Journal of Infrastructure, Policy and Development 2024, 8(7), 5125.
https://doi.org/10.24294/jipd.v8i7.5125

Article

Research on the impact of high-speed railway on urban-rural income gap in China

Mengxin Wang[1], Lang Li[2,*], Kaiming Zhen[3], Hanyong Lan[4]

[1] Institute of Finance, Guangzhou University, Guangzhou 510405, China
[2] Guangfa Bank Credit Card Center, Guangzhou 510623, China
[3] School of Economics, Beijing Technology and Business University, Beijing 100048, China
[4] Jiangmen Xinhui District Statistics Bureau, Jiangmen 529100, China
*** Corresponding author:** Lang Li, 15521191789@126.com

CITATION

Wang M, Li L, Zhen K, Lan H. (2024). Research on the impact of high-speed railway on urban-rural income gap in China. Journal of Infrastructure, Policy and Development. 8(7): 5125. https://doi.org/10.24294/jipd.v8i7.5125

ARTICLE INFO

Received: 9 March 2024
Accepted: 8 April 2024
Available online: 29 July 2024

Abstract: Using data from 31 provinces, municipalities, and autonomous regions in mainland China from 2006 to 2019, we employ a double difference (DID) model and a spatial double difference (SDID) model to estimate the impact of the High-speed Railway (HSR) on the income gap between urban and rural residents, as well as its spatial spillover effects. Our research reveals several key findings. Firstly, the introduction of high-speed railways helps to narrow the income gap between urban and rural residents within local areas, but its spatial effects can lead to an increase in the income gap in neighboring provinces. Secondly, from a spatial perspective, intermediate variables such as industrial structure, education, science and technology, and foreign trade can also contribute to balancing the income gap between urban and rural residents, although the impact of population mobility is not significant. Thirdly, further analysis of the spatial effects demonstrates that education plays a significant role in balancing the income gap both within the local province and neighboring provinces. Additionally, adjustments in industrial structure, advancements in science and technology, and foreign trade have stronger spillover effects in reducing the income gap among neighboring provinces compared to their impact at a local level.

Keywords: high-speed railways; income gap between urban and rural residents; double difference model (DID); spatial double difference model (SDID)

1. Introduction

Since the initiation of the reform and opening-up, China's economy has sustained rapid growth for more than 40 years, transforming from a poverty-stricken, agrarian nation to the world's second-largest economy. However, beneath this rapid economic development lies an undeniable reality: The substantial income gap between urban and rural residents in China. According to the data from the National Bureau of Statistics, the income disparity between urban and rural residents has remained significant, escalating from 2.57 times in 1978 to 3.28 times in 2009. Though it decreased to 2.64 times in 2019, it remains pronounced. Additionally, China's Gini coefficient has consistently exceeded the international warning threshold of 0.4. Researched conducted by Zhang et al. (2020) estimated that the urban-rural income gap accounted for 38.67% of China's domestic income disparity in 2007, a figure that has remained above 50% since 2013. Addressing the income disparity between urban and rural residents is an urgent social challenge in China, and devising solutions to this issue has become the focal point of academic research and discourse.

Previous research indicates that investing in transportation infrastructure can boost the income of residents in less-developed areas, thereby reducing the urban-rural income gap (Alder, 2016; Jiang, 2016; Liu et al., 2013; Luo, 2020). In recent years, China's high-speed railway (HSR) network has expanded rapidly. The Beijing-Tianjin intercity railway, operational since 2008 with a design speed of 350 km/h, marked the inauguration of China's modern HSR era. By the end of 2013, China boasted the world's longest operational HSR network, surpassing 10,000 km. By the close of 2020, China's national HSR network extended over 38,000 km. In recent years, HSR has become a preferred mode of transportation in China, with its construction scale continuing to expand. Against this backdrop, the impact of HSR on income disparities among residents has garnered widespread academic attention.

2. Literature review

As a high-tech product adapted to modern civilization and social progress, HSR not only promotes economic exchanges and facilitates the flow of resources between regions by reducing transportation costs, but also brings about new development opportunities for HSR stations and their surrounding areas. However, with China's HSR network operational for only 13 years since its inception in 2008, research into its impact on Chinese residents' incomes remains limited.

Based on existing research, scholars' opinions are primarily divided into two categories. Some believe that the construction of high-speed rail has a positive impact on narrowing the income gap between urban and rural residents in China. Research by Mo et al. (2018) found that the introduction of HSR has prompted individuals to migrate from resource-limited large cities to underdeveloped areas, consequently influencing the reduction of the income gap between urban and rural regions in those underdeveloped areas. Chen et al.'s (2018) finding indicate that the development of HSR generally contributes to narrowing the income gap between urban and rural residents in China. However, there are notable variations among different city types, with the positive effect being more pronounced in larger cities in the central and eastern regions. Bao et al.'s (2019) research demonstrates that the inauguration of HSR has notably diminished the income gap between urban and rural residents in the central and western regions. Furthermore, Lu et al.'s (2022) study, along with that of other scholars, reveals that the accessibility of HSR stations significantly positive effect on narrowing the income gap between urban and rural areas in the central and western regions.

On the contrary, another group of scholars holds a contrasting view, suggesting that HSR have widened the income gap between urban and rural residents. Fang et al. (2016) argue that while HSR may bolster the economic growth of the Yangtze River Delta metropolitan agglomeration, it widened the urban-rural income gap. Yu et al. (2020) believe that in comparison to cities without HSR, the advent of HSR in China has widened the urban-rural income gap, particularly pronounced in the eastern region, the Beijing-Tianjin-Hebei region, the Yangtze River Delta, and the Pearl River Delta economic circle. Zhu's (2020) research suggests that the accelerated speed of railways has curbed the income growth of both urban and rural residents in peripheral cities, especially rural residents.

This paper aims to expand existing research in three aspects: Firstly, given the nationwide network formation of HSR in China, this study will comprehensively investigate the overall effect of HSR on the income gap between urban and rural residents. Secondly, a spatial double difference model will be constructed that combines economics, geography, physics, and other disciplines to examine the spatial spillover effect of HSR on the income gap. Finally, by incorporating intermediary variables, it will explore their influence on the income gap between urban and rural residents and their spatial spillover effects within the context of HSR expansion.

The structure of the following sections is as follows: Section 3 presents the research hypotheses of this paper. Section 4 introduces the data sources and the selection of variables. Section 5 measures the impact of the opening of high-speed rail on the income gap between urban and rural residents by constructing a multi-period double-difference model. Section 6 assesses the spatial spillover effects of the opening of high-speed rail on the income gap between urban and rural residents in China through the construction of a spatial double-difference model. Finally, Section 7 summarizes the research and provides prospects for future studies along with corresponding policy recommendations.

3. Research hypotheses

3.1. The impact of HSR on the urban-rural income gap in China

Since China implemented its reform and opening-up policy, there has been a fluctuating income gap between urban and rural residents. This gap is measured by the per capita disposable income ratio. Initially, from 1978 to 2003, the gap increased from 2.57 times to 3.28 times. From 2003 to 2009, it stabilized at around 3.1, indicating a significant yet stable income disparity. However, since 2009, there has been a consistent reduction in the urban-rural income gap. The ratio decreased from 3.1 in 2009 to 2.64 in 2019. This narrowing of the income gap can be attributed to various factors, including national welfare policies and initiatives aimed at rural development. These efforts have collectively contributed to enhancing rural residents' incomes and, consequently, mitigating the urban-rural income disparity.

From the perspective of high-speed rail development, the decline in the urban-rural income gap since 2009 coincides with the opening of China's first high-speed rail line, typically occurring within 1–2 years. This temporal correlation, along with the findings of Mo et al. (2018), supports the idea that the introduction of high-speed rail infrastructure in China has played a significant role in reducing the income gap between urban and rural residents. Based on these observations, this paper proposes the following research hypothesis:

- H1: The inauguration of HSR systems significantly contributes to the reduction of the urban-rural income disparity.

3.2. The impact of high-speed rail opening on the urban-rural income gap through multiple pathways

The impact of the opening of HSR on the income disparity between urban and rural residents is not a simple process, but rather a result of the combined effects of multiple factors. In addition to its direct effects, high-speed rail also influences the urban-rural income gap through intermediary factors such as industrial structure upgrading, rural labor mobility, education levels, technological advancement, commodity trade, and others.

Firstly, the improvement in people's living standards has resulted in the growth of the tertiary sector, particularly the service industry, which has created numerous job opportunities. This expansion not only promotes regional economic growth but also benefits rural residents by providing additional income and reducing the urban-rural income gap (Huang, 2016; Karampela, 2018; Kizos, 2016). Secondly, since the beginning of reform and opening up, there has been a significant influx of rural labor into urban areas over a prolonged period. However, China's urban-rural household registration system has placed restrictions on the access of migrant workers to favorable development conditions. Additionally, urban areas have faced the issue of local employment opportunities being taken by migrant workers, leading to competition. Scholarly research emphasizes the substantial impact of labor mobility on the income disparity between local and non-local residents (Ottaviano, 2006; Peri, 2007; Shen, 2006; Zhou, 2011). Thirdly, there is a significant disparity in educational levels between urban and rural residents in China. This disparity acts as a barrier for rural individuals, especially those with lower technical skills, limiting their access to high-tech professions in urban areas. As a result, it hampers the upward mobility of rural residents' income levels, further exacerbating the urban-rural income gap (Chen et al., 2010; Yang et al., 2015). Therefore, educational attainment plays a crucial role in the urban-rural income gap. Fourthly, the advancement of science and technology has increased the skill requirements for employment, leading to a loss of job opportunities for low-skilled workers, particularly those from rural areas. This phenomenon contributes to the widening income gap (Agovino, 2018; Batóg, 2008; Galvis, 2010; Pellegrino, 2019). Lastly, with the progress of economic globalization, there has been a diversification in trading channels for global goods, which has expanded income acquisition opportunities for individuals. As a result, these dynamic influences the distribution structures of residents' incomes. Numerous scholarly studies highlight the significant impact of international trade in balancing income distribution (Chakrabarti, 2000; Gourdon, 2007).

The factors mentioned above not only directly affect the income gap between urban and rural areas but also indirectly influence it through their interaction with the opening of high-speed rail. With a comprehensive analysis in mind, the text proposes the following two hypotheses:

- H2: Intermediate factors, such as upgrading industrial structure, rural labor mobility, educational levels, technological advancement, and commodity trade, can impact the income disparity between urban and rural residents.
- H3: The opening of high-speed rail can indirectly affect the income gap between urban and rural residents through intermediate factors such as upgrading industrial structure, rural labor mobility, educational levels, technological advancement, and commodity trade.

3.3. Spatial spillover effects of HSR opening on regional urban-rural income gap

There is an ongoing debate among scholars regarding the impact of the opening of HSR on the income disparity between urban and rural areas in China. One significant challenge is the relatively short timeframe since the introduction of HSR in China, along with limited operating mileage and sparse distribution of stations during its initial phases. Most railway lines operated independently after the inauguration of China's first HSR line in 2008, until around 2016 when a preliminary interconnected HSR network began to emerge. Due to the time lag in the impact of transportation infrastructure on socio-economic development, existing studies often rely on early-stage data, making it difficult to comprehensively assess the influence of HSR on residents' income.

However, recent years have witnessed significant expansion, with HSR lines extending to all mainland provinces except Tibet, covering over three-quarters of the national territory and connecting more than 80% of major cities. The spatial economic effects of HSR operations are becoming increasingly evident, facilitating the rapid circulation and integration of goods and people among different regions through the HSR network. Consequently, it is expected that the opening of HSR will result in certain spatial spillover effects among different regions.

Based on these observations, the paper proposes the fourth hypothesis:

- H4: The economic effects of HSR opening demonstrate spatial spillover effects, contributing to the regional urban-rural income disparity.

4. Methods and data sources

4.1. Multi-period DID models

The difference-in-differences model is often used to evaluate the impact of a specific policy or measure on indicators because it can better address endogenous problems. This allows researchers to identify the advantages and disadvantages of policy implementation and explore the underlying reasons behind the impact indicators.

The basic idea of the model is to divide the sample into the experimental group and the control group, and calculate the net effect of the policy intervention through two differences. The first difference is within the experimental group and the control group, which indicates the changing relationship before and after the intervention. The second difference is between the two groups after the first difference, which eliminates the gap between the experimental group and the control group in order to obtain a more realistic intervention effect. The general representation of the double DID model is as follows:

$$Y_{it} = \alpha + \beta D_i + \gamma T_t + \rho D_i T_t + \epsilon_{it} \tag{1}$$

The formula above includes variables D_i and T_t, which both range from 0 to 1. These variables indicate whether the subject has been intervened and whether the subject is in the post-intervention period, respectively. If the subject has been intervened, the value is 1, otherwise, it is 0. Variable $G_i \times T_t$ indicates whether the subject has been intervened in a certain period of time.

After the first difference,

$$\Delta Y_0 = \overline{\mathrm{Y}_{0,1}} - \overline{\mathrm{Y}_{0,0}} = \gamma T_1 \tag{2}$$

$$\Delta Y_1 = \overline{\mathrm{Y}_{1,1}} - \overline{\mathrm{Y}_{1,0}} = \gamma T_1 + \rho D_1 T_1 \tag{3}$$

After the second difference,

$$\rho = \mathrm{E}(\Delta Y_1 - \Delta Y_0) \tag{4}$$

It is evident from the derivation that the coefficient ρ of $G_i \times T_t$ serves as an indicator for assessing the overall impact of policy intervention. In order to address the progressive expansion of China's HSR, which initially focused on densely populated and economically developed regions before extending to economically disadvantaged and remote areas, a multi-period DID model becomes essential. In accordance with the research needs, the following multi-period DID model has been formulated:

$$Y_{it} = \beta_0 + \varphi G_i T_t + \sum_{j}^{5} \beta_j Z_{it,j} + \delta_i + \upsilon_t + \varepsilon_{it} \tag{5}$$

In Equation (5), the subscript *i* represents the province for each variable, and *t* represents the time. The dependent variable Y is the ratio of per capita disposable income of urban households to per capita net income of rural households. G is a dummy variable, that takes the value of 1 if the province has opened HSR and 0 otherwise. T is a time dummy variable that indicates whether the HSR policy has been implemented in the current period, with 1 indicating it has been opened and 0 indicating it has not. Z represents control variables, with the subscript j representing IND, EDU, SCI, LAB, and TRA. The meaning of each symbol is shown in **Table 1**. β_0 is the constant term in the model, and β_j is the exogenous parameter corresponding to each control variable, δ_i represents the city fixed effect, υ_t represents the time fixed effect, and ε_{it} represents random disturbance.

Table 1. Descriptive statistics.

Variable	Symbol	Unit	Number of observations	Average	SD	Min	Max
Income gap between urban and rural residents	Y	-	434	2.821	0.535	1.845	4.594
HSR openness	GT	-	434	0.412	0.493	0	1
The proportion of the tertiary industry in GDP	IND	%	434	0.441	0.097	0.286	0.835
The proportion of the education expenditure in GDP	EDU	%	434	0.04	0.023	0.013	0.173
The proportion of the science and technology expenditure in GDP	SCI	%	434	0.004	0.003	0.0003	0.014
Proportion of floating population	LAB	%	434	0.172	0.123	0.001	0.644
The proportion of the total import and export in GDP	TRA	%	434	0.292	0.354	0.013	1.721

The variable $G_i T_t$ is the core variable in the model, representing whether a region has introduced HSR in a specific year. The coefficient φ reflects the extent to which the introduction of HSR affects the income gap between urban and rural residents. A significantly negative coefficient φ suggests that the introduction of HSR has a positive impact on narrowing the urban-rural income gap. Moreover, to consider possible indirect effects between the introduction of HSR and the control variables, the model incorporates interaction terms between HSR and the control variables.

4.2. Variable selection and data interpretation

This paper aims to examine the impact of high-speed railway construction on the income gap between urban and rural residents. To accomplish this, we have selected indicators and gathered data from two main areas: High-speed railway construction and urban-rural income. Furthermore, we have incorporated control variables that reflect the level of regional development to ensure a precise reflection of the relationship between the two.

4.2.1. Income gap between urban and rural residents

The dependent variable in this article is the income gap between urban and rural residents. There are various methods that can be used to measure the income gap, including the difference method for calculating the absolute gap, the Gini coefficient, the coefficient of variation, the generalized entropy method, and the ratio method for calculating the relative gap. Among these methods, the ratio method is commonly used by academic circles and government departments due to its simplicity and interpretability. In this study, we use the ratio of per capita disposable income between urban and rural residents as the indicator for assessing the income gap between these two groups.

4.2.2. High-speed railway

The main feature of high-speed rail is its ability to travel at high speeds. However, the definition of 'high-speed' is subjective and changes over time as technology advances. Presently China's most recent standard for high-speed rail construction is outlined in the 'High-Speed Railway Design Specifications' (TB10621-2014), issued by the National Railway Administration in December 2014. This standard defines high-speed rail as 'newly-built passenger dedicated railways with standard gauge and designed for speeds of 250–350 km/h using electric multiple units (EMUs)'. In China, there are three types of high-speed EMU trains: those with a '*D*' prefix that have a maximum operating speed of 250 km/h, intercity EMUs with a '*C*' prefix that have an average operating speed of 200–250 km/h, and high-speed EMUs with a '*G*' prefix that have an average operating speed of over 250 km/h and can reach speeds of up to 350 km/h. In this study, we will focus on collecting data from the '*G*' prefixed HSR trains, as they align with the current definition of high-speed rail.

In this paper, the key variable being examined is the opening of HSR. Due to the significant differences between urban and rural areas in China's provinces and municipalities, the analysis in this study is conducted at the provincial level. The experimental group is comprised of provinces that opened HSR in a specific year, while the control group consist of provinces that did not. It is worth noting that the effect of HSR opening is not immediate, and most openings occur towards the end of the year. Therefore, to accurately capture the impact, the year of the first HSR opening in each province is shifted by one year.

4.2.3. Control variable

Combined with the previous analysis, this paper introduces several control variables including industrial structure, rural labor force, education, technology, and international trade. These factors are measured by specific indicators such as the proportion of the tertiary industry in GDP, the proportion of floating population in the total local population, the proportion of education expenditure in fiscal expenditure, the proportion of science and technology expenditure in GDP, and the proportion of total foreign trade import and export in GDP.

4.2.4. Descriptive statistics of variables

This paper has selected variables for the analysis, which are presented in **Table 1**. To capture long-term trend changes and assess the effects before and after the opening of the HSR, data collection for this study covers the years 2006–2019. The data used for the variables in the table are sourced from publications such as the 'China Statistical Yearbook', 'China Macroeconomic Database', and 'China Urban and Rural Development Database'. Missing data for some variables in certain years are replaced using the average growth rate or mean interpolation method. The per capita GDP data is adjusted for inflation using 1978 prices to eliminate the price factor.

4.3. Parallel trend test

The first prerequisite for employing the DID model is ensuring that both the experimental group and the control group exhibit a similar development trend before the policy implementation. To verify this, the parallel trend test conducted on the data. **Figure 1** presents the results of the parallel trend test, considering the three years prior to the policy implementation year. On the horizontal axis, '0' denotes the year of policy implementation, with negative values indicating the pre-implementation period and positive values indicating post-implementation. The test results indicate that the coefficients before the opening of HSR are not significant, suggesting a common trend between the experimental and the control groups. Hence, the DID model is deemed appropriate for analyzing the impact of HSR opening on the income gap between urban and rural residents.

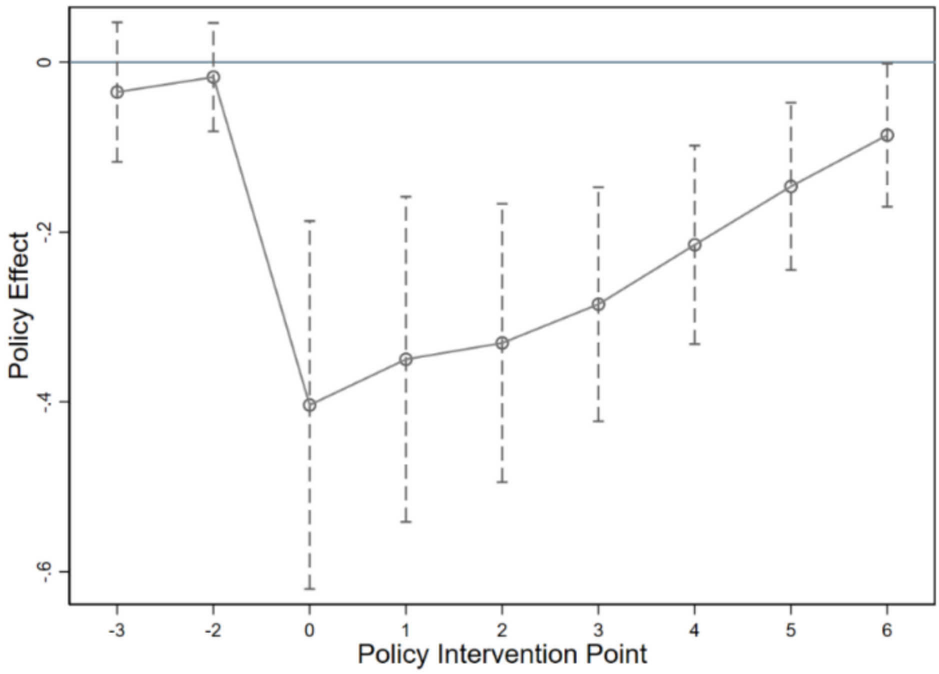

Figure 1. The results of parallel trend test.

5. Horizontal effect of HSR on residents' income gap

This section examines the horizontal effect of opening HSR on the income gap between urban and rural residents from two perspectives. Firstly, it investigates the direct impact of HSR on the income gap. Secondly, it explores the influence of HSR on the income gap through intermediary variables. This approach aims to further analyze the roles of these intermediary variables in mediating the relationship between HSR and the income gap.

5.1. Model estimation results analysis

Estimation results of multi-period DID model

Table 2 presents the results of the multi-period DID model, which examines the effect of the opening of HSR on the income gap. Models 1 to 6 display regression results with include step-by-step control variables. All six models utilize the fixed effect model, and the sample size is 434. The simulation results indicate that each model fits well. As the number of control variables increases, the R^2 value significantly improves, enhancing the model's overall fit.

Table 2. Stepwise regression results of multiple periods DID.

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
GT	−0.407***	−0.255***	−0.210***	−0.199***	−0.160***	−0.175***
	(0.050)	(0.060)	(0.054)	(0.052)	(0.052)	(0.048)
IND	-	−1.761***	−1.267***	−1.180***	−0.897***	−1.206***
	-	(0.393)	(0.350)	(0.349)	(0.312)	(0.336)
EDU	-	-	−13.538***	−12.861***	−10.753***	−9.973***
	-	-	(1.499)	(1.416)	(1.703)	(1.521)
SCI	-	-	-	−13.134	−4.134	−22.369
	-	-	-	(16.100)	(15.148)	(14.629)
LAB	-	-	-	-	−1.124**	−1.204***
	-	-	-	-	(0.464)	(0.435)
TRA	-	-	-	-	-	−0.759***
	-	-	-	-	-	(0.169)
$CONS$	2.989***	3.703***	4.002***	3.984***	3.918***	4.337***
	(0.021)	(0.166)	(0.149)	(0.144)	(0.124)	(0.156)
Fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
N	434	434	434	434	434	434
R^2	0.344	0.412	0.522	0.525	0.554	0.615

Notes: The values in the table are the coefficients of the corresponding model variables. The values in the parentheses are the corresponding standard errors, '*', '**' and '***' respectively representing significant at the levels of 0.9, 0.95 and 0.99.

The regression results of the multi-period DID model in **Table 2** demonstrate the effect of the opening of the HSR on the income gap between urban and rural residents. It was found that the core variable GT had a significantly negative impact at the 1% level in all six models. This indicates that the HSR can effectively narrow the income gap between urban and rural residents and that the results are robust. The accuracy of the model in measuring the HSR effect improves with the inclusion of more variables. This is shown in Model 6, where the coefficient of GT was highly significant at −0.175 when all control variables were included. The results indicate that the HSR has played a positive role in narrowing the income gap after several years of operation. This also verifies hypothesis 1. In the early stages, the 'Matthew Effect' led to rural capital mainly flowing mainly in one direction, resulting in the hollowing out of rural capital, weakening rural productivity, and widening the income gap between urban and rural areas. However, the HSR's positive impact on narrowing the income gap has become increasingly apparent over the years, playing an important role in adjusting the income imbalance between urban and rural areas.

In terms of control variables, the model shows that the industrial structure (IND), education (EDU), floating population (LAB), and foreign trade (TRA) all have a significant impact with negative coefficients and high stability. This implies that an increase in the proportion of the tertiary industry, investment in education, the proportion of the floating population, and the proportion of foreign trade have a significant and positive effect on narrowing of the income gap between urban and rural residents. Among the control variables, education has the most significant effect on narrowing the income gap, with a coefficient of −9.973, which is significantly higher than the other control variables. The effects of the industrial structure and floating population are similar, with coefficients of around −1.2. The model also reveals that although the coefficient of science and technology (SCI) is negative, it fails the test, indicating that investment in science and technology does not have a significant impact on the income gap. This may be because there is a lower level of investment in science and technology compared to education and other areas, as well as a slow popularization and application of technology. Overall, the test results of the model also confirm Hypothesis 2. In addition to technological factors, other factors such as industrial structure, education, population mobility, foreign trade, also have a significant influence on the urban-rural income gap.

5.2. Interaction results of HSR

Based on the results of the multi-period DID model, this section aims to analyze the indirect impact of the HSR on the income gap between urban and rural areas through other control variables. The HSR has a significant impact on the social economy, and its effect on the income gap may be influenced by other factors. Therefore, an interaction term of the core variable GT and the control variables was introduced to estimate the indirect effect of intermediary variables. **Table 3** presents the estimation results of the intermediary variables.

The results presented in **Table 3** indicate that all the intermediary variables have passed the significance test at a significance level of 0.1. However, the direction of effect varies significantly among these variables. The interaction coefficient between the opening of HSR and education is negative, indicating that HSR can effectively improve the quality and work ability of rural residents by delivering educational resources. This abundance of educational resources has also led to a significant narrowing of the income gap. On the other hand, the intermediary variables between the opening of HSR and factors such as industrial structure, technological input, floating population, and foreign trade are all significantly positive. This implies that HSR has widened the income gap through these intermediary factors. The opening of HSR has connected urban and rural areas, accelerated the flow of people and information, and transported more surplus rural labor to cities. This has promoted the adjustment of industrial structure and the growth of foreign trade, which are more conducive to short-term urban economic development. Thus, indirectly, HSR has widened the income gap between urban and rural residents through these intermediary factors. Overall, the test results of the model also support Hypothesis 3, indicating that the introduction of high-speed rail not only directly affects the income gap between urban and rural areas, but also indirectly impacts factors such as industrial structure, education, population mobility, and foreign trade.

Table 3. The DID result of intermediary variables.

Variable	Model 7	Model 8	Model 9	Model 10	Model 11
$GT \times IND$	1.177***	-	-	-	-
	(0.307)	-	-	-	-
$GT \times EDU$	-	−6.918*	-	-	-
	-	(3.618)	-	-	-
$GT \times SCI$	-	-	29.640**	-	-
	-	-	(13.847)	-	-
$GT \times LAB$	-	-	-	1.201***	-
	-	-	-	(0.312)	-
$GT \times TRA$	-	-	-	-	0.373***
	-	-	-	-	(0.103)
control variable	Yes	Yes	Yes	Yes	Yes
fixed effect	Yes	Yes	Yes	Yes	Yes
N	434	434	434	434	434
R^2	0.633	0.628	0.624	0.646	0.644

Note: Standard errors are in parentheses, *$p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

6. Spatial effect of HSR on residents' income gap

This section describes the use of a spatial difference-in-differences (SDID) model to assess the impact of high-speed rail (HSR) on the income gap. This approach was proposed by Fan et al. (2018) and Mi (2017) to address the limitations of traditional econometric models. They introduced a spatial weight matrix and constructed a spatial data panel to overcome these limitations. By integrating the multi-period DID model with the double fixed effect spatial Durbin model, the SDID model can effectively measure the spatial dimension that panel data fails to capture. The objective is to accurately measure the inter-regional effects of HSR on labor and information flow, which are crucial factors contributing to the income gap.

6.1. Construction of the SDID model

There are three spatial econometric models: the spatial lag model (SLM), the spatial error model (SEM), and the Spatial Durbin Model (SDM). The SLM, also known as the spatial autoregressive model (SAR), incorporates the spatial effect of the dependent variable. On the other hand, the SEM considers the spatial effect of the random error on other spaces. The SDM is the most comprehensive model as it captures the effects of both the dependent and independent variables. These three spatial econometric models can be expressed as follows:

$$SLM: \ Q_{it} = \rho W Q_{it} + X_{it}\beta + \varepsilon_{it} \tag{6}$$

$$SEM: \ Q_{it} = X_{it}\beta + \varepsilon_{it}, \quad \varepsilon_{it} = \theta W \varepsilon_{it} + \mu_{it} \tag{7}$$

$$SDM: \ Q_{it} = \rho W Q_{it} + X_{it}\beta + W X_{it}\tau + \varepsilon_{it} \tag{8}$$

In these models, Q_{it} and X_{it} represent the dependent and explanatory variables, respectively. The spatial weight matrix, W, is an $n \times n$ matrix that captures the spatial relationships among the observations. The spatial effect coefficient matrix of the dependent variable is denoted by ρ, while the spatial autocorrelation coefficient matrix of the random disturbance item is denoted by θ. Finally, τ represents the spatial effect coefficient matrices of each independent variable. Together, these elements allow for the measurement of spatial effects on the dependent variable, as well as the relationships among the explanatory variables and spatial autocorrelation.

To make the model more applicable, this study has constructed an SDID model based on the Spatial Durbin Model. The equation for the SDID model is as follows:

$$Y_{it} = \alpha + X_{it}\beta + \lambda W Y_{it} + W X_{it}\gamma + \delta_i + v_t + \varepsilon_{it} \tag{9}$$

The equation above represents the SDID model used in this study to measure the spatial impact of HSR on the income gap between urban and rural residents. In this model, Y represents the income gap, while α is the constant term. The spatial weight matrix, W, includes normalized factors such as the geographical distance of each province, economic strength, and the opening of the HSR. The explanatory variable, X, includes both core and control variables from the previous Equation (5), and β represents the corresponding regression coefficient. λ is the spatial autoregressive coefficient of the dependent variable, which ranges from −1 to 1. A significantly negative λ indicates that the urban-rural income gap in neighboring provinces will cause the local urban-rural income gap to narrow, while a positive λ indicates the opposite spatial spillover effect. The spatial lag coefficient, γ, measures the spatial effect of the opening of HSR or other control variables in adjacent areas, causing the local urban-rural income gap to narrow. δ represents the space fixed effect, v represents the time fixed effect, and ε is the random disturbance item.

In spatial measurement, the space-time sequence has a feedback effect (Anselin, 2010). This means that if a variable changes, it not only affects the corresponding dependent variable in its own space, but also affects other spatial dependent variables. According to the spatial measurement dichotomy proposed by LeSage and Pace (2009), the former influence is referred to as the direct effect, while the latter is referred to as the indirect effect. In addition to exploring the overall spatial effect of the opening of HSR on the income gap between urban and rural residents, this section will also examine the direct and indirect spatial effects of the opening of HSR. Following LeSage's method, Equation (9) can be rewritten as:

$$Y = (I - \lambda W)^{-1}(X\beta + WX\gamma) + P \tag{10}$$

Among these variables, P includes the intercept term, error term, and space-time fixed effects after a form transformation. The remaining coefficients have the same meaning as in Equation (9). The spatial spillover effect can be decomposed by taking the partial derivative of X with respect to the expected value of Y, and the expression is as follows:

$$\left(\frac{\partial E(Y)}{\partial X_{1k}} \cdots \frac{\partial E(Y)}{\partial X_{nk}}\right) = S(W_{a,b}) = \begin{pmatrix} \frac{\partial E(y_1)}{\partial X_{1k}} & \cdots & \frac{\partial E(y_1)}{\partial X_{nk}} \\ \vdots & \ddots & \vdots \\ \frac{\partial E(y_n)}{\partial X_{1k}} & \cdots & \frac{\partial E(y_n)}{\partial X_{nk}} \end{pmatrix} = (I_n - \lambda W)^{-1} \begin{pmatrix} \beta_k & w_{12}\gamma_k & \cdots & w_{1n}\gamma_k \\ w_{21}\gamma_k & \beta_k & \cdots & w_{2n}\gamma_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1}\gamma_k & w_{n2}\gamma_k & \cdots & \beta_k \end{pmatrix} \quad (10)$$

Here, *n* represents the number of spatial observations, *k* represents the number of explanatory variables, and $w_{a,b}$ represents the element in position (a, b) of the spatial weight matrix, W. The matrix $S(w_{a,b})$ contains both the direct and indirect effects of the independent variable on the dependent variable. In this matrix, the diagonal elements represent the direct effect of the observation, while each off-diagonal element represents the indirect effect. The average value of the diagonal elements in the $S(w_{a,b})$ matrix represents the direct effect of the explanatory variables, while the mean of the rows or columns of the off-diagonal elements represents the spatial indirect effects of the explanatory variables.

6.2. Selection and setting of space matrix

When using Equation (9) to measure spatial effects, the first step is to select an appropriate spatial weight matrix. Common spatial matrices include the 0–1 matrix, geographic distance matrix, inverse distance matrix, and others. However, when measuring measure, the spatial effect brought by the HSR, it is necessary to consider not only the geographical distance but also other relevant factors, such as regional economic development. Therefore, economic factors have been incorporated into the weight matrix. Additionally, this study uses the number of HSR stopping schedules in each province as an indicator of HSR development level, which is also included in the spatial weight matrix. This allows for the construction of an economic spatial weight matrix with HSR attributes. The specific process for constructing the weight matrix is as follows:

$$W = \begin{bmatrix} W_{11} & \cdots & W_{1j} \\ \vdots & \ddots & \vdots \\ W_{i1} & \cdots & W_{ij} \end{bmatrix} = W^* \cdot E_{ij} \cdot H_{ij} \quad (12)$$

$$W^* = \frac{W^* - minW_i^*}{maxW_i^* - minW_i^*} \quad (13)$$

$$\mathrm{E}_{ij} = \begin{cases} \dfrac{\overline{RGDP_i}}{\overline{RGDP_j}}, & i \neq j \\ 1 \quad , & i = j \end{cases} \quad (14)$$

$$\overline{RGDP}_i = \frac{1}{n}\sum_{t=1}^{n} RGDP_i \quad (15)$$

$$H_{ij} = \begin{cases} \dfrac{h_i}{h_j}, & i \neq j \\ 1, & i = j \end{cases} \quad (16)$$

To construct the weight matrix of economic space with HSR attributes, this paper considers economic development, geographical location, and HSR development level between provinces. The resulting matrix, denoted as W, captures the relationships between each province in these dimensions. To normalize the anti-geographical distance, the matrix is then multiplied by W^*, which is obtained by measuring the geographical distance using ArcGIS software. Specifically, the degree of economic connection between provinces i and j (E_{ij}) is calculated as the ratio of their actual per capita gross product (RGDP), while the development level of HSR (H_{ij}) is calculated as the ratio of the number of HSR stopping schedules in province i to that in province j.

6.3. Estimated result of SDID model

To begin the analysis, a spatial autocorrelation test is performed. The global Moran index is commonly used to assess the overall spatial correlation within the dataset. The formula for calculating the Moran index is as follows:

$$\mathrm{MoranI} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\,(Z_i-\bar{Z})(Z_j-\bar{Z})}{S^2\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}}, \quad \mathrm{i,j} = 1,2,3\ldots,\mathrm{n} \tag{17}$$

where $S^2 = \frac{\sum_{i=1}^{n}(Z_i-\bar{Z})^2}{n}$, $\bar{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i$, Z_i and Z_j represent the studied object in the area i and nearby areas j respectively. W_{ij} is the spatial weight matrix. The Moran's index I has a value range of [−1, 1]. If $\mathrm{I} > 0$ and it is significant, it suggests a positive spatial correlation among the objects being studied. On the other hand, if $\mathrm{I} < 0$ and it is significant, it indicates a negative spatial correlation. In this study, the global Moran index is used to measure the spatial correlation of the income gap between urban and rural residents from 2006 to 2019, as shown in **Table 4**. The results demonstrate that the index passes the significance test at the 1% level for each year, implying the existence of a significant spatial spillover effect in the urban-rural income gap between provinces.

Table 4. Global Moran's index of explained variable.

Year	Moran's I index	Year	Moran's I index
2006	0.197***	2013	0.177***
2007	0.201***	2014	0.151***
2008	0.202***	2015	0.151***
2009	0.194***	2016	0.150***
2010	0.200***	2017	0.147***
2011	0.197***	2018	0.144***
2012	0.193***	2019	0.145***

Note: *** $p < 0.01$.

Table 5 presents the estimated results of the SDID model. The results indicate that the impact coefficient of the main variable "HSR opening (GT)" is −0.071, which is significantly negative and consistent with the findings of the multiple-period DID model. The coefficient 0.809 in the table represents the spatial lag coefficient "$W \times GT$" of the HSR opening. This coefficient is significantly positive, indicating that the opening of HSR in a region has a spatial spillover effect on neighboring provinces.

Consequently, the income gap in neighboring provinces widens.

The regression results in **Table 5** demonstrate the indirect effects of each intermediary variable resulting from the opening of the HSR. Under the influence of the HSR, industrial structure, education, science and technology, and foreign trade all had significant negative spatial effects on the income gap in neighboring provinces. This also partially supports hypothesis 4. Additionally, the spatial lag coefficient of population mobility (LAB) showed a negative correlation, although the result was not significant.

Table 5. The results of SDID.

Variable	SDID
GT	−0.071***
	(0.022)
$W \times GT$	0.809***
	(0.131)
$W \times GT \times IND$	−6.050***
	(1.225)
$W \times GT \times EDU$	−57.917***
	(7.522)
$W \times GT \times SCI$	−255.897***
	(58.658)
$W \times GT \times LAB$	−0.556
	(1.339)
$W \times GT \times TRA$	−0.737*
	(0.387)

Note: Standard errors are in parentheses, *$p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

6.4. Decomposition of space effects

Through decomposing the spatial effect, we can enhance our understanding of the direct impact of HSR opening and intermediary variables on the income gap in the local area, as well as their indirect effects in neighboring provinces through spatial spillover effects. The results of the spatial effect decomposition for each variable are presented in **Table 6**.

The coefficient of the core variable GT in **Table 6** represents the direct effect, which is significantly negative with a value of −0.068. This result is consistent with the influence in the DID model, confirming the effect of HSR opening in narrowing the income gap. The indirect effect of GT, which represents the spatial spillover effect, is significantly positive with a coefficient of 0.872. This indicates that HSR opening has an expanding effect on the income gap in neighboring provinces. This further supports hypothesis 4. From a direct effects perspective, areas with preferred HSR opening have a higher economic level, a greater degree of openness, and a more widespread education. Opening HSR can quickly address the issue of low income among local rural residents. and reduce the income gap. Furthermore, from an indirect perspective, HSR opening can have a compression effect on time and space. This can reduce the costs of medium and long-distance travel and attract talents from neighboring areas. It also leads to the gathering of information advantages in provinces with a higher degree of railway development. As a result, neighboring areas experience constrained economic development and a widening of the income gap to some extent.

Table 6. Decomposition of SDID spatial effects.

Variable	Direct effect	Indirect effect
GT	−0.068***	0.872***
	(0.023)	(0.196)
GT × IND	−0.046	−6.241***
	(0.190)	(1.584)
GT × EDU	−4.055***	−51.561***
	(1.244)	(9.867)
GT × SCI	5.700	−260.113**
	(6.868)	(69.351)
GT × LAB	0.107	−0.531
	(0.152)	(1.302)
GT × TRA	−0.042	−0.753*
	(0.047)	(0.389)

Note: Standard errors are in parentheses, *$p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

The spatial regression results of the intermediary variables show that the effects of the control variables on the income gap vary greatly after the opening of HSR. With the presence of HSR, the industrial structure, education, science and technology, and foreign trade all have a significant impact in reducing the urban-rural income gap in neighboring provinces. The spillover effect of the intermediary variable is more noticeable than the direct effect. Among the intermediary variables, education has a clear influence on adjusting the income disparity between urban and rural areas, both locally and in neighboring provinces, while technology has the most significant effect in reducing the urban-rural income gap in neighboring provinces.

7. Conclusions, recommendations and research prospects

Based on data from China's HSR and regional economic development spanning from 2006 to 2019, this study used the DID and SDID methods to measure the impact of HSR opening on income gap between urban and rural residents. The research findings can be summarized as follows: Firstly, HSR opening helps to narrow the income gap between local urban and rural residents, but it widens the income gap between neighboring provinces. Secondly, when it comes to spatial effects, intermediary variables such as industrial structure, education, science and technology, and foreign trade also play a role in balancing the income gap between urban and rural residents, but the impact of population mobility is less significant. Thirdly, the decomposition of spatial effects reveals that education significantly contributes to balancing the urban-rural income gap in both local and neighboring provinces. Furthermore, the spatial spillover effect of industrial structure, technology, and foreign trade on narrowing the income gap is stronger in neighboring provinces than in local provinces.

Compared to previous research, this study supports the narrowing effect of the introduction of high-speed rail on the income gap between urban and rural areas, which is consistent with the findings of Mo et al. (2018) and Chen et al. (2018). Furthermore, this study also confirms the spatial spillover effect, as suggested by scholars such as Bao et al. (2019) and Lu et al. (2022). However, there are certain limitations to the research presented in this article. For instance, due to the relatively short period of time that China's high-speed rail has been operating, the study only used data from 2006 to 2019, resulting in a small sample size. Additionally, as time goes on, the long-term impact of high-speed rail on economic growth will become more apparent. Therefore, future research on related topics may reveal further insights. Based on the research results, this paper offers several recommendations to exploit the benefits of HSR in narrowing the income gap between urban and rural residents. Firstly, expedite the expansion of HSR to counties and townships to maximize its positive impact on reducing the income gap. Additionally, enhance the connectivity of transportation infrastructure surrounding HSR stations to further extend the spillover effect of HSR on local rural residents. Secondly, support the planning and development of characteristic rural projects, and provide training program such as service ability, and management courses for rural residents. The rapid growth of the tertiary industry can attract both travelers brought by the HSR and capital to rural areas. Thirdly, enhance the HSR operation plan, including appropriately lowering the ticket prices, in order to reduce the obstacles faced by rural residents when taking the HSR and encourage rural population mobility. Fourthly, facilitate the flow of resources such as education, technology, and foreign trade information into rural areas through HSR, thereby enhancing the overall quality and capabilities of rural residents and increasing their income.

Author contributions: Conceptualization, MW and LL; methodology, LL; software, LL; validation, MW, KZ and HL; formal analysis, MW; investigation, KZ; resources, LL; data curation, LL; writing—original draft preparation, MW; writing—review and editing, LL; visualization, HL; supervision, MW; project administration, LL and KZ; funding acquisition, MW. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Research Projects of Guangzhou University (YJ2023017) and Major Projects of National Social Science Fund (23&ZD127).

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

References

Agovino, M., Aldieri, L., Garofalo, A., et al. (2016). R&D spillovers and employment: evidence from European patent data. Empirica, 45(2), 247–260. https://doi.org/10.1007/s10663-016-9359-x

Alder, S. (2016). Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2856050

Anselin, L. (2010). Thirty years of spatial econometrics. Papers in Regional Science, 89(1), 3–26. https://doi.org/10.1111/j.1435-5957.2010.00279.x

Bao, H., & Wang, Y. (2019). The Impact of High-speed Railway Opening on the Income Gap between Urban and Rural Areas.

Proceedings of the 2019 3rd International Conference on Education, Economics and Management Research (ICEEMR 2019). https://doi.org/10.2991/assehr.k.191221.128

Batóg, J. (2008). Income Inequality and Innovations in the Processes of Economic Growth and Real Convergence. Folia Oeconomica Stetinensia, 7(1). https://doi.org/10.2478/v10031-008-0011-1

Chakrabarti, A. (2000). Does Trade Cause Inequality? Journal of Economic Development, 2(25), 1–21.

Chen, B., Zhang, P., Yang, R. (2010). Government investment in education, human capital investment and China's urban-rural income gap. Journal of Management World, (01), 36–43.

Chen, F., Xu, K., Wang, M. (2018). High Speed Railway Development and Urban-Rural Income Inequality: Evidence from Chinese Cities. Economic Review, 210(2), 61–75.

Diao, M., Zhu, Y., & Zhu, J. (2017). Intra-city access to inter-city transport nodes: The implications of high-speed-rail station locations for the urban development of Chinese cities. Urban Studies, 54(10), 2249–2267. https://doi.org/10.1177/0042098016646686

Fan, Q., Darren, H. (2018). A New Endogenous Spatial Temporal Weight Matrix Based on Ratios of Global Moran's I. Journal of Quantitative & Technological Economics, 35(01), 131–149.

Fang, D., Sun, M. (2016). An Evaluation on the Impact of High-speed Rail on Economic Development of City Clusters in the Yangtze River Delta—An Empirical Study Based on Difference in Difference Model. East China Economic Management, 30(02), 42–47.

Galvis, L. A. (2010). Essays in economic geography, convergence, inequalities, and innovations in the knowledge economy. Dissertations & Theses—Gradworks.

Gómez, L. A. (2021). The Kuznets Curve in Brazil 1850–2010. Journal of Iberian and Latin American Economic History, 39(1), 37–61.

Gourdon, J. (2006). Openness and inequality in developing countries: a new look at the evidence. MPRA Paper.

Huang, W., Li, Z., Wu, M. (2016). Empirical study of the impact of Chinese tertiary industry development on the urban-rural income gap. China Forestry Economics, (05), 40–43.

Jiang, M., & Kim, E. (2016). Impact of high-speed railroad on regional income inequalities in China and Korea. International Journal of Urban Sciences, 20(3), 393–406. https://doi.org/10.1080/12265934.2016.1213139

Karampela, S., & Kizos, T. (2018). Agritourism and local development: Evidence from two case studies in Greece. International Journal of Tourism Research, 20(5), 566–577. Portico. https://doi.org/10.1002/jtr.2206

Kizos, T., Karampela, S. & Spilanis, I. (2016). Evaluating the impact of agritourism on local development in small islands. Island Studies Journal, 11(1), 161–176.

LeSage, J., Pace, R. K. (2009). Introduction to Spatial Econometrics. CRC Press.

Liu, C., Zhou, L., Xu, L. (2013). The Impact of Highway Access on Urban-Rural Income Inequality: County-Level Evidence from China. In: Proceedings of the 2013 China Youth Economists Forum.

Lu, H., Zhao, P., Dong, Y., et al. (2022). Accessibility of high-speed rail stations and spatial disparity of urban-rural income gaps. Progress in Geography, 41(1), 131–142. https://doi.org/10.18306/dlkxjz.2022.01.012

Luo, S., & Fu, Q. (2020). Spatial Effects of Traffic Infrastructure on Income Disparity. Journal of Coastal Research, 104(sp1), 705–710.

Mo, Y., Huang, X., Liu, X. (2018). Study on the Economic Impact of High-speed Railway on Guangdong and Guangxi-The Empirical Analysis Based on DID Econometric Model. Reformation & Strategy, 34(07), 84–89.

Ottaviano, G. I. P., & Peri, G. (2012). Rethinking the effect of immigration on wages. Journal of the European Economic Association, 10(1), 152–197. https://doi.org/10.1111/j.1542-4774.2011.01052.x

Pellegrino, G., Piva, M., & Vivarelli, M. (2019). Beyond R&D: the role of embodied technological change in affecting employment. Journal of Evolutionary Economics, 29(4), 1151–1171. https://doi.org/10.1007/s00191-019-00635-w

Peri, G., & Davis, U. (2006). Wages, rents and prices: The effects of immigration on us natives. Center for Research & Analysis of Migration.

Shen, K., Yu, J. (2011). The impact of rural labor flow on the income of urban residents in China. Journal of Management World, (3), 58–65.

Yang, J., Lai, D. (2015). How to alleviate income inequality through education? Economic Research Journal, 50(09), 86–99.

Yu, F, Tang, Y, Zhang, M. (2020). The Effect of High-speed Railway on Urban-rural Income Disparity in China. International Business, (04), 129–143.

Zhang, Y., Xu, Y. (2020). Financial Development, Labor Income Distribution and Urban-Rural Income Gap: An Empirical Analysis by Using the Provincial Panel data. Reform, 321(11), 135–146.

Zhou, L., Biswas, B., Bowles, T., et al. (2011). Impact of Globalization on Income Distribution Inequality in 60 Countries. Global Economy Journal, 11(1), 1850216. https://doi.org/10.2202/1524-5861.1628

Zhu, W., Cai, L., Zhu, H. (2020). Does the Speed Increase of Railways Inhibit the Income Growth of Residents in Peripheral Cities? Modern Economic Science, 42(01), 60–70.