The effect of the construction of high speed railway on urban-rural income disparity in China

WANG YING

This paper, based on the data of China City Statistical Yearbook and information of the high-speed railway（HSR）, explores the effect of the accessibility HSR on urban-rural income disparity during 2013-2017 by using the method of Diff-in-Diff and triple difference. I find that compared with areas without HSR, the construction of HSR has expanded the average income disparity between urban and rural residents. Thus, some measures should be taken by the government to narrow the urban-rural income gap enlarged by the use of HSR.

Keywords: China's HSR; Urban-rural income disparity; Diff-in-Diff; Triple Difference

Ⅰ. Introduction

Income disparity has been one of the most pressing and prominent social problems in China for a long time. Some researches show that the widening urban-rural income gap is the main source of the China's income disparity (**LU Ming and CHEN Zhao 2004; LI Shi 2012**). Nowadays, the infrastructures of the transport have already become the key to break the established market segmentation as well as to facilitate the flow of the economic factors. However, many relevant studies has been mostly restricted to other alternative transportations like cars and airplanes, and furthermore only few of them discussed the time predate HSR is well-constructed in most areas of China. Therefore, in terms of the impact caused by HSR on the income gap in China, there is a lack of research focusing on the period after HSR has been basically popularized in China.

Ⅱ. Experimental Design

Ⅱ.A. Model and Variable Design

For the urban area, it has:

$$\mathrm{In}_u = \alpha_0 + \alpha_1 D_p + \alpha_2 HSR_u + \alpha_3 HSR_u \times D_p + e_u \quad (1)$$

For the rural area, it has:

$$\mathrm{In}_r = \beta_0 + \beta_1 D_p + \beta_2 HSR_r + \beta_3 HSR_r \times D_p + e_r \quad (2)$$

For the fully-saturated regression model, it has:

$$\mathrm{In}_m = \gamma_0 + \gamma_1 D_p + \gamma_2 HSR + \gamma_3 Urban + \gamma_4 HSR \times D_p + \gamma_5 Urban \times D_p + \gamma_6 HSR \times Urban + \gamma_7 HSR \times D_p \times Urban + e_m \quad (3)$$

In these equations, the subscript of "u" and "r" represent urban area and rural area separately, and "m" is the symbol of combined group. Then the "In" represents the income of people, while HSR is the primary dummy variable for if the people were from the city that has a HSR station and D_p, the time dummy, will equal to 1 if the year the data were collected was after the year HSR station opened (2014) otherwise it will be 0 and e is the variable for regression error.

As for the coefficient, α_1 and β_1 are the time trends for control groups, α_2 and β_2 are the cross-section differences between treated and control group before the time HSR opened and α_3 and β_3 are equaling to the difference-in -difference for the two kinds of territories numerically.

In the third fully-saturated model, the first three coefficients γ_1, γ_2 and γ_3 have the similar meanings as the previous models, and γ_4 is the Diff-in-Diff estimate for rural people. And γ_7 represents the triple difference and equals to the estimate of the effect of opening the HSR on urban-rural income gap numerically.

Ⅱ.B. Data sources

(ⅰ)Data of City:

Primarily are the data about characteristics of its economy, including population、GDP、resident income and so on. Data are collected from《China Statistical Yearbook》and supplementary data from《China Statistical Yearbook For Regional Economy》

(ⅱ)Data of HSR：

I get the information about high-speed railway, mainly is the opening time of the HSR line, from CRAD database of Chinese Research Data Services (CNRDS) Platform

Ⅱ.C. Data processing

Because《China Statistical Yearbook》and《China Statistical Yearbook For Regional Economy》did not collect the data that can directly show the urban and rural residents' income, I decided to use the increment of financial institutions' deposit balances per person divided by the national average savings rate for that year as an indicator for the residents' income, that is:

$$In = \frac{increment\ of\ financial\ institutions'\ deposit\ balances}{population \times \text{national average savings rate that year}}$$

It is reasonable since the ratios of this indicator and the real income collected in the very early years 'yearbooks are very similar, meaning that there exists a nearly paralleled trend between this indicator and the real income.

Ⅲ. Experimental Procedure

Ⅲ.A. data pre-processing

I use excel to simply process the data, filter and integrate the obtained data, and integrate data into one sheet. Variables include the per capita income, fiscal revenue, fiscal expenditure, population, labor population, amount of industrial enterprises.

Ⅲ.B. regression analysis

Use R to perform regression analysis on the opening of high-speed rail and the urban-rural income gap, and investigate the impact of the opening of HSR on the urban-rural income gap.

