NEREUS
Núcleo de Economia Regional e Urbana
da Universidade de São Paulo
The University of São Paulo
Regional and Urban Economics Lab

THE IMPACT OF HIGH-SPEED RAILWAY ON LABOR SPATIAL MISALLOCATION – BASED ON SPATIAL DIERENCE-IN-DIERENCES ANALYSIS

Linnan Yan
Menger Tu
André L. S. Chagas
Lufeng Tai

TD Nereus 04-2022
São Paulo
2022

The Impact of High-speed Railway on Labor Spatial Misallocation– Based on Spatial Difference-in-Differences Analysis

Linnan Yan, Menger Tu*, André L.S. Chagas, Lufeng Tai

[a] School of Economics, Shanghai University, 200444 Shanghai, China
[b] Regional Economics Applications Laboratory (REAL), University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA
[c] Department of Economics, University of São Paulo, Av. Luciano Gualberto, 908, São Paulo, SP, Brazil

URL: tme18tc@gmail.com (Menger Tu*)

Preprint submitted to Elsevier *December 15, 2021*

The Impact of High-speed Railway on Labor Spatial Misallocation– Based on Spatial Difference-in-Differences Analysis

Abstract

Existing studies neglected to assess the resource allocation effect of high-speed railway (HSR). This paper examines the impact of HSR on labor spatial misallocation in China by applying a modified spatial difference-in-differences approach, which identify local treatment effect, spillover effect on treated and untreated regions. The study finds: (1) Opening HSR alleviates not only the local labor misallocation but also the misallocation in surrounding areas to a greater extent, including cities with HSR (treatment group) and without HSR (control group), which contributes to the overall productivity increase. The spillover effect of HSR is larger than the direct effect. (2) The largest spillover effect occurs in adjacent areas near 350 km apart, while the spillover effect disappears beyond 500 km. (3) The direction and magnitude of HSR effect depend on the urban scale. For large-scale cities, the impact of opening HSR is greater versus small-scale ones.

Keywords: *high-speed railway, spatial difference-in-differences, labor spatial misallocation*

JEL Classification: *R40;R15*

1. Introduction

At the beginning of the 21st century, China embarked on an ambitious program of high-speed rail (HSR) network construction, aiming to promote regional economic convergence and balance efficiency and equity (Chen and Haynes, 2015). By the end of 2018, China had more than 29,000 km of HSR, ranking the first in the world. Compared to traditional transportation modes, the HSR has narrowed the spatial distance, which could more effectively enhance the connectivity between cities and improve the

Preprint submitted to Elsevier *December 14, 2021*

free flow of labor across regions. However, some scholars believe that the opening HSR has intensified the ”siphon effect”, which benefits the central cities at the sacrifice of the peripheral ones (Hall, 2009; Cao et al., 2013; Vickerman, 2015; Zhang and Tao, 2016) . The main reason for the controversy is that the existing studies focus only on the local economic effect and neglect the impact of HSR on the aggregate outcome of the nation. Moreover, the connectivity of HSR undoubtedly makes its impact cross-regional. Therefore, ignoring the spatial spillover effects would also underestimate the impact of HSR.

Given the relationship between resource allocation and overall economic, we believe that it is most appropriate to assess the overall economic effect of HSR by studying its impact on resource misallocation. Under the ideal market mechanism, production factors would shift to the higher efficient department and prompt the equivalence of marginal revenue across diverse sectors, thus attaining Pareto Optimality(Hsieh and Klenow, 2009; Banerjee and Moll, 2010). Once the obstacle appears in the free movement, the effective allocation of factors will be lost, which will lead to a decline in overall economic growth (Hsieh and Moretti, 2019). In China, many factors hinder labor mobility across regions, leading to market segmentation and severe labor misallocation at space, such as government intervention, administrative monopoly, household registration system, and high housing price. In addition, the extent of labor spatial misallocation determines aggregate output loss (Hsieh and Moretti, 2019). Therefore, it is essential to evaluate the HSR's effect on the aggregate economy from the perspective of labor misallocation.

The aim of this paper is to contribute, at a methodological level, to evaluating the aggregate economic impact of the HSR program. First, we identify the impact of HSR on aggregate economic output from the perspective of labor spatial misallocation, thus adding a new understanding about HSR's policy effect. Second, compared to the previous studies, this article takes the spillover effect of HSR into account. We distinguish the spillover effect of HSR on the treated region and the non-treated region by adopting a modified spatial difference-in-differences (DID) approach. This allows for a more accurate and complete assessment of the direct and indirect effects of HSR than traditional spatial specifications. The few studies using spatial DID only emphasize their own treatment effect and spillover effect within the treatment group, ignoring the spillover effect on the surrounding non-treated region (Delgado and Florax, 2015). In fact, the results may also be biased if this part of the effect is not controlled for. Third, we estimate the scope of the spatial spillover effect by setting different distance cut-offs as the definition standard of the spatial weight matrix cells. This provides a practical reference for the layout of HSR.

This article is organized in seven sections, including this introduction. Section 2 includes a literature review and presents the study's theoretical background. In section 3, we present the empirical dataset and strategies used in the analysis. In section 4, we show the empirical results of the regression. In sections 5 and 6, we process the heterogeneity analysis and robustness check, respectively. The final section presents concluding remarks.

2. Related literature and mechanism analysis

2.1. Literature Review

Research on resource misallocation can be divided into two categories: one set of studies mainly focus on the measurement of the degree of resource misallocation and the impact of resource misallocation on economic growth (Olley and Pakes, 1996; Hsieh and Klenow, 2009; Aoki, 2012; Bartelsman et al., 2013). The most representative is the research of Hsieh and Klenow (2009). By setting the C-D production function with the constant return to scale, they construct an analysis framework of exogenous output and capital distortion to measure the degree of resource misallocation accurately. They find that China's overall economic efficiency would increase by 30% to 50% if there were no resource misallocation. Another set of studies highlight the influencing factors of resource misallocation. The existing literature mainly attributes the causes of resource misallocation to market factors (imperfect financial market and distorted labor market price) and institutional factors (policy intervention and policy distortion). For example, Moll (2014) believes that the credit constraint caused by the imperfect financial markets in backward countries is the main reason for enterprise resource misallocation. Bernard et al. (2019) find that the state-owned enterprises with low productivity can get more credit, which is a critical factor distorting China's manufacturing resource allocation. It can be seen that many studies focus on resource misallocation across enterprises or industries, while the research on misallocation across regions (spatial misallocation) is still scarce. In China, due to the local protection and market segmentation, the resource misallocation at space is more severe. Therefore, the study of labor misallocation from the spatial perspective is of more significance to China's overall economic growth.

Besides, we also note that considerably fewer studies are concerned about how to alleviate resource misallocation. Some scholars have studied the mitigation effect of industrial agglomeration on resource misallocation based on new economic geography (NEG) (Ji et al., 2016). However, according to NEG, transportation cost and accessibility could also influence the factor movement across regions and spatial distribution of economic activities (Helpman, 1998; Krugman, 1991). Asturias et al. (2014) find that the transportation infrastructure can affect resource allocation by influencing enterprise market control and industry concentration in India. Ghani et al. (2016) suggest that India's golden quadrant highway upgrading project improves the resource allocation efficiency across industries by facilitating sharper industrial sorting between the major core cities and the areas along the highway. However, most previous studies have focused on the impact of the highway on resource misallocation across industries or enterprises. Exploration into the role of HSR remains scarce, especially with reference to spatial misallocation.

