Economic Modelling 126 (2023) 106402

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.journals.elsevier.com/economic-modelling

Does geographic proximity affect firms' cross-regional development? Evidence from high-speed rail construction in China☆

Hongwen Chen [a], Ken Cheng [b], Meiyang Zhang [b,1,*]

[a] School of Tourism, Nanchang University, No. 999 Xuefu Avenue, Nanchang, Jiangxi, 330031, China
[b] School of Economics, Xiamen University, No. 422 Siming South Road, Xiamen, 361005, China

ARTICLE INFO

Handling Editor: Sushanta Mallick

JEL classification:
D21
L11

Keywords:
Geographic proximity
Information friction
Cross-regional development
High-speed rail
Difference-in-Differences

ABSTRACT

Firms' cross-regional development is essential for their growth, with communication costs arising from geographic distance becoming a widely discussed factor. However, the causal relationship between geographic proximity and firms' cross-regional development remains inadequately underexplored. By leveraging firm registration data from 2003 to 2014, and exploiting the staggered construction of China's high-speed rail (HSR) as a quasi-natural experiment, this paper studies the effects of falling communication costs on firms' cross-regional development. HSR construction leads to an 8.6% increase in the number of firms' affiliates and a 68.6% rise in the weighted average distance between new affiliates and their headquarters. The effects are more pronounced for firms with higher performance heterogeneity and in cities with limited Internet infrastructure, suggesting that HSR construction can mitigate information asymmetry and improve monitoring. This study provides new micro-evidence for how geographic proximity promotes firms' cross-regional expansion through communication cost channels, offering implications for multiregional firms and policymakers.

1. Introduction

Firms' cross-regional development (CRD) is essential for their growth and economic performance. There are two potential benefits for firms' CRD. First, firms can take advantage of different areas and achieve economies of scale (Krugman, 1980). Second, firms' geographic expansion could help firms become better in response to local shocks (Gervais, 2018).[2] Despite these potential benefits, firms' CRD also faces substantial costs. Among those costs, the costs of communication are important (Gokan et al., 2019; Gumpert et al., 2022).

To operate distant affiliates well, firms have to monitor and communicate with their workers of distant affiliates. This often necessitates a greater reliance on frequent physical and face-to-face interactions (Lin et al., 2022).[3] Consequently, if the geographic distance between the firms' headquarters and affiliates is large, the communication costs may be high and firms might not choose to operate affiliates in distant areas. Notably, the improvement in passenger transportation could help reduce communication costs between different areas. By reducing personal travel time, improved passenger transportation facilitates more frequent physical and face-to-face interactions (Dong et al., 2020; Gao and Zheng, 2020; Shang et al., 2023). However, limited research investigates how the improvement of passenger transportation affects firms' CRD. Therefore, we attempt to fill this research gap by exploiting an exogenous shock to the communication costs of firms' geographic expansion.

One difficulty in examining this relationship is the need for data on the locations of affiliates and headquarters. To tackle this challenge, we use the data of firm registration from the State Administration for Industry and Commerce (SAIC) from 2003 to 2014. For any firm registered during this period, we can identify its affiliates with affiliates' locations

☆ This paper has benefited from the comments and suggestions of the editor, Prof. Sushanta Mallick, and three anonymous referees. We are grateful to the reviewers and the editor for their valuable input.

* Corresponding author.
E-mail addresses: hongwen_chen@ncu.edu.cn (H. Chen), kencheng0717@foxmail.com (K. Cheng), daisy_zhang0623@foxmail.com (M. Zhang).

[1] The authors are listed in alphabetical order and contributed equally to this manuscript.

[2] For example, firms' CRD can reduce their exposure to any one specific local economic shock by bringing in a new set of potential customers who may not be affected by these localized shocks.

[3] Although the information technology (e.g., Internet and video conferencing) advances, there may still need personal travel to get information, such as soft information, which cannot be credibly transmitted.

https://doi.org/10.1016/j.econmod.2023.106402
Received 16 October 2022; Received in revised form 5 June 2023; Accepted 6 June 2023
Available online 8 June 2023
0264-9993/

and registration dates. This set of firm registrations encompasses all firms' affiliates in China, which allows us to identify how firms' geographic distribution changes.

Another challenge is the endogeneity issue. The locations of affiliates and headquarters are choice variables. Accordingly, commonly used proxies for proximity—such as the physical distance between affiliates and headquarters—are likely to be endogenous, making it difficult to establish causality. To alleviate this issue, we exploit China's high-speed rail (HSR) construction as an exogenous shock to identify the causal effect of geographical proximity on firms' CRD. As passenger transportation, China's high-speed rail (HSR) could reduce travel times between firms' headquarters and their distant affiliates. HSR construction can result in a time–space compression effect, which does not alter "spatial distance" but reduces "time distance". It enhances the flow of information and increases the information transparency (Cheng et al., 2016). Most notably, it reduces the cost of obtaining soft information[4] while lowering the monitoring and information costs for firms' headquarters (Charnoz et al., 2018). Thus, HSR construction enhances the information flow within firms and facilitates firms' management and monitoring of their cross-regional affiliates. As a result, we can speculate that geographic proximity promotes firms' cross-regional development.

Using the difference-in-differences (DID) approach, we find that HSR construction promotes firms to operate a greater number of cross-regional affiliates. Furthermore, HSR construction also leads to an increase in the weighted average distance between these newly opened affiliates and the headquarters, consistent with the findings of Gumpert et al. (2022). The results are robust after conducting a battery of tests, including dynamic DID test, placebo test, and other robustness checks. Regarding the heterogeneous effects, the HSR effect is larger for smaller firms and firms with less cross-regional development experience. In addition, we find that firms in the service industry are more likely to open new cross-regional affiliates after HSR construction, which can be attributed to the service industry requiring more face-to-face communications and physical interactions.

To further alleviate endogeneity concerns, we first employ an instrumental variables (IV) approach. We select IV according to the rule that the IV is correlated with the HSR construction variable and is uncorrelated with the outcome variables. We use the historical approach to build our IV, that is, we use the 1961 train station as the IV of the HSR construction dummy variable (Dong et al., 2020; Turner, 2014). The IV estimation results are also robust. Second, we also extend our analysis to the city-pairs level, enabling us to incorporate high-dimensional fixed effects. We add bilateral city-pair fixed effects to control for the time-invariant omitted variables, such as geographical and cultural distance within city pairs. We also add origin city-year and destination city-year fixed effects to control for time-varying omitted factors at the city level that affect both the HSR construction and firms' CRD.[5] The city-pairs level estimation results remain positive and statistically significant.

To shed light on the underlying mechanism, we provide auxiliary evidence of the heterogeneity analyses at the city and industry levels. First, we show that the HSR effects on firms' CRD are higher for industries with greater firm-performance heterogeneity, which is positively correlated with communication costs. Second, the HSR construction effects on firms' CRD are stronger for cities with less developed internet that benefit more from the reduction of travel time. These results suggest that HSR plays a role in mitigating information asymmetry and improving monitoring.

Our study contributes to several strands of literature. First, our paper relates to the literature studying the effect of communication costs on firms' spatial organizations. Several scholars theoretically show that falling communication costs lead to firms' geographic expansion (Cao et al., 2019; Gokan et al., 2019). However, there is a lack of rigorous empirical evidence on how falling communications costs promote firms' geographic expansion. Most empirical studies focus on how falling communication costs affect the relationship between headquarters and their existing distant affiliates (Giroud, 2013; Gumpert et al., 2022; Kalnins and Lafontaine, 2013). For example, Charnoz et al. (2018) study the effect of falling communication costs between headquarters and affiliates and find that falling communication costs foster the functional specialization of distant affiliates. Giroud (2013) find that falling communication costs increase plant-level investments and total factor productivity. Compared to these studies, our contribution lies in providing causal evidence of how falling communication costs promote firms' geographic expansions.

Second, our study also contributes to the literature focusing on the firms' CRD in the extensive margin (Aghion and Tirole, 1997; Cao et al., 2019; Hsieh and Rossi-Hansberg, 2023). It is acknowledged that geographic distance plays an important role in explaining the extension of the firms' boundary. To the best of our knowledge, however, there is a dearth of existing evidence on the causal relationship between geographic proximity and firms' CRD in the extensive margin. Therefore, this paper sheds new light on providing compelling evidence that HSR construction plays an important role in fostering firms' CRD in the extensive margin.

Finally, our study also relates to the literature that investigates the economic consequences of geographical frictions through HSR construction (Bernard et al., 2018; Lin et al., 2022; Wan and Wan, 2022; Wu et al., 2022). Previous research mainly focus on the macroeconomic consequences of HSR construction (Baum-Snow et al., 2017; Donaldson and Hornbeck, 2016), such as economic growth (Ahlfeldt and Feddersen, 2018; Ke et al., 2017), market integration (Zheng and Kahn, 2013), reducing migrants' wage (Kong et al., 2021) and knowledge spillover (Dong et al., 2020). More recently, there is a growing body of literature examining the impact of HSR on micro units, such as firms or individuals. They find that HSR construction affects innovation activities (Gao and Zheng, 2020), firms' value (A. Zhang et al., 2020a), environmental efficiency (Kong et al., 2021), and investments (Lin et al., 2022). While these papers either treat firms as a whole or focus on the role of proximity between firms (Lin et al., 2022) and innovators (Dong et al., 2020), they ignore that HSR construction could also affect the firms' internal spatial distributions, which is a vital aspect of firms' growth (Cao et al., 2019). As a result, this paper complements the existing literature by investigating the relationship between HSR construction and firms' CRD, that is, whether firms establish new cross-regional affiliates due to HSR construction.