Ⅲ.C. Analysis and hypothesis of regression outcomes

I find that the opening of HSR has increased the income gap between urban and rural areas after analyzing the regression outcomes. Then I assume that the reason for the result is that the opening of HSR has led to the loss of rural labor.

Ⅲ.D. Further data analysis proves the hypothesis

Obtain more relevant data, use data of urban areas as the sample to analyze the relationship between hypothesized variables and per capita income. Then try to improve the model and prove the hypothesis.

Ⅳ. Preliminary outcomes and Summary

Ⅳ.A. Regression Estimates of Main Effects of HSR

The summary statistics and regression estimates presented in the table blow, and then I will analyze the effect of HSR on urban and rural people's income as well as the income disparity between the two areas.

I selected 946 data of the income from 2013 to 2017 for all cities, and the data of cities that opened HSR for the first time in 2014 are divided into treatment group and data that have not opened HSR until 2017 are used as control group. Eliminating those are less than 0 or missing, then I have 492 for urban and 409 for rural.

Table 1

Influence Analysis

	Dependent variable:		
	urban VS rural VS triple		
	(1)	(2)	(3)
POST	4.900**	9.476***	9.476***
	(2.388)	(2.986)	(2.710)
HSR	2.345	1.581	1.581
	(3.245)	(4.736)	(4.299)
urban			2.788
			(2.886)
HSR x Post	2.157	-4.830	-4.830
	(4.260)	(6.299)	(5.718)
Post x Urban			-4.576
			(3.778)
HSR x Urban			0.764
			(5.593)
HSR x Post x Urban			6.988
			(7.399)
Constant	8.846***	6.058***	6.058***
	(1.827)	(2.277)	(2.067)
Observations	492	409	901
F Statistic	3.691** (df = 3; 488)	3.652** (df = 3; 405)	3.336*** (df = 7; 893)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 1 provides regression estimates of the main effects of HSR and urban-rural classification on personal income. Column (1) shows results for the urban group, as noted in the ***Ⅱ.A.***, the third row is the Diff-in-Diff estimate of the treatment effect, which is positive, meaning that the construction of HSR benefits the income of urban people. By the contrast, column (2) that illustrates results for rural groups has a negative Diff-in-Diff estimate, which means that the HSR had a negative impact on the rural residents' income. Column (3) combines the previous two regressions and

makes a mixed regression. γ_7, the seventh row, represents the triple difference for effect of the construction of HSR, and the positive value of it justifies that the income disparity of urban-rural areas is expanded by the newly constructed transport.

Table 2

Effect of HSR for urban people:

	Pre	Post	Difference
Control	8.846411	13.746605	4.900194
Treatment	11.191388	18.249042	7.057654
Difference	2.344977	4.502437	2.157460

Table 3

Effect of HSR for rural people:

	Pre	Post	Difference
Control	6.058475	15.534197	9.475723
Treatment	7.639698	12.285197	4.645499
Difference	1.581223	-3.249001	-4.830224

Table 2 and 3 are the difference and Diff-in-Diff estimates of personal income in urban and rural regions, and it is same as the regression model.

For the rural group, it is remarkable that although the before-after estimator shown in Table 3 is positive, meaning that the income of rural people has increased these years. However, the negative cross-section estimate for the post-period (-3.2) demonstrates the income of rural areas that did not have a HSR station increased more than that of those have built HSR station.

Ⅳ.B. Summary

According to the results above, we can conclude that the construction of HSR enlarged the income disparity between urban and rural areas. And we can deduce the expanded income gap is primarily due to the negative influence of HSR on rural population, maybe the opening of HSR encourages young people to migrate to cities, leaving the olds and teenagers in the suburb region.

Ⅴ. Assumption Verifying

Ⅴ.A. Data source and pro-processing

(ⅰ) Data source:

Data are selected by screening the economic variables common to urban and rural areas in the statistical yearbooks. And the chosen variables are those have significant correlation coefficients between per capita income.

(ii) Data pre-processing

Three indicators representing fiscal expenditure: general fiscal expenditure, scientific fiscal expenditure and educational fiscal expenditure have multicollinearity according to the test. So the principal component analysis is carried out on them in order to combine them into one variable to keep the regression as effective as possible.

The cumulative contribution rate of the first principal component obtained reached 92.1%, so the standardized data of the first principal component can be used as the indicator of general fiscal expenditure.