In fact, the research on the economic effect of HSR has been concerned by many scholars, but the controversy about HSR has never stopped. This mainly stems from the impact of opening HSR on the economy to small cities along the line. Some studies argue that the operation of HSR improves the accessibility between cities, driving the production factors of peripheral cities along the line to flow to central cities, which promotes the economic growth of central cities but inhibits the economic development of peripheral cities in turn (Sasaki et al., 1997; Givoni, 2006; Hall, 2009). Taking China's sixth railroad speed increase as an example, Qin (2017) finds that HSR reshapes the economic landscape of cities along HSR lines by reducing transportation costs. The empirical results show that the lower transportation costs introduce the shift of economic activities from the counties along the route and concentrate on the central cities, which results in a 3%-5% decrease in GDP and GDP per capita in the counties along the route. However, some studies have found that the opening HSR drives the diffusion effect from central cities to surrounding places, which is conducive to regional economic growth (Jedwab, 2016; Baum-Snow et al., 2017). Ahlfeldt and Feddersen (2017) find that HSR in Germany has contributed to regional economic growth along the route by increasing inter-city accessibility, strengthening economic agents links between cities, promoting market access and market potential, and facilitating the diffusion of knowledge & technology. There are two reasons for this argument: first, most current studies have primarily focused on the impact of HSR on regional economic development or regional net welfare increases (Rietveld and Bruinsma, 1998) while ignoring the role of HSR on overall economic growth. Second, the spatial spillover effect of HSR on the economies has been neglected, thus underestimating the economic effects generated by HSR. Resource allocation efficiency is one of aggregate economic growth paths. We attempt to fill this gap by exploring the impact of HSR on resource allocation and HSR's spillover effect.

2.2. Mechanism Analysis

As a fast transportation infrastructure, the impact of HSR on cities along the route is mainly in the form of increased accessibility, connectivity, and transitivity (Liu et al., 2020). However, most studies only focus on improved accessibility on regional economic development (Hall, 2009; Kim and Sultana, 2015) while ignoring the connectivity and the transitivity. In fact, the role of connectivity and transitivity has gradually emerged with the expansion of the HSR network. Based on these three characteristics, this paper analyzes the mechanism of the impact of HSR on labor spatial misallocation.

First, the most direct effect of opening HSR is to reduce transportation costs and break the labor market segmentation, which increases urban accessibility. The cities with HSR expand the scope of the labor market and improve the matching efficiency of supply and demand with regard to the labor force, thus alleviating the local labor misallocation. In addition, the construction of HSR stations in peripheral cities improves their inherent location disadvantages, enabling them to undertake the industrial transfer and diffusion of the core area better, which realizes the complementary and reconfiguration of intercity factors. At the same time, the continuous rational allocation of intercity factors promotes the improvement of labor productivity in peripheral cities and eventually forms an overall convergent development pattern.

Second, the commuting convenience of HSR drives urban integration, which has weakened the role of border barrier between cities. The way people live and work has gradually changed as a result. For example, the job-dwelling separation mode has emerged, in which people work in one city and live in another, with a ”pendulum” commute between cities via HSR (Chen et al., 2016; Heuermann and Schmieder, 2014). By this mode, labor can enter the high-productivity cities without the impediment of high housing prices, thereby improving the efficiency of labor allocation between cities.

Third, the rapid expansion of the HSR network strengthens the connections between nodes, between nodes and domain faces, and between domain faces, which generates more prominent indirect effect than direct effect. For cities connected to the HSR network, HSR significantly improves the interconnectivity between high-productivity cities and low-productivity ones by considerably reducing time distance, which mitigate the labor misallocation for both types of cities. For cities that are not connected to HSR network, on the one hand, the transitivity of HSR and other transportation infrastructure allows the labor force in those cities to move to destinations through nearby HSR hubs, thus improving labor allocation (Xu, 2017). On the other hand, HSR increases the speed and scope of information diffusion and dissemination (Dong et al., 2019; Lin, 2017), which facilitates the knowledge and technology spillover from core regions to peripheral areas. This also helps peripheral areas to improve production and resource allocation efficiency.

3. Data and Research Method

3.1. Data description

The information on the opening date of HSR is mainly extracted from China Railway Statistical Yearbook. China's HSR era began in 2008, which opened the first HSR line with a speed of 350 kilometers per hour (218 mph). Since then, according to the medium and long-term railway development plan, China's HSR network has gradually formed the spatial layout of "four north-south corridors and four East-West corridors". By the end of 2016, the network had covered 177 cities associated with a mileage of more than 23,000 kilometers (14,291 miles). Chinese cities with HSR in the given years are presented in Figure 1.

Figure 1. Cities with high-speed rail in China from 2008 to 2016

To ensure the integrity and availability of data, we mainly use the data of 280 prefecture-level cities from 2004 to 2016. The socio-economic variables are derived from China City Statistical Yearbook, such as employment, capital, local financial revenue, and government expenditure.

3.2. *Method and Variable Description*

3.2.1. *Method*

The root of our work is the difference-in-differences approach (Ashenfelter, 1978), a basic tool for evaluating treatment effect, which has been widely applied to study the policy influence. In its benchmark form, the model considers that the operation of HSR affects the labor misallocation through the interaction between time dummy and treatment term. The equation is set as follow:

$$M_{it} = \alpha_0 + \beta HSR_{it} + \sigma \mathbf{X}_{it} + \lambda_t + \mu_i + \xi_{it} \tag{1}$$

Where M_{it} is the dependent variable representing the degree of labor misallocation. The HSR_{it} is a dummy variable, which represents a city had operated HSR in the year t or not. $\mathbf{X}_{it}$ is the vector of a set of control variables at the city level, and ξ_{it} is an i.i.d error term. In the meantime, we control for the time and individual fixed effect with λ_t and μ_i, respectively.

The traditional DID must follow the SUTVA (stable value of the treatment unit assumption) hypothesis, which means that the treatment units do not affect other non-treated units (Rubin, 1978). However, there is no obvious physical barrier in policy spillover, and the implement of a program in a region might influence its neighbors (Verbitsky-Savitz, 2012). In other words, using traditional DID would violate the SUTVA, which leads to the biased estimation of the policy effect for the treatment region. We follow Chagas et al. (2016) by introducing the spatial DID approach to address this problem. Hence, the spatial lag term, which consists of a weight matrix and policy variable, is added to the model:

$$M_{it} = \alpha_0 + \beta HSR_{it} + \theta \mathbf{W} HSR_{it} + \sigma \mathbf{X}_{it} + \lambda_t + \mu_i + \xi_{it} \tag{2}$$

Where the new term $\theta \boldsymbol{W} HSR_{it}$ captures the spillover effect of opening HSR. $\boldsymbol{W}$ is a spatial weight matrix, which is constructed by the geographical distance across cities[1].

The term $\theta \boldsymbol{W} HSR_{it}$ indicates the average spillover effect of the opening HSR on neighboring cities, but the spillover effect might be different between treated and non-treated regions (Chagas et al., 2016). So we try to measure the spillover effect on treated and non-treated respectively by decomposing the spatial weight matrix[2]. Thus, the modified specification is as follows:

$$M_{it} = \alpha_0 + \beta HSR_{it} + \theta_1 \boldsymbol{W_{T,T}} HSR_{it} + \theta_2 \boldsymbol{W_{NT,T}} HSR_{it} + \sigma \boldsymbol{X}_{it} + \lambda_t + \mu_i + \xi_{it} \tag{3}$$

Where $\boldsymbol{W_{T,T}} HSR_{it}$ represents the spillover effect of opening HSR on neighboring cities with HSR. As such, $\boldsymbol{W_{NT,T}} HSR_{it}$ means the spillover effect on cities without HSR. Table 1 shows the policy effect of classical DID and two spatial DID approaches.

Table 1. The policy effects of classical DID and spatial DID

Policy effect	Classical DID	Spatial DID	Decomposed Spatial DID
Direct effect	a	a	a
Indirect effect	—	b	b1/b2
Neighboring treated group	a	a+b	a+b1
Neighboring control group	0	b	b2

3.2.2. *Variable Description*

The dependent variable is labor spatial misallocation, which represents the deviation degree between the optimal labor input proportion and the actual labor input proportion (Hsieh and Klenow, 2009). If the deviation degree is greater than 1, it means that the urban labor input is insufficient. Otherwise, the urban labor input is excessive

[1] The elements of W are shown: $\mathbf{W}_{ij} = \begin{cases} 1/d_{ij} & if\ i \neq j \\ 0 & if\ i = j \end{cases}$
where d_{ij} is the geographical distance between city and calculated by their latitude and longitude.

[2] The spatial weight matrix $\boldsymbol{W} = \boldsymbol{W_{T,T}} + \boldsymbol{W_{NT,T}} + \boldsymbol{W_{T,NT}} + \boldsymbol{W_{NT,NT}}$. While $\boldsymbol{W_{NT,T}} HSR = 0$ and $\boldsymbol{W_{NT,NT}} HSR = 0$, so they are removed from the decomposed specification.

if the value is less than 1. In this paper, we use the absolute value of deviation minus 1 to measure the degree of urban labor spatial misallocation (M_{it}). If the value of M_{it} is large, it means that the spatial labor misallocation of the target city is severe. See Appendix 1 for the specific calculation process.

The explained variable, HSR_{it}, is a dummy variable that takes the value of one if the city had operated HSR in the year t, and zero otherwise. According to the definition of the European Union in 1996, HSR refers to a rail service with trains that the speed of the new-built lines exceeds 250 km/h (155 mph) or 200 km/h (124 mph) on the upgraded lines (Givoni, 2006). This rule is consistent with China's definition of the Administration of Railway Safety, and we choose the standard of 200 km/h (124 mph).

We also control a series of variables that affect the allocation of labor resources, such as urban productivity, FDI, industrial structure, government intervention, financial development, housing price. Table 2 reports the statistical description of variables.

Table 2. Variables Description

Variables	Variable name	Index to explain
M	Degree of labor misallocation	Absolute value of(labor input ratio in effective state/labor input ratio in distorted state -1)
HSR	High-speed rail	1=high-speed rail; 0=no high-speed rail
FDI	Foreign direct investment	Foreign direct investment in nominal GDP
Ind	The industrial structure	The ratio of output value of the tertiary industry to that of the secondary industry
AGDP	Real GDP per capita	The logarithm of real GDP per capita
Fd	Level of financial development	The ratio of outstanding loans to GDP
Gov	Government intervention	Proportion of government budget expenditure in GDP
Hp	House price	The logarithm of the city house price after the CPI index adjusted

4. Empirical Result

4.1. Classical DID model result

4.1.1. Effect of HSR

Based on Equation (1), we first estimate the impact of opening HSR on labor spatial misallocation by the traditional DID approach. In Table 3, column (1) is a panel data regression with time and individual fixed effect. Column (2) adds several control variables in the specification on the basis of column (1). Both regression results indicate that the opening HSR could reduce the urban labor misallocation, which is in agreement with Hu et al. (2020).

Table 3. Classical DID estimation result

	M	M
HSR	-0.0725***	-0.0643***
	(-5.236)	(-4.7102)
control	no	yes
City FE	yes	yes
Year FE	yes	yes
Observation	3640	3640

Notes:t statistics reported in parentheses; *** p <0.01, **p <0.05, * p <0.1.

4.1.2. *Solution to endogeneity*

When using the DID approach, the operation of HSR for cities is supposed to be an exogenous quasi-natural experiment. However, some scholars believe that the policymakers tend to choose cities that were expected a higher economic growth rate or developed ones, which results in endogenous placement problems (Hodgson, 2018).

To deal with the endogenous issue, we construct the instrument variable based on the principle which Faber (2014) points out that the placement of HSR depends on geographical development cost after eliminating the issue that considers the economic condition. Terrain, geology, hydrogeology are the main factors that influence geographical development cost, and these geographical factors are exogenous strictly. We firstly refer to Faber (2014) and Zhang (2018) by using geographic information data[3] and the minimum spanning tree rule to construct the least cost path spanning tree network. Then, we get the dummy variable about whether the prefecture-level city should have HSR. Since the dummy variable is cross-sectional data, we multiply the variable with the time dummy to construct a set of instrument variables (Duflo and Pande, 2007).

[3]The data are sourced from a Geo-spatial cloud dataset founded by Computer Network Information Center of Chinese Academy of Sciences. http://www.gscloud.cn/

Table 4. IV Regression of HSR's impact on labor misallocation

VARIABLES	M (second-stage)	HSR (first-stage)
HSR	-0.131*	
	(1.926)	
IV *year2008		-0.013
		(-1.49)
IV *year2009		0.010
		(0.25)
IV *year2010		0.118**
		(2.26)
IV *year2011		0.229***
		(3.87)
IV *year2012		0.195***
		(3.06)
IV *year2013		0.335***
		(5.12)
IV *year2014		0.280***
		(4.32)
IV *year2015		0.267***
		(4.15)
IV *year2016		0.266***
		(4.18)
control	yes	yes
City FE	yes	yes
Year FE	yes	yes
Observation	3185	3185
Kleibergen-Paap Wald rk F statistic	16.412	
Hansen J statistic	Chi-sq(8)=4.491 P-val = 0.8104	

Notes:t statistics reported in parentheses; *** p <0.01, **p <0.05, * p <0.1.

Table 4 presents the result of instrument regression. Column 3 shows the first-stage regression of IV on HSR. We find that the Kleibergen-Paap Wald rk F statistic is 16.412, which is larger than the critical value 10 that Staiger and Stock (1997) propose. This result rejects the assumption of weak IV, indicating that the IV we employ is appropriate.

Column 1 presents the result of the second stage, which shows that the HSR could mitigate the misallocation. It can be seen that after dealing with endogeneity, the impact of HSR on alleviating labor spatial misallocation still exists, indicating the robustness of the result. On this basis, we further consider the spatial effect of the HSR.

