Research Article

Does High-Speed Rail Promote Agglomeration in China?

Shanlang Lin and Tian Gan

School of Economy and Management, Tongji University, Shanghai 200092, China

Correspondence should be addressed to Tian Gan; 1910028@tongji.edu.cn

Received 1 May 2022; Accepted 6 July 2022; Published 25 July 2022

Academic Editor: Kuruva Lakshmanna

Copyright © 2022 Shanlang Lin and Tian Gan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper constructs the agglomeration model of crossregional labor flow and market accessibility based on the new economic geography model. Using panel data in 2000-2013, this paper examines the impact of high-speed railways on industrial agglomeration. The results show that high-speed railway significantly promotes the spatial agglomeration of industrial firms. The high-speed railway promotes industrial agglomeration by promoting crossregional mobility of industrial labor and improving market accessibility. We also construct the instrument variable based on HSR construction cost and apply the TSLS model to solve the potential endogenous problems. At the same time, PSM results also get the net effect caused by the repair of a high-speed railway, which verifies the robustness of the results.

1. Introduction

Since the first high-speed rail opened in 2008, China’s high-speed rail development has grown rapidly (see Figure 1). By 2018, China is the leading nation in the world, with 29,000 kilometers of HSR operating mileage [1]. As a vital transportation infrastructure, high-speed rail can play a significant role in economic development and industrial agglomeration [2, 3]. Based on the evidence from developed countries, there is no consensus in the literature on HSR’s impact on industrial agglomeration. High-speed railways play a role in accelerating the transfer of resources from surrounding towns to regional centers and increasing the agglomeration effect of centers over their surrounding counterparts. However, some scholars found that the high-speed railway brings job opportunities and economic growth to the less-developed areas, thus leading to the decentralization of regional economics [4–7]. These considerations raise questions: what is the impact of China’s HSR system on industrial agglomeration? What mechanisms do high-speed railways rely upon to influence agglomeration? Limited studies have addressed these questions. Mechanisms such as reducing transportation costs are insufficient to explain why HSR has different effects on agglomeration depending on the region.

Based on the new economic geography model, we construct a decision model of industrial agglomeration for crossregional labor mobility and market accessibility [8–10]. Utilizing microdata of industrial firm coordinates, this paper analyzes the impact of high-speed railways on industrial agglomeration. The empirical results find that high-speed railway significantly promotes spatial agglomeration by 13-32%. Firms in HSR cities are 1.74-6.30 km closer to the city centers. In the heterogeneous analysis, we find that high-speed railways have a greater impact on industrial agglomeration in regions with greater economic development and a higher ratio of manufacturing. Moreover, we decompose HSR effects into an extensive margin effect from new entrants and exits and an intensive margin effect from incumbents. Specifically, we find that high-speed railways tend to attract incumbents and new entrants to be located near city centers.

According to our empirical analysis, the opening of the HSR may be endogenous. Stations are assigned to cities based upon the unobserved factors associated with economic activity. In order to deal with the conventional endogeneity problem, we use DEM digital elevation data to calculate the HSR construction costs as the instrument variables. The construction costs of high-speed railway lines play an important role in determining their route. Moreover, the
construction cost does not influence industrial agglomeration via other channels. In addition, we offer comprehensive robustness checks to deal with possible confounders which may have slightly different effects. Our study investigates the dynamic relationship between HSR construction and industrial agglomeration. The parallel-trend test shows no effect of agglomeration changes before the HSR station’s construction. Moreover, this paper uses the propensity score matching method (PSM) to determine the net impact of opening high-speed railways on industrial agglomeration, thus demonstrating the validity of its findings. In addition, we show that results are unlikely to be driven by random factors and robust to the alternative measurements of agglomeration and other contemporaneous economic shocks (such as the 2008 financial crisis).

What is the underlying mechanism of HSR’s impact on industrial agglomerations? Firstly, since migration remains very costly between regions, China has low population mobility rates [11, 12]. The construction of the high-speed rail network contributed to travel time savings and reduce transportation costs, thereby promoting labor mobility [13]. Due to the convenience of mobility, HSR’s construction facilitates crossregional face-to-face communication and field investigations [14]. In order to better understand the mechanism of labor mobility at work, we examine whether the opening of HSR has a significant impact on labor mobility using city-level data. We find that the HSR construction significantly facilitates labor mobility by around 7%-29%, thereby promoting industrial agglomeration. As the HSR system facilitates better access to regional markets, cities can capture larger markets from nearby cities. Having improved access to local markets, HSR can contribute positively to the local economy. Therefore, we estimate the dynamic parameter for HSR connections in several cities based on the “market potential” measurement of Baum-Snow et al. [15]. In a mechanism analysis, we examine how the construction of HSR affects industrial agglomeration through market potential.

An increasing body of empirical literature focuses on the evaluation of transportation infrastructures. In recent studies, scholars have examined the economic effects of transportation systems [2, 16–18], skill, and innovation (Michaels, 2008; [1, 19]), as well as gains from trade [20]. As a major transport infrastructure, HSR networks can narrow the time distance between regions, decrease the cost of transportation, and increase accessibility [21, 22]. Based on the home market effect, the decrease in trade costs may promote the agglomeration of production resources towards the core city and strengthen the existing core-periphery structure [8, 18]. Compared to previous works, this paper develops a multiregional model of industrial agglomeration and considers high-speed railway costs as information costs for labor rather than transportation costs for freight [14]. Benefiting from lower commuting costs, the expansion of the transportation network motivates tacit knowledge of face-to-face communication and facilitates technological information sharing [23]. The empirical results also demonstrate that high-speed rail improves regional industrial agglomeration through labor mobility and market accessibility. The spatial employment effect of the high-speed railway was analyzed in a study by Liu et al. [24]. The study implemented the difference-in-differences (DID) model and
showed that HSR enhanced spatial employment agglomeration. Similar to this study, the impact of HSR operation on resident income from 2008 to 2018 using propensity score matching-based DID methods. The results revealed that rural and urban residents of Central China were impacted by HSR [25].

The unique contributions of the paper include the following:

(i) Impact analysis of high-speed railways on industrial agglomeration

(ii) Construction of an instrumental variable based on HSR construction cost

(iii) Application of the TSLS model to relieve the potential endogenous problem

(iv) Use of propensity score matching (PSM) method to analyze the impact of HSR on industrial agglomeration

This paper is arranged as follows. Section 2 discusses the theoretical model of the potential effects of HSR construction on industrial agglomeration. Section 3 introduces the variable setting and discusses empirical analysis methods. The empirical results of the model are presented in Section 4. Lastly, Section 5 concludes the paper and offers directions for future research.

