (intensive margin). We observe similar patterns using the indirect connection measure in Panel B. This last set of results illustrates the potential impact of passenger transportation infrastructure on reducing frictions in cross-regional capital flows and improving the efficiency of capital allocation across regions.

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 implications of physical distances and travel costs for economic integration and development. The current study focuses on how the reduction in frictions associated with passenger travels helps firms to expand their geographical footprints in investment. Conceptually, the reductions in travel cost should help controlling shareholders gain better control of subsidiaries and non-controlling shareholders overcome information obstacles.

We examine these potential implications by examining shocks to passenger travel costs due to the staggered expansions of the HSR system in China. With the focus on the passenger rail network, our article is notably distinct from prior studies on the effects of goods-shipping transportation infrastructure such as highways or traditional railroads. The HSR system is a convenient mode of passenger transportation between cities, facilitating easier physical, 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 a crucial step in evaluating the general economic impacts of HSR and other large travel infrastructure projects.

26 When calculating the difference in ROC for each city pair, we focus on each city's firms in the same industry. This means that we compare a particular industry in the destination city with the same industry in the source city. For cases in which the firm inspection data for an industry does not exist in either the destination or the source city, this industry is dropped for the city pair. The specifications of the regression models in Table VIII are the same as Equation (1) and (3) except that (i) the dependent variable is investment (number or amount) from source city i to industry k in destination city j within year t—that is, we use (city-pair $\times$ industry $\times$ year) observations, instead of (city-pair $\times$ month) observations as in all other tables and (ii) industry-fixed effects are controlled for in each regression. Only ever-connected city pairs are included in the analyses. We aggregate the observations at annual frequency because the firm inspection information is reported annually; and aggregating at yearly frequency facilitates computational convergence at the remote server where the SAIC database is maintained.

Table VIII. The impact of HSR connections, conditional on ROC

The table reports difference-in-differences estimation results from Equation (1) (Panel A) and Equation (3) (Panel B). The main dependent variables are: $\text{Number}_{i,j,t}$, which is the number of unique investments from city i to city j within month t and $\text{Investment}_{i,j,t}$, which is the total investment flow from city i to city j within month t. Each of these y_{ijt} dependent variables is transformed using the inverse hyperbolic sine transformation (Burbidge, Magee, and Robb, 1988): $\log(y_{ijt} + \sqrt{y_{ijt}^2 + 1})$, to handle observations with zero y_{ijt} values. $\text{Connect}_{i,j,t}$ is an indicator variable, taking the value of 1 if a city pair (i, j) is directly connected by HSR at month t. ROCgap is the difference in the industry ROC measure between the city pair (i, j). The industry ROC measure is the average of ROC (=firm profit divided by total registered capital) for all firms operating in that industry in the city. ROCgap is positive for a particular industry if the industry firms in the destination city (j) have a higher average ROC than the corresponding industry firms in the source city (i), and negative otherwise. Panel A includes 410 city pairs that are ever connected by HSR from 2005 to 2012 at yearly frequency, with nineteen industries defined at the CIC industry code one-digit level. Panel B includes 42,329 city pairs that are never directly connected by HSR from 2005 to 2012 at yearly frequency, with nineteen industries. Robust standard errors are clustered in two dimensions: city-pair and monthly levels. The corresponding t-statistics are reported in parentheses. The 1%, 5%, and 10% statistical significance are denoted by asterisks ***, **, and *, respectively.

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

	(1)	(2)	(3)	(4)
Variables	Number		Investment	
Connect	−0.023	0.094***	−0.058	0.517**
	(−1.055)	(3.964)	(−0.477)	(2.107)
Connect × ROCgap	0.363*	0.646***	1.706*	2.837***
	(1.751)	(2.598)	(1.858)	(2.649)
Observations	18,336	18,336	18,336	18,336
R^2	0.48	0.50	0.43	0.45
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Industry FE	✓	✓	✓	✓
Origin (i) × year-month FE		✓		✓
Destination (j) × year-month FE		✓		✓

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

	(1)	(2)	(3)	(4)
Variables	Number		Investment	
Indirect Connect	0.091***	0.057***	0.398***	0.225***
	(4.634)	(3.546)	(4.401)	(2.970)
Indirect Connect × ROCgap	1.021***	1.104***	4.080**	4.505**
	(2.650)	(2.632)	(0.249)	(2.301)
Observations	1,187,317	1,187,317	1,187,317	1,187,317
R^2	0.33	0.34	0.27	0.28
City-pair FE	✓	✓	✓	✓
Year-month FE	✓		✓	
Industry FE	✓	✓	✓	✓
Origin (i) × year-month FE		✓		✓
Destination (j) × year-month FE		✓		✓

We find that direct HSR connection between a pair of cities increases the number of investments between the city pair by 8%, and the 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. To address the potentially endogenous formation of HSR network, we exploit the indirect HSR connections of non-nodal cities on vertical and horizontal railway lines. Our results are robust when we consider only city pairs that are indirectly connected by the HSR system, which are unlikely to be connected on purpose.