The remaining sections of our study are organized as follows. Section 2 discusses the conceptual framework. The data and empirical design are described in Section 3. Section 4 provides the baseline results, robustness tests, and economic consequences. Section 5 presents the analyses at the city-pair level. Section 6 discusses the channels. Section 7 concludes the paper with a discussion of policy implications.

2. Conceptual framework

Theoretically, whether firms expand their geographic boundary depends on whether they can enter the distant local market successfully and the expectation of development in the future, and these two factors are partly determined by the information flows (Gordon and Bovenberg, 1996). How much the headquarters is privy to local information depends on the distance between the headquarters and its affiliates (Giroud, 2013). With a shorter distance, the time-constrained firms' managers can take personal travels to have first-hand knowledge about how the cross-regional affiliates perform and monitor whether the employees are diligent or competent, which further improves the quality of

[4] Soft information is information that "cannot be credibly transmitted" and "cannot be directly verified by anyone other than the agent who produces it".

[5] For example, cities doing economically better are more likely to have HSR construction, and such city level economic factors can also contribute to the expansion of firms located in the city.

knowledge received by the headquarters (Alonso et al., 2008). A shorter distance should increase the likelihood of opening cross-regional affiliates.

Since the opening of the first HSR in 2008, the Chinese HSR has developed rapidly and made remarkable achievements, which has changed China, with people's travel, economy, culture, social customs, and many other aspects to varying degrees. HSR has dramatically improved the convenience of transportation, shortened the spatial and temporal distances between cities, and alleviated the negative impact of geographic distance on information transmission. Benefiting from the opening of HSR, firms have better access to information on their cross-regional affiliates. Under such circumstances, firms would find it more profitable to open new distant affiliates.

Specifically, the HSR construction helps headquarters obtain affiliates' soft information (Giroud, 2013). Compared to traditional trains and cars, HSRs are faster and safer as passenger transportation; furthermore, unlike air service, they allow individuals to avoid long waiting times and the risk of flight delays. Therefore, HSR optimizes the traditional transportation network and significantly shortens the travel time between two places. HSR connection between two cities can save firms' managers considerable time and transportation costs, providing them the convenience of personally visiting the affiliates for communication and meetings. This situation makes it easier for long-distance firms to obtain important "overtones" (Zhang et al., 2020b) and then significantly reduces information asymmetry and improves monitoring. In summary, we propose hypothesis H1 in the following.

H1. HSR construction promotes firms' cross-regional development.

Geographical distance serves as one measure of information asymmetry, and in addition to HSR construction and distance, the extent of information asymmetry is influenced by other communication costs (Acemoglu et al., 2007; Bloom et al., 2014). Two types of communication costs are particularly relevant. The first type pertains to firm-performance heterogeneity (Huang et al., 2017). When the firms face higher heterogeneity in affiliates' performance, the importance of affiliates-specific local information increases. Consequently, firms may require more face-to-face communication, leading to owners traveling more frequently to acquire such information. Therefore, in cases of higher performance heterogeneity, the reduction of communication costs resulting from HSR construction may yield greater benefits for firms. The second type of communication cost is related to information technology (Petersen and Rajan, 2002). Throughout our sample period (2003–2014), information technology, including the Internet and video conferencing, experienced rapid growth. Such advancements facilitate information flows within firm units, reducing the necessity of personal travel to distant affiliates. In summary, when other communication costs are higher, the effects of HSR construction on firms' cross-regional development (CRD) may be more pronounced. Thus, we propose the following hypothesis H2.

H2. The positive effects of HSR construction on firms' CRD are stronger for firms with greater firm-performance heterogeneity and firms located in cities with less developed internet.

3. Empirical models and data

3.1. Empirical models

This paper first calculates the number of new affiliates opened in other cities by all firms of city c in year t ($num_{c,t}$) using SAIC's data to investigate whether HSR construction will promote firms' cross-regional development. Accordingly, the regression model is set up as follows:

$$num_{c,t} = \alpha + \beta \bullet HSR_{c,t} + \gamma \bullet X_{c,t} + \delta_{pt} + \theta_c + \varepsilon_{c,t} \tag{1}$$

where the key explanatory variable ($HSR_{c,t}$) is a dummy variable. The value is 1 if city c has opened an HSR in year t; otherwise, it is 0. θ_c represents the city fixed effect, which controls the factors at the city level that does not change over time, such as the geographical location, history, and culture of the city c. Using province-year fixed effects, we are able to control for macro-shocks that are common to cities within the same province (δ_{pt}). $\varepsilon_{c,t}$ is the error term, and the standard errors are clustered at the city level.

The identifying assumption of the difference-in-differences method is the parallel trends of firms' CRD between cities with HSR construction and cities without HSR construction before HSR construction. However, if HSR constructions are primarily determined by the cities' economic conditions, this assumption might be violated. Hence, considering that the HSR construction decisions are correlated with regional shocks and other shocks are correlated with firms' geographic expansions, we add time-varying controls at the city level, including *per capita* GDP, employment in the tertiary industry, foreign investment, and population density, which are calculated as logarithms (Lin, 2017).[6]

Based on the city-level panel data used in Equation (1), the number of newly opened cross-regional affiliates of firms in city c is summed across this city. However, such an aggregation comes with certain disadvantages. Multiregional firms may exist in city c, and each firm may have a different industry, affiliate year, and development trend. Unobservable factors at the firm level cannot be controlled if we use the aggregation number. HSR construction may be correlated with these unobservable factors, resulting in endogeneity problems. Additionally, the regression model in Equation (1) fails to capture the heterogeneous impact of HSR construction on different types of firms cannot be accounted for. To address these concerns, this paper constructs panel data at the firm-city-year level and calculates the number of newly opened cross-regional affiliates of firms i (located in city c) in year t, which is referred to as $num_{i,c,t}$, to eliminate the error caused by this summation. As a result, we obtain the following regression model:

$$num_{i,c,t} = \alpha + \beta \bullet HSR_{c,t} + \gamma \bullet X_{c,t} + \delta_{p,t} + \theta_i + \varepsilon_{i,t} \tag{2}$$

where $HSR_{c,t}$ is a dummy variable. $HSR_{c,t}$ takes a value of 1 if firms are located in a city with existing HSR or 0 otherwise. θ_i represents the firm fixed effects, which control for factors relevant to a particular firm that are constant over time, such as firms' time-invariant characteristics. The province-year fixed effects ($\delta_{p,t}$) control for macro-shocks that may be common to different cities within the same province. Similar to Equation (1), the time-varying economic and social factors are controlled at the city level, including *per capita* GDP, employment in the tertiary industry, foreign investment, and population density, which are calculated as logarithms. Equation (2) differs from Equation (1) by controlling fixed effects at the firm level. In this way, factors related to HSR construction at the firm level that do not change over time are controlled. $\varepsilon_{i,t}$ represents the random error term, and we cluster the standard errors at the city level. Therefore, we use the firm-level data to conduct the main analysis, and the results of city-level data are regarded as a reference.

The construction of HSR may depend on other time-varying omitted variables, such as local shocks. If these variables are unrelated to firms' CRD, their presence poses no concern. However, if there are (omitted) factors that are correlated with both the HSR construction and firms' geographic expansion, any relationship between the two could be spurious. To illustrate, consider the scenario where city A experiences strong economic growth, increasing the likelihood of having HSR connections. Simultaneously, the favorable economic conditions may enhance firms' capacity for cross-regional expansion. In addition, it is worth noting that HSR construction can also be influenced by large corporations seeking to expand their geographic reach. To mitigate

[6] We also provide the correlation analysis about the HSR construction variables and city-level economic conditions. The results are available by request from the corresponding author. The results table is not included in the text due to space limitations.

these endogeneity concerns, we try to employ multiple methods.

First, we use an instrumental variables (IV) approach to alleviate the endogeneity concerns. At the city level, we seek instrumental variables that are correlated with the likelihood of HSR connectivity for a city while being unlikely to be correlated with the unobserved determinants of firms' CRD. Our approach is built on past work by Duranton and Turner (2012). The instrument variable we employ is an indicator denoting whether city A was connected by railway at a historical time point (the year 1962), based on the nation's historical railway network (Dong et al., 2020). As shown in the results, the coefficients are still statistically positive using the IV method.

Second, we alleviate the endogeneity concerns by using a sample only including peripheral cities. The network structure of China's HSR system consists of four main horizontal lines and four vertical lines, and whether the peripheral cities are connected by HSR along these lines is usually unplanned (Lin et al., 2022). As shown in the robustness analysis, the results are similar when we use a sample only including peripheral cities.

Third, we consider the dynamic effects of the HSR construction. If HSR construction is the endogenous outcome of a firm-level or city-level shock, we should find the HSR effects before the HSR construction. However, we find no significant effect before HSR construction, which is consistent with the parallel trend assumption.

Fourth, we note that our results are still robust when we consider only small firms. Arguably, it is less likely that small firms can lobby for HSR construction. In the meantime, HSR constructions are less likely to respond to shocks affecting small firms' CRD.

Finally, we also conduct robustness checks at the city-pair level. Under this specification, we could capture firms' cross-regional affiliates in a directional aspect, which would be more intuitional. In addition, we could also add origin city-year and destination city-year fixed effects to control for the contemporaneous local shocks.

In addition to accounting for the sample selection and potential local or firm shocks, we carefully consider several other endogeneity concerns. First, some firms may not depend on HSR to communicate with and monitor their affiliates, then the HSR construction should not matter. Although this is unlikely to introduce any systematic bias, it introduces noise into the regressions, making it only harder for me to find any significant results. Second, we change the sample period for our regression analysis and find consistent results, further strengthening the reliability of our findings. Third, an important concern —especially about difference-in-differences estimations—is that serial correlation of the error term can lead to underestimated standard errors. In our specification, we cluster standard errors at the city level. This clustering not only accounts for the presence of serial correlation within the same firms, but it also accounts for potential correlation of the error terms across firms in the same city in any given year as well as over time. Our results are still statistically significant when clustering standard errors at the province or firm level.