2. Theoretical Research

This paper presents a multiregional model of the high-speed railway that follows Redding and Turner [10]. This model shows the impact of the high-speed railway on regional industrial agglomeration. It introduces high-speed railways to the theoretical foundations and structures our review of the empirical evidence below.

2.1. Consumer Behaviors. Suppose the economic system consists of a series of cities with industrial and agricultural sectors, \( L \) denotes a mass of representative labors in the industrial sector, and they can travel freely across regions. There is imperfect competition in the industrial sector, increasing returns to scale. The agricultural sector produces homogeneous goods with constant marginal returns. The effective labor supply in each region depends on its population \( L_i \) and the level of transportation technology serving the workers \( (b_i + r_i) \). \( b_i \) represents the transportation technology without high-speed railways, and \( r_i \) represents the improvement of transportation level in region \( i \) due to the construction of high-speed railways. The construction of high-speed railway has reduced the time cost of labor supply, especially the cost of crosscity meetings and remote research activities has been greatly reduced. For each region \( i, (b_i + K_i) \) unit of labor is available for production, and the remaining \( [1- (b_i + K_i)] \) is the transportation cost of labor. It depends on the principle that higher population density increases congestion costs [10].

It is assumed that the total income of region \( n \) is equal to the total expenditure, which contains a consumption index of tradeable industrial products \( C_{ni} \), and the consumption of nontradable agricultural products \( H_{ni} \). The tradeable consumption index is in a form of the constant elasticity of substitution (CES). Consumers use a constant proportion of income \( \mu \) to buy tradable goods. The utility function is assumed to be Cobb-Douglas form and listed as follows:

\[
U_n = C_{n}^\sigma H_{n}^{1-\mu},
\]

(1)

where \( \sigma \) refers to the elasticity of substitution among variables, and this paper assumes that the variables are substitutable \( (\sigma > 1) \); \( \mu \) represents the proportion of industrial products in total consumption, where \( 0 < \mu < 1 \); \( c_{ni} \) refers to the products produced in region \( i \) in the amount of consumption in the region \( n \). \( M_i \) units of products produced in region \( i \) are equivalent to the amount of consumption \( c_{ni} \) in region \( n \). The budget constraint of consumers is listed as follows.

\[
P_{mn}c_{ni} + P_{hn}H_{ni} = I_n,
\]

(2)

where \( P_{mn} \) and \( P_{hn} \) refer to the price of industrial tradable products and the price of agricultural products, respectively, and \( I_n \) refers to the income of consumer \( n \).

We assume that products are subject to iceberg trade costs. Therefore, trade costs are affected by transportation costs and transaction information costs. One unit of product produced in region \( i \) and shipped to region \( n \), needs to deliver the quantity \( d_{ni} - \text{info}_n \) of products to region \( n \), \( d_{ni} - \text{info}_n > 1 \). \( [(d_{ni} - \text{info}_n) - 1] \) is consumed as transportation cost, \( d_{ni} \) represents transportation cost, and \( \text{info}_n \) represents transaction information cost. The construction of high-speed railways can greatly facilitate crossregional face-to-face communication and field investigations ([14, 26], Gundluru, Nagaraja, et al. 2022). Due to effectively connected in a transportation network, tacit knowledge of face-to-face communication becomes more effective, thereby reducing the cost caused by information asymmetry. If high-speed railways connect region \( n \) with region \( i \), \( \text{info}_{ni} > 0 \); otherwise, \( \text{info}_{ni} = 0 \). Therefore, the price \( P_{ni} \) of the product produced in region \( i \) shipped to region \( n \) is

\[
P_{ni} = p_{i}(d_{ni} - \text{info}_n),
\]

(3)

where \( p_{i} \) denotes the free on board price produced in region \( i \). We have used the fact that the varieties of products produced in region \( i \) face the same urban demand elasticity and the same equilibrium price for consumers in region \( n \). The price index \( P_{ni} \) derived from the consumption index \( C_{n} \) of tradable products is

\[
P_n = \left[ \sum_{i \in N} M_{i} p_{i}^{1-\sigma} \right]^{1/1-\sigma} = \left\{ \sum_{i \in N} M_{i} [p_{i}(d_{ni} - \text{info}_n)]^{1-\sigma} \right\}^{1/1-\sigma},
\]

(4)

Since each region \( n \) has a constant consumption
proportion and an inelastic supply of nontradable products, the equilibrium price $y_n$ of nontradable products depends entirely on the consumption proportion $1-\mu$, the total income $v_nL_n$, and the nontradable products $H_n$:

$$y_n = \frac{(1-\mu)v_nL_n}{H_n}.$$  \hfill (5)

On this basis, applying Shepard’s lemma to the price index $P_{ni}$ of tradable products, the equilibrium demand of region $n$ for tradable products made in region $i$ is derived as

$$x_{ni} = p_{i}^{\sigma}(d_{ni} - \text{info}_{ni})^{1-\sigma}(\mu v_{n}L_{n})(P_{ni})^{\sigma-1},$$  \hfill (6)

where $v_n$ denotes the per capita income of city $n$, and $v_nL_n$ denotes the total income of consumers in region $n$. The consumer utility function follows the Cobb-Douglas form, which spends a fixed expenditure share of income on tradable products ($\mu$).

It also indicates that tradable industrial products have lower prices in cities with high-speed rail; so, the residents of high-speed rail cities can obtain this commodity at a lower price. In addition, with equal nominal wages, lower prices also mean higher real incomes for consumers in HSR-connected regions.

2.2. Firm Behaviors. For firms in region $i$, production and operating costs consist of fixed costs in terms of labor of producing tradable varieties ($F > 0$) and a constant variable cost that depends on the productivity in region $i$ ($A_i$). Assume that all kinds produced within a region have the same fixed and variable costs. According to the principle of economies of scale, consumer preferences for different products, and firms will opt to manufacture products that are different from those of other firms [27, 28]. Following Redding and Turner [10], the labor input $l_i$ required to make $x_i$ unit of products in region $i$ is

$$l_i = F + \frac{x_i}{A_i},$$  \hfill (7)

where $F$ denotes fixed cost for firms, and $A_i$ denotes the productivity of region $i$, which varies across regions to account for heterogeneity in production fundamentals. To produce $x_i$ units of products, firms demand $l_i$ labor force and pay the wages ($w$) to them. We also define the price of selling the products as $p_{ni}$. Then, we find the profit function ($\pi$) for firms equals:

$$\pi = p_{ni}x_i - w\left(F + \frac{x_i}{A_i}\right).$$  \hfill (8)

The demand function of the products in region $i$ can be written as $x_i = kp_i^{-\delta}$. Then, we construct the Lagrange function by deriving $p_i$ and $x_i$ and get the profit maximizing price for firm. The profit maximization implies that equilibrium prices are a constant over marginal cost:

$$p_{ni} = \left(\frac{\sigma}{\sigma - 1}\right)\left(\frac{d_{ni} - \text{info}_{ni}}{A_i}\right).$$  \hfill (9)

We combine profit maximization and production function and get the equilibrium output of each tradable variety that is

$$\bar{x} = x_i = \sum_{i} x_{ni} = A_i F (\sigma - 1).$$  \hfill (10)

Labor market clearing for each region signifies that labor demand equals the effective labor supply in a region determined by labor mobility. Using the constant equilibrium output of each product together with the production technology, we can express the labor market clearing condition with the following equation:

$$b_i L_i = M_i l_i = M_i F \sigma_i,$$  \hfill (11)

where $T_i$ each product’s constant equilibrium labor demand. This relationship establishes the quantity of tradable products produced in each region as a function of labor, transportation technology (such as high-speed trains), and the economic system model’s fundamental parameters.

2.3. The Impact of HSR on Agglomeration. This paper assumes that labor can migrate freely across regions so that they can arbitrage away in real income differences. The real income of each region $n$ depends on tradable products price index ($P_n$), per capita income ($v_n$), and nontradable products’ price ($y_n$). Thence, for labor mobility reaches general equilibrium, real incomes are equal across regions, which implies

$$V_n = \frac{v_n}{(P_n)^\mu(y_n)^{1-\mu}} = \bar{V}.$$  \hfill (12)

The region has a share of workers $a L_i = L_i/L^w$, where $L^w$ represents the total number of workers. Then, the labor mobility equation is

$$a L_i = (V_n - \bar{V})a L_i (1 - a L_i).$$  \hfill (13)

Since we assume that firms have a constant production technology and free access to a market, they will manufacture the same scale in equilibrium. Therefore, the proportion of fixed factor input equals variable factor input of each firm; thus, it can be considered that labor share owned by region ($a L_i$) equals the share of firms owned by region $n$ ($a L_i$):

$$a L_i = a L_i = L_i / L^w.$$  \hfill (14)

The price index depends on the varieties of the tradable product ($M_i$), the free on board price of the product produced in region $i$ ($p_i$), and the transportation and information cost of the product from region $i$ to $n$ ($d_{ni} - \text{info}_{ni}$).
Following Redding and Turner [10], it defines the consumers’ access to tradeable products using the accessibility index \( cma_n \):
\[
cma_n = \sum_{i \in N} M_i \left[ p_i (d_{mi} - info_m) \right]^{1-\sigma}.
\] (15)

Therefore, the price index \( P_n \) can be expressed as
\[
P_n = [cma_n]^{1/1-\sigma}.
\] (16)

Substituting for \( v_n, P_n, \) and \( y_n \), the labor condition equation can be rewritten into the following form:
\[
L_n = \mu^{\mu-1} V^{\mu-1} (b_i + K_n)^{\mu-1} \delta_{n} w_n^{\mu-1} (cma_n)^{\mu(1-\mu)(\sigma-1)}.
\] (17)

The labor force \( L_n \) is expressed by a function of the level of transportation technology \((b_i + K_n)\) and the consumer market accessibility \(cma_n\). The following labor condition equation \( L_n \) will take the \((b_i + K_n)\) derivation of the transportation skill level:
\[
\frac{\partial L_n}{\partial (b_i + K_n)} = \mu^{\mu-1} V^{\mu-1} \left( b_i + K_n \right)^{\mu-1} \delta_{n} w_n^{\mu-1} (cma_n)^{\mu(1-\mu)(\sigma-1)}.
\] (18)

As \( (\partial L_n) / (\partial (b_i + K_n)) > 0 \), so equilibrium workforce \( L_n \) is increasing in the transportation technology. The improvement of transportation infrastructure is conducive to the agglomeration of industrial labor. High-speed railway cities have shortened the time and cost of communication between regions, thereby promoting population agglomeration to high-speed railway cities.

**Hypothesis 1.** The opening of high-speed railway has a positive effect on regional industrial agglomeration.

We assume that each firm produces only one type of industrial product, and the share of firms in a region represents the regional industrial agglomeration:
\[
al_c = al_c = \frac{\mu^{\mu-1} V^{\mu-1} (b_i + K_n)^{\mu-1} \delta_{n} w_n^{\mu-1} (cma_n)^{\mu(1-\mu)(\sigma-1)} }{ L^\mu }.
\] (19)

Next, we regard the level of transportation \((b_i + K_n)\) as a whole and take the derivative of \( s_c \) to get
\[
\frac{\partial al_c}{\partial (b_i + K_n)} = \mu^{\mu-1} V^{\mu-1} \left( b_i + K_n \right)^{\mu-1} \delta_{n} w_n^{\mu-1} (cma_n)^{\mu(1-\mu)(\sigma-1)}.
\] (20)

Since \( 0 < \mu < 1 \), then \( (\partial al_c) / (\partial (b_i + K_n)) > 0 \). It indicates that the level of transportation infrastructure has a positive impact industrial agglomeration. We obtain by derivation of high-speed railway \((K_n)\) by the industrial agglomeration (al_c):
\[
\frac{\partial al_c}{\partial K_n} = \frac{\mu^{\mu-1} V^{\mu-1} (b_i + K_n)^{\mu-1} \delta_{n} w_n^{\mu-1} (cma_n)^{\mu(1-\mu)(\sigma-1)} }{ (1 - \mu) L^\mu }.
\] (21)

Similarly, \( 0 < \mu < 1 \), and then \( (\partial al_c) / (\partial K_n) > 0 \). High-speed railways promote the industrial agglomeration through the flow of industrial labor, demonstrating that changes in employment brought about by high-speed railways would reshape the layout of industry in cities. The market potential is further extended as the workers increase, and customers have a more differentiated need for products, further attracting more workers to agglomeration. This cycle’s cumulative effect eventually results in an unequal allocation of enterprises and labor. As a result, it impacts both inflow and outflow labor and commodities markets. In addition, industries will relocate to cities along the high-speed railway route, generating industrial agglomerations.

Substituting \( cma_n \) into the labor condition:
\[
L_n = \mu^{\mu-1} V^{\mu-1} (b_i + K_n)^{\mu-1} \delta_{n} w_n^{\mu-1} \left\{ \sum_{i \in N} M_i (p_i (d_{mi} + info_m)) \right\}^{\mu(1-\mu)(\sigma-1)}
\] (22)

Therefore, the industrial agglomeration can be expressed as
\[
al_c = al_c = .
\] (23)

Based on the above mechanism research, this paper proposes the following hypotheses:

**Hypothesis 2.** High-speed railway changes the agglomeration of industrial firms between regions by changing the flow of industrial labor between regions.