The introduction of HSR connection increases intercity investment flow 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. Although quantifying the welfare effect of the HSR system is not the main focus of the article, we also document that the incremental intercity 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 perspective of capital allocation efficiency.

Data Availability

The data underlying this article were provided by the State Administration for Market Regulation, China, under license. Data will be shared on request to the corresponding author with the permission of the State Administration for Market Regulation, China.

Supplementary Material

Supplementary data are available at *Review of Finance* online.

Funding

The Funding was provided by the Sustainable and Green Finance Institute (WBS A-0006413-00-00), and the Ministry of Education of Singapore (Grant RG84/17).

Conflict of interest: The authors declare no conflict of interests.

References

Ahlfeldt, G. M. and Feddersen, A. (2018): From periphery to core: measuring agglomeration effects using high-speed rail, *Journal of Economic Geography* 18, 355–390.

Alcácer, J. and Delgado, M. (2016): Spatial organization of firms and location choices through the value chain, *Management Science* 62, 3213–3234.

Behrens, C. and Pels, E. (2012): Intermodal competition in the London–Paris passenger market: high-speed rail and air transport, *Journal of Urban Economics* 71, 278–288.

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

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., and Zhang, Q. (2017): Roads, Railroads, and Decentralization of Chinese Cities, *The Review of Economics and Statistics* 99, 435–448.

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

Bernard, A. B., Moxnes, A., and Saito, Y. U. (2019): Production networks, geography and firm performance, *Journal of Political Economy* 127, 639–688.

Bernile, G., Kumar, A., and Sulaeman, J. (2015): Home away from Home: Geography of Information and Local Investors, *The Review of Financial Studies* 28, 2009–2049.

Bernstein, S., Giroud, X., and Townsend, R. (2016): The impact of venture capital monitoring, *Journal of Finance* 71, 1591–1622.

Bullock, R., Salzberg, A., and Jin, Y. (2012): High-speed rail—the first three years: taking the pulse of China's emerging program. The World Bank Report: China Transport Topics No. 04.

Burbidge, J., Magee, L., and Robb, A. L. (1988): Alternative transformations to handle extreme values of the dependent variable, *Journal of the American Statistical Association* 83, 123–127.

Carvalho, V., and Gabaix, X. (2013): The Great Diversification and its Undoing, *American Economic Review* 103, 1697–1727.

Coval, J. D. and Moskowitz, T. J. (1999): Home bias at home: local equity preference in domestic portfolios, *The Journal of Finance* 54, 2045–2072.

Donaldson, D. (2018): Railroads of the Raj: estimating the impact of transportation infrastructure, *American Economic Review* 108, 899–934.

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

Duranton, G., and Turner, M. A. (2012): Urban Growth and Transportation, *The Review of Economic Studies* 79, 1407–1440.

Faber, B. (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 Poterba, J. M. (1991): Investor diversification and international equity markets, *American Economic Review* 81, 222–226.

Giroud, X., and Mueller, H. M. (2019): Firms' Internal Networks and Local Economic Shocks, *American Economic Review* 109, 3617–3649.

Giroud, X. (2013): Proximity and investment: evidence from plant-level data, *Quarterly Journal of Economics* 128, 861–915.

Glaeser, E. L. and Resseger, M. G. (2010): The complementarity between cities and skills, *Journal of Regional Science* 50, 221–244.

Ivković, Z. and Weisbenner, S. (2005): Local does as local is: information content of the geography of individual investors' common stock investments, *Journal of Finance* 60, 267–306.

Lin, Y. (2017): Travel costs and urban specialization patterns: evidence from China high speed railway system, *Journal of Urban Economics* 98, 98–123.

Lucas, R. (1990): Why doesn't capital flow from rich to poor countries?, *American Economic Review* 80, 92–96.

Michaels, G. (2008): The effect of trade on the demand for skill—evidence from the interstate highway system, *Review of Economics and Statistics* 90, 683–701.

MRC. (2008): Mid-to-long term railway network plan. Report of the Ministry of Railways of China.

Qin, Y. (2017): No county left behind? The distributional impact of high-speed rail upgrades in China, *Journal of Economic Geography* 17, lbw013.

Stein, J. C. (2002): Information production and capital allocation: decentralized versus hierarchical firms, *Journal of Finance* 57, 1891–1921.

Storper, M. and Venables, A. J. (2004): Buzz: face-to-face contact and the urban economy, *Journal of Economic Geography* 4, 351–370.

Van Nieuwerburgh, S., and Veldkamp, L. (2010): Information Acquisition and Under-Diversification, *Review of Economic Studies* 77, 779–805.

Zheng, S. and Kahn, M. E. (2013): China's bullet trains facilitate market integration and mitigate the cost of megacity growth, *PNAS* 110, 1248–1253.