3.2. Data

We use three datasets in our research. First, we utilize the Firm Registration Database to capture our dependent variable, namely the number of firms' cross-regional affiliates. This comprehensive database encompasses administrative information for the entire population of Chinese firms, spanning over 60 million registered entities from 1949 onwards, including those that have been de-registered or revoked. The dataset provides valuable details such as firm names, addresses, registration dates, firm types, industry categories, registered capital, cancellation or revocation dates, and other variables pertaining to firm registration. Obtained from the State Administration for Industry and Commerce (SAIC), this database represents the most extensive and complete source of information on business activities across all regions and industries in China. Moreover, the dataset includes a variable indicating membership that enables us to gather information about each affiliate and its corresponding owner firm. Consequently, we can accurately determine the number of newly established cross-regional affiliates for all firms during the period from 2003 to 2014. To address concerns regarding the reliability of registration data, we limit our sample to the end of 2014. Prior to this date, SAIC implemented stricter regulations, minimizing the likelihood of opportunistic behaviours such as registering affiliates for non-legitimate business purposes, including capturing government subsidies.

Second, we collect HSR data from the website of the China National Railway Administration and China Railway Corporation. Firms operating in a city with an established HSR throughout our sample period are defined as the treatment group in the DID framework.

Third, the remaining city-level control variables are derived from city statistical yearbooks. Descriptive statistics for the main variables are presented in Table 1. A total of 4220 observations at the city level and 3,527,857 observations at the firm level are obtained from 2003 to 2014.

4. Results

4.1. Benchmark results at the city level

Table 2 presents the results of city-level regressions that are controlled for city fixed effects and year fixed effects. Column (1) shows the results of the regression without including control variables, whereas the results in columns (2)–(4) add control variables. The coefficients of $HSR_{c,t}$ are significantly positive, indicating that HSR construction promotes firms' cross-regional development. In terms of economic significance, the regression coefficient for $HSR_{c,t}$ in column 4 is 0.086, indicating that the new affiliates opened by firms located in the city with HSR construction increase by 8.6%. Furthermore, it implies that if a city has 100 firms, the number of newly cross-regional affiliates of all firms in that city increases by 8.

4.2. Benchmark results at the firm level

Table 3 displays the outcomes of firm-level regressions with all regressions adding firm fixed effects and year fixed effects. The regression results in columns (1) and (2) show that firms open more cross-regional affiliates after HSR construction. The results align with the findings obtained at the city level. We also use the logarithm of $num_{i,c,t}$ as the dependent variable and report the regression results in columns (3) and (4). The coefficient of $HSR_{i,c,t}$ is 0.005, which is significant at the 5% level. These results are consistent with the evidence that firms geographically expand when the communication costs decrease (Charnoz et al., 2018).

4.3. Dynamic effect test

Satisfying the assumption of a parallel trend between the treatment and control groups in Equation (2) is necessary to ensure unbiased estimation results. Specifically, the distribution of treatment and control groups should have the same change trend before HSR construction; otherwise, the DID method may overestimate or underestimate the effects of HSR construction. This paper examines the dynamic effect of HSR construction on firms' cross-regional development by applying the two-way fixed effect estimation method to test the parallel trend assumption. The regression model is constructed as follows:

$$num_{it} = \beta_k \sum_{k=-5}^{6} D_{t_{c0}+k} + \gamma \bullet X_{ct} + \delta_t + \theta_i + \varepsilon_{i,t} \quad (3)$$

where t_{c0} represents the year when city c constructs HSR. $D_{t_{c0}+k}$ is a dummy variable and set as 1 after HSR construction in city c or 0 otherwise. k is negative before HSR construction and positive after HSR construction. Following Beck et al. (2010), $D_{t_{c0}+6}$ is defined as 1 for

Table 1
Summary statistics.

Variable	N	Mean	Std. Dev	Min	Med	Max
$num_{c,t}$ (city level)	4220	226.2156	844.8349	1	40	17013
$num_{i,t}$ (firm level)	3,527,857	0.2258	4.0446	0	0	5252
High-speed rail	4220	0.1829	0.3867	0	0	1
GDP per capita	4220	10.0217	0.8158	4.6052	10.0441	13.0557
Employment in service industries	4220	11.9853	0.7737	9.0241	11.9592	15.6439
Foreign-directed investment (FDI)	4220	9.5146	1.9981	0	9.6022	14.5638
Population intensity	4220	5.7263	0.9064	1.7405	5.8752	7.8870

Table 2
Benchmark results at the city level.

Dependent Variable	(1)	(2)	(3)	(4)
	log(1+num)	*log(1+num)*	*log(1+num)*	*log(1+num)*
HSR	0.085**	0.087***	0.086***	0.086**
	(0.03)	(0.03)	(0.03)	(0.03)
GDP per capita		0.042***	0.037**	0.037**
		(0.02)	(0.02)	(0.02)
Employment in service industries		0.004	0.001	−0.000
		(0.02)	(0.02)	(0.02)
FDI			0.015**	0.015**
			(0.01)	(0.01)
Population intensity				0.002
				(0.04)
Observations	4220	4220	4220	4220
City FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

Notes: This table reports regression results using Equation (1). ***p < 0.01, **p < 0.05, *p < 0.1. Robust standard errors clustered at the city level are reported in parentheses. The dependent variable is the cross-regional affiliates of all firms in cities.

Table 3
Benchmark results at the firm level.

Dependent Variable	(1)	(2)	(3)	(4)
	num	*num*	*log(1+num)*	*log(1+num)*
HSR	0.013**	0.013**	0.005**	0.005**
	(0.01)	(0.01)	(0.00)	(0.00)
GDP per capita		−0.003		0.000
		(0.00)		(0.00)
Employment in service industries		0.003		0.001
		(0.00)		(0.00)
FDI		−0.003		−0.001
		(0.01)		(0.00)
Population intensity		0.002		0.001**
		(0.00)		(0.00)
Observations	3,527,857	3,527,857	3,527,857	3,527,857
Controls	NO	YES	NO	YES
Firm FE	YES	YES	YES	YES
Year#Province FE	YES	YES	YES	YES

Notes: This table reports regression results using Equation (2). ***p < 0.01, **p < 0.05, *p < 0.1. Robust standard errors clustered at the city level are reported in parentheses. The dependent variable is firms' cross-regional affiliates and the following tables all report results using firm-level dependent variables without taking logarithm as the dependent variables.

6 years or more after HSR construction or 0 otherwise. $D_{t_{c0}-5}$ is defined as 1 for 5 years or more before HSR construction or 0 otherwise. Our base year is the year before HSR construction; therefore, $D_{t_{c0}-1}$ is not included in Equation (3). In the same way as Equation (2), X_{ct} represents the time-varying economic and social factors at the city level. δ_t represents the year fixed effect. θ_i represents the firm fixed effects. $\varepsilon_{i,t}$ represents the random error term, and we cluster the standard errors at the city level.

Fig. 1 illustrates the trend of coefficient β_k over time. The dotted line represents the 95% confidence interval of coefficient β_k, whereas the standard errors are clustered at the city level. According to Fig. 1, coefficient β_k is insignificant before the event but significantly greater than 0 after the event, which satisfies the parallel trend assumption.

4.4. Placebo test

We conduct a placebo test using a randomly generated dummy variable to verify that the baseline regression results are not driven by chance, the method is similar to Li et al. (2016). We randomly chose the year and number of cities for the treatment group based on the amount of HSR constructions in each year. Specifically, we select 12 years from 2003 to 2014 at random (namely *t1, t2, …, t12*). a1 cities randomly selected from all cities are regarded as the opening of HSR in year *t1*; a2 cities from the remaining city pairs are regarded as the opening of HSR in year *t2*; and so on. Then, we generate treatment and control groups randomly. We re-estimate column (4) of Table 3 with randomly generated core independent variables to obtain the estimated coefficients. The above steps are repeated 500 times using Monte Carlo simulations, and

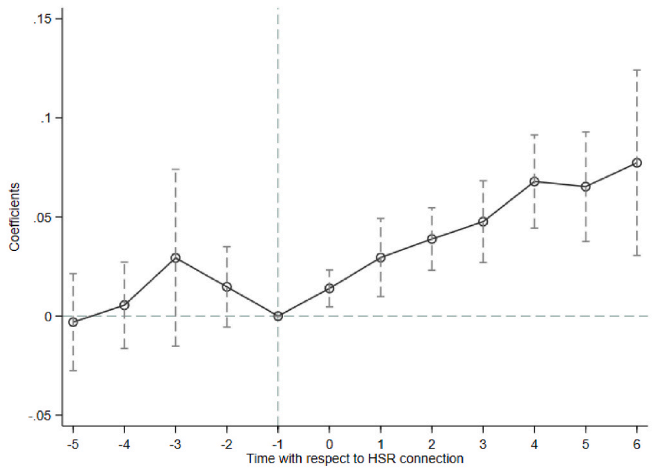

Fig. 1. Dynamic effects of HSR construction.