In order to explore the impact of high-speed railway on industrial agglomeration, we obtain by derivation of the industrial agglomeration degree \( al_c \) to \( K_n \):
\[
\frac{\partial al_c}{\partial K_n} = \frac{\mu^{\mu-1} V^{\mu-1} (b_i + K_n)^{\mu-1} \delta_{n} w_n^{\mu-1} (cma_n)^{\mu(1-\mu)(\sigma-1)} }{ (1 - \mu) L^\mu }.
\] (24)

where
\[
\sigma = \frac{\mu^{\mu-1} V^{\mu-1} (b_i + K_n)^{\mu-1} \delta_{n} w_n^{\mu-1} (cma_n)^{\mu(1-\mu)(\sigma-1)} }{ (1 - \mu) L^\mu }.
\] (25)

Since \( \sigma > 1 \), then \( 1 - \sigma < 0 \). Thus \( (\partial al_c) / (\partial K_n) > 0 \). From the theoretical model, it is proved that the improvement of the service intensity \( K_n \) of the high-speed railway can promote industrial agglomeration. As high-speed railway lines...
are built, travel times between cities will shorten, improving the accessibility between cities along the line, affecting the market potential. An increase in market potential will result in a greater flow of various production factors between cities, further promoting industrial agglomeration. Ahlfeldt and Feddersen [29] examined the economic development along the high-speed railway between Cologne and Frankfurt, Germany. They find that high-speed rail contributes significantly to urban accessibility, urban economic potential, and improved urban productivity, which further affects the degree of urban economic agglomeration.

Hypothesis 3. HSR can promote the spatial agglomeration of industrial firms by improving the accessibility of the consumer market.

3. Data and Empirical Model

3.1. Sample Description

3.1.1. Industrial Agglomeration. We construct spatial agglomeration index according to the following steps. The first step is to connect the Baidu API geographic database (Baidu Map)(Baidu API geographic database have a similar geocoding function of Google Map) with the location information (such as the province, city, street, and address) of each manufacturing firm from 2000 to 2013. Then, the latitude and longitude coordinates are matched with each firm. In the second step, we use python to correct the longitude and latitude and obtain the precise coordinates. In the third step, this paper calculates the coefficient of variation of longitude ($CV_{\text{longitude}}$) and the coefficient of variation of latitude ($CV_{\text{latitude}}$) of firms in each city, respectively, and uses the above two coefficients of variation to construct an index of spatial agglomeration degree.

\[
\text{spatial agglomeration} = - \ln \left( CV_{\text{longitude}} \times CV_{\text{latitude}} \right). \tag{26}
\]

The reason why we choose spatial agglomeration is that the coefficient of variation reflects the discrete degree of the spatial distribution of firms and eliminates the influence of measurement scale and dimension caused by China’s vast territory and large differences in longitude and latitude. The larger the spatial agglomeration index is, the higher degree of spatial agglomeration.

The second measurement of industrial agglomeration is also based on the coordinates of manufacturing firms. Then, the distance to city centers for each firm is calculated based on the coordinates of firms. We calculate each city’s average distance to the city center to measure agglomeration.

3.1.2. High-Speed Railway. To construct the national high-speed railway network, we manually collected detailed information on each HSR line (the G-, C-, and D-series lines) from China Railway from 2007 to 2013, published by China Railway Publishing House. The book contains detailed information on each HSR line’s stations, such as the location and HSR opening time. For further study, geocode HSR stations at the city level were computed.

3.1.3. Control Variables. Population density (pop_density) is as follows: compared to smaller cities, large cities with large population density promote the concentration of economic activity [30, 31]. Therefore, we introduce population density into the model as a control variable.

Government spending (government) is as follows: the increase in local government’s capital expenditure plays an important part in industrial development and promotes agglomeration [32]. We take the ratio of government expenditure to GDP to measure government spending.

Industrial structure (ind_ratio) is as follows: regional industrial structure positively affects localization agglomeration [33–35]. We choose the proportion of the manufacturing industry as the proxy variable of industrial structure.

Foreign direct investment (FDI) is as follows: FDI can improve agglomeration by learning technology from foreign markets and increasing market competition [36]. We select regional foreign direct investment as the measurement of FDI.

The city-level data comes from the China City Statistical Yearbook and China Regional Statistical Yearbook. The descriptive statistics of the variables in this paper are shown in Table 1.

3.2. Empirical Model. To examine whether an HSR station is conducive to counties’ economic development, it is reasonable to compare the counties GDPs before and after the HSR station is set up. The baseline model is listed as follows:

\[
Y_{it} = \beta_0 + \beta_1 \text{HSR city}_i \times \text{post}_t + \beta_2 \text{Control}_{it} + \mu_i + \mu_t + \epsilon_{it}. \tag{27}
\]

$Y_{it}$ is the degree of spatial agglomeration of industrial firms in city $i$. Two measurements of industrial agglomeration are used: (spatial agglomeration) and (Decenter). HSR city$_i$ represents whether city $i$ have high-speed-rail station, and post, indicates after the opening of high-speed railway. Its coefficient $\beta_1$ represents the impact of each high-speed railway opening on the spatial agglomeration of industrial firms. Control$_i$ represents control variables, including population density, government spending, industrial structure, and foreign direct investment. City and year fixed effects are $\mu_i$ and $\mu_t$, respectively. $\epsilon_{it}$ is the residual.

4. Empirical Results

4.1. Baseline Results of HSR on Industrial Agglomeration. Table 2 reports our baseline DID results. Columns (1) and (3) control for city and year fixed effects, and columns (2) and (4) add city characteristics. The dependent variables in columns (1) and (2) are spatial agglomeration.

In Table 2, the estimated coefficient of HSR city ‘post’ is significant and positive with the value of 0.13–0.32, suggesting that firms in HRS cities have a 13–32% higher spatial concentration relative to ones in non-HSR cities. The dependent variables in columns (3) and (4) are the average distance to city centers. The results show that firms in HSR cities are 1.74–6.30 km closer to the city centers. The results
in Table 2 verify Hypothesis 1, indicating the construction of HSR promotes the agglomeration of industrial firms.

In line with the findings of Shao et al. [18], our results indicate a positive impact of high-speed rail networks on industrial agglomeration. The exchange of information and technology is a vital component of modern manufacturing. HSR network accelerates population flow and facilitates knowledge exchange, thus changing the spatial agglomeration of cities.

4.2. Dynamic Effects of High-Speed Railway. A concern with the baseline results of Table 2 is that cities construing HSR might already have been on a higher growth rate of industrial agglomeration. To address this concern, we investigate whether the performance of HSR constructed cities was different from that of control cities before the construction of the HSR station. We examine the dynamics of the relation between HSR construction and industrial agglomeration by including a series of dummy variables in the baseline regression. The dynamic model is presented as follows:

$$Y_{it} = \beta_0 + \sum \beta_{i} HSR\; city_{i} \ast dummy_{j} + \beta_2 Control_{it} + \mu_i + \mu_t + \epsilon_{it},$$

(28)

where dummy_{j} equals 1 for firm i in the j-th year after the construction of HSR, while dummy_{j} equals 1 for firm i in the j-th year before the establishment of SEZ s; otherwise, they each equals 0. The black circles in Figure 2 represent the coefficients of HSR city_{i} \ast dummy_{j} in a set of regressions of equation (2). The solid lines represent 95% confidence intervals clustered at the city level. We consider an 11-year window, ranging from 6 years before HSR construction until 5 years after construction.

### Table 1: Descriptive statistics of main variables.

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial agglomeration</td>
<td>Log of variation of longitude and latitude</td>
<td>6.3908</td>
<td>0.9177</td>
</tr>
<tr>
<td>Dcenter</td>
<td>Average distance to city centers</td>
<td>35.3524</td>
<td>18.1123</td>
</tr>
<tr>
<td>HSR city</td>
<td>Whether the city connects to the HSR system</td>
<td>0.4368</td>
<td>0.4961</td>
</tr>
<tr>
<td>Post</td>
<td>After the opening of HSR</td>
<td>0.1580</td>
<td>0.3648</td>
</tr>
<tr>
<td>pop_density</td>
<td>Population density</td>
<td>102.1999</td>
<td>157.6307</td>
</tr>
<tr>
<td>Government</td>
<td>The ratio of government expenditure to GDP</td>
<td>0.1227</td>
<td>0.0765</td>
</tr>
<tr>
<td>ind_ratio</td>
<td>The proportion of secondary industry</td>
<td>40.7367</td>
<td>15.6482</td>
</tr>
<tr>
<td>FDI</td>
<td>Log of foreign direct investment</td>
<td>9.1191</td>
<td>2.1357</td>
</tr>
<tr>
<td>Industrial labor mobility</td>
<td>Rate of change in the industrial labor</td>
<td>0.3162</td>
<td>1.5633</td>
</tr>
<tr>
<td>Railway passenger</td>
<td>Log of passengers transporting by railway</td>
<td>5.4582</td>
<td>1.5826</td>
</tr>
<tr>
<td>Total passenger</td>
<td>Log of total passengers</td>
<td>8.6068</td>
<td>0.9210</td>
</tr>
<tr>
<td>MA(population)</td>
<td>Market access constructed by the population</td>
<td>5.8403</td>
<td>1.0399</td>
</tr>
<tr>
<td>Ma(GDP)</td>
<td>Market access constructed by GDP</td>
<td>15.8542</td>
<td>1.5578</td>
</tr>
<tr>
<td>MA(GDP per capita)</td>
<td>Market access constructed by GDP per capita</td>
<td>8.9852</td>
<td>3.0152</td>
</tr>
<tr>
<td>HSR cost</td>
<td>The construction cost of HSR</td>
<td>3.2493</td>
<td>0.5824</td>
</tr>
</tbody>
</table>

### Table 2: Baseline results of HSR on industrial agglomeration.

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>(1) Spatial agglomeration</th>
<th>(2) Spatial agglomeration</th>
<th>(3) Dcenter</th>
<th>(4) Dcenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSR city$^{\ast}$post</td>
<td>0.3229*** (0.0510)</td>
<td>0.1256*** (0.0471)</td>
<td>-6.3098*** (0.9704)</td>
<td>-1.7465*** (0.8315)</td>
</tr>
<tr>
<td>pop_density</td>
<td>0.0002$^{*}$ (0.0001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Government</td>
<td>-1.1895*** (0.3655)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ind_ratio</td>
<td>0.0259*** (0.0018)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDI</td>
<td>0.0100 (0.0108)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>3,676</td>
<td>3,184</td>
<td>3,676</td>
<td>3,184</td>
</tr>
<tr>
<td>$R$-squared</td>
<td>0.0120</td>
<td>0.1947</td>
<td>0.0122</td>
<td>0.1757</td>
</tr>
<tr>
<td>City fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Table 2 reports the baseline results of HSR on industrial agglomeration. The dependent variables of columns (1) and (2) and (3) and (4) are spatial agglomeration and average distance to city center, respectively. Robust standard errors are clustered at the city level in the parentheses. City fixed effects and year fixed effects are included in all results. Significance levels: $^*p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. 

Wireless Communications and Mobile Computing
Figure 2 presents the pretrends and the lagged effects estimated from equation (2). These coefficients should be positive and significant if cities were selected to construct HSR stations due to the predetermined trends, but the coefficients of the interaction term are insignificant before HSR construction. These results confirm that HSR and non-HSR cities had similar trends before the HSR construction. After the HSR construction, the estimate of spatial agglomeration is positive and significant, and that of average distance to city center is negative. Overall, changes in agglomeration have no effect before the construction of the HSR station, which satisfies the parallel-trend test.

4.3. IV Regression Results. Due to the reverse-causality problem between HSR location and industrial agglomeration, we use instrumental variables to conduct a two-stage least square (2SLS) regression. The selection of instrumental variables should satisfy two assumptions. The first one is that instrumental variables correlate with the construction of HSR stations. The second requirement is that the instrumental variables must be exogenous. Therefore, we calculate the average construction cost of each city as the instrumental variable.

Following previous literature [17, 37, 38], we construct HSR construction cost as an instrument. Geographical development cost is an essential basis for determining the direction of high-speed railway lines. In addition, the construction cost cannot affect industrial agglomeration through other channels. It can only affect industrial agglomeration through HSR construction, which satisfies the exogeneity of the instrumental variable.

We use DEM digital elevation data from China Geospatial Data Cloud (CGDC) to construct an elevation map. We extracted the slope, gradient, and hydrology using ArcGIS software from the map. We reclassify this geographic variable into ten levels with equal intervals to obtain the cost data, slope cost (slope), gradient cost (gradient), and hydrological cost (hydr). The minimum level is 1, and the maximum level is 10. According to the following equation, we calculate the geographic development cost for each cell on the map.

\[ \text{cost} = 0.4 \times \text{slope} + 0.3 \times \text{gradient} + 0.3 \times \text{hydr}. \]  

Figure 3 shows the distribution of HSR construction costs. The darker the color on the map, the higher the cost of high-speed railway construction.