V.B. Model and Variable Design

For the urban area, it has:

$$Labor_u = \alpha_0' + \alpha_1' D_p + \alpha_2' HSR_u + \alpha_3' HSR_u \times D_p + e_u' \quad (4)$$

$$In_u' = \alpha_0'' + \alpha_1'' D_p + \alpha_2'' HSR_u + \alpha_4 Labor_u + \alpha_5 Exp_u + e_u'' \quad (5)$$

For the rural area, it has:

$$Labor_r = \beta_0' + \beta_1' D_p + \beta_2' HSR_r + \beta_3' HSR_r \times D_p + e_r' \quad (6)$$

$$In_r' = \beta_0'' + \beta_1'' D_p + \beta_2'' HSR_r + \beta_4 Labor_r + \beta_5 Exp_r + e_r'' \quad (7)$$

V.C. Regression Estimates

(i) Main effect of HSR on labor force

As can be seen from the data in the third row of Table 4 below, the opening of HSR has a negative impact on the number of urban labor force (regression coefficient is -1), while a positive effect on that of rural labor force (coefficient is 2.05).

Table 4

Reason Analtsis1

	Dependent variable:	
	urban VS rural	
	(1)	(2)
POST	1.711	6.487***
	(1.429)	(1.870)
HSR	3.314*	8.813***
	(1.940)	(2.959)
HSR X Post	-1.003	2.053
	(2.547)	(3.935)
Constant	13.807***	10.180***
	(1.092)	(1.427)
Observations	491	407
F Statistic	2.080 (df = 3; 487)	14.491*** (df = 3; 403)
Note:		*p<0.1; **p<0.05; ***p<0.01

(ii) Relationship between labor force and per capita income

According to the data in the third row of Table 5 below, the number of urban labor force population is positively correlated with its per capita income (regression coefficient is 0.57), while the number of rural labor force population is negatively correlated with that (regression coefficient is -0.24). And the regression results are significant.

Table 5

Reason Analtsis2

	Dependent variable:	
	urban vs rural	
	(1)	(2)
POST	5.133***	8.202***
	(1.911)	(2.583)
HSR	2.218	-0.261
	(2.026)	(3.096)
labor	0.569***	-0.238**
	(0.126)	(0.095)
exp_ind	-0.003	0.017**
	(0.003)	(0.008)
Constant	0.847	4.237
	(2.145)	(2.736)
Observations	491	406
F Statistic	14.138*** (df = 4; 486)	4.162*** (df = 4; 401)

Note: *p<0.1; **p<0.05; ***p<0.01

(iii) Summary

The results above show that the opening of HSR does not lead to the decrease of rural labor force population as I suppose, but has a positive effect on rural labor force. However, as the rural labor force index is negatively correlated with its per capita income, making the opening of HSR actually leads to the decrease of rural per capita income.

With regard to the conclusion, I believe that due to the accelerated urbanization process in rural areas, a large number of migrant workers have returned to their hometown. But because of the imbalance of the quality of rural labor force, the increased labor force has brought about a decrease in per capita income.

Ⅵ. Inspiration and Further Improvements

Ⅵ.A. Inspiration

The effect of HSR on personal income in certain areas is not as good as I imagined, so it does not mean that the more accessible to HSR, the better development for some rural areas. And thus, the HSR is not a potent tool to reduce

the reginal imbalances in economic growth due to the imbalanced labor force in rural areas. More effective measures should be taken by the local and national government to improve the quality of the labor force and the selection of the HSR station site should consider not only the needs of the local people, but also the continuous impact of transportation infrastructure on the local economy

VI.B. Further Improvements

(i) Considering there are part of income invested in the financial market and other markets, the indicator I choose may not be the perfect one. For further study, the money invested out of bank can be added into the indictor, making it closer to the real income. Fortunately, it does not influence the final outcome too much since urban people generally invest more and deposit less fraction of their income than the rural ones.

(ii) Only considering the variable of the opening of HSR, but HSR is not the first transportation infrastructure built in many areas. This makes the opening of HSR unable to represent a significant improvement in traffic, and is not convincing enough to investigate the impact of traffic infrastructure. Variables such as road opening can be added to further research.

(iii) More data are needed. Although there were most HSR stations opened in 2014, the amount of the data is not significantly enough, I only collected 975 data and some of them are useless, which leads to the regression results maybe not persuasive enough.

(iv)For further analysis, I can examine the parallel trend between the treatment and control group very systematically and make an additional study to explore the mechanism and reasons underlying this phenomenon.

Ⅶ. Reference

[1] LuMing, ChenZhao. Urbanization,. Urban-Biased Economic Policies and Urban-Rural Inequality [J]. Economic Research Journal, 2004(06):50-58.(In Chinese)
[2] LiShi. The disparity between urban and rural areas is the biggest inequity in income distribution [J]. News about Work in China Rural Areas, 2012(20):43. (In Chinese)