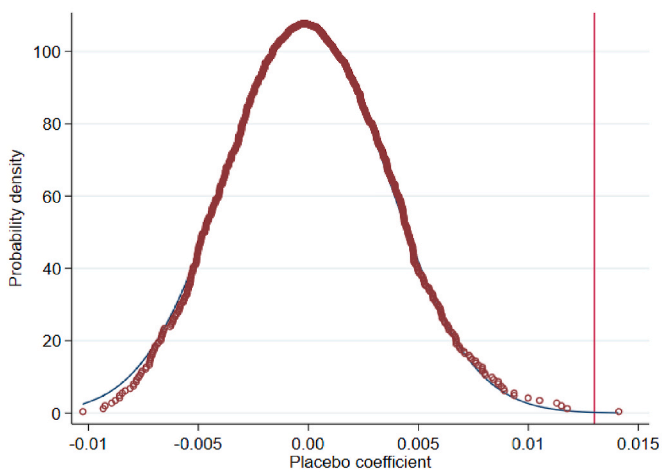

Fig. 2. Distribution of the estimated coefficients of falsification test. Note: The figure shows the cumulative distribution density of the estimated coefficients from 500 simulations randomly assigning the HSR construction to cities. The red vertical line presents the result of column (2) in Table 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the results are plotted in Fig. 2. As a result of 500 Monte Carlo simulations, the estimated coefficients are almost all less than 0.013 derived from real data and are normally distributed with 0 as the mean. As a result, the benchmark regression results cannot be attributed to some accidental causes.

Table 4
Two-stage least squares regressions.

	(1) *First-stage estimation*	(2) *Second-stage estimation*
IV	0.295***	
	(0.05)	
HSR		0.152***
		(0.05)
Observations	3,539,179	3,539,179
Controls	YES	YES
Firm FE	YES	YES
Year#Province FE	YES	NO

Notes: This table reports the results of two-stage least squares regression using the historical route (1961 train station and Ming Dynasty post station) as the instrumental variables. ***p < 0.01, **p < 0.05, *p < 0.1. Robust standard errors clustered at the city level are reported in parentheses.

Table 5
Robustness checks.

	(1)	(2)	(3)
HSR	0.018**	0.013*	0.014**
	(0.01)	(0.01)	(0.01)
Observations	1,195,315	3,527,857	1,774,389
Controls	YES	YES	YES
Firm FE	YES	YES	YES
Year#Province FE	YES	NO	YES

Notes: This table reports regression results using Equation (2). In column 1 we used the sub-sample excluding node cities. In column (2), we cluster standard errors at the province level. In column (3), we use the sub-sample of small firms. ***p < 0.01, **p < 0.05, *p < 0.1.

4.5. Other robustness checks

Although our prior analysis provides plausible exogenous variations in proximity, there are still some concerns regarding the endogenous problem, mainly because of the non-randomness of HSR construction. It has been well established in the transport economics literature that a bidirectional causal relationship may exist between investment in transportation and related economic activities (Donaldson and Hornbeck, 2016). We then conduct several robustness checks to address these concerns.

First, we use IV estimations to alleviate the endogenous issues (Wu et al., 2022). Similar to Dong et al. (2020), we use the 1961 train station based on China's historical railway network as a source of quasi-random variation in HSR construction. Table 4 reports the first-stage result that regresses the HSR dummy variable on the instrumental variables while controlling for fixed effects. It confirms that our instruments are strong predictors of HSR connection, which is consistent with Wu et al. (2022). The coefficient of the IV are positive and significant at the 5% level. The instrumental variables also pass identification tests.

Second, we use the sample excluding nodal cities. The Chinese State Council adopted the Medium- and Long-Term Railway Network Plan in 2004, which proposed the construction of "four vertical" and "four horizontal" special passenger lines. Considering that the 28 node cities in the plan are artificially selected, HSR construction in these cities is highly dependent on the administrative level, economic development level, and population of the cities themselves rather than the geographical location of node cities, resulting in endogeneity concerns. Therefore, this study removes the observed data of firms located in node cities to eliminate the interference of non-random factors. According to column (1) of Table 5, the regression coefficients are still significantly positive, and the magnitude of the coefficients is similar to those in Table 3.

Third, we cluster the standard errors at different levels. In the baseline specification, we cluster the standard errors at the city level, ignoring the correlation between the number of newly opened cross-regional affiliates of firms in different cities within the same province and the error term may be underestimated. Hence, we further cluster our standard errors at the province level, and the results in column (2) are similar to the results in Table 3.

Forth, we examine whether small firms are affected by the HSR construction. We use the sample only including small firms, and the results are still statistically significant (column (3)). According to Giroud (2013), small firms are unlikely to lobby for airline route construction. Similarly, small firms are also unlikely to affect the decisions of HSR

Table 6
Bacon decomposition.

Type	Weighted Average	Total weight
Timing groups	0.003	0.545
Always treated vs Timing groups	0.049	0.074
Never treated vs Timing groups	0.022	0.381

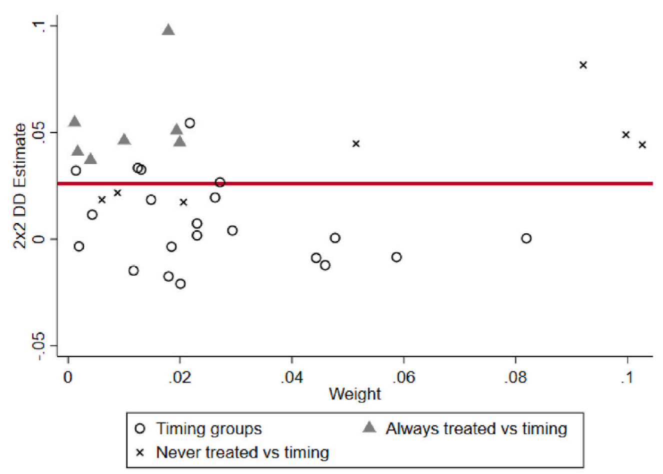

Fig. 3. Bacon decomposition.

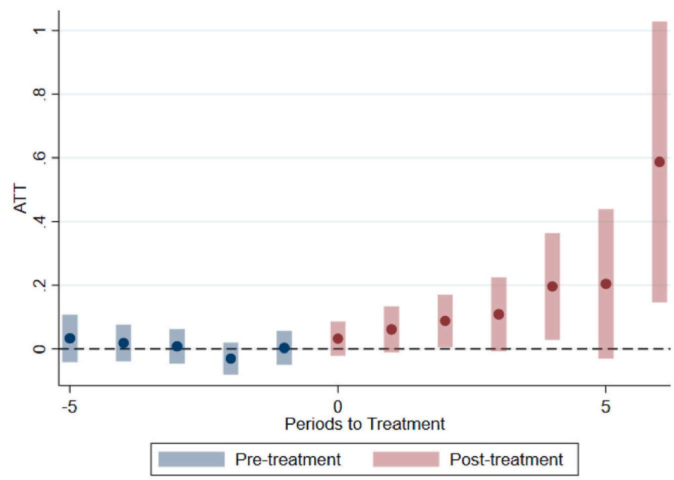

Fig. 4. Callaway and Sant'Anna estimator.

construction in China. We find that small firms' CRD is also affected by HSR construction, which further consolidates our estimation.

Finally, we conduct alternative DID estimators. Our baseline specification is staggered DID, we provide further evidence to make our results more robust according to a recent discussion by theoretical econometricians (Callaway and Sant'Anna, 2021; De Chaisemartin and D'Haultfoeuille, 2020; Goodman-Bacon, 2021). They argue that when policy effects change over time, the result estimated by the staggered DID method (i.e., Equation (2)) is not a well-defined average treatment effect but a weighted average of multiple standard differential methods that estimate average treatment effects. Moreover, the weights can be negative (De Chaisemartin and D'Haultfoeuille, 2020). The dynamic effect test method (event study method) of the staggered difference-in-differences method also has problems (Sun and Abraham, 2021). Thus, they argue that we should provide more evidence to make estimation robust when using staggered DID estimators.

In this study, we first provide the Bacon decomposition developed by Goodman-Bacon (2021). The decomposition results are presented in Table 6 and Fig. 3. In the Timing groups, the control groups are either earlier treated or later treated groups. In the Always treated vs Timing groups, the control groups are those treated in the first period of our sample. In the Never treated vs Timing groups, the control groups are those never treated groups. The decomposition results show that the aggregate treatment effect is mainly driven by unbiased samples, and the estimates in the biased sample are positive.

We then provide robust evidence introduced by Callaway and Sant'Anna (2021). The average treated effect is found to be 0.099, which is significant at a 1% confidence level. Furthermore, the dynamic event study estimates and confidence intervals are reported in Fig. 4, in which the coefficient is insignificant before the event and significantly greater than 0 after the event, indicating that our baseline results are still valid.

4.6. Economic consequences

Based on the results of the above analysis, HSR construction promotes firms' cross-regional development. After HSR construction, face-to-face contact can be easier, and the communication costs are reduced. Hence, firms are inclined to open affiliates in more distant regions, which could result in a greater distance between firms and their affiliates. Consequently, we then examine whether HSR construction have this effect. Furthermore, two opposite effects may occur as firms expand. On the one hand, firms can conduct cross-regional operations in distant markets, improve their competitiveness, and establish economies of scale by taking advantage of the new market, improving their performance (Luo and Tung, 2007; Tallman and Li, 1996). On the other hand, firms may overexpand, which increases the likelihood of bankruptcy (Higgins and Schall, 1975) and puts them at greater risk (Mueller, 1977). Lacking detailed firm-level information (for instance, productivity cannot be estimated without information on investment and profit), This study investigates the effect of HSR construction on the performance of firms from the perspective of firms' duration.

4.6.1. Average distance between affiliates and headquarters

Geographic distance is used as a measure of information asymmetry and regulatory costs (Devos and Rahman, 2014; John et al., 2011). In a sense, HSR construction corresponds to the artificial compression of "geographic distance". As a result, geographical proximity can help firms to overcome the challenge of high information costs, enable more frequent and timely communication with remote affiliates, and facilitate easier field visits. The distance between firms' headquarters and their affiliates eventually increase through better supervision and guidance of daily operations and the production decisions of cross-regional affiliates. We use the weighted distance index ($Wdis_{i,t}$) to calculate the distance between firms and their affiliates based on existing research (Petersen and Rajan, 2002) that assesses the distribution of firms' affiliates:

$$Wdis_{i,t} = ln\left(1 + \sum_{c} Dis_{i,c',t} \times weight_{c'}\right) \tag{4}$$

where $Wdis_{i,t}$ represents the weighted distance between firm i and all its affiliates in year t. $Dis_{i,c',t}$ denotes the geographical distance between city c where the firms' headquarters are located and city c' where its affiliates are located. $Weight_{c'}$ denotes the ratio of the number of affiliates that firm i opens in city c' in year t to the number of affiliates that firm i

Table 7
Economic consequences of HSR.

Dependent Variable	(1)	(2)	(3)	(4)
	$Wdis_{i,t}$	$Wdis_{i,t}$	*Hazard rate*	
HSR	0.641***	0.686***	0.053***	0.039**
	(0.08)	(0.08)	(0.02)	(0.02)
Observations	9,773,970	9,773,970	455,019	439,010
Controls	NO	YES	NO	YES
Firm FE	YES	YES	YES	YES
Year#Province FE	YES	YES	YES	YES

Notes: Columns (1)–(2) in this table reports regression results using Equation (2), and dependent variables are the weighted distance between firms' headquarters and all its newly opened affiliates in each year from Equation (4). Columns (3)–(4) in this table present regression results using Equation (5). ***p < 0.01, **p < 0.05, *p < 0.1. Robust standard errors clustered at the city level are reported in parentheses.

opens in all cities in year t. The regression results after replacing the explained variable in Equation (2) with $Wdis_{i,t}$ are presented in columns (1) and (2) of Table 6. Accordingly, the coefficient of $HSR_{c,t}$ is significantly positive, regardless of whether control variables are included or not, indicating that HSR construction encourages firms to open affiliates in farther places. The coefficient of $HSR_{c,t}$ in column (2) is 0.686, which indicates that the weighted distance between firms and their affiliates increases by 68.6% after HSR construction in the location where firms register, which is also important in the economic sense.

4.6.2. Firms' survival time

We construct a Cox proportional hazards model as depicted in Equation (5) to assess the impact of HSR construction on the duration of firms. The period of each firm's presence on the market is observed from the registration and cancellation dates in the SAIC data.

$$h_{i,c}(t) = h_0(t)exp\left(\alpha + \beta \bullet HSR_{c,t} + \gamma \bullet X_{c,t} + \delta_t + \theta_i + \varepsilon_{i,t}\right) \quad (5)$$

The hazard of firm i located in city c died in time t is given by $h_{i,c}(t)$. $h_0(t)$ is the nonparametric baseline hazard of exit. The independent variable ($HSR_{c,t}$) is a dummy variable. As soon as city c opens its HSR in year t, $HSR_{c,t}$ takes the value of 1; otherwise, it takes the value of 0. δ_t is the year fixed effect, θ_i is the firm fixed effect, and $\varepsilon_{i,t}$ is the error term. The regression results are reported in columns (3) and (4) of Table 7. The coefficients of $HSR_{c,t}$ are all significantly positive; therefore, the effect of the firms' cross-regional expansion brought by the construction of the HSR on the headquarter is more negative. According to column (4), the coefficient of $HSR_{c,t}$ is 0.039, which indicates that firms in cities with HSR construction are more likely to exit the market by 3.9% compared with those in cities without HSR construction.

According to the above analysis, there are two opposite effects of firms' expansion on firms' duration. The results show that HSR construction shortens firms' survival time, which could imply that the negative effect of HSR effects dominates. For example, firms' expansion also creates new demand for capital and capital structure at the same time. Firms' capital often cannot fully meet the expansion needs, and firms often expand with high debt. In these cases, when the capital chain breaks, firms would face bankruptcy risk.

However, we cannot conclude that HSR construction is bad for firms. Since we cannot observe other firms' level economic variables, further analysis is unavailable. In addition, we use firms' presence and cancellations in the SAIC's database to proxy for firms' duration, which is a coarse measure (Shi et al., 2021).

5. City-pair level analysis

5.1. Main specification

We provide empirical evidence at the city-pair level in this section. Our regression model in Equations (1) and (2) only controls for city/firm fixed effects, which could only control time-invariant factors, and there are still several concerns about the estimation results. On the one hand, HSR constructions may be correlated with time-varying factors at the city or firm level, for example, cities doing economically better are more likely to have HSR connections and city-level economic factors could also contribute to the expansion of firms located in the city. On the other hand, in city-level regression, we do not observe how firms expand. Hence, we also conduct our analysis at the headquarter-affiliate pair level, the model is as follows:

$$log\left(num_{ijt} + 1\right) = \alpha + \beta connect_{ijt} + \theta_{ij} + \delta_{it} + \gamma_{jt} + \varepsilon_{ijt} \quad (6)$$

Where i indicates the city where headquarter is located, j indicates the city where affiliates are located. num_{ijt} is the total number of new affiliates opened in city j by all firms of city i in year t. $connect_{i,j,t}$ is a dummy variable indicating whether a city pair (i,j) is connected by HSR at year t, which equals to 1 if city i and city j are connected by HSR in year t or 0 otherwise. The advantage of this model is that it can control for high-dimensional fixed effects. We add city-pair fixed effects (θ_{ij}) to control for time-invariant omitted unobservable variables at the city-pair level, including pre-existing pairwise economic linkages. We also add origin city-year and destination city-year fixed effects (δ_{it} and γ_{jt}), these two fixed effects can control for time-varying omitted unobservable variables that could affect both the HSR construction and firms' geographic expansion. In addition, we cluster the standard errors at the city-pair level.

The regression results of Equation (4) are presented in Table 8. The dependent variable is the total number of new affiliates opened in city j by all firms of city i in year t. Column (1) includes city-pair and year-fixed effects. Column (2) adds origin city-year and destination city-year fixed effects, allowing for a highly flexible functional form of origin- and destination-city time trend. The coefficient estimates in columns (1) and (2) are 0.168 and 0.085 respectively and are both statistically significant (1% level). These results indicate that the connection between two cities brought by HSR construction promotes firms' CRD, and increase the number of cross-regional affiliates by 8.5%, compared with the control city pairs that are not connected by HSR or connected in later years.

Table 8
City-pair level results.

Dependent Variable: $log\left(num_{ijt} + 1\right)$	(1)	(2)	(3)	(4)
HSR	0.168***	0.085***	0.143***	0.070***
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	1,366,872	1,366,872	1,336,296	1,336,296
City-Pair FE	YES	YES	YES	YES
year FE	YES	NO	YES	NO
Origin-Year FE	NO	YES	NO	YES
Destination-Year FE	NO	YES	NO	YES

Notes: Columns (1)–(2) in this table reports regression results using Equation (4). Columns (3)–(4) in this table present regression results using Equation (4), and the HSR connection are indirect connection. ***p < 0.01, **p < 0.05, *p < 0.1. Robust standard errors clustered at the city level are reported in parentheses.

5.2. Dynamic effect test

In Equation (4), we include high-dimensional fixed effects to address the endogeneity concerns. However, there are still other concerns for our identification, such as the time-varying unobservable factors at the city-pair level that are correlated with HSR construction. For example, when two cities become more connected in economic linkages and these

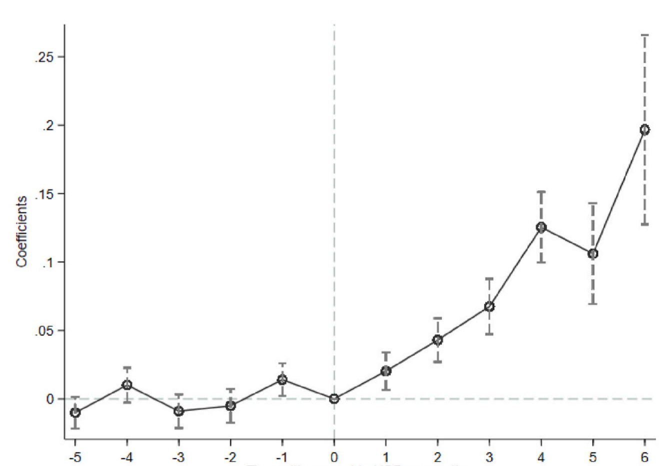

Fig. 5. Dynamic effects of HSR construction at city-pair level.

linkages could correlate with HSR construction and firms' CRD, we should observe the increase in the number of affiliates between the two cities even before the HSR connection.

To check whether HSR effects exist before HSR construction, we estimate the dynamic effects using the following model:

$$\log\left(num_{ijt}+1\right)=\beta_k\sum_{k=-5}^{6}connected_{t_{c0}+k}+\theta_{ij}+\delta_{it}+\gamma_{jt}+\varepsilon_{ijt} \quad (7)$$

where t_{ij0} represents the year when city i and j are connected by HSR. $D_{t_{ij0}+k}$ is a dummy variable and equals 1 after city i and j are connected by HSR or 0 otherwise. k is negative before the HSR connection and positive after the HSR connection. As in Equation (3), $D_{t_{ij0}+6}$ is defined as 1 for 6 years or more after HSR connection or 0 otherwise. $D_{t_{ij0}-5}$ is defined as 1 for 5 years or more before HSR connection or 0 otherwise. The base year is one year before the HSR connection.

The coefficients are plotted in Fig. 5, the vertical dotted lines represent the 95% confidence intervals. From Fig. 5, we can see that compared to the baseline period, there is no significant increase in the number of affiliates before the actual HSR connection. This result is consistent with the parallel trend assumption of the staggered difference-in-differences method and alleviates the concern of omitted variables and sample selection. After the HSR connection, we can see that the number of affiliates in city j opened by firms in city i increased by 8.5% and are significant at 1% level.

5.3. Indirect connection

To further alleviate the endogeneity concern on the HSR construction, we exploit the network structure of China's HSR system and check whether the indirect connection by HSR could also affect firms' CRD. The system is mainly constructed by four vertical rail lines and four horizontal rail lines. When a vertical line is connected with a horizontal line, two non-nodal cities on each line can also become connected, which we called an indirect connection. As noted, these indirect connections are largely unplanned, especially when we add city-pair fixed effects (θ_{ij}), origin city and destination city fixed effect (δ_{it}, γ_{jt}) as in Equation (2). Hence, the threat to identification due to potential endogenous selection can be ruled out more explicitly.

We drop a sample consisting of city pairs that are directly connected by HSR construction. Our treatment group is the city pairs that are directly connected by HSR expansion, and the control group is the city pairs that are not connected or connected at a later date. We then employ the method used by Lin et al. (2022) to construct the indirect connection of the HSR variable $Indirect_{ijt}$, which equals to 1 if city i and city j are indirectly connected after year t (include year t) and 0 otherwise. We then replace the variable $Indirect_{ijt}$ with $connect_{ijt}$ in Equation (4) and run the regression again.

The estimation results are presented in columns (3)–(4) in Table 8, the coefficients of the variable $Indirect_{ijt}$ are 0.143 and 0.07, respectively, and significant at 1% level, which implies that HSR connection increases the number of affiliates in city j opened by firms in city i.

6. Discussion

6.1. Discussions on channels

In this section, we further try to clarify the communication costs channel behind the HSR effect on firms' CRD. First, we conduct a heterogeneity analysis based on the level of Internet development in each city. During our sample period (2003–2014), China has witnessed rapid advancements in Internet technology. The proliferation and wide adoption of the Internet facilitate the exchange of information within firms' spatial units, leading to a reduction in their internal communication costs (Giroud, 2013). The Internet enables efficient information flow within firms' units, thereby diminishing the necessity for personal travel to oversee cross-regional affiliates (Giroud, 2013). Conversely, in cities with less developed Internet, firm owners may need to engage in more frequent personal travel to effectively manage their geographically dispersed affiliates. Consequently, if the construction of HSR promotes firms' cross-regional development primarily through the information channel, we would expect to observe stronger effects of HSR in cities with relatively less developed Internet.

Specifically, we divide the cities into two groups based on their Internet penetration rates: cities with a high Internet penetration rate and cities with a low Internet penetration rate. This division allows us to test the above conjecture and perform regression analyzes separately for each group. The regression results are presented in Table 9. Columns (1) and (2) show the regression results of cities with higher internet penetration rates, and the coefficients of $HSR_{c,t}$ are 0.016, which are significant at the 10% and 5% levels, respectively. As shown in columns (3) and (4), the regression results for cities with low internet penetration have a coefficient of 0.032, which is significant at the 1% level. Notably, the regression results for the sample of cities with low internet penetration rates exhibit higher levels of significance and coefficient magnitudes compared to those for cities with high internet penetration rates. In addition, we also interact the internet penetration rate with the HSR construction variable in Equation (2). The results are reported in columns (5) and (6), the coefficients of the interaction term are both negative and statistically significant at the 5% level, which means that when the Internet is more developed, the HSR effects on firms' CRD are weaker. Therefore, these results confirm, to some extent, that HSR construction promotes firms' geographic expansion via the channel of information costs. Our results are consistent with the findings in Giroud (2013), which find that new airline routes increase distant affiliates' investments, and the effects are stronger in the earlier years of the sample period when other nonpersonal means of exchanging information (e.g., Internet) are either unavailable or less developed.

Second, we provide empirical evidence of the firm-performance heterogeneity analysis. According to Huang et al. (2017), firm-performance heterogeneity is positively correlated with communication costs. When firms operate cross-regional affiliates and face higher performance heterogeneity, firms acquire more affiliates' specific local information and it is more difficult to acquire such information without face-to-face communication. Firms need to conduct frequent personal travels to acquire such information and improve monitoring. Hence, if HSR construction promotes firms' CRD through the communication costs channel, then the effect of geographic proximity on firms' geographic expansion should be stronger with greater firm-performance heterogeneity.

Specifically, we adopt the method from Huang et al. (2017) to measure the performance heterogeneity, that is, the dispersion of TFP.[7] We interact the performance heterogeneity with the HSR construction variable and add it in Equation (2), the results are reported in Table 10. The coefficients of the interaction term in columns (1)–(3) show that the HSR effects are stronger when firm performance heterogeneity is greater. Industries characterized by higher performance heterogeneity, relative to other industries, typically necessitate increased face-to-face communication and on-site and physical monitoring. Consequently, the impact of HSR on these industries is stronger. This suggests that enhanced passenger transportation, facilitating smoother mobility, plays a crucial role in supporting the operation of cross-regional affiliates for firms in such industries. Therefore, the results in Table 10 are consistent with our hypothesis that HSR reduces travel time, and promotes firms' geographic expansion by reducing communication costs. Furthermore, these results are also consistent with Huang et al. (2017),

[7] The standard deviation of the firm's TFP in the three-digit-industry-year level, we use three method to calculate TFP: OLS, Olley-Pakes and Index Number methods.

Table 9
Internet penetration rate.

Sample	*Internet penetration > median*		*Internet penetration < median*		*Interaction*	
	(1)	(2)	(3)	(4)	(5)	(6)
*Treat*Internet*					−0.021**	−0.021**
					(0.01)	(0.01)
Treat	0.016*	0.016**	0.032***	0.032***	0.03***	0.03***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Observations	3,055,328	3,055,328	417,277	417,277	3,472,605	3,472,605
Controls	NO	YES	NO	YES	NO	YES
Firm FE	YES	YES	YES	YES	YES	YES
Year#Province FE	YES	YES	YES	YES	YES	YES

Notes: This table reports regression results using Equation (2). ***p < 0.01, **p < 0.05, *p < 0.1. Robust standard errors clustered at the city level are reported in parentheses.

Table 10
Performance heterogeneity.

	TFP OLS dispersion	*TFP Olley-Pakes dispersion*	*TFP Index Nummber dispersion*
	(1)	(2)	(3)
*Treat*Performance Heterogeneity*	0.022***	0.041***	0.034***
	(0.00)	(0.00)	(0.00)
Treat	0.027***	0.055***	0.042***
	(0.00)	(0.00)	(0.00)
Observations	3,472,605	3,472,605	3,472,605
Controls	YES	YES	YES
Firm FE	YES	YES	YES
Year#Province FE	YES	YES	YES

Notes: This table reports regression results using Equation (2). ***p < 0.01, **p < 0.05, *p < 0.1. Robust standard errors clustered at the city level are reported in parentheses.

Table 11
Commuting cost.

Sample	*Number of affiliates > 1*		*Number of affiliates > 2*	
	(1)	(2)	(3)	(4)
Treat	0.035*	0.033*	0.056*	0.053*
	(0.02)	(0.02)	(0.03)	(0.03)
Observations	1,157,529	1,157,529	667,748	667,748
Controls	NO	YES	NO	YES
Firm FE	YES	YES	YES	YES
Year#Province FE	YES	YES	YES	YES

Notes: This table reports regression results using Equation (2). ***p < 0.01, **p < 0.05, *p < 0.1. Robust standard errors clustered at the city level are reported in parentheses.

in which the positive distance-decentralization relationship is stronger for state-owned enterprises (SOE) with greater firm-performance heterogeneity.

Next, we try to provide empirical evidence to exclude the alternative channel, that is, the reduction of commuting costs. For the sample of sole proprietors, firms may mainly benefit from the reduction of commuting costs rather than internal communication costs. Sole proprietors usually manage their affiliates personally and then have to commute daily. Similar to Kalnins and Lafontaine (2013), to distinguish communication costs channel from pure commuting costs, we look for evidence that the effects of HSR on firms with multi-establishments. Such firms are required to hire employees to operate and manage at least some of their businesses, making communication and monitoring costs as a crucial mechanism through which the HSR construction effects operate. If the channel is the reduction of commuting costs, we would not find the HSR effects for multiregional firms. Table 11 shows the regression results. Columns (1) and (2) only retain samples when the number of cross-regional affiliates open by firms is greater than 1 prior to HSR construction. As a result of adding the control variable to column (2), the coefficient of $HSR_{c,t}$ is 0.033, which is significant at the 10% level. Moreover, only firms that open more than two cross-regional affiliates are included in columns (3) and (4). In column (4) where control variables are added, the coefficient of $HSR_{c,t}$ is 0.053, which is statistically significant at the 10% level. Hence, we reinforce the communication costs channel by confirming positive and statistically significant HSR effects in a sample that excludes sole proprietors.

6.2. *Other heterogeneity analysis*

6.2.1. *Cross-regional development experience*

Intuitively, firms with previous cross-regional development experience are more likely to overcome geographic barriers. In particular, firms with previous experience in cross-regional development can accumulate corresponding knowledge and skills based on previous experience (McDonald et al., 2008). Such knowledge and skills are an important component of firm-specific advantages (Clarke et al., 2013). As a result, firms can effectively distinguish between relevant and irrelevant information when making cross-regional development decisions and make high-quality business decisions. Firms with experience in cross-regional operations are better equipped to identify and take advantage of opportunities in the institutional environment of a target market and to expand successfully in that market (Henisz, 2003). Accordingly, firms with rich experience in cross-regional operations are more likely to open cross-regional affiliates even without HSR construction. Hence, the effect of HSR construction on firms' cross-regional development may be less for firms with previous cross-regional development experience. Based on the idea of Heimeriks and Duysters (2007), the study measures the cross-regional development experience of firms using the existing number of cross-regional affiliates that are open before HSR construction and interacted with $HSR_{c,t}$. The regression results are presented in columns (1) and (2) of Table 12. The coefficient of the triple interaction term is significantly negative, suggesting that HSR construction has a stronger promoting effect on the cross-regional development of firms with less cross-regional development experience. In addition, this paper defines another cross-regional development experience variable to enhance the robustness of the results. A firm with more than 0 cross-regional affiliates before HSR construction is considered experienced (experience = 1); otherwise, it is considered inexperienced (experience = 0). In columns (3) and (4), the coefficients of triple interaction terms are still significantly negative, which strengthens the results.