4.2. Spatial DID model result

4.2.1. Average spillover effect of HSR

In this section, we compare the results of three types of models. Column (1) in Table 5 is the panel data regression with the traditional DID approach. Column (2) shows the spatial lag of independent variable specification (SLX) suggested by Vega and Elhorst (2015). This is our baseline case to perform the DID model in spatial econometrics. Column (3) considers another spatial correlation (SDEM), including the lag of error term. We choose the appropriate specification based on Lagrange Multiplier (LM), LM robust tests, Akakike Information Criteria(AIC), Bayesian Information Criterion(BIC), and Loglikelihood Value. The tests above indicate that we should employ SDEM model (Anselin et al., 1996; Elhorst, 2014).

Table 5. Spatial DID estimation result

	Classical DID M	SLX-DID M	SDEM-DID M
HSR	-0.0643*** (-4.710)	-0.0468*** (-3.2505)	-0.0497*** (-3.7490)
$\boldsymbol{W} * HSR$		-0.418*** (-3.4538)	-0.3165** (-2.2629)
control	yes	yes	yes
$\boldsymbol{W} * control$	yes	yes	yes
lambda			0.8734*** (36.259)
Moran's I	0.0437	0.0417	
P-Value	0.001	0.001	
LM-lag	450.71	413.81	
P-Value	0.000	0.00	
LM-error	370.64	338.43	
P-Value	0.000	0.000	
Robust LM-lag	135.65	146.42	
P-Value	2.2e-16	2.2e-16	
Robust LM-error	55.577	71.054	
P-Value	8.988e-14	0.000	
AIC	-298.1988	-308.2912	-489.451
BIC	1561.723	1570.23	1395.27
Loglik	449.099	457.146	548.726
City FE	yes	yes	yes
Year FE	yes	yes	yes
Observation	3640	3640	3640

Notes:t statistics reported in parentheses; *** p <0.01, **p <0.05, * p <0.1.

In column (2), we find that the coefficient of HSR is -0.0468, and it is significant at the 5% level, which means that the opening HSR could improve urban local labor misallocation. The coefficient of $\boldsymbol{W} * HSR$ is -0.418, and it is significant at the 1% level, which indicates that the opening HSR could mitigate the labor misallocation of the vicinity.

The spatial correlation could happen through other channels. Column (3) shows the empirical result of SDEM specification. In this situation, the treatment effect is -0.0497, which means that the opening HSR could decrease the extent of misallocation.

The coefficient of treatment effect has no noticeable difference with the column (2). However, the spillover effect of HSR decreases to -0.3165 because the spatial correlation of unobserved heterogeneity in error term absorbs the impact in part. The parameter lambda is positive at the 1% significant level which controls for random shocks to the dependent variable.

4.2.2. Spillover effect to Treated and Non-treated Neighbors

The spillover effect in the previous analysis is an average effect, which ignores the differences between non-treated and treated regions (Chagas et al., 2016). Drawing on their work, we test the indirect effect of HSR on the treatment group and control group, respectively. Table 6 shows the result that we distinguish the two sets of indirect effects. Column (2) and column (4) present the decomposed case for each specification. The result of SLX-DID shows that the coefficient of the direct effect is -0.0607 after decomposing the average effect, 30% larger than the average model. The indirect impact on the neighboring non-treated region (-0.4289) is slightly larger than that of the adjacent treated region (-0.3975). This might be because the spillover effect of HSR is the only effect influencing the region (Chagas et al., 2016).

Table 6. Modified spatial DID estimation result

	SLX-DID		SDEM-DID	
	(1)	(2)	(3)	(4)
HSR	-0.0468**	-0.0607*	-0.0497***	-0.0854**
	(-3.2505)	(-1.9181)	(-3.7490)	(-2.8032)
$\boldsymbol{W} * HSR$	-0.418***		-0.3165*	
	(-3.4538)		(-2.2629)	
$\boldsymbol{W_{T,T}} * HSR$		-0.3975**		-0.2656*
		(-3.1164)		(-1.8287)
$\boldsymbol{W_{NT,T}} * HSR$		-0.4289***		-0.3466**
		(-3.4842)		(-2.444)
control	yes	yes	yes	
W*control	yes	yes	yes	
lambda			0.8734***	0.8747***
			(36.259)	(36.648)
City FE	yes	yes	yes	yes
Year FE	yes	yes	yes	yes
Observation	3640	3640	3640	3640

Notes:t statistics reported in parentheses; *** p <0.01, **p <0.05, * p <0.1.

As for column (4), the direct effect becomes -0.0854 after decomposing the average spillover effect, almost doubled versus column (3). At the same time, the coefficient of

$\boldsymbol{W_{T,T}} * HSR$ and $\boldsymbol{W_{NT,T}} * HSR$ are -0.2656 and -0.3466, respectively. Both of them are statistically significant. Combined with the direct effect, the total effect of the treatment group reaches -0.3510. It is worth noting that the indirect effect on the non-treated region is 97% of the treatment group (-0.3466/-0.3510). This result indicates that the opening HSR not only mitigates the labor misallocation of cities with HSR but also improves the extent of misallocation to cities without it. For treated regions, the indirect effect account for 75.67% of the total effect, which demonstrates that the improvement effect of opening HSR on misallocation derived from indirect effect mainly and direct effect in part. With regard to untreated regions, the influence comes from the spillover effect of treated regions only.

In summary, whether it is local, neighboring treated cities, or neighboring non-treated cities, the HSR plays a key role in alleviating the labor spatial misallocation. We also find that the indirect effect of HSR is greater than the direct effect. For treated regions, the role of HSR in alleviating labor misallocation derived from the location advantages brought by the improvement of the local infrastructure and the connection effect through linking other cities with HSR. For non-treated regions, they are mainly affected by the siphon effect and transitivity of vicinities opened HSR. Consequently, the labor force could reconfigure by means of HSR, which alleviates the spatial misallocation.

5. Heterogeneity analysis

5.1. Distance heterogeneity

According to the First Law of Geography (Tobler, 1970), the spatial spillover effect of HSR on labor misallocation may diminish with increasing distance between cities. Therefore, we try to figure out the variation of the spillover effect by introducing different distance cut-offs as the definition of the weight matrix cells:

$$\boldsymbol{W}_{ij} = \begin{cases} 1/d_{ij} \;\; if \; d_{ij} \geq d \\ 0 \qquad if \; d_{ij} < d \end{cases}$$

Where d_{ij} is the distance between cities, is different distance cut-offs with the interval of 50 km, such as d=(50km, 100km, 150km, ...).

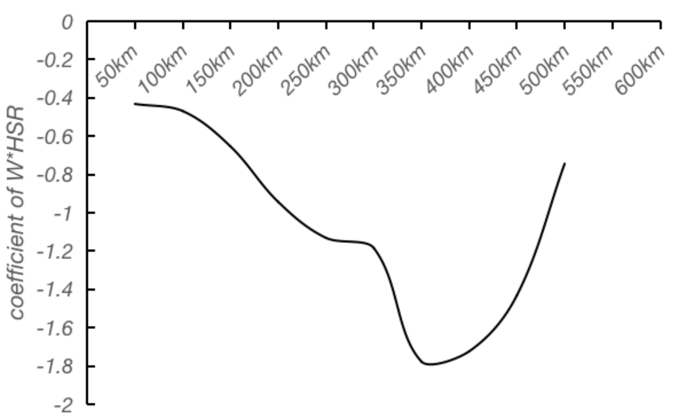

Figure 2. The spatial attenuation of spillover effect

Figure 2 shows the coefficients of the $\boldsymbol{W} * HSR$ at different distance cut-offs. It reveals that the coefficient becomes larger until 350 km (218 miles) and then declines, presenting a U-shape. The largest absolute value of the coefficient is approximately 1.77 when the distance reaches 350 km, while it becomes statistically insignificant when the distance exceeds 500 km (311 miles). This result suggests that the HSR would influence neighboring regions within 500 km, and the largest spillover effect appears near 350 km. That is to say, labor mobility by high-speed rail is more convenient if two cities are within 500 km of each other. In the meantime, the labor force living in the neighboring cities without HSR can also migrate to other cities with high productivity through the HSR station of the current city so as to alleviate the misallocation. When distance is larger than 500 km, there is no obvious spillover effect. The reason might be that when the distance is long, people tend to choose another transportation such as airplane. The optimal distance is 350 km, which proves the one-hour commuting circle on account of the fastest speed of HSR is 350 km/h (218 mph).

5.2. Urban scale heterogeneity

Due to the difference in city scale, the diffusion and the backflow of production factors are diverse between cities (Ureña et al., 2009). Whether the impact of opening HSR on misallocation varies with urban scale? We employ the population and the ratio of non-agricultural sector to represent the urban scale, which describes the supply and the demand side of the labor force, respectively. Specifically, on the one hand, the larger the urban population, the greater the market potential. The operation of HSR is more conducive to attracting factor inflow, which benefits the improvement of resource allocation efficiency of local and adjacent cities. On the other hand, the passenger transport function of HSR mainly acts on the industry and service sectors. Therefore, the larger scale of the non-agricultural sector is, the greater return of HSR will be.

Table 7 reports the result of heterogeneity in terms of population scale and industry scale, respectively. Column (1) and column (3) are based on the regular spatial DID approach. Column (2) and column (4) are based on spatial DID with decomposed weight matrix.

Column (1) and column (3) show that the coefficient of $HSR*pop$ and $HSR*ind$ are -0.113 and -0.944, respectively, and both are significant at the 1% level. Meanwhile, the coefficient of $\boldsymbol{W}*HSR*pop$ and $\boldsymbol{W}*HSR*ind$ are negative significantly. The result demonstrates that for cities with a large population or low agricultural sector share, the improvement of labor misallocation in the local and surrounding areas is greater on account of opening HSR, which is consistent with the theoretical expectation.

It is worth noting that the corresponding coefficient of HSR is significantly positive, which indicates that opening HSR can alleviate the spatial misallocation of labor only when the urban population scale and non-agricultural sectors scale reach a certain level. The same situation happens on the indirect effect as well.

As such, the coefficients of $HSR*pop$ and $HSR*ind$ are still significantly negative when we decompose the weight matrix, which column (2) and column (4) show. This result is in line with regular spatial DID specification. In addition, the interaction terms of the decomposed weight matrix with heterogeneous factors are negative at the 1%

Table 7. Heterogeneity analysis results

	population		non-agriculture sector	
	(1)	(2)	(3)	(4)
HSR	0.638*** (6.6541)	0.625*** (5.9502)	4.190*** (7.3669)	3.3674*** (5.6460)
$HSR * pop$	-0.113*** (-7.2637)	-0.114*** (-7.2287)		
$HSR * ind$			-0.944*** (-7.4450)	-0.770*** (-5.7870)
$\boldsymbol{W} * HSR$	5.363*** (6.0819)		59.129*** (11.1391)	
$\boldsymbol{W} * HSR * pop$	-0.956*** (-6.5815)			
$\boldsymbol{W} * HSR * ind$			-13.246*** (-11.1975)	
$\boldsymbol{W_{T,T}} * HSR$		5.254*** (5.8530)		61.590*** (9.5484)
$\boldsymbol{W_{NT,T}} * HSR$		6.228*** (3.8901)		51.455*** (7.9702)
$\boldsymbol{W_{T,T}} * HSR * pop$		-0.932*** (-6.2232)		
$\boldsymbol{W_{NT,T}} * HSR * pop$		-1.097*** (-4.1986)		
$\boldsymbol{W_{T,T}} * HSR * ind$				-13.779*** (-9.5988)
$\boldsymbol{W_{NT,T}} * HSR * ind$				-11.549*** (-8.0216)
control	yes	yes	yes	yes
$\boldsymbol{W} * control$	yes	yes	yes	yes
lambda	0.853*** (30.625)	0.853*** (30.707)	0.820*** (24.407)	0.818*** (24.026))
City FE	yes	yes	yes	yes
Year FE	yes	yes	yes	yes
Observation	3640	3640	3640	3640

Notes:t statistics reported in parentheses; *** p <0.01, **p <0.05, * p <0.1.

significant level as well, either for the treated or untreated regions, which indicates that the indirect effect of opening HSR is larger for cities with large scale.

Specifically, by computing, we find that cities with a population scale larger than 2.68 million could improve their own labor misallocation. The spillover effect between cities with HSR and the spillover effect from treated to non-treated cities could be generated as the population scale larger than 2.81 and 2.94 million, respectively. As for the demand side (column 4), cities that the proportion of non-agricultural sectors reaches 79.45% could alleviate their own labor misallocation. The mitigating effect on neighboring cities that HSR connected occurs when the non-agricultural share of cities exceeds 87.37% compared to 86.25% for neighboring cities without HSR.

From the analysis above, we find that the threshold of occurrence for direct effect is lower than indirect effect. For both cases, about 80% of cities could mitigate their own misallocation through HSR operation, whereas approximately 60%-70% of cities affect their surrounding areas only. Thus, cities with a large scale could mitigate their labor misallocation on account of opening HSR, while just part of them could improve that of the vicinity.

6. Robust test

To confirm the validity of the estimation results, we employ four strategies to process the robustness test. Firstly, We perform parallel trend test. Secondly, in order to eliminate the interference of other random factors on the results, we process the placebo test by assigning HSR to cities randomly (Fu and Gu, 2017). Thirdly, we remove the core cities to avoid biased sample selection. The last strategy is to change the spatial weight matrix.

6.1. Event study

The premise of the validity of the DID approach is to meet the common trend hypothesis. We follow Lin (2017) by employing the event study method to test the parallel trend assumption. The equation is as follows:

$$M_{it} = \alpha_0 + \sum_{m=1}^{5} \beta_m FirstHSR_{i,t-m} + \sum_{n=0}^{5} \beta_n FirstHSR_{i,t+n} + \delta \mathbf{X}_{it} + \lambda_t + \mu_i + \xi_{it} \tag{4}$$

$$\begin{aligned} M_{it} = \alpha_0 &+ \sum_{m=1}^{5} \beta_m FirstHSR_{i,t-m} + \sum_{n=0}^{5} \beta_n FirstHSR_{i,t+n} + \sum_{m=1}^{5} \beta_m \boldsymbol{W} * FirstHSR_{i,t-m} \\ &+ \sum_{n=0}^{5} \beta_n \boldsymbol{W} * FirstHSR_{i,t+n} + \delta \mathbf{X}_{it} + \lambda_t + \mu_i + \xi_{it} \end{aligned} \tag{5}$$

where is $FirstHSR_{i,t}$ a dummy variable which represents that whether the city opened HSR in year first time. $FirstHSR_{i,t-m}$ and $FirstHSR_{i,t+n}$ are m-th lag and n-th ahead respectively, which are dummy variable also. The coefficient of these terms represent the difference between the treated group and the non-treated group in a given year.

Table 8. Event Study result

VARIABLES	M (without spatial effect)	HSR (with spatial effect)
pre5	0.0147	0.0153
	(-0.6682)	(-0.765)
pre4	0.0164	0.0269
	(-0.9111)	(-1.345)
pre3	-0.0012	-0.0032
	(0.06316)	(0.16)
pre2	-0.0100	0.0042
	(0.5)	(-0.21)
current	-0.0365	-0.0363*
	(1.5208)	(1.7286)
post1	-0.0545*	-0.0517**
	(1.9464)	(2.35)
post2	-0.0735**	-0.0647***
	(2.2272)	(2.8130)
post3	-0.1019**	-0.0969***
	(2.4262)	(3.7269)
post4	-0.0998**	-0.0788***
	(2.0792)	(2.7172)
post5	-0.0859	-0.0513*
	(1.5339)	(1.9)
$\boldsymbol{W} * pre_i$	no	yes
$\boldsymbol{W} * post_i$	no	yes
lambda		0.8801***
		(27.5031)
control	yes	yes
$\boldsymbol{W} * control$	yes	yes
City FE	yes	yes
Year FE	yes	yes
Observation	3640	3640

Notes:t statistics reported in parentheses; *** p <0.01, **p <0.05, * p <0.1.

Based on table 8, we find that the coefficients of the m-th lag term are not significant in event study, neither with spatial effect nor without. It means that there is no heterogeneous temporal trend of the labor misallocation changes between the treated and non-treated group before opening HSR. These results prove that the common trend assumption is tenable.

6.2. Placebo test

We draw on Li et al. (2016) by randomly generating experimental groups to implement the placebo test. First, we employ random simulation 500 times to generate 500 sets of fake variables, HSR_{fake}. Next, we replace the original variable HSR with HSR_{fake} to generate 500 simulation samples. Third, we conduct empirical analysis based on benchmark specification to produce the direct effect and spillover effect.

Figure 3 shows the distribution of the estimated coefficient based on 500 simulation samples. Panel A for direct effect and Panel B is spillover effect. It can be found

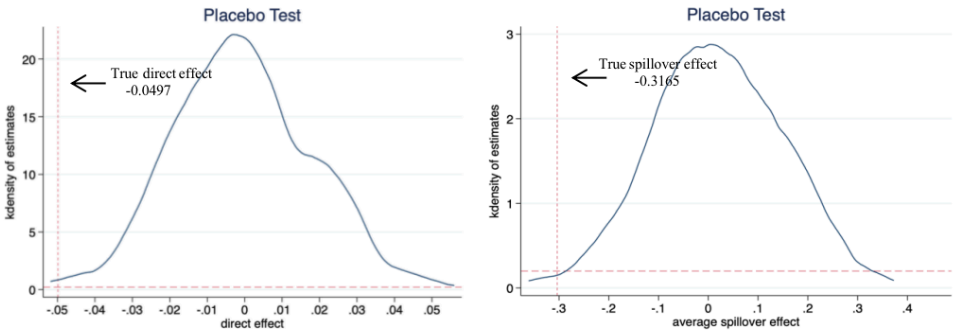

Figure 3. Placebo test

that the distribution of the direct and indirect effect of HSR_{fake} is concentrated in zero, and the true value place at the end of the left tail of the simulation parameter distribution, which reflects that the results of spatial DID regression are not caused by some accidental factors.

6.3. *Changing Sample*

As the first-tier cities in China, Beijing, Shanghai, Guangzhou, and Shenzhen significantly influence the economy and society of their vicinity. We remove these cities from our samples to eliminate the effects of these samples themselves. The regression result is attached in Table 9. It shows that the direct, average indirect, and the decomposed indirect effects of HSR are all significantly negative, which indicates that our benchmark result is robust.

Table 9. Changing Sample

	SDEM-DID	
	(1)	(2)
HSR	-0.0435** (-3.2648)	-0.0744** (-2.4363)
$\boldsymbol{W} * HSR$	-0.34815** (-2.4723)	
$\boldsymbol{W_{T,T}} * HSR$		-0.3028* (-2.0671)
$\boldsymbol{W_{NT,T}} * HSR$		-0.37355** (-2.6193)
control	yes	yes
$\boldsymbol{W} * control$	yes	yes
lambda	0.8786*** (37.946)	0.8790*** (38.11)
City FE	yes	yes
Year FE	yes	yes
Observation	3592	3592

Notes:t statistics reported in parentheses; *** p <0.01, **p <0.05, * p <0.1.

6.4. *Using different spatial weight matrix*

We replace the inverse geographic distance weight matrix with k-nearest weighted by the inverse Euclidean distance matrix. The result is displayed in Table 10. All the specifications demonstrate that the opening HSR could mitigate the labor misallocation for their own and vicinity, which verifies the robustness of the spatial DID result.

Table 10. Using different spatial weight matrix

	SDEM-DID (5k-nearest weighted)	
HSR	-0.0495*** (-3.7741)	-0.04174* (-1.9563)
$\boldsymbol{W} * HSR$	-2.333*** (-3.0537)	
$\boldsymbol{W_{T,T}} * HSR$		-2.652** (-2.5752)
$\boldsymbol{W_{NT,T}} * HSR$		-2.0925** (-2.2623)
control	yes	yes
$\boldsymbol{W} * control$	yes	yes
lambda	0.37956*** (19.161)	0.379681*** (19.169)
City FE	yes	yes
Year FE	yes	yes
Observation	3640	3640

Notes:t statistics reported in parentheses; *** p <0.01, **p <0.05, * p <0.1.

7. Conclusion

This article attempts to answer whether HSR alleviates the labor spatial misallocation to assess the overall economic effect of HSR. Meantime, we also emphasize the spatial spillover effect of HSR. Using China's HSR program as a quasi-natural experiment, the paper introduces a modified spatial DID approach to quantify three different types of treatment effects, including own treatment effect, spillover effect within the treatment group, and spillover effect from the treatment group to the control group. The conclusions and policy implications are as follows.

Firstly, the urban labor misallocation could be alleviated through opening HSR. In other words, the operation of HSR facilitates the flow of labor factors migrating to high-productivity cities, where labor is more likely to find matching jobs. Previous studies suggest that the siphon effect of HSR leads to the agglomeration of resources in large cities, which is detrimental to the development of small cities. In fact, from the perspective of the resource allocation effect, the siphon phenomenon caused by HSR is not necessarily harmful. On the contrary, it facilitates the efficient allocation of labor among cities and contributes to the overall productivity. Policymakers should take the resource allocation effect into account when planning the construction of HSR stations.

Secondly, the result of the spillover effect suggests that opening HSR alleviates the labor misallocation in the surrounding cities (including the treated and non-treated). And the spillover effect is significantly larger than the local effect. This spillover mechanism may occur due to the connectivity and transitivity of HSR. For cities connected to the HSR network, HSR shortens inter-city travel costs, which generate the cross-regional impact. For cities that are excluded from the HSR network, the labor can transfer to destinations through neighboring cities that opened HSR, thus obtaining the effect of opening HSR. This result implies that the HSR project may have a broader impact. Planners also need to consider the spillover effect of HSR and its impact on the ”non-treated group”. Cities without HSR should improve their supporting transportation infrastructure to connect with HSR cities and facilitate the movement of their labor force and other factors.

Thirdly, our study also finds that the spillover effect of HSR exists only in a certain range, which presents a U-shape curve with the increase of distance. Specifically, the mitigation effect of opening HSR on the labor misallocation to neighboring cities increases at first when distance below 350 km, then decreases, and disappears beyond 500 km. Therefore, planners need to consider the boundary of the affected city to optimize the layout of the HSR network. In general, the government should guarantee that each city is located no more than 500 km from the nearest HSR station in order to facilitate the free flow of labor, to mitigate the labor misallocation further. In addition, the optimal decision of the distance is 350 km.

Finally, the direction and magnitude of HSR effect depend on the urban scale. In our research, the local treatment effect, spillover effect within the treatment group, and spillover effect from the treatment group to the control group occur simultaneously when city’s population is larger than 2.94 or non-agricultural sector share exceeds

87.37% which account for about 60% of cities opened HSR. The scale of the population or non-agricultural sectors reflects the supply or demand of cities' labor force. Cities with larger labor supply or demand are better positioned to benefit from the HSR project. When considering the HSR station, planners also need to carefully evaluate and study potentially relevant factors, such as the city's population and industry scale.

Acknowledgments

Funding

Declarations of interest

Appendix Measurement of labor spatial misallocation

This paper consider a static allocation problem. The aggregate economy is composed of n cities,indexed by i = 1,...,n. We assume Cobb–Douglas production technologies with the same factor elasticities in all cities:

$$Y_i = A_i {K_i}^{\alpha} {L_i}^{1-\alpha} \tag{A.1}$$

Here Y_i , L_i , K_i and A_i are the real GDP, employment, capital stock and TFP in city i. We assume that the aggregate GDP is a CES function of the city's GDPs:

$$Y = (\sum_{i=1}^{n} \theta_i {Y_i}^{\sigma})^{\frac{1}{\sigma}} = AK^{\alpha}L^{1-\alpha} \tag{A.2}$$

θ_i represents the weight of urban output in the overall economic production, and $\sum_{i=1}^{n} \theta = 1$. It shows that the total economic output of the whole economy is generated by the coordination between urban output. σ is the elasticity of output substitution between cities. α is capital output elasticity.

Let $K = \sum_{i=1}^{n} K_i$ and $L = \sum_{i=1}^{n} L_i$ be the total capital stock and total employment. Let $\mathcal{K} = \frac{K_i}{K}$ and $\mathcal{L} = \frac{L_i}{L}$ be the shares of capital and employment. According to the

Eq.(A.1) and Eq.(A.2), we can get aggregate TFP, which is affected by the proportion of factor inputs.

$$A = \frac{Y}{K^{1-\alpha}L^{\alpha}} = [\sum_{i=1}^{n} \theta_i (A_i \mathcal{K}^{1-\alpha} \mathcal{L}^{\alpha})^{1-\sigma}]^{\frac{1}{1-\sigma}} \tag{A.3}$$

We need ensure that both aggregate output and urban output are maximised.

$$\pi_1 = \max_{Y_i}[PY - \textstyle\sum_{i=1}^{n} P_i Y_i]$$
$$\pi_2 = \max_{K_i, L_i}[P_i Y_i - \tau_i^l w L_i - r K_i]$$

w and r are the price of labor and the price of capital in an ideal state without market friction respectively. But Labor markets are not perfect mobile, and prices accepted by cities can be distorted. τ_i^l is labor wedges. P_i is the output price index of the city.

The first-order condition of total output profit maximization in π_1 is:

$$pY^{1-\sigma} \theta_i Y_i^{\sigma-1} = p_i \tag{A.4}$$

The first-order condition of city output maximization in π_2 is:

$$(1-\alpha) P_i A_i K^{1-\alpha} L^{-\alpha} = \tau_i^l w \tag{A.5}$$

$$\alpha P_i A_i K^{\alpha-1} L^{1-\alpha} = r \tag{A.6}$$

By maximizing $\pi_1$and π_2, We can get the proportion of urban labor input in a distorted state:

$$\mathcal{L} = \frac{L_i}{L} = \frac{\theta_i^{\frac{1}{1-\sigma}} \widetilde{A_i}^{\frac{\sigma}{1-\sigma}} \tau_i^{l-1}}{\sum_{i=1}^{n} \theta_i^{\frac{1}{1-\sigma}} \widetilde{A_i}^{\frac{\sigma}{1-\sigma}} \tau_i^{l-1}} \tag{A.7}$$

Here $\widetilde{A_i} = \frac{A_i}{\tau^{l_i (1-\alpha)}}$.

when τ^{l_i} is equal to 1, we can get the proportion of urban labor input in the effective state:

$$\mathcal{L}^{\star} = \frac{L_i^{\star}}{L^{\star}} = \frac{\theta_i^{\frac{1}{1-\sigma}} A_i^{\frac{\sigma}{1-\sigma}}}{\sum_{i=1}^{n} \theta_i^{\frac{1}{1-\sigma}} A_i^{\frac{\sigma}{1-\sigma}}} \tag{A.8}$$

By the ratio of labor input in the effective state to labor input in the distorted state, we can get the distortion degree of labor input in city i:

$$m_i = \frac{\mathcal{L}^\star}{\mathcal{L}} \tag{A.9}$$

If m_i is closer to 1, it means that the actual labor input in the city is more ideal; if m_i is greater than 1, it means that the labor input in the city is insufficient; if m_i is less than 1, it means that the labor input in the city is excessive. so we use M_i to represent the degree of labor misallocation in city i:

$$M_i = \left| \frac{\mathcal{L}^\star}{\mathcal{L}} - 1 \right| \tag{A.10}$$

The larger M_i is, the more serious the labor misallocation is.

To measure M_i, we need to have measures of city-level TFP(A_i), city weights(θ_i), labour price wedge(τ_i^l).

In Eq. (A.4), θ_i can be obtained. Following Brandt et al. (2013), we choose average of θ_i over the entire period of 2004–2016 as city weights. θ_i is set as follows:

$$\theta_i = \frac{1}{13} \sum_{t=2004}^{2016} \frac{P_i Y_i^{1-\sigma}}{\sum_{i=1}^{n} P_i Y_i^{1-\sigma}}$$

we can identify the wedge and TFP from Eq.(A.5) and Eq.(A.6).

$$\tau_i^l \propto \frac{Y_i}{P_i L_i}$$

$$A_i = \frac{Y_i}{P_i K_i^\alpha L_i^{1-\alpha}}$$

The elasticity of output substitution between cities ($\sigma = \frac{1}{3}$) and output elasticity ($\alpha = 0.45$) were able to obtain directly from Brandt et al. (2013). GDP and capital stock and employment are obtained from the China city Statistical Yearbook, respectively. Employment is measured in terms of the number of urban workers. Real GDP is calculated using 2003 as the base period GDP through the GDP index of each city. The capital stock is obtained by referring to Zhang et al. (2004).

