SPATIAL SPILLOVER EFFECTS OF HIGH-SPEED RAIL (HSR) INFRASTRUCTURE ON ECONOMIC GROWTH --- EVIDENCE FROM SOUTH CHINA

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Qing Yang

May 2021

ABSTRACT

The High-Speed Rail (HSR) as trains with an average speed of more than 200 kilometers per hour, can save travel time and improve regional linkage to a large extent. Now, there is emerging consensus that the HSR has significant impacts on regional economic growth, but the exact effects may differ across time and space.

This thesis presents research concerning the effects of HSR development on regional economic growth in South China. Using prefectural panel data from 2009, 2011, 2013, 2015 and 2017, a multi-factor spatial panel model has been specified and estimated with several methodologies leading to the following conclusions: (1) The construction of the HSR system has enhanced the overall economic development in South China. (2) There are heterogeneity and interaction effects among geographical units. If these effects are not considered, the impact of the HSR system on economic growth may be overestimated. (3) There are, moreover, positive spillover effects between cities. (4) The coefficients of the market potential in the spatial panel model for both developed and backward areas are positive and significant. However, the coefficient for the developed area is larger than the coefficient for the developing area, which indicates that the effects of HSR on economic growth in developed and backward regions are unbalanced. The developed region would benefit more from the construction of the HSR, and this may exacerbate regional development inequality.

BIOGRAPHICAL SKETCH

Qing Yang is a second-year Master candidate in the Regional Science program at Cornell University. Qing received her bachelor's degree in Architecture from Tianjin University in China, with a focus on architecture theory.

With an interest in urban and regional development, Qing has developed related theories and skills during her two-year study at Cornell. In the course "NTRES 6200 Spatial Modelling and Analysis", she gained more skills and tools on spatial econometrics and benefited a lot from the interdisciplinary study. The preliminary research ("Spatial Spillover Effects of HSR on Economic Growth ——Evidence from Pan-Pearl River Delta Region) has been used as the basis of this thesis.

Qing is also interested in urban development theory and international finance. She will keep learning and develop her skills on seeking the internal factors affecting societal and regional development.

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my chair advisor Professor Kieran Patrick Donaghy, who has given me valuable instructions and help for this thesis. Besides, he helped me a lot with my academic writing, improving my writing skills.

I also would like to thank my minor advisor Professor David Rossiter for his suggestions and help in developing this thesis. He inspired me and made my work more rigorous. He provided me with the required knowledge and methods for doing the complicated study in this thesis, helping me seek the logic behind the development of urban economic and the connection of cities.

In addition, I would like to thank all faculties and staff I met during the years at Cornell. They have shared their knowledge generously with me and helped me in my campus life.

Finally, I would like to say thank you to my entire family, for their unconditional love, support and understanding.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

HSR: High-Speed Rail

OLS: Ordinary least squares method of solving linear equations

GDP: Gross domestic product

VIF : Variance inflation factor

CHAPTER 1. INTRODUCTION

Economists have a continuous focus on the impacts of transportation infrastructure on economic growth. Now, it is commonly accepted that the HSR has significant impacts on regional economic growth through both direct and indirect effects, but the exact effects may differ across time and space. Tangible measurements of effects on regional economic growth and inequality are crucial for policy decision-making.

This study will focus on South China which including seven southernmost provinces in China as its research object. South China is one of the most diverse regions in China, both in geography and economic growth. Its diverse geography has led to poor land connectivity and unbalanced economic development. With an ambitious HSR plan initiated in the early 2000s in China, the link between cities in South China has been greatly improved.

The overall purpose of this research is to analyze the spatial spillover effects of HSR on economic development in south China. This study intends to explore whether the construction of the HSR system enhances the overall regional economic growth in south China, whether the effects of HSR on economic growth are balanced in developed and backward regions, and whether the effects (coefficients) to be different using different regression models.

1.1 Transportation infrastructure and regional economic development

Economists have analyzed the relationship between transportation infrastructures and regional economic growth since the early 1940s. As they wanted to determine if the huge investment in these infrastructures is cost-effective, and they wanted to measure the effects on economic growth and regional inequality. Now it is commonly accepted that infrastructure investments do act as a stimulus on economic growth through two mechanisms: multiplier effects and externality effects. But the exact coefficients may differ across time and space.

Multiplier effect: Infrastructure investment drives the development of interrelated industries through the multiplier effect, which has a direct stimulating influence on regional economic growth. