Are cities with HSR more attractive for labour settlement? Micro-evidence from China Migrants Dynamic Survey *

By Jingwen Tan and Lishuang Huang

This paper empirically analyses the impact of cities with high-speed railways(HSR) on the rural migrants' willingness to settle in the working city using data from the 2016 and 2017 China Migrants Dynamic Survey. The results of the linear probability model (LPM) show that cities with HSR can increase the intention to settle of the migrants by 5.45%. The counterfactuals were constructed using three matching models in this paper: Entropy Balancing Matching (EBM), Coarsened Exact Matching (CEM) and Propensity Score Matching (PSM), all of which supported the baseline results. Two geographical variables, elevation and slope, were also used as instruments for the HSR cities to avoid potential endogeneity affecting the robustness. In addition, the mechanism analysis shows that the greater the number of relatives accompanying the migrants, the less willingness to settle is affected by HSR, while the greater the distance they move or the more they own land assets in their hometown, the greater their willingness to settle is affected by HSR. The analysis of regional heterogeneity shows that the HSR in cities in Northeast China does not have a significant effect on migrants' settlement intentions, while this mechanism is significant in all other regions of China. This paper provides micro-level evidence of this mechanism by examining the impact of transport infrastructure on migrants settlement intentions in China in both theoretical and empirical dimensions.
JEL: J60, J61, J68, R41, R53
Keywords: Migrants, High-speed Railways, Settlement Intention, China

I. Introduction

Urbanization is a necessary path for developing countries to transform into developed countries (Chen et al., 2013). The concentration of industries and economic activities in areas with easy access to transportation has led to a concentration of rural migrants in large cities (Gakenheimer, 1999). In 2022, China's urban migrant population reach 283 million, with the migrant population making a huge contribution to urban development as the main workforce in cities (CNBS, 2023). Studying the willingness of the migrant population to settle can transform the labour force that stays in the city for a short period of time into a permanent one, providing a more stable human resource for economic growth.

Why do large cities attract large migrant populations? Global development experience shows that transport is a prerequisite for economic and social development and population

* Jingwen Tan: University of Copenhagen (tjw@henu.edu.cn). Lishuang Huang: Jinan University (lishuanghuanggz@163.com). The authors acknowledge the data support provided by the National Health Care Commission of China.

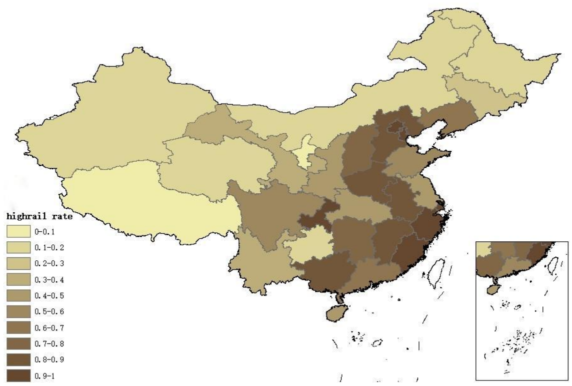

Figure 1. : Proportion of cities with HSR

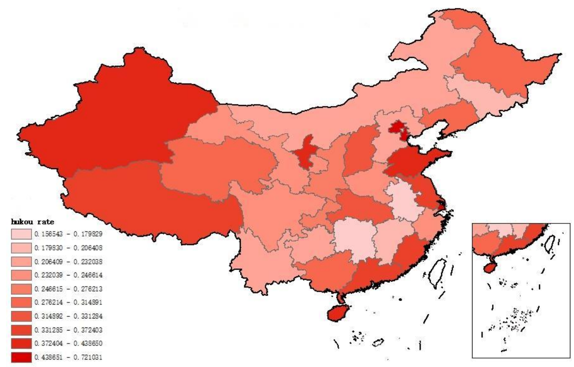

Figure 2. : Migrants with settlement intention

mobility. From traditional maritime transport (Clark et al., 2004), rail transport (Donaldson, 2018), road transport (Faber, 2014), to modern air transport (Zhang and Zhang, 2016) and HSR transport (Qin, 2017), the level of transport in a city not only affects the level of economic development, but also influences the migration choices of the migrant population (Bernard et al., 2019). Transport infrastructure has a catalytic effect on the economic growth of cities (Gospodini, 2005). A common view is that a well-developed transport infrastructure can indirectly contribute to local economic growth through mechanisms such as improved investment and savings (Aschauer, 1989). As a city's transport infrastructure builds economic growth as well as environmental changes, there will be an influx of more labour into the city, and building up human capital for the city (Desmet and Rossi-Hansberg, 2013). Based on this, this paper explores the factors influencing the settlement intentions of the migrant population from the perspective of urban transport measures.

Over the past decade, the Chinese government has invested heavily in the HSR network. The maximum design speed of China's HSR has reached 350 kilometres per hour, with an average operating speed of around 250 kilometres per hour (Yuan et al., 2023). The construction of HSR has effectively increased the convenience of travelling long distances and reduced the time cost of travelling. By 2022, the operating mileage of China's railway will reach 155,000 kilometres, of which the operating mileage of HSR will reach 42,000 kilometres, accounting for 27.10% of the total railway traffic (CNRB, 2023). By 2035, China plans to achieve a national railway network of 200,000 km, including about 70,000 km of HSR. cities with a population of more than 200,000 will be covered by railways, and cities with a population of more than 500,000 will be covered by HSR (CNRG, 2020). The speed, convenience and reliability of HSR provide more choices and opportunities for population mobility, and reduce the economic distance between migrants travelling to and from their work cities and hometowns(Hanley et al., 2022). In Figure 3, it can be seen that there is a positive correlation between the access rate of HSR at the provincial level and the willingness of the migrant population to settlement.

This paper empirically analyses the impact of cities with HSR on the rural migrants' willing-

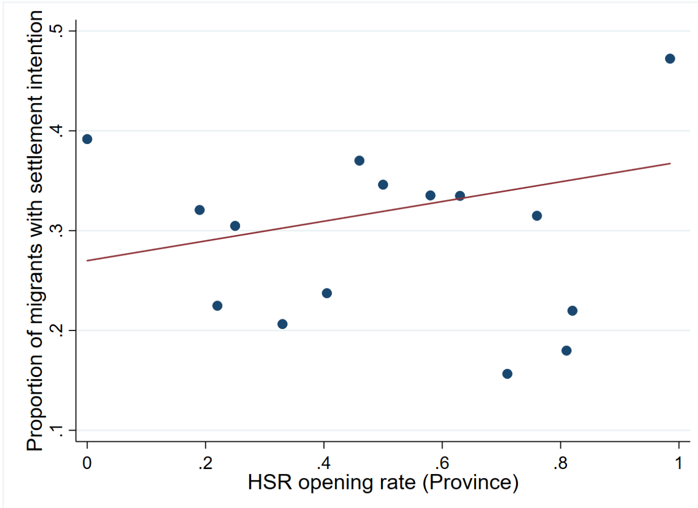

Figure 3. : HSR access rate and migrant settlement intention

ness to settle in the working city using data from the 2016 and 2017 China Migrants Dynamic Survey. The results of the linear probability model show that cities with HSR can increase the intention to settle of the migrants by 5.45%. In addition, the mechanism analysis shows that the greater the number of relatives accompanying the migrants, the less willingness to settle is affected by HSR, while the greater the distance they move or the more they own land assets in their hometown, the greater their willingness to settle is affected by HSR. The analysis of regional heterogeneity shows that the HSR in cities in Northeast China does not have a significant effect on migrants' settlement intentions, while this mechanism is significant in all other regions of China.

The value and marginal contributions of this paper are: Firstly, while most studies focus on the impact of HSR on macroeconomic indicators, this paper focuses on China's large migrant population and uses micro-survey data to enrich the research on the micro-behaviour of the population due to HSR construction, concluding that the development of transport infrastructure construction is conducive to enhancing the settlement intentions of the migrant population, which can transform the migrant population into a long-term urban labour force and This will facilitate the urbanisation process. Secondly, this paper investigates the mechanism of the impact of China's HSR construction on the settlement intentions of the migrant population from both theoretical and empirical perspectives, identifying two mechanisms: homesickness and hometown land assets. Additionally, the paper finds that there is heterogeneity in the impact of transport infrastructure on the settlement intentions of the migrant population across different groups, such as regional differences, providing more refined empirical evidence for policy evaluation.

The paper is structured as follows: Part I introduces the background related to urban basic transport facilities and the settlement intention of the migrant population, Part II conducts a relevant literature review, Part III details the research design, including theoretical model setting and data sources, Part IV introduces the basic data variables, Part V presents the empirical results, and Part VI contains conclusions and recommendations.

II. Review of the Literature

Previous studies have shown that the access of HSR reduces the cost of inter-city commuting, which facilitates information dissemination and knowledge spillovers at the level of face-to-face communication, and reduces the social costs and investment risks of business innovation (Donaldson and Hornbeck, 2016). Not only that, HSR significantly reduces the cost of inter-city trade (Bernard et al., 2019). HSR improves regional innovation levels and widens the innovation gap between HSR cities and non-HSR cities by enhancing the flow effect of innovation factors, and talent mobility is the most important pathway for inter-regional knowledge spillovers. The migrant population faces the obstacles of urban-rural dualistic structure and household registration system in the development of cities, and the trade-off between returning to settle in their hometowns and moving to big cities is affected by the external infrastructure, among which the city's basic transport facilities are an important factor (Liu, 2005; Bosker et al., 2012). Most of the existing literature has been studied from the development of urban economy and its impact on human capital, and less literature has been studied from the impact of the construction of HSR on the willingness of the labour force to migrate, especially based on the current background of China's huge migrants.

A large number of studies have been conducted to examine the factors influencing the willingness of migrant populations to settle. Among these, economic factors are the main ones that influence the willingness of migrants to settle. The 'share, match and learn' effect of large cities results in higher productivity, higher skills matching, and higher accumulation of human capital and wage premiums (Glaeser and Mare, 2001; Duranton and Puga, 2019). Under the urban-rural binary system, there is a large income gap between rural and urban areas (Au and Henderson, 2006; Zhang et al., 2016), and migrants can have better employment opportunities and higher wages in big cities. Secondly, superior public services in cities (Tiebout, 1956)and a more comfortable urban environment (Banzhaf and Walsh, 2008)are also factors that influence the migration of migrants. Among them, education, as an important route to class mobility, is closely related to labour migration decisions (Dustmann and Glitz, 2011; Xiong, 2015), and in order to gain access to better educational resources, migrant populations move to big cities and hope to obtain urban hukou (Chan and Buckingham, 2008; Chen and Feng, 2013). The (Vortherms and Liu, 2022) study found that migrant workers in Beijing were willing to give up between 9% and 14% of their income within five years in order to obtain a local hukou. This shows that better development opportunities and quality public services in the city have a greater impact on the willingness of the migrant population to settle.

Under China's special household registration system, a large body of literature has explained the motivations for the willingness of migrant populations to settle in terms of the value of urban household registration(Song, 2014; Chan, 2015; Zhang et al., 2020; Vortherms and Liu, 2022). The household registration system poses many barriers to migration(Chan and Buckingham, 2008; Bosker et al., 2012), for example, the existence of household registration discrimination in cities affects family reunification, and the problem of children left behind is highlighted(Fan et al., 2011; Fang and Shi, 2018). Family reunification is closely related to commuting costs and spatial distances between urban and rural areas, with ease of access to transport between cities facilitating population mobility, reducing commuting costs (Hodgson, 2018), and leading to talent pooling and skill partitioning (Fretz et al., 2017). At the same time, the land rights and benefits associated with rural hukou are seen as increasingly valuable. As a result, many migrant workers choose to move back and forth between urban and rural areas rather than give up their rural hukou (Chen and Fan, 2016). Transport facilities can improve the cost of commuting to and from their residence and workplace when the migrants is separated (Monte et al., 2018). However, the existing literature ignores another important characteristic fact in the process of population migration in China, namely the fact that the fixed costs of population migration differ between regions and that basic transport facilities are an important factor affecting the fixed costs of migration (Ma et al., 2021).

Based on the current duality of China's household registration system, public services in large cities differ in terms of transport, education and healthcare. Such disparities limit equal rights and opportunities for the migrants in accessing public services (Zhou and Cheung, 2017). The construction of transport infrastructure between cities can, to a certain extent, alleviate the geographical segregation caused by the household registration system. In a spatial equilibrium model, an individual's decision to move depends on the equilibrium between public services and the cost of residence, with labour moving between regions until the individual's utility reaches equilibrium (Rosen, 1979; Roback, 1982). As an important part of public services, transport facilities affect the willingness of the migrants to settle depending on the individual utility. This paper explores the mechanism of the impact of transport infrastructure construction on the willingness of the population to settle from the perspective of two aspects of individual utility of the migrants, family reunion and hometown land assets, which can fill the gap of the current study.

HSR plays an important role in China's transport facilities. Studies have been conducted to examine the effects of HSR construction from different perspectives, including urban transformation(Ahlfeldt and Feddersen, 2018; Fang et al., 2020), economic spillover and innovation(Wang and Cai, 2020; Miao and Zhang, 2019). In terms of human capital, the construction of HSR is conducive to increasing the local spirit of innovation(Ma et al., 2021), knowledge diffusion and idea spillover(Dong et al., 2020). Some scholars have concluded that HSR has a significant role in promoting regional economic growth(Ahlfeldt and Feddersen, 2018; Chen and Silva, 2013), while some literature has pointed out that HSR construction has exacerbated the phenomenon of uneven regional development(Faber, 2014). From the perspective of labour supply and demand, HSR infrastructure construction affects population migration and regional employment distribution(Kim, 2000) and has a significant driving effect on regional teamwork and technological innovation(Dong et al., 2020; Wang and Cai, 2020) .

Currently, there is more research on the impact of HSR construction on macroeconomic indicators, and less literature on the impact of HSR construction on the micro-behaviour of the population, and even less research based on the context of migrants. This paper clarifying the relationship between the construction of HSR and the willingness of migrants to settle, which can help to promote the government's rational allocation of urban basic transport resources and the coordinated development of regional population mobility and economy.

III. Methodology

A. Theoretical Model

The theoretical framework of this paper is based on the dynamic spatial structure model of Desmet and Rossi-Hansberg (2013). The model incorporates spatial geographical factors and investigates the impact of population clustering, infrastructure development situation on economic development. The model is based on a dynamic system and has a higher degree of flexibility compared to Allen and Arkolakis (2014)'s static spatial economic model. In the model, the consumer's utility is determined by both consumption and the quality of life in the place of residence, and the consumer's utility function can be expressed as

$$u_t^i\left(\bar{r}_-, r\right) = a_t(r)\left[\int_0^1 c_t^\omega(r)^\rho d\omega\right]^{\frac{1}{\rho}} \varepsilon_t^i(r) \prod_{s=1}^{t} m\left(r_{s-1}, r_s\right)^{-1} \tag{1}$$

In the equation, $1/(1-p)$ is the commodity elasticity of substitution, where $0 < p < 1$. $c(r)$ denotes the consumption of the commodity ω by consumers in region r in period t. $m\left(r_{s-1}, r_s\right)$ denotes the cost of moving consumers from region r_1 to r.

$$a_t(r) = \bar{a}_t(r)\bar{L}_t(r)^{-\lambda} \tag{2}$$

where $a_t(r)$ denotes the livability of region r, and $a_t(r) > 0$ is an exogenously given continuous function. $L_t(r)$ denotes the population density of region r in period t. λ is a fixed parameter and $\lambda > 0$.

$$Pr\left[\varepsilon_t^i(r)z\right] = exp^{-z^{-\frac{1}{\Omega}}} \tag{3}$$

The $\varepsilon_t^i(r)$ indicates the attitude of consumer i towards their region r in period t, which may be either positive or negative. The parameter $\Omega < 1$, with larger values of Ω implying larger differences in consumers' likes and dislikes towards region r.

Transport infrastructure development has an important impact on the assumptions of the dynamic spatial structure model. The development of transport infrastructure facilitates travel for local residents, the migrant population the city then has a higher quality of life (Klinger et al., 2013), livability $a_t(r)$ goes up or remains in equilibrium; consumer attitudes towards the local area $\varepsilon_t^i(r)$ become more positive; at the same time the cost of moving between consumer regions $m\left(r_{s-1}, r_s\right)$ decreases. In the short run, consumer consumption is not directly affected by the development of transport infrastructure and $c_t^\omega(r)$ does not change, but in the long run it may be affected by economic growth and increase in income and consumption. In summary, the level of utility of consumers rises and budget constraints increase. The impact of transport infrastructure development on the consumer's utility function is as follows:

$$(4) \qquad \overbrace{u_t^i(\bar{r}_-, r)}^{increase} = \overbrace{a_t(r)}^{increase\ or\ equilibrium} \overbrace{\left[\int_0^1 c_t^{\omega}(r)^{\rho} d\omega\right]^{\frac{1}{\rho}}}^{unchange} \overbrace{\varepsilon_t^i(r)}^{increase} \overbrace{\prod_{s=1}^{t} m(r_{s-1}, r_s)^{-1}}^{increase}$$

The development of transport infrastructure can improve the livability of the local area and attract more consumers; at the same time, the utility level of consumers will increase, the total local consumption will expand and there will be a positive impact on the local economic development. Within this research framework, this paper proposes the research hypothesis that there is a positive relationship between the willingness of the migrant population to stay in the incoming city and the situation of transport infrastructure (the access of HSR).

B. Data

This paper uses data from the 2016 and 2017 China Migrants Dynamics Survey (CMDS), organised by the National Health Planning Commission for empirical analysis. The survey covered 31 provinces (municipalities and autonomous regions) in mainland China, and the sample was selected from the migrant population who stayed in the area for more than one month. The average age of the sample ranged from 15-59 years. A stratified, multi-stage, large-scale PPS (Probability Proportionate to Size Sampling) sampling method was used to investigate in detail the developmental status, individual characteristics, social integration and employment of the migrant population in China. Because the questionnaire on intention to settle was only available in the CMDS 2016 and 2017 data, only data from these two years were selected for analysis in this paper. The data relating to the urban control variables were obtained from the statistical yearbooks of the cities.

In this paper, the data were cleaned before empirical analysis. First, we excluded all samples under the age of 18 as well as the sample of women over 50 and the sample of men over 55. The core explanatory variable in this paper is the household settlement willingness, according to the household settlement policy in some regions of China, household migration requires adult citizens who are over 18 years old as well as five years before retirement, which in China is 60 years old for men and 55 years old for women. Secondly, the population studied in this paper is the migrants in city, and this paper retains a sample of rural hukou (household registration). Finally, we matched the micro-questionnaire data of the sample with the data from the urban statistical yearbook and retained the successfully matched sample. The processed sample size was distributed among 292 municipalities, with a total of 238,406.

C. Baseline Model

This paper focuses on the impact of the availability of HSR in the cities to which the migrant population flows on their willingness to settle, with the explanatory variable being their willingness to settle. The question "Are you willing to move your household registration to the local area" in the CMDS questionnaire, is used to measure the willingness of the migrant population to settle. This indicator is a dummy variable (1=willing, 0=unwilling).

China's household registration system, also known as the "hukou system", is a system of population management based on family ties. Every Chinese citizen has a household register (also known as a "hukou"), which is used to record basic personal information about that citizen. Under the hukou system, there are two types of hukou for Chinese citizens: rural hukou and urban hukou. These two types of hukou have different entitlements and policies in terms of social welfare, education and healthcare. Moving the hukou requires a certain application and approval process, as well as a certain amount of time, effort and money, so being willing to move the hukou also indicates that the person is more willing and determined to settle in that city.

With both the dependent and core independent variables as dummy variables, models that can be used include linear probability models based on least squares estimation and Logit, Probit models based on likelihood estimation. This paper chooses to use a linear probability model in the baseline regression to estimate the effect of city ownership of HSR on the intention to settle among the migrant population, while the estimation results of the Logit and Probit models are reported in the robustness check section. The interpretation of the coefficients of the linear probability model is more intuitive than that of the likelihood estimation. And with a reasonable model setup, the direction and significance of the coefficients obtained using least squares estimation are not significantly different from that of the likelihood estimation (Ferrer-i-Carbonell and Frijters, 2004; Angrist and Pischke, 2009). The linear probability model in this paper can be expressed as

$$settle_{it} = \beta_0 + \beta_1 hsr_{it} + \beta_2 X_{it} + \mu_i + \lambda_t + (\mu_i \times \lambda_t) + u_{it} \tag{5}$$

where $settle_{it}$ is the willingness of the migrant population to settle, and this variable is a dummy variable. hsr_{it} is the HSR access dummy variable, which equals 1 if city i opens HSR in year t and 0 otherwise. X_{it} are other control variables. μ_i is the regional fixed effect, λ_t is the time fixed effect, and $\mu_i \times \lambda_t$ is the cross result of the time fixed effect and the regional fixed effect, i.e. the interaction fixed effect. As the interaction fixed effects generate more dummy variables in the estimation, the use of municipal-level interaction fixed effects may generate more serious multicollinearity problems. Therefore, this paper chooses to control for the interaction fixed effects at the provincial level, while putting in control variables at the local municipality level for estimation.

D. Matching Model

Linear probability model may not address the problem of estimation bias due to 'sample self-selection'. In addition, the core explanatory and explanatory variables in this paper are dummy variables, and this data structure is more suitable to use matching models to statistically construct a 'counterfactual' environment to obtain more accurate estimates. In this paper, the entropy-balanced matching model(EBM) is chosen to establish a counterfactual framework: for a rural-urban migrant labourer with a rural household registration, how the migrant's propensity to settle would change, assuming that the incoming city has HSR, while balanced other personal characteristics and the characteristics of the inflow city.

Among the commonly used matching models, in addition to EBM, the more common models are Propensity Score Matching (PSM) and Coarse Exact Matching (CEM). Compared to PSM and CEM, there are three main advantages of EBM **?**. First, EBM ensures that the experimental and control groups are similarly distributed over the higher order moments of each covariate, whereas PSM methods that focus on balancing the propensity scores of the experimental and matched control groups do not ensure this effect. Secondly, in the implementation of EBM, the only parameters that need to be set subjectively by the researcher are almost the tolerance of the results of the iterative algorithm, which circumvents the disadvantages of having too many subjective settings in the PSM and CEM implementations, including but not limited to: variable selection and model setting for propensity score calculation in the PSM phase, and whether observations are put back to matching. In addition, EBM is able to retain all samples for estimation, and a larger number of samples are lost in the PSM and CEM matching process.

The total average treatment effect of interest when using EBM to construct a counterfactual framework equation can be represented by

$$\tau = E[settle(1) \mid hsr = 1] - E[settle(0) \mid hsr = 1] \tag{6}$$

$E[settle(1) \mid hsr = 1]$ indicates the size of the effect of having HSR in the cities of the treatment group on the willingness of the migrant population to settle. $E[settle(0) \mid hsr = 1]$ denotes the size of the effect of the absence of HSR in the cities of the treatment group on the willingness of the migrant population to settle, and is a counterfactual indicator. the expression for CEM is

$$E[settle(0) \mid hsr = 1] = \frac{\sum_{i|hsr=0} settle_i \omega_i}{\sum_{i|hsr=0} \omega_i} \tag{7}$$

$$\min H(\omega_i) = \sum_{i|hsr=0} \omega_i \log(\omega_i / q_i) \tag{8}$$

where $q_i = 1/n$ and n denotes the number of samples in the control group. The core idea of ω_i with certain constraints is that the control covariates are given an additional set of moment constraints to balance with the covariates in the treatment group. To ensure the rigour of the conclusions, the PSM, CEM estimates are also reported in this paper when the entropy balance matching results are reported.

E. Instrumental Variable

The core explanatory variable for measuring the impact of HSR construction on the willingness of the migrants to settle down is likely to be an endogenous variable. When the willingness of migrants to settle down in the economically developed region is high, the potential movement of people may increase local passenger traffic and prompt the local government to open HSR, thus creating a possible two-way causality endogeneity problem in the prediction and affecting the unbiasedness of the estimates. Instrumental variables are widely used in the treatment of two-way causality problems.

The use of geographical instrumental variables is well established in empirical studies targeting transport infrastructure. In this paper, we choose a city's average slope and elevation as the instrumental variables for the access of a city's HSR. The average slope and elevation of a city affect the construction of HSR, which is consistent with the correlation hypothesis of the instrumental variable; at the same time, the average slope and elevation of a city are not considered in the integration process of the migrants in the inflow area, which is consistent with the exogeneity hypothesis of the instrumental variable. Based on this instrumental variable, this paper performs regression analysis in two stages OLS. The first stage expression is

$$hsr_{it} = \varphi_0 + \varphi_1 altitude_i + \varphi_2 slope_i + \psi_{it} \tag{9}$$

where hsr_{it} is the city's highrail access, $altitude_i$ is the average elevation of the city, $slope_i$ is the average slope of the city, and ψ_{it} is a random error term.

IV. Variables and Descriptive Statistics

The dependent variable in this paper is a dummy variable in the form of willingness to stay of the migrants (settle), and the core explanatory variable in this paper is the dummy variable of whether the city with HSR (highrail). The control variables are selected along two dimensions: individual and city.

The demographic characteristics of the sample are often used as control variables. In human capital theory, factors such as education level and work experience can help individuals to obtain better jobs and income, thus enabling them to integrate into the inflow cities. In this paper, gender (gender), years of education (edu), age (age) and marital status (marriage), whether or not one has a job (job), and monthly household income (lnincome)) are chosen as individual control variables.

City characteristics may affect the intention of the migrants to stay in the inflow area. The urban control variables selected in this paper mainly reflect the level of economic development and other infrastructure development in the inflowing city. They mainly include the number of tertiary sector employees (tpeople), a measure of the city's industrial establishment; logarithmic gross domestic product per capita (lnpgdp), a measure of the city's economic situation; the number of hospital beds per 1,000 population (pbed), a measure of the city's health care resources; and the number of primary school teachers per 1,000 population (pteacher), a measure of the city's educational resources.

Table 1 reports the results of t-tests for differences in means between groups grouped according to the dummy variables for the access of the HSR. It can be seen that all of the control variables in this paper are significant in the t-test of differences in means between groups, demonstrating that the control variables can better control for the effects of factors other than the access of the HSR on the willingness of the migrants to settle in the city.

Table 1—: t-test for Mean Between Groups.

Variables	N	Mean	N	Mean	Difference	p-value
	hsr=0		hsr=1			
gethukou	39689	0.25	198717	0.35	0.10	0.00***
gender	39689	0.53	198717	0.53	0.00	0.09*
age	39689	35.03	198717	34.32	0.71	0.00***
edu	39689	9.23	198717	10.03	0.81	0.00***
marriage	39689	0.82	198717	0.81	0.01	0.00***
job	39689	0.81	198717	0.86	0.06	0.00***
lnincome	39689	8.5	198717	8.69	0.20	0.00***
pteacher	39689	4.51	198717	4.53	0.02	0.09*
pbed	39689	5.06	198717	6.62	1.56	0.00***
lnpgdp	39689	10.37	198717	11.14	0.77	0.00***
tpeople	39689	56.77	198717	52.36	4.41	0.00***

Note: p-values are t-test results, $***p < 0.01, **p < 0.05, *p < 0.1$..

V. Empirical Analysis and Discussion

In the first part of the empirical study, this paper reports on the impact of HSR cities on the intention to settle of the migrants. According to the regression results in Table 2, there is a significant positive effect of city with HSR on the settlement intention of the migrants in all cases. Specifically, without control variables, the treatment effect of HSR cities on the intention to settle of the migrants is 9.66% (Model 1). After adding the city and individual control variables, the treatment effect of HSR cities on the willingness of the migrants to settle decreases to 5.37% (Model 4). The coefficient rebounded slightly after adding the interaction fixed effects. Following the results of model 5 with the inclusion of control variables and interaction fixed effects, the proportion of migrants having the intention to settle is on average 5.45% higher in cities with inflowing cities with HSR than in cities without HSR. All of the above results reached a confidence level of 99%.

A city's well-developed transport infrastructure can improve travel efficiency. People usually consider employment, education and healthcare when choosing where to live, and cities with convenient and efficient transport infrastructure can offer more options, such as relying on easy access to the resources of neighbouring cities. At the same time, the construction of transport infrastructure can drive economic development, attract more talent and capital inflow, and promote the upgrading and transformation of the city's industrial structure. As a result, the migrants is more willing to settle in cities with developed transport infrastructure.

Linear probability model may not address the problem of estimation bias due to 'sample self-selection'. In addition, the core explanatory and explanatory variables in this paper are dummy variables and this data structure lends itself to the use of matching models to statistically construct a 'counterfactual' environment to obtain more accurate estimates.

Table 2—: Linear Probability Model Regression Results.

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) OLS
hsr	0.0966***	0.0762***	0.0692***	0.0537***	0.0545***
	(0.00243)	(0.00247)	(0.00261)	(0.00263)	(0.00317)
gender		-0.0193***		-0.0186***	-0.0176***
		(0.00202)		(0.00199)	(0.00193)
age		0.000823***		0.000575***	0.000801***
		(0.000131)		(0.000130)	(0.000127)
edu		0.0149***		0.0130***	0.0143***
		(0.000370)		(0.000365)	(0.000361)
marriage		0.0125***		0.0183***	0.0153***
		(0.00290)		(0.00287)	(0.00285)
job		-0.0346***		-0.0246***	-0.0260***
		(0.00287)		(0.00284)	(0.00279)
lnincome		0.0566***		0.0514***	0.0264***
		(0.00191)		(0.00187)	(0.00187)
pteacher			0.00699***	0.00779***	0.00542***
			(0.000607)	(0.000604)	(0.000915)
pbed			0.0116***	0.0106***	0.00799***
			(0.000575)	(0.000569)	(0.000798)
lnpgdp			0.0372***	0.0345***	0.0112***
			(0.00138)	(0.00135)	(0.00157)
tpeople			0.00439***	0.00427***	0.00248***
			(0.0000725)	(0.0000723)	(0.000107)
constant	0.252***	-0.367***	-0.473***	-1.000***	278.9***
	(0.00218)	(0.0165)	(0.0147)	(0.0206)	(37.77)
Individual Variable	No	Yes	No	Yes	Yes
City Variable	No	No	Yes	Yes	Yes
Interaction FE	No	No	No	No	Yes
N	238406	238406	238406	238406	238406
R^2	0.006	0.020	0.037	0.048	0.103

Note: $***p < 0.01, **p < 0.05, *p < 0.1$.

Model (6) in Table 3 shows the results of the EBM estimation, with the coefficient increasing slightly from 5.45% to 6.84% compared to the results of the baseline regression. The coefficients and significance of model (7) PSM and model (8) CEM also not change significantly from the baseline regression. PSM is using the 1:1 nearest neighbour matching method with twice the sample size of the control group and the nearest neighbour matching lost some samples due to the larger sample size in the treatment group. Similarly, the number of samples successfully matched by CEM is only 8150, giving up most of the samples. EBM adjusts the first-order moments of the treatment and control groups by weighting them and does not lose samples, making the conclusion more robust based on the perspective of sample integrity. The quality of the data after matching for all three methods has improved considerably compared with that before matching, please see the appdendix table for the relevant tests.

Table 3 also reports the results of regressions that include urban slope and elevation as instrumental variables for HSR cities. The results of the two-stage least squares (2SLS) method of model (9) indicate a treatment effect of 19% for HSR cities, with the potential endogeneity issue possibly underestimating the effect of city with HSR on the willingness of the migrants to settle. The estimation results of model (10) limited instrumental variable method of information (LIML) are the same as those of the 2SLS model, proving that there is no weak instrumental variable problem in the model. Model (11) generalised moment estimation (GMM) estimates also yielded results similar to those of 2SLS, proving that the results are robust. The exogeneity of the instrumental variables needs to be argued before proceeding with the instrumental variable regressions. There is no research that demonstrates that urban slope and elevation have an effect on the intention to settle among migrants, and this paper also performs over-identification tests for these two instrumental variables. The results of the over-identification test are reported in the Appendix, indicating that the two instrumental variables are exogenous (Sargan statistic=0.958, p-value=0.3276¿0.05).

Table 3—: Results of Matching Models and Instrumental Variables

	(6) EBM_OLS	(7) PSM_OLS	(8) CEM_OLS	(9) TSLS	(10) LIML	(11) GMM
highrail	0.0684***	0.0470***	0.0550***	0.190***	0.190***	0.188***
	(0.00494)	(0.00741)	(0.0169)	(0.00701)	(0.00701)	(0.00688)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Interaction FE	Yes	Yes	Yes	Yes	Yes	Yes
N	238406	62320	8150	238406	238406	238406
R^2	0.087	0.057	0.036	0.037	0.037	0.037

Note: $***p < 0.01, **p < 0.05, *p < 0.1$.

A. Robustness

To increase the rigour of the paper's findings, robustness tests were conducted in the following three areas: (1) Non-linear regression models. This paper replaces the linear probability model based on OLS for the main regression, and uses the Logit and Probit model based on the likelihood estimation for robustness testing. Probit model assumes that the random disturbance terms of the model obey a normal distribution, and the corresponding random disturbance terms of the Logit model are assumed to obey a Logistic distribution, so this paper reports the results of both the Logit and Probit. (2) Robust standard errors for clustering using prefecture-level municipalities. Since the perturbation terms of the regression model may have heteroskedasticity and the perturbation terms of the same prefecture-level administrative region may be autocorrelated at different periods, cluster robust standard errors clustered at the prefecture-level administrative region are used for further estimation. (3) Replacement of explanatory variables. In order to avoid potential bias in the question "intention to get the household registertion", the dependent variable is replaced with the questionnaire's intention to long stay (whether or not they plan to live in the local area for more than 5 years). As with the main regression, the Logit, Probit, Cluster Robust Standard Error OLS models and the OLS model using intention to long stay as the dependent variable all include control variables and interaction fixed effects.

Table 4—: Robustness Tests

	(12)	(13)	(14)	(15)
	Logit	Probit	OLS_clustercity	OLS_longliving
highrail	0.311***	0.180***	0.054***	0.0163***
	(0.0185)	(0.0107)	(0.0134)	(0.00347)
Control Variables	Yes	Yes	Yes	Yes
Interaction FE	Yes	Yes	Yes	Yes
N	238406	238406	238406	238406
$(Pseudo)R^2$	0.081	0.0.103	0.193	0.193

Note: $***p < 0.01, **p < 0.05, *p < 0.1$.

As can be seen from the regression results in Table 4 for models (12) (13) using the Logit and Probit models for robustness, the sign of the core explanatory variables and significance unchanged. The results of model (14) using the cluster robustness standard errors for the prefecture-level cities also do not change significantly from the baseline regression. The above suggests that the findings of this paper are robust.

B. Excluding Competing Hypotheses

In order to accurately identify the impact of HSR on the social integration of migrant population, a competing hypothesis needs to be falsified: it is HSR, not other modes of transport affect the willingness of migrant population to settle. Cities with HSR tend to be economically developed and land flat, so the dummy variable for HSR access does not only reflect the construction of HSR, but may also contain 'information' about other transport infrastructure. For example, cities with HSR have a developed economy and are likely to have an airport; cities with HSR have flat terrain and good road infrastructure, etc. In addition, China's provincial capitals are generally more developed in terms of HSR lines, and the distance of a city from the provincial capital may also influence the willingness of the migrant population to settle.

To address both of these issues, this paper reports in Table 5 the results of putting in variables measuring water transport infrastructure (Model 16), air infrastructure (Model 17), general railway infrastructure (Model 18), and road infrastructure (Model 19), respectively, and controlling for the distance from each city to the provincial capital (Model 20). It can be seen that the coefficients on the HSR variable (hsr) do not change significantly, either by putting in the other transport infrastructure and the distance from the provincial capital variables separately (models 16-20) or by controlling for all of these variables (model 21). As passenger transport by water, air and general rail is public transport, this paper uses logarithmic passenger traffic to measure the level of infrastructure for these three modes of transport. For road transport, total road mileage is used to measure the level of infrastructure.

Table 5—: Excluding Competing Hypotheses

	(16) OLS	(17) OLS	(18) OLS	(19) OLS	(20) OLS	(21) OLS
highrail	0.0538***	0.0495***	0.0548***	0.0538***	0.0552***	0.0481***
	(0.00317)	(0.00316)	(0.00317)	(0.00326)	(0.00318)	(0.00333)
lnwater	0.0118***					0.0108***
	(0.000584)					(0.000613)
lnair		0.00979***				0.00538***
		(0.000449)				(0.000498)
lnrail			0.000794*			-0.00188***
			(0.000461)			(0.000490)
lnroad				-0.000879		-0.00250**
				(0.000894)		(0.00103)
lndistance					0.00130	-0.00146
					(0.00102)	(0.00115)
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Interaction FE	Yes	Yes	Yes	Yes	Yes	Yes
N	238406	238406	238406	238406	238406	238406
R^2	0.105	0.105	0.103	0.104	0.105	0.110

Note: $***p < 0.01, **p < 0.05, *p < 0.1$.

C. Mechanism Analysis

The mechanism of the effect of transport infrastructure on the willingness of the migrant population to settle down is influenced by a variety of factors, and this paper focuses on two perspectives: homesickness and hometown assets. The first is the homesickness of the migrant population, which can be reflected by two factors: the number of relatives accompanying mobility and the distance mobility. When the migrant population has more relatives

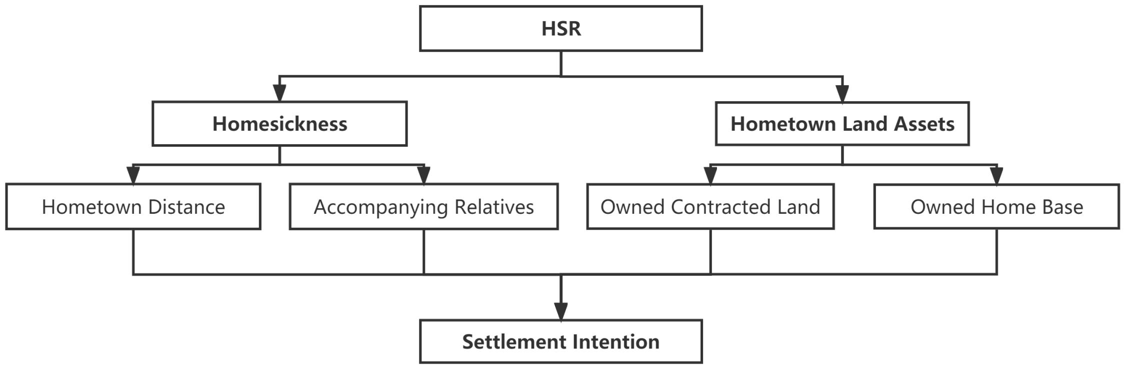

Figure 4. : Mechanism analysis

accompanying them, the need to return hometown to visit relatives is less strong because their relatives are close by, and the need for transport infrastructure is reduced. The distance of mobility can also have an impact on the demand for transport infrastructure. When the migrant population is further away from home, there is likely to be a greater emphasis on long-distance rapid transport infrastructure. Secondly, the ownership of land assets in the home town of the migrant population. When the migrant population owns housing (homestead) or land (contracted land) in their hometown, they will have more demand (e.g. for agricultural production) to return home by transport. Therefore, the willingness to settle of the migrant population with land assets in their hometowns should be more sensitive to HSR. Figure 1 illustrates the two mechanisms analysed in this paper.

Table 6 models (22) (23) report the regression results for the interaction term of the HSR city dummy variable and the mobility range dummy variable (1 = cross-provincial mobility, 0 = intra-provincial mobility). The coefficients on the interaction terms are positive and significant in both the models without and with the control variables, indicating that cross-provincial mobility increases the effect of city ownership of HSR on the willingness of migrants to settle, with a positive moderating effect. The coefficient on mobility range is negative and significant, indicating that the willingness to settle is weaker for migrant population who move across provinces at the place of inflow and stronger for those who move within provinces. As an important cross-province rapid transport mode, HSR provides convenient and fast transport services for the cross-province migrant population. The intra-provincial migrant population, on the other hand, does not have a higher demand for long-distance travel compared to the inter-provincial migrant population, so their attitude towards HSR is not as sensitive as that of the inter-provincial migrant population.

Table 6 models (24) (25) report the regression results of the interaction term between the dummy variable for the city with HSR and the number of accompanying relatives. The coefficients on the interaction terms are negative and significant in both the models without and with the control variables, indicating that the number of accompanying relatives weakens or suppresses the effect of city ownership of HSR on the willingness of the migrant population

Table 6—: Interaction Term Results

	(22) mobility range	(23) mobility range	(24) accompany relative	(25) accompany relative
highrail	0.0535***	0.0483***	0.0993***	0.0677***
	(0.00316)	(0.00362)	(0.00750)	(0.00794)
scoperail	0.0967***	0.0232***		
	(0.00493)	(0.00511)		
scope	-0.0434***	-0.0878***		
	(0.00445)	(0.00455)		
familiesrail			-0.000623	-0.00377*
			(0.00216)	(0.00218)
families			0.00459**	0.0108***
			(0.00196)	(0.00204)
Control Variables	No	Yes	No	Yes
Interaction FE	No	Yes	No	Yes
N	238406	238406	238406	238406
R^2	0.009	0.107	0.006	0.103

Note: $***p < 0.01, **p < 0.05, *p < 0.1$.

to settle in the city, with a negative moderating effect. The coefficient on the number of accompanying relatives is positive and significant, indicating that the greater the number of accompanying relatives, the stronger the willingness of the migrant population to settle in the inflow area. For the migrant population group, the HSR facilitates their return to their hometown. When the migrant population has more relatives travelling with them, the need to return home to visit relatives decreases and the reliance on the HSR decreases.

Table 7 reports the regression results for the sample grouped by whether or not they owned a home base, and whether or not they owned contracted land. Model (26) (27) shows that the treatment effect of having a home base for the sample HSR cities on the intention to settle is 6.80 %, while this mechanism is not significant for the sample without a home base. Models (28)(29) show that the treatment effect of having contracted land for the sample HSR cities on the intention to settle down is 6.74%, while the treatment effect for the sample without contracted land is negative but the coefficient is small and not have economically significant. The above results demonstrate that the willingness to settle is more likely to be influenced by the construction of HSR in the inflow city for migrants who own land assets in their hometown.

This paper also performs group difference test based on a seemingly uncorrelated regression (SUR) for subgroup regressions. When regressing subgroups on a sample, direct comparison of the magnitude of the coefficients can be biased, and a significance test for differences in coefficients between groups is necessary. The seemingly uncorrelated model can be estimated more accurately by estimating the two subgroup equations jointly and taking into account the correlation of the disturbance terms between the equations. The results in Table 7 show that the tests for the between-group coefficients based on the grouping of all cases of homestead and contracted land are all significant at the 99% confidence level, proving that the groupings are statistically significant.

Table 7—: Homestead, Contracted Land Ownership Subgroup Results

	(26)	(27)	(28)	(29)
	homestead=0	homestead=1	contractedland=0	contractedland=1
highrail	-0.00779	0.0680***	-0.0166***	0.0674***
	(0.00699)	(0.00296)	(0.00610)	(0.00303)
Control Variables	Yes	Yes	Yes	Yes
Interaction FE	Yes	Yes	Yes	Yes
N	35718	202688	54965	183441
R^2	0.040	0.050	0.052	0.047
SUR difference test				
difference	0.076		0.084	
Chi^2	98.99		149.95	
P-value	0.000		0.000	

Note: $***p < 0.01, **p < 0.05, *p < 0.1$.

D. *Regional Heterogeneity Analysis*

Regional heterogeneity allows provinces to have different environments, which may lead to regional heterogeneity in the impact of HSR cities on the willingness of the migrant population to settlement. In this paper, all samples were divided into four groups for group regressions based on the four economic regions delineated by the 2005 China Development Research Centre of the State Council's Strategies and Policies for Coordinated Regional Development. The results (Table 8) show that the effect of HSR cities in the northeast region on the willingness of the migrant population to settle is not significant; the effect of HSR cities in the eastern region on the willingness of the migrant population to settle is higher, with a treatment effect of about 5.76%, and significant at the 99% confidence level. The effect of HSR cities in the central region on the willingness of the migrant population to settle down is low, with a treatment effect of about 2.50%. The treatment effect in the western region is about 6.78%, the highest among all regions.

The main reasons for the differences in the impact of HSR cities on the willingness of the migrant population to settle down in each region are: the different geographical locations of the provinces, differences in climatic environments and economic development patterns, resulting in regional heterogeneity in the impact of HSR cities on the willingness of the migrant population to settle down. Additionally, the northeast region of China has shown a net outflow of population in recent years, and the construction of HSR has facilitated the movement of labour across the region, which may lead to a more pronounced outflow of population, maybe resulting in a less significant effect of HSR cities on the willingness of the migrant population to settle.

Table 8—: Regional Heterogeneity Analysis Results

	(30) North East	(31) East Region	(32) Central Region	(33) West Region
highrail	-0.0134	0.0567***	0.0250***	0.0678***
	(0.0171)	(0.00770)	(0.00590)	(0.00580)
Control Variables	Yes	Yes	Yes	Yes
Interaction FE	Yes	Yes	Yes	Yes
N	14044	108243	45515	70604
R^2	0.045	0.138	0.044	0.036

Note: $***p < 0.01, **p < 0.05, *p < 0.1$.

VI. Conclusion

Based on data from the 2017 and 2018 China Migrants Dynamic Survey (CMDS), this paper empirically analyses the impact of HSR cities on the willingness of the migrant population to settle down. The results show that the proportion of migrants willing to have the intention to settle in cities with HSR is 5.45% higher than that in cities without HSR. The results of the mechanism analysis indicate that the impact of transport infrastructure on the willingness of migrants to settle is influenced by the number of accompanying relatives and the distance of movements. The number of relatives accompanying the migrant population has a negative moderating effect on this mechanism. The more distant the migrants leave from their hometowns, the greater the impact of transport infrastructure on the migrants' willingness to settle. The willingness to settle is more likely to be influenced by the construction of HSR for migrants who own land assets in their hometowns. Except for Northeast China, the effect of whether cities in other regions of China have HSR on the willingness of migrants to settle is significant. The findings of this paper can provide empirical evidence for policy evaluation that the development of transport infrastructure can attract more human resources to stay in cities and provide human capital for economic development. Based on the empirical results of this paper, this paper gives policy recommendations from three perspectives:

First, the development of transport infrastructure can attract labour to settle in cities, and cities still have both short- and long-term benefits from building HSR. Local governments can continue to invest and develop transport infrastructure, especially HSR networks. This will help to attract more labour to settle in cities and increase urbanisation. At the same time, local governments need to balance the short-term costs and long-term benefits of investment in the HSR network. The short-term benefits of HSR construction include stimulating economic growth and promoting tourism, while the long-term benefits can improve the quality of life of the local population and improve the balance of regional development.

Secondly, China's land ownership system is still an important force hindering urbanisation, and the construction of HSR network can mitigate the negative impact of land ownership on the willingness of the migrants to settle. The results of this paper demonstrate that the willingness to settle of the migrants with land assets is more influenced by the construction of HSR, but land assets themselves can negatively affect the willingness of the migrants to settle. China's collective land system restricts land transfer, and Chinese farmers only have the right to use land, but not the right to transfer it. This means that farmers are unable to sell or lease their land to others at market prices, limiting the efficient allocation and mobility of land. If farmers choose to settle in cities and obtain an urban hukou, their land assets in their hometowns are reclaimed collectively, resulting in a loss of assets, which affects their willingness to settle in places of inflow. When cities have a more widespread and convenient HSR network, migrants can quickly travel between their hometowns and the city where they work and better maintain their land hometown assets. Thus, the construction of HSR plays an important role in urbanisation. On the one hand, it can facilitate the settlement of migrants in cities and promote urban mobility. On the other hand, it can improve the convenience of travelling between home and work cities for the migrants, taking into account home town land assets and reducing the negative economic costs of settling in cities due to home town land assets.

Thirdly, the construction of HSR does not have the effect of attracting migrants to settle in Northeast China. The economic development of Northeast China is relatively lagging behind, especially when compared to the eastern coastal region. This has resulted in relatively low employment opportunities and economic dynamism in the region, which has not been able to attract a large number of mobile people to settle there. Even though the construction of HSR has provided faster transportation, the lack of adequate employment opportunities and development prospects in parts of the region has limited interest in settling in the region. In areas of net population outflow, the development of transport infrastructure does not work well to attract labour to settle there. Attracting labour requires a combination of measures to promote economic development, adjust the industrial structure and improve living conditions, in order to enhance the attractiveness and competitiveness of the region and thus attract more labour to settle in the region.

REFERENCES

Ahlfeldt, G. M. and Feddersen, A. (2018). From periphery to core: measuring agglomeration effects using high-speed rail. *Journal of Economic Geography*, 18(2):355–390.

Allen, T. and Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, 129(3):1085–1140.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Aschauer, D. A. (1989). Is public expenditure productive? *Journal of monetary economics*, 23(2):177–200.

Au, C.-C. and Henderson, J. V. (2006). Are chinese cities too small? *The Review of Economic Studies*, 73(3):549–576.

Banzhaf, H. S. and Walsh, R. P. (2008). Do people vote with their feet? an empirical test of tiebout's mechanism. *American economic review*, 98(3):843–863.

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

Bosker, M., Brakman, S., Garretsen, H., and Schramm, M. (2012). Relaxing hukou: Increased labor mobility and china's economic geography. *Journal of Urban Economics*, 72(2-3):252–266.

Chan, K. W. (2015). Five decades of the chinese hukou system. In *Handbook of Chinese migration*, pages 23–47. Edward Elgar Publishing.

Chan, K. W. and Buckingham, W. (2008). Is china abolishing the hukou system? *The China Quarterly*, 195:582–606.

Chen, C. and Fan, C. C. (2016). China's hukou puzzle: Why don't rural migrants want urban hukou? *China Review*, 16(3):9–39.

Chen, G. and Silva, J. d. A. e. (2013). Regional impacts of high-speed rail: A review of methods and models. *Transportation Letters*, 5(3):131–143.

Chen, M., Liu, W., and Tao, X. (2013). Evolution and assessment on china's urbanization 1960–2010: under-urbanization or over-urbanization? *Habitat International*, 38:25–33.

Chen, Y. and Feng, S. (2013). Access to public schools and the education of migrant children in china. *China Economic Review*, 26:75–88.

Clark, X., Dollar, D., and Micco, A. (2004). Port efficiency, maritime transport costs, and bilateral trade. *Journal of development economics*, 75(2):417–450.

CNBS (2023). In *China Statistical Yearbook 2022, by China National Bureau of Statistics (CNBS).*

CNRB (2023). In *Statistical Bulletin of National Railways of China 2022, by China National Railway Bureau (CNRB).*

CNRG (2020). In *Outline of Planning for Railway Pioneering in the New Era of a Strong Transportation State, by China National Railway Group (CNRG).*

Desmet, K. and Rossi-Hansberg, E. (2013). Urban accounting and welfare. *American Economic Review*, 103(6):2296–2327.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858.

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

Duranton, G. and Puga, D. (2019). Urban growth and its aggregate implications. Technical report, National Bureau of Economic Research.

Dustmann, C. and Glitz, A. (2011). Migration and education. In *Handbook of the Economics of Education*, volume 4, pages 327–439. Elsevier.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies*, 81(3):1046–1070.

Fan, C. C., Sun, M., and Zheng, S. (2011). Migration and split households: A comparison of sole, couple, and family migrants in beijing, china. *Environment and Planning A*, 43(9):2164–2185.

Fang, H., Gu, Q., and Zhou, L. (2020). High-speed rail and urban transformation: Evidence from china. *American Economic Journal: Applied Economics*, 12(1):186–223.

Fang, Y. and Shi, Z. (2018). Children of migrant parents: Migrating together or left behind. *Habitat international*, 76:62–68.

Ferrer-i-Carbonell, A. and Frijters, P. (2004). How Important is Methodology for the Estimates of the Determinants of Happiness? *The Economic Journal*, 114(497):641–659.

Fretz, S., Parchet, R., and Robert-Nicoud, F. (2017). Highways, market access, and spatial sorting.

Gakenheimer, R. (1999). Urban mobility in the developing world. *Transportation Research Part A: Policy and Practice*, 33(7-8):671–689.

Glaeser, E. L. and Mare, D. C. (2001). Cities and skills. *Journal of labor economics*, 19(2):316–342.

Gospodini, A. (2005). Urban development, redevelopment and regeneration encouraged by transport infrastructure projects: The case study of 12 european cities. *European Planning Studies*, 13(7):1083–1111.