Due to the mountains and basins in the central and western regions, the construction cost of high-speed railway is higher than that in eastern China. It also indicates a strong correlation between high-speed railway stations and HSR construction costs. We compute the average construction cost for each city. Since the geographic information is time invariance, we construct an interaction term between the average construction cost and the year dummy variable after 2007 to construct an instrumental variable (the first high-speed rail in China was launched in 2007).

We estimate the effects of HSR on agglomeration externalities using a two-stage least square (TSLS) model. The two-stage least square method is used to analyze model having endogenous explanatory variables in a linear regression framework. This methodology uses uncorrelated instrumental variables with error terms that enable the estimation of model parameters. The study by Taylor et al. [39] analyzed the determinants of transit ridership implementing a two-stage least squared (TSLS) model. This model helped find simultaneity between supply and demand of the variables concerned. We start by estimating the validity of the instrument using a fixed-effects model.

\[ \text{HSR}_{it} = \beta_0 + \beta_1 \cos t_i \ast \text{post}_i + \beta_2 \text{Control}_{it} + \mu_i + \mu_t + \epsilon_{it}. \]  

Specifically, the constructing cost can be treated as the instrumental variable (IV) for HSR. The second stage...
estimation uses the following equation:

$$Y_{it} = \beta_0 + \beta_1 \text{HSR}_{it} + \beta_2 \text{Control}_{it} + \mu_i + \mu_t + \epsilon_{it},$$  \hspace{1cm} (31)

where HSR_{it} indicates the interaction term of HRS city and after construction dummy. Table 3 shows the results of the two-stage least square model. A set of fixed effects and control variables are all included in both stages of the model.

As we can see in Table 3, column (1) is the result of the first stage, and columns (2) and (3) are the results of the second stage. In column (1), HSR construction cost and time interaction items are significantly negative, suggesting that high construction costs reduce the possibility of high-speed railway construction. It verifies that the instrumental variable of HSR construction cost meets the first assumption of IV. Kleibergen-Paap rk Wald $F$ statistics in the first stage is 163.9, larger than the critical value of 10 [40]. Thus, it rejects the hypothesis of the weak instrumental variable. The second stage results show that HSR construction promotes spatial agglomeration and reduces the average distance to city centers. The IV results are consistent with the benchmark regression, indicating that the HSR cities have a higher industrial agglomeration level after considering endogeneity.

4.4. PSM-DID Results. Based on the DID analysis, we use the propensity score matching (PSM) method to analyze the impact of HSR on industrial agglomeration [41, 42]. The advantages of the PSM method are to eliminate sample selection bias by controlling for observed and unchangeable differences between groups [28, 43]. The use of PSM removes biases when estimating treatment effect compared to other approaches; hence, it is extremely popular, especially in healthcare. Due to the matching with the propensity score, a balanced dataset is created, allowing simple and direct comparisons among the baseline covariates. PSM has been used to study acute heart failure related mortality in the project AHEAD (2006-2009) [44]. Thus, the HSR cities and non-HSR cities are similar in socioeconomic status.

First, we conduct logit regression on the control variables to achieve the conditional probability (a propensity score value) of whether an HSR line is constructed in the city [45]. Then, five non-HSR cities with the closest propensity scores are matched with high-speed railway cities. We also matched the treatment group with a control group within a radius of 0.1. These PSM methods can effectively solve sample selection bias and reduce DID estimation bias, but they cannot avoid inherent problems caused by missing variables. To derive the average treatment effect, we estimate the difference in agglomeration before and after the construction of HSR stations. Table 4 lists the PSM-DID results of HSR on agglomeration.

Table 4 presents the impact of HSR stations on industrial agglomeration using the PSM-DID model. Columns (1) and (3) use the radius matching method, while columns (2) and (4) apply the nearest neighbor matching method. The dependent variables of columns (1) and (2) and (3) and (4) are spatial agglomeration and average distance to city centers, respectively. The results in Table 4 show that both the magnitude and signal are similar to the baseline regression results.

We perform balancing tests to examine whether the economic status significantly differs between the control and treated groups. As shown in Table 5, differences between prematching treatment and control groups imply potential sample selection bias. After matching, the average values of all control variables are not significantly different for the treated and control groups.
4.5. Spillover Effects of HSR. As a major transportation system, HSR may have a spillover effect on local economy [46, 47]. By saving time, the HSR system improves supply chain efficiency and customer service quality, attracts firms to more accessible locations, and generates agglomeration externalities such as knowledge spillovers and input sharing among neighboring firms. In order to test the spillover effects of HSR, we developed a set of dummy variables based on the distance of HSR stations (the dummy variables are based on the distance to high-speed railway stations, which is below 50 km, 50-100 km, 100-150 km, and 150-200 km).

We construct the following empirical model using firm samples from non-HSR cities:

\[ Y_{it} = \beta_0 + \sum_k \beta_k^{\text{dist}_\text{HSR}} + \beta_2 \text{Control}_a + \mu_i + \mu_t + \epsilon_{it}, \quad (32) \]

where \( Y_{it} \) is the degree of spatial agglomeration of industrial firms in city \( i \), which is spatial agglomeration and Dcenter. \( \text{dist}_\text{HSR} \) is distance bins containing that the distance to the nearest HSR station is below 50 km, 50-100 km, 100-150 km, and 150-200 km.

Table 3: IV regression results.

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>(1) First stage</th>
<th>(2) Second stage</th>
<th>(3) Second stage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HSR city * post</td>
<td>Spatial agglomeration</td>
<td>Dcenter</td>
</tr>
<tr>
<td>Cost * post</td>
<td>-0.2303*** (0.0180)</td>
<td>1.2857*** (0.1696)</td>
<td>-14.1523*** (2.8563)</td>
</tr>
<tr>
<td>HSR city * post</td>
<td>1.2857*** (0.1696)</td>
<td>1.2857*** (0.1696)</td>
<td>-14.1523*** (2.8563)</td>
</tr>
<tr>
<td>Observations</td>
<td>3,676</td>
<td>3,676</td>
<td>3,676</td>
</tr>
<tr>
<td>Kleibergen-Paap rk Wald</td>
<td>163.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>City fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Table 3 reports the IV results of HSR on industrial agglomeration. Column (1) is the result of first stage. Columns (2) and (3) represent the second stage results of spatial agglomeration and average distance to city center. Robust standard errors are clustered at the city level in the parentheses. City fixed effects and year fixed effects are included in all results. Significance levels: *** \( p < 0.01 \).

Table 4: PSM-DID results.

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>(1) Spatial agglomeration</th>
<th>(2) Spatial agglomeration</th>
<th>(3) Dcenter</th>
<th>(4) Dcenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>PSM method</td>
<td>Radius</td>
<td>Nearest neighbors</td>
<td>Radius</td>
<td>Nearest neighbors</td>
</tr>
<tr>
<td>HSR city * post</td>
<td>0.2070*** (0.0402)</td>
<td>0.1999*** (0.0465)</td>
<td>-3.2897*** (0.7279)</td>
<td>-3.0564*** (0.8061)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,524</td>
<td>1,164</td>
<td>1,524</td>
<td>1,164</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.1135</td>
<td>0.1153</td>
<td>0.1275</td>
<td>0.1361</td>
</tr>
<tr>
<td>City fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Table 4 reports the results of propensity score matching. Columns (1) and (3) adopt the radius matching method, while columns (2) and (4) conduct the nearest neighbors matching method. Robust standard errors are clustered at the city level in the parentheses. City fixed effects and year fixed effects are included in all results. Significance levels: *** \( p < 0.01 \).

Table 5: Balance tests results.

| Variable | (1) Unmatched | (2) Matched | (3) Mean | (4) % reduct | (5) % bias | (6) t-test | (7) p > |t| |
|----------|---------------|-------------|----------|--------------|------------|------------|--------|-----|
|          | Unmatched     | Treated     | Control  | % reduct     | [bias]     | t          | p      |    |
| pop_den  | Unmatched     | 209.24      | 92.48    | 54.8         | 93         | 11.84      | 0.00   |    |
|          | Matched       | 204.11      | 212.27   | -3.8         | -0.45      | 0.65       |        |    |
| Government | Unmatched   | 0.13        | 0.17     | -49.6        | 96.9       | 8.32       | 0.00   |    |
|          | Matched       | 0.13        | 0.13     | 1.6          | 0.38       | 0.70       |        |    |
| ind_ratio | Unmatched     | 48.54       | 42.44    | 46.7         | 96.9       | 8.52       | 0.00   |    |
|          | Matched       | 48.53       | 48.34    | 1.4          | 0.24       | 0.81       |        |    |
| FDI      | Unmatched     | 10.88       | 9.34     | 97           | 96.4       | 17.63      | 0.00   |    |
|          | Matched       | 10.89       | 10.83    | 3.5          | 0.59       | 0.56       |        |    |

Notes: Table 5 reports the results of the balance test of propensity score matching.
150 km, and 150-200 km. Control variables are consistent with the baseline model in equation (1). The spillover results are shown in Figure 4.

Figure 4 shows the spillover effects of HSR at various distances (with confidence intervals). On the left, the below 50 km, 50-100 km, and 100-150 km are all significant negatives, whereas the coefficients of 150-200 km are insignificant and relatively small. On the right, we also see the similar pattern on the average distance to city centers. It indicates that HSR has a spillover effect of industrial agglomeration with a radiation range of 150 km.

### 4.6. Intensive and Extensive Margin Analysis

Firms could have responded to the construction of HSR initiatives by incumbent along the intensive margin and by a new entrant or exit firms along the extensive margins. The HSR impacts are divided into an extensive margin effect created by new entrants and exits and an intense margin effect related to continuing firms to elucidate further such effects [48]. As dependent variables, we determine the average distance to city centers for incumbents, entrants, and exits companies. Table 6 shows the outcomes of the intense and extensive margins.

The decomposition results for continuing firms are reported in column (1). It shows that HSR has an intensive margin effect on agglomeration, which indicates incumbents may move to the city center after constructing high-speed railways. The net entry effects in column (2) are statistically and economically significant. Newly established firms will be closer to the city center, triggering industrial agglomeration, but it shows no extensive margin effect from exiting firms in column (3). Overall, this decomposition indicates that the high-speed railway primarily promotes incumbents and entrants.

### 5. Mechanisms, Heterogeneity, and Robustness Checks

#### 5.1. Mechanisms

The empirical results in the previous section show that the construction of the HSR has extensively promoted industrial agglomeration. The mechanisms to industrial agglomeration driven by HSR construction are discussed in the theoretical analysis. This section will use empirical methods to verify the potential mechanisms of population mobility and market potential.
5.1.1. The Mechanism of Labor Mobility. High-speed railway shortens time distance and reduces travel costs, thus facilitating the flow of workers and resources across regions [1]. By facilitating the movement of people and factors between regions, HSR network changes the spatial connections between cities and the emergence of their social economies. The more effective transfer and flow of labor and resources in peripheral areas are transferred to central areas, resulting in a siphon effect. We define three variables to construct labor mobility: industrial labor mobility (column (1)), the volume of railway passengers (column (2)), and the volume of total passengers (column (3)).

As shown in Table 7, the mechanism results show that HSR construction significantly promotes labor mobility. In column (1), the coefficient is significantly positive. It means that HSR cities have 15.27% more labor mobility than non-HSR cities, thus demonstrating the mechanism of population mobility. The results in columns (2) and (3) indicate that the HSR construction can promote agglomeration by improving market potential.

5.1.2. The Mechanism of Market Accessibility. The construction of the high-speed railway improves the spatial accessibility of the cities and brings the benefits of reducing transportation costs [49]. To explore the mechanisms, we further examine the impact of HSR-induced market potential on industrial agglomeration. This paper uses market access (MA) as a proxy for market accessibility, which is used to measure a regional market’s size and the market’s proximity. We use the market potential function method proposed by Harris [50] to estimate the market potential of each region:

\[ MA_{xy} = \sum \frac{\text{economic}_{xy}}{d_{xy}}, \]  

(33)

where \( MA_{xy} \) represents the market access of city \( x \) in year \( t \), economic\(_{xy} \) represents the economics condition (such as population, GDP, and GDP per capita) in city \( y \) in year \( t \). \( d_{xy} \) represents the distance between city \( x \) and city \( y \). \( id_{x} \) is a HSR accessibility index between two places. If both cities \( x \) and \( y \) have high-speed railways, \( id_{x} \) equals 0.5; if city \( x \) or city \( y \) has high-speed railways, \( id_{x} \) equals 0.75; if neither city \( x \) nor \( y \) has high-speed railways, \( id_{x} \) equals to 1.

In order to examine how the market access mechanism responds to high-speed railways, we conduct empirical regressions in a fashion analogous to equation (1). The results are presented in Table 8 with MA constructed by population (column (1)), GDP (column (2)), and GDP per capita (column (3)).

In column (1), the coefficient of HSR city’s post is positive and significant, indicating that cities with HSR have a higher level of market accessibility relative to non-HSR cities. Columns (2) and (3) also tell a similar story: high-speed railways promote market access for firms, thus increasing industrial agglomeration.