References

Ahlfeldt, G. M. and Feddersen, A. (2017). From periphery to core: Measuring agglomeration effects using high-speed rail. *Serc Discussion Papers*, 18(2):355–390.

Anselin, L., Bera, A. K., Florax, R., and Yoon, M. J. (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, 26(1):77–104.

Aoki, S. (2012). A simple accounting framework for the effect of resource misallocation on aggregate productivity. *Journal of the Japanese and International Economies*, 26(4):473–494.

Ashenfelter, O. (1978). Union relative wage effects: New evidence and a survey of their implications for wage inflation. *International Economic Association Series*.

Asturias, J., García-Santana, M., and Ramos, R. (2014). Misallocation, internal trade, and the role of transportation infrastructure. *PEDL Research*.

Banerjee, A. V. and Moll, B. (2010). Why does misallocation persist? *American Economic Journal Macroeconomics*, 2(1):189–206.

Bartelsman, E., Dobbelaere, S., and Peters, B. (2013). Allocation of human capital and innovation at the frontier: Firm-level evidence on germany and the netherlands. *Tinbergen Institute Discussion Papers*.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., and Zhang, Q. (2017). Roads, railroads, and decentralization of chinese cities. *Review of Economics and Statistics*, 99(3):435–448.

Bernard, A. B., Moxnes, A., and Saito, Y. U. (2019). Production networks, geography, and firm performance. *Journal of Political Economy*, 127(2):639–688.

Brandt, L., Tombe, T., and Zhu, X. (2013). Factor market distortions across time, space and sectors in china. *Review of Economic Dynamics*, 16:39–58.

Cao, J., Liu, X. C., Wang, Y., and Li, Q. (2013). Accessibility impacts of china's high-speed rail network. *Journal of Transport Geography*, 28:12–21.

Chagas, A. L., Azzoni, C. R., and Almeida, A. N. (2016). A spatial difference-in-differences analysis of the impact of sugarcane production on respiratory diseases. *Regional Science and Urban Economics*, 59:24–36.

Chen, Z. and Haynes, K. E. (2015). *Chinese Railways in the Era of High-speed*. Emerald Group Publishing Limited Bingley.

Chen, Z., Xue, J., Rose, A. Z., and Haynes, K. E. (2016). The impact of high-speed rail investment on economic and environmental change in china: A dynamic cge analysis. *Transportation Research Part A: Policy and Practice*, 92:232–245.

Delgado, M. S. and Florax, R. J. (2015). Difference-in-differences techniques for spatial data: Local autocorrelation and spatial interaction. *Economics Letters*, 137:123–126.

Dong, X., Zheng, S., and Kahn, M. E. (2019). The role of transportation speed in facilitating high skilled teamwork across cities. *Journal of Urban Economics*, 115:103–212.

Duflo, E. and Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, 65(3):557–586.

Elhorst, J. P. (2014). *Spatial econometrics: from cross-sectional data to spatial panels*. Springer.

Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from china's national trunk highway system. *Review of Economic Studies*, 3:1046–1070.

Fu, S. and Gu, Y. (2017). Highway toll and air pollution: evidence from chinese cities. *Social Science Electronic Publishing*, 83:32–49.

Ghani, E., Goswami, A. G., and Kerr, W. R. (2016). Highway to success: The impact of the golden quadrilateral project for the location and performance of indian manufacturing. *The Economic Journal*, 126(591):317–357.

Givoni, M. (2006). Development and impact of the modern high speed train: A review. *Transport Reviews*, 26(5):593–611.

Hall, P. (2009). Tricky dilemmas for high speed rail. *Regeneration and Renewal*, page 14.

Helpman, E., editor (1998). *The size of regions. In: Topics in Public Economics: Theoretical and Applied Analysis*. Cambridge University Press.

Heuermann, D. F. and Schmieder, J. F. (2014). Warping space: High-speed rail and returns to scale in local labor markets. In *Annual Conference 2014 (Hamburg): Evidence-based Economic Policy*.

Hodgson, C. (2018). The effect of transport infrastructure on the location of economic activity: Railroads and post offices in the american west. *Journal of Urban Economics*, 104(MAR.):59–76.

Hsieh, C. T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics*, 4:1403–1448.

Hsieh, C. T. and Moretti, E. (2019). Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics*, 11(2):1–39.

Hu, Y., Deng, T., and Zhang, J. (2020). Can commuting facilitation relieve spatial misallocation of labor? *Habitat International*, 106(2):102–136.

Jedwab, R.and Moradi, A. (2016). The permanent effects of transportation revolutions in poor countries: evidence from africa. *Alexander Moradi*, 98(2):533–540.

Ji, S. H., Zhu, Y., and Zhang, X. (2016). The study of industrial agglomeration improving misallocation of resources. *China Industrial Economics*, 6(2):73–89.

Kim, H. and Sultana, S. (2015). The impacts of high-speed rail extensions on accessibility and spatial equity changes in south korea from 2004 to 2018. *Journal of Transport Geography*, 45:48–61.

Krugman (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3):483–499.

Li, P., Lu, Y., and Wang, J. (2016). Does flattening government improve economic performance? evidence from china. *Journal of Development Economics*, 123:18–37.

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

Liu, S., Wan, Y., and Zhang, A. (2020). Does china's high-speed rail development lead to regional disparities? a network perspective. *Transportation Research Part A: Policy and Practice*, 138:299–321.

Moll, B. (2014). Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, 104(10):3186–3221.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Qin, Y. (2017). 'no county left behind?' the distributional impact of high-speed rail upgrades in china. *Journal of Economic Geography*, 17(3):489–520.

Rietveld, P. and Bruinsma, F. (1998). The accessibility of cities in european infrastructure networks. *Springer Berlin Heidelberg*, pages 115–139.

Rubin, D. B. (1978). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(05):688–701.

Sasaki, K., Ohashi, T., and Ando, A. (1997). High-speed rail transit impact on regional systems: does the shinkansen contribute to dispersion? *Annals of Regional Science*, 31(1):77–98.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.

Tobler, W. R. (1970). A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240.

Ureña, J. M., Menerault, P., and Garmendia, M. (2009). The high-speed rail challenge for big intermediate cities: A national, regional and local perspective. *Cities*, 26(5):266–279.

Vega, S. H. and Elhorst, J. P. (2015). The slx model. *Journal of Regional Science*, 55(3):339–363.

Verbitsky-Savitz, N.and Raudenbush, S. (2012). Causal inference under interference in spatial settings: A case study evaluating community policing program in chicago. *Epidemiologic Methods*, 1(1).

Vickerman, R. (2015). High-speed rail and regional development: the case of intermediate stations. *Journal of Transport Geography*, 42:157–165.

Xu, M. Z. (2017). *Riding on the new silk road: Quantifying the welfare gains from high-speed railways*. PhD thesis, University of California Davis.

Zhang, J., Wu, G., and Zhang, J. (2004). The estimation of china's provincial capital stock: 1952—2000. *Economic Research Journal*, 10:35–44.

Zhang, K. Z. and Tao, D. J. (2016). Economic distribution effect of transportation infrastructure – evidence from the opening of high-speed railway. *Economic dynamics*, 6:62–73.

Zhang, M. T. (2018). High-speed railways, market access and enterprises' productivity. *China Industrial Economics*, 5:137–155.