Hanley, D., Li, J., and Wu, M. (2022). High-speed railways and collaborative innovation. *Regional Science and Urban Economics*, 93:103717.

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*, 104:59–76.

Kim, K. S. (2000). High-speed rail developments and spatial restructuring: A case study of the capital region in south korea. *Cities*, 17(4):251–262.

Klinger, T., Kenworthy, J. R., and Lanzendorf, M. (2013). Dimensions of urban mobility cultures–a comparison of german cities. *Journal of Transport Geography*, 31:18–29.

Liu, Z. (2005). Institution and inequality: the hukou system in china. *Journal of comparative economics*, 33(1):133–157.

Ma, L., Niu, D., and Sun, W. (2021). Transportation infrastructure and entrepreneurship: Evidence from high-speed railway in china. *China Economic Review*, 65:101577.

Miao, Q. and Zhang, H. (2019). The effect of high-speed rail on china's regional innovation efficiency. *Journal of Comparative Economics*, 47(2):406–420.

Monte, F., Redding, S. J., and Rossi-Hansberg, E. (2018). Commuting, migration, and local employment elasticities. *American Economic Review*, 108(12):3855–90.

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.

Roback, S. S. (1982). The tanypodinae (diptera: Chironomidae) of australia ii. *Proceedings of the Academy of Natural Sciences of Philadelphia*, pages 80–112.

Rosen, H. S. (1979). Housing decisions and the us income tax: An econometric analysis. *Journal of Public Economics*, 11(1):1–23.

Song, Y. (2014). What should economists know about the current chinese hukou system? *China Economic Review*, 29:200–212.

Tiebout, C. M. (1956). A pure theory of local expenditures. *Journal of political economy*, 64(5):416–424.

Vortherms, S. A. and Liu, G. G. (2022). Hukou as benefits: Demand for hukou and wages in china. *Urban Studies*, 59(15):3167–3183.

Wang, J. and Cai, S. (2020). The construction of high-speed railway and urban innovation capacity: Based on the perspective of knowledge spillover. *China Economic Review*, 63:101539.

Xiong, Y. (2015). The broken ladder: Why education provides no upward mobility for migrant children in china. *The China Quarterly*, 221:161–184.

Yuan, H., Guan, Z., Jiang, C., and Yang, H. (2023). Peer effect in the construction of china's high-speed rail stations: Empirical evidence from spatial econometric analysis. *Transportation Research Part A: Policy and Practice*, 173:103720.

Zhang, J., Huang, J., Wang, J., and Guo, L. (2020). Return migration and hukou registration constraints in chinese cities. *China Economic Review*, 63:101498.

Zhang, L., Sharpe, R. V., Li, S., and Darity Jr, W. A. (2016). Wage differentials between urban and rural-urban migrant workers in china. *China Economic Review*, 41:222–233.

Zhang, Y. and Zhang, A. (2016). Determinants of air passenger flows in china and gravity model: deregulation, lccs, and high-speed rail. *Journal of Transport Economics and Policy (JTEP)*, 50(3):287–303.

Zhou, S. and Cheung, M. (2017). Hukou system effects on migrant children's education in china: Learning from past disparities. *International social work*, 60(6):1327–1342.

Appendix

Table A1—: Variable Description.

Variables	Variable Description
gethukou	Willingness to settle = 0, no willingness to settle = 1
gender	Male = 1, Female = 0
age	Age of the sample
edu	Years of education of the sample
marriage	Married = 1, unmarried = 0
job	With work = 1, without work = 0
lnincome	Logarised monthly household income
pteacher	Number of primary school teachers per 1,000 population
pbed	Number of hospital beds per 1,000 population
lnpgdp	Logarised GDP per capita, unit of measure: RMB.
tpeople	Proportion of urban tertiary sector employees to total number of employees

Table A2—: Detailed Descriptive Statistics.

Variable	N	Mean	P50	SD	Min	Max
settle	238406	0.332	0	0.471	0	1
hsr	238406	0.834	1	0.373	0	1
gender	238406	0.527	1	0.499	0	1
age	238406	34.43	33	8.785	18	55
edu	238406	9.899	9	2.932	0	19
marriage	238406	0.813	1	0.390	0	1
job	238406	0.854	1	0.353	0	1
lnincome	238406	8.658	8.700	0.547	0.693	13.82
pteacher	238406	4.528	4.136	1.889	1.977	16.30
pbed	238406	6.359	5.981	2.269	1.425	13.77
lnpgdp	238406	11.01	11.14	1.030	3.969	13.16
tpeople	238406	53.09	53.43	14.26	17.88	91.30

Table A3—: Entropy Balance Matching Equilibrium Test.

variables	mean	treat variance	skewness	mean	control variance	skewness
gender	0.5259	0.2493	-0. 1039	0.5259	0.2493	-0. 1039
age	34.32	76.8	0.2809	34.32	75.66	0.3348
edu	10.03	8.312	0.03702	10.03	9.444	-0.004226
marriage	0.8112	0. 1531	- 1.591	0.8112	0. 1531	- 1.591
job	0.864	0. 1 175	-2. 124	0.864	0. 1 175	-2. 124
lnincome	8.691	0.2929	0.3524	8.69	0.3078	0.6692
pteacher	4.531	4.049	4.028	4.531	1.017	0.8148
pbed	6.619	5.237	0.3353	6.617	2.324	-0.4984
lnpgdp	1 1. 14	1.02	-3.386	1 1. 14	0.421	-0.4931
tpeople	52.36	208.5	-0.05725	52.36	1 12.3	-0.8375

Table A4—: Instrumental Variable Test.

Underidentification test (Anderson canon. corr. LM statistic):	3.90E+04
Chi-sq(2) P-value	0.000
Weak identification test (Cragg-Donald Wald F statistic):	2.30E+04
Stock-Yogo weak ID test critical values: 10% maximal IV size	19.93
15% maximal IV size	1 1.59
20% maximal IV size	8.75
25% maximal IV size	7.25
Source: Stock-Yogo (2005). Reproduced by permission.	
Sargan statistic (overidentification test of all instruments):	0.958
Chi-sq(1) P-value	0.3276

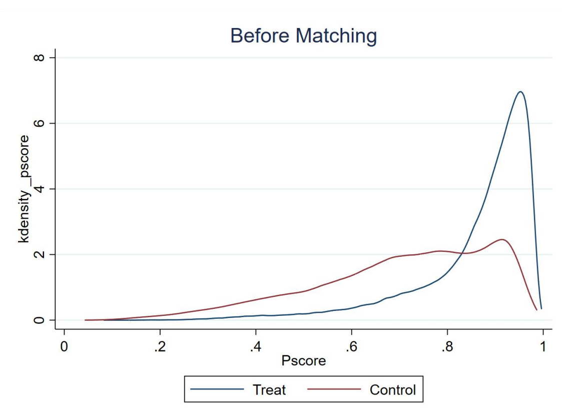

Figure A1. : Pre-matching kernel density

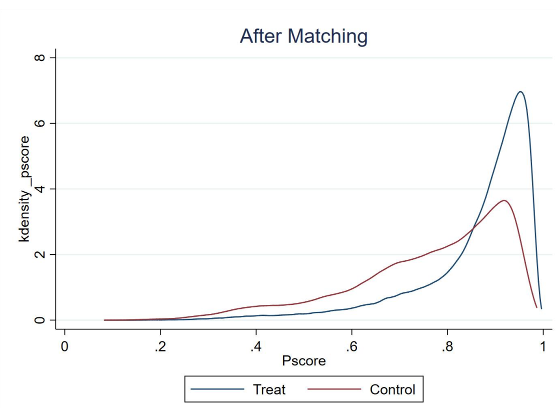

Figure A2. : Post-matching kernel density