6.2.2. *Heterogeneity of scales*

The theory of foreign investment emphasizes the importance of explaining foreign direct investment (FDI) behaviour from the perspective of firm heterogeneity (Helpman et al., 2004). The productivity and scale of FDI firms tend to be higher than those of non-FDI firms. Only high-productivity firms can bear the market entry costs caused by FDI and therefore profit from it. In the domestic market, more

Table 12
Heterogeneity analysis: Cross-regional development experience and scale.

	(1)	(2)	(3)	(4)	(5)	(6)
*HSR*experience*	−0.084***	−0.084***	−0.458***	−0.458***		
	(0.00)	(0.00)	(0.01)	(0.01)		
*HSR*size*					−0.020**	−0.020**
					(0.01)	(0.01)
HSR	0.161***	0.160***	0.343***	0.343***	0.014	0.004
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Observations	3,269,156	3,269,156	3,527,857	3,527,857	3,555,353	3,555,353
Controls	NO	YES	NO	YES	NO	YES
Firm FE	YES	YES	YES	YES	YES	YES
Year#Province FE	YES	YES	YES	YES	YES	YES

Notes: This table reports regression results using Equation (2). ***p < 0.01, **p < 0.05, *p < 0.1. Robust standard errors clustered at the city level are reported in parentheses. Note: The results in columns (1)–(4) are the heterogeneity regression of firms' cross-regional development experience. The results in columns (5)–(6) are the heterogeneity regression of firms' scales according to registered capital.

Table 13
Heterogeneity analysis: Industry.

Sample	*Agricultural industry*		*Manufacturing industry*		*Service industry*		*interaction*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
*HSR*Industry*							0.397***	0.397***
							(0.01)	(0.01)
HSR	0.060	0.057	0.015	0.014	0.016*	0.015*	0.235***	0.234***
	(0.08)	(0.08)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Observations	263,534	263,534	898,981	898,981	2,330,361	2,330,361	3,988,360	3,988,360
Controls	NO	YES	NO	YES	NO	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
Year#Province FE	YES	YES	YES	YES	YES	YES	YES	YES

Notes: This table reports regression results using Equation (2). ***p < 0.01, **p < 0.05, *p < 0.1. Robust standard errors clustered at the city level are reported in parentheses.

productive and larger companies are more likely to break geographical distances and establish affiliates in other cities. Therefore, HSR construction may have less impact on bigger firms' cross-regional development. We use the variable "registered capital" in SAIC data to distinguish the size of firms. As shown in columns (5) and (6) in Table 12, the triple interaction terms are all statistically negative, indicating that HSR construction has a greater impact on the cross-regional development of smaller firms.

6.2.3. Heterogeneity in sectors

The opening of affiliates in different industries may have different effects as a result of HSR construction. The tertiary industry has a greater demand for talent, information, and knowledge compared with primary and secondary industries, which are highly dependent on raw materials and goods. If HSR construction promotes firms' cross-regional development through the communication costs channel, we could speculate that HSR effects on firms' geographic expansion are stronger in the service industry. In this study, we divide the sample into three categories: agricultural industry, manufacturing industry, and service industry, and regression is carried out based on Equation (2). As shown in Table 13, the regression results are presented with and without control variables. For primary and secondary industries, the results in columns (1)–(4) demonstrate that the effects of HSR construction on the number of firms' cross-regional affiliates is positive but insignificant. According to the results in columns (5) and (6), HSR construction has a positive impact on the number of cross-regional affiliates for the tertiary industry with a coefficient of 0.015, which is statistically significant at 10%. In addition, we also interact with the industry dummy variable[8] in Equation (2) with the HSR construction variable. The results are presented in columns (7) and (8), the coefficients of the interaction term are both negative and statistically significant at 1% level, which means that the HSR effects on firms' CRD are strongest in the service sector and weakest in the agricultural sector. Therefore, the results confirm the results that HSR construction affects firms' CRD through the communication cost channel, which is consistent with expectations.

7. Conclusion

By exploiting the staggered HSR construction in China as a quasi-natural experiment, we analyze the effect of internal communication costs, specifically travel times, on firms' cross-regional development. HSR represents a novel transportation technology that facilitates convenient and efficient physical connectivity between cities. It not only enables easier access for passenger transportation but also fosters enhanced communication and interaction within firms. Consequently, HSR construction has the potential to spur the establishment of new cross-regional businesses. Understanding the extent to which the reduction in passenger travel costs promotes firms' cross-regional development is crucial for comprehending the overall economic implications of HSR construction and other significant travel infrastructure projects.

As a result of HSR construction, firms respond at an extensive margin by establishing more cross-regional affiliates, and the distance between headquarters and affiliates increases. According to the results at the city level, HSR construction increases the number of cross-regional affiliates by 8.6%. After conducting several tests to alleviate endogeneity concerns, the results remain robust. And our results imply that firms with less cross-regional operation experience and smaller sizes are more affected. Furthermore, the estimated effects are greatest in the service industries, where the information being transmitted is arguably softer. Finally, we provide auxiliary evidence suggesting that communication costs may play a remarkable role in determining the location of affiliates. Specifically, we demonstrate that the HSR effect is greater for firms

[8] We define this variable equals to 1 if firms belong to the agricultural industry, 2 if firms belong to the manufacturing industry, and 3 if firms belong to the service industry.

whose headquarters are located in areas with a lower Internet penetration rate and industries with higher-performance heterogeneity.

The results of our study have several implications. First, HSR construction effectively reduces communication costs within firms and enhances information flows between distant places. Thus, HSR construction enables firms to operate their cross-regional affiliates at lower costs compared to the absence of HSR construction, thereby facilitating the establishment of new affiliates in remote areas. In general, HSR connection is beneficial for firms' geographic expansion through information channels. Second, transportation plays an essential role in economic activities. Yet, relatively little attention has been paid to the implications of travel costs for economic integration. Our study focuses on how the reduction in travel costs promotes firms' geographic expansion. As a result, we offer a novel perspective on the economic integration and development brought about by HSR construction. In this regard, policymakers should further improve the HSR connections. Third, the current study also helps us understand how firms' geographic footprints are shaped. Such geographic footprints have potentially important implications on the propagation of economic shocks across regions, and even foretelling how future local specific shocks would propagate to other regions.

Further research could examine how geographical proximity affects the economic performance (such as total factor productivity) of newly established cross-regional affiliates or their organizational structures in the future. Given the limitations of the current dataset, we can only examine how the reduction of communication costs by HSR construction affects firms' cross-regional development at an extensive margin. In the future, these research lines may deepen our understanding of firms' spatial organization and their evolution in response to changes in travel time with more sophisticated datasets, such as identifying the performance data of affiliates and the employment data of affiliates and firms' headquarters.

Ethical approval

This is an original article that did not use other information that requires ethical approval.

Consent to participate

All authors participated in this article. The authors are listed in alphabetical order and contributed equally to this manuscript.

Consent to publish

All authors have given consent to the publication of this article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

We sincerely thank the handling Editor, Prof. Sushanta Mallick, and the three anonymous referees for their insightful suggestions and constructive comments that have significantly improved the paper.

References

Acemoglu, D., Aghion, P., Lelarge, C., Van Reenen, J., Zilibotti, F., 2007. Technology, information, and the decentralization of the firm. Q. J. Econ. 122 (4), 1759–1799. https://doi.org/10.1162/qjec.2007.122.4.1759.

Aghion, P., Tirole, J., 1997. Formal and real authority in organizations. J. Polit. Econ. 105 (1), 1–29. https://doi.org/10.1086/262063.

Ahlfeldt, G.M., Feddersen, A., 2018. From periphery to core: measuring agglomeration effects using high-speed rail. J. Econ. Geogr. 18 (2), 355–390. https://doi.org/10.1093/jeg/lbx005.

Alonso, R., Dessein, W., Matouschek, N., 2008. When does coordination require centralization? Am. Econ. Rev. 98 (1), 145–179. https://doi.org/10.1257/aer.98.1.145.

Baum-Snow, N., Brandt, L., Henderson, J.V., Turner, M.A., Zhang, Q., 2017. Roads, railroads, and decentralization of Chinese cities. Rev. Econ. Stat. 99 (3), 435–448. https://doi.org/10.1162/REST_a_00660.

Beck, T., Levine, R., Levkov, A., 2010. Big bad banks? The winners and losers from bank deregulation in the United States. J. Finance 65 (5), 1637–1667. https://doi.org/10.1111/j.1540-6261.2010.01589.x.

Bernard, A.B., Moxnes, A., Saito, Y.U., 2018. Production networks, geography, and firm performance. J. Polit. Econ. 127 (2), 639–688. https://doi.org/10.1086/700764.

Bloom, N., Garicano, L., Sadun, R., Van Reenen, J., 2014. The distinct effects of information technology and communication technology on firm organization. Manag. Sci. 60 (12), 2859–2885. https://doi.org/10.1287/mnsc.2014.2013.

Callaway, B., Sant'Anna, P.H.C., 2021. Difference-in-differences with multiple time periods. J. Econom. 225 (2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001.