5.2. Heterogeneity Analysis. In this section, we provide estimated results of HSR on industrial agglomeration using subgroup samples to explore the heterogeneous effects.

We first show the heterogeneous effects by agglomeration level. We extract two subgroups of cities based on the initial
level of GDP and explore the impact of high-speed railways for these two subgroups. One subgroup of cities has a higher GDP, above the medium. Cities with initial high-GDP levels significantly affect HSR on industrial agglomeration. The other subgroup of cities with lower GDP has a more minor and less substantial impact on spatial agglomeration, indicating that economic activity is more dispersed. This finding supports the new economic theory that there are core-periphery effects of trade integration between ex-ante asymmetric markets [17]. The heterogeneous results of HSR are shown in Table 9.

To examine the heterogeneity of the effect across the industrial structure, we divide the sample into high-industrialized and low-industrialized cities, according to the initial proportion of the secondary industry. We note that the effect for high-industrialized cities appears larger and statistically more significant than the effect for low-industrialized ones. The construction of HSR promotes agglomeration by around 21%, and the average distance from firms to the city centers decreases by 2 km for high-industrialized cities. These effects are equivalent to 1.3-1.7 times the baseline estimate.

### 5.3. Robustness Checks

#### 5.3.1. Placebo Tests of High-Speed Railway

To check the extent to which the results are affected by omitted variables, placebo tests are conducted by randomly assigning the opening of HSR to cities (La [51, 52]). We first randomly assign city samples to artificially opening of high-speed railway

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>(1) Spatial agglomeration</th>
<th>(2) Spatial agglomeration</th>
<th>(3) Dcity</th>
<th>(4) Dcity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Panel A: initial GDP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>HSR city*post</td>
<td>0.1095*** (0.0352)</td>
<td>0.0182 (0.0473)</td>
<td>-2.2455** (1.0906)</td>
<td>0.9541 (0.5836)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,575</td>
<td>1,609</td>
<td>1,575</td>
<td>1,609</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.1179</td>
<td>0.0394</td>
<td>0.0725</td>
<td>0.0175</td>
</tr>
<tr>
<td>Panel B: initial industrial structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>HSR city*post</td>
<td>0.2137*** (0.0738)</td>
<td>-0.0107 (0.0570)</td>
<td>-2.2594* (1.2040)</td>
<td>0.9996 (0.8796)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,505</td>
<td>1,485</td>
<td>1,505</td>
<td>1,485</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0306</td>
<td>0.1575</td>
<td>0.0525</td>
<td>0.1849</td>
</tr>
<tr>
<td>City fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Table 9 reports the results of heterogeneity analysis. Robust standard errors are clustered at the city level in the parentheses. City fixed effects and year fixed effects are included in all results. Significance levels: *p<0.10, **p<0.05, ***p<0.01.

Figure 5: Placebo tests on HSR. Notes: Figure 5 shows that the kernel density of the t-statistics of coefficients is from 500 simulations randomly assigning false HSR to cities. The solid line presents the kernel density of the t-statistics of false HSR. The dashed line is the normal distribution fitted with the t-statistics density distribution. The vertical lines are the baseline results of spatial agglomeration and the average distance to the city center in Table 2.
Table 10: Robustness checks.

<table>
<thead>
<tr>
<th>Dependent variables</th>
<th>(1) Space Gini measured by labor</th>
<th>(2) Entropy measured by labor</th>
<th>(3) Space Gini measured by capital</th>
<th>(4) Entropy measured by labor</th>
<th>(5) Spatial agglomeration</th>
<th>(6) Dcenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSR city * post</td>
<td>0.5361*** (0.1711)</td>
<td>0.1637*** (0.0177)</td>
<td>1.0123*** (0.2447)</td>
<td>0.1969*** (0.0229)</td>
<td>0.1148** (0.0475)</td>
<td>-1.7465** (0.8315)</td>
</tr>
<tr>
<td>Observations</td>
<td>3,676</td>
<td>3,676</td>
<td>3,676</td>
<td>3,676</td>
<td>3,178</td>
<td>3,184</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0090</td>
<td>0.0895</td>
<td>0.0159</td>
<td>0.0735</td>
<td>0.2042</td>
<td>0.1757</td>
</tr>
<tr>
<td>City fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Notes: Table 10 reports a series of robustness checks for the baseline results. Robust standard errors are clustered at the city level in the parentheses. City fixed effects and year fixed effects are included in all results. Significance levels: *p<0.10; **p<0.05; ***p<0.01.

As reported in Table 10, the independent variable coefficients are significantly positive in columns (1)-(4), indicating the alternative measurements will not affect the results. The magnitude and signal of columns (5) and (6) are similar to the baseline regression results. All regression results in Table 10 further prove the robustness of our findings.

6. Conclusion

In this paper, we construct a crossregional model of industrial agglomeration based on the new economic geography model. Using the panel data in 2000-2013, this paper estimates the impact of the high-speed railway on the agglomeration of industrial firms. The results show that high-speed railways significantly enhance the spatial agglomeration of industrial firms. To solve the endogenous problem, we construct an instrument variable based on HSR construction costs and estimate the effects of HSR using a two-stage least squared (TSLS) model. At the same time, PSM results also get the positive effect caused by high-speed railways, which proves the validity of the results. It is shown that results are robust to the alternative measurements of agglomeration and other contemporaneous economic shocks.

As a major transportation system, HSR may have a positive space spillover effect. The paper examines the spillover effects at various distances and finds HSR stations have a radiation range of 150 km on neighbor regions. Also, we investigate how the high-speed rail system impacts industrial agglomerations. By reducing travel time and transportation costs, high-speed rail facilitates crossregional labor mobility and boosts regional industrial agglomeration. Better access to markets provided by the HSR system can contribute to local economies and agglomerations. The results of the heterogeneous analysis indicate that high-speed railways have a greater impact on industrial agglomeration in regions with greater economic development and a higher ratio of manufacturing. The findings are consistent with the new economic theory which have core-periphery effects of trade integration between ex-ante asymmetric markets [17].

This study tries to enrich the literature on the economic impact of HSR in developing countries. Nevertheless, there is still a need for more research on this topic. Due to a lack of data, this study can only evaluate the short-term effects of high-speed railways. Further research should more precisely assess the long-term impacts of HSR with better data.
and a more complicated model. It would be interesting to explore the relationship between transportation and economics [3, 53]. These analyses would be of value in determining the best way for developing countries to implement a transportation system. In planning and construction of HSR, the government should allocate more capital and resource in the less developing regions with backward transportation. It is conducive to facilitating the accessibility and economy and narrowing the regional economic development gap [54].

Data Availability

The datasets used during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References