Cao, D., Hyatt, H.R., Mukoyama, T., Sager, E., 2019. Firm Growth through New Establishments. https://doi.org/10.2139/ssrn.3361451. ***working paper***.

Charnoz, P., Lelarge, C., Trevien, C., 2018. Communication costs and the internal organisation of multi-plant businesses: evidence from the impact of the French high-speed rail. Econ. J. 128 (610), 949–994. https://doi.org/10.1111/ecoj.12592.

Cheng, Q., Du, F., Wang, X., Wang, Y., 2016. Seeing is believing: analysts' corporate site visits. Rev. Account. Stud. 21 (4), 1–42. https://doi.org/10.1007/s11142-016-9368-9.

Clarke, J.E., Tamaschke, R., Liesch, P.W., 2013. International experience in international business research: a conceptualization and exploration of key themes. Int. J. Manag. Rev. 15 (3), 265–279. https://doi.org/10.1111/j.1468-2370.2012.00338.x.

De Chaisemartin, C., D'Haultfoeuille, X., 2020. Two-way fixed effects estimators with heterogeneous treatment effects. Am. Econ. Rev. 110 (9), 2964–2996. https://doi.org/10.1257/aer.20181169.

Devos, E., Rahman, S., 2014. Location and lease intensity. J. Corp. Finance 29, 20–36. https://doi.org/10.1016/j.jcorpfin.2014.08.005.

Donaldson, D., Hornbeck, R., 2016. Railroads and american economic growth: a "Market access" approach. Q. J. Econ. 131 (2), 799–858. https://doi.org/10.1093/qje/qjw002.

Dong, X., Zheng, S., Kahn, M.E., 2020. The role of transportation speed in facilitating high skilled teamwork across cities. J. Urban Econ. 115, 103212 https://doi.org/10.1016/j.jue.2019.103212.

Duranton, G., Turner, M.A., 2012. Urban growth and transportation. Rev. Econ. Stud. 79 (4), 1407–1440. https://doi.org/10.1093/restud/rds010.

Gao, Y., Zheng, J., 2020. The impact of high-speed rail on innovation: an empirical test of the companion innovation hypothesis of transportation improvement with China's manufacturing firms. World Dev. 127, 104838 https://doi.org/10.1016/j.worlddev.2019.104838.

Gervais, A., 2018. Multiregional firms and region switching in the U.S. manufacturing sector. Econ. Inq. 56 (2), 955–982. https://doi.org/10.1111/ecin.12532.

Giroud, X., 2013. Proximity and investment: evidence from plant-level data. Q. J. Econ. 128 (2), 861–915. https://doi.org/10.1093/qje/qjs073.

Gokan, T., Kichko, S., Thisse, J.-F., 2019. How do trade and communication costs shape the spatial organization of firms? J. Urban Econ. 113, 103191 https://doi.org/10.1016/j.jue.2019.103191.

Goodman-Bacon, A., 2021. Difference-in-differences with variation in treatment timing. J. Econom. 225 (2), 254–277. https://doi.org/10.1016/j.jeconom.2021.03.014.

Gordon, R.H., Bovenberg, A.L., 1996. Why is capital so immobile internationally?: possible explanations and implications for capital income taxation. Am. Econ. Rev. 86 (5), 1057–1075. https://www.jstor.org/stable/2118278.

Gumpert, A., Steimer, H., Antoni, M., 2022. Firm organization with multiple establishments. Q. J. Econ. 137 (2), 1091–1138. https://doi.org/10.1093/qje/qjab049.

Heimeriks, K.H., Duysters, G., 2007. Alliance capability as a mediator between experience and alliance performance: an empirical investigation into the alliance capability development process. J. Manag. Stud. 44 (1), 25–49. https://doi.org/10.1111/j.1467-6486.2006.00639.x.

Helpman, E., Melitz, M.J., Yeaple, S.R., 2004. Export versus FDI with heterogeneous firms. Am. Econ. Rev. 94 (1), 300–316. https://doi.org/10.1257/000282804322970814.

Henisz, W.J., 2003. The power of the Buckley and Casson thesis: the ability to manage institutional idiosyncrasies. J. Int. Bus. Stud. 34 (2), 173–184. https://doi.org/10.1057/palgrave.jibs.8400015.

Higgins, R.C., Schall, L.D., 1975. Corporate bankruptcy and conglomerate merger. J. Finance 30 (1), 93–113. https://doi.org/10.1111/j.1540-6261.1975.tb03162.x.

Hsieh, C.-T., Rossi-Hansberg, E., 2023. The industrial revolution in services. J. Polit. Econ. Macroecon. 1 (1), 3–42. https://doi.org/10.1086/723009.

Huang, Z., Li, L., Ma, G., Xu, L.C., 2017. Hayek, local information, and commanding heights: decentralizing state-owned enterprises in China. Am. Econ. Rev. 107 (8), 2455–2478. https://doi.org/10.1257/aer.20150592.

John, K., Knyazeva, A., Knyazeva, D., 2011. Does geography matter? Firm location and corporate payout policy. J. Financ. Econ. 101 (3), 533–551. https://doi.org/10.1016/j.jfineco.2011.03.014.

Kalnins, A., Lafontaine, F., 2013. Too far away? The effect of distance to headquarters on business establishment performance. Am. Econ. J. Microecon. 5 (3), 157–179. https://doi.org/10.1257/mic.5.3.157.

Ke, X., Chen, H., Hong, Y., Hsiao, C., 2017. Do China's high-speed-rail projects promote local economy?—new evidence from a panel data approach. China Econ. Rev. 44, 203–226. https://doi.org/10.1016/j.chieco.2017.02.008.

Kong, D., Liu, L., Yang, Z., 2021. High-speed rails and rural-urban migrants' wages. Econ. Modell. 94, 1030–1042. https://doi.org/10.1016/j.econmod.2020.02.043.

Krugman, P., 1980. Scale economies, product differentiation, and the pattern of trade. Am. Econ. Rev. 70 (5), 950–959. https://www.jstor.org/stable/1805774.

Li, P., Lu, Y., Wang, J., 2016. Does flattening government improve economic performance? Evidence from China. J. Dev. Econ. 123, 18–37. https://doi.org/10.1016/j.jdeveco.2016.07.002.

Lin, Y., 2017. Travel costs and urban specialization patterns: evidence from China's high speed railway system. J. Urban Econ. 98, 98–123. https://doi.org/10.1016/j.jue.2016.11.002.

Lin, Y., Qin, Y., Sulaeman, J., Yan, J., Zhang, J., 2022. Expanding footprints: the impact of passenger transportation on corporate locations. Rev. Finance. https://doi.org/10.1093/rof/rfac049.

Luo, Y., Tung, R.L., 2007. International expansion of emerging market enterprises: a springboard perspective. J. Int. Bus. Stud. 38 (4), 481–498. https://doi.org/10.1057/palgrave.jibs.8400275.

McDonald, M.L., Westphal, J.D., Graebner, M.E., 2008. What do they? The effects of outside director acquisition experience on firm acquisition performance. Strat. Manag. J. 29 (11), 1155–1177. https://doi.org/10.1002/smj.704.

Mueller, D.C., 1977. The effects of conglomerate mergers: a survey of the empirical evidence. J. Bank. Finance 1 (4), 315–347. https://doi.org/10.1016/0378-4266(77)90008-5.

Petersen, M.A., Rajan, R.G., 2002. Does distance still matter? The information revolution in small business lending. J. Finance 57 (6), 2533–2570. https://doi.org/10.1111/1540-6261.00505.

Shang, D., Yuan, D., Li, D., Lin, Q., 2023. Independent directors' geographic distance, high-speed railway, and corporate cash holdings. Econ. Modell. 121, 106213 https://doi.org/10.1016/j.econmod.2023.106213.

Shi, X., Xi, T., Zhang, X., Zhang, Y., 2021. "Moving Umbrella": bureaucratic transfers and the comovement of interregional investments in China. J. Dev. Econ. 153, 102717 https://doi.org/10.1016/j.jdeveco.2021.102717.

Sun, L., Abraham, S., 2021. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. J. Econom. 225 (2), 175–199. https://doi.org/10.1016/j.jeconom.2020.09.006.

Tallman, S., Li, J.T., 1996. Effects of international diversity and product diversity on the performance of multinational firms. Acad. Manag. J. 39 (1), 179–196. https://doi.org/10.2307/256635.

Turner, G.D.M.M.A., 2014. Roads and trade: evidence from the US. Rev. Econ. Stud. 81 (2), 681–724. https://doi.org/10.1093/restud/rdt039.

Wan, L., Wan, Q., 2022. High-speed railway and the intercity transmission of epidemics: evidence from COVID-19 in China. Econ. Modell. 114, 105934 https://doi.org/10.1016/j.econmod.2022.105934.

Wu, Y., Lee, C.C., Lee, C.C., Peng, D., 2022. Geographic proximity and corporate investment efficiency: evidence from high-speed rail construction in China. J. Bank. Finance. https://doi.org/10.1016/j.jbankfin.2022.106510.

Zhang, A., Liu, L., Liu, G., 2020a. High-speed rail, tourist mobility, and firm value. Econ. Modell. 90, 108–116. https://doi.org/10.1016/j.econmod.2020.05.004.

Zhang, X., Wu, W.-L., Zhou, Z., Yuan, L., 2020b. Geographic proximity, information flows and corporate innovation: evidence from the high-speed rail construction in China. Pac. Basin Finance J. 61, 101342 https://doi.org/10.1016/j.pacfin.2020.101342.

Zheng, S., Kahn, M.E., 2013. China's bullet trains facilitate market integration and mitigate the cost of megacity growth. Proc. Natl. Acad. Sci. USA 110 (14), E1248–E1253. https://doi.org/10.1073/pnas.1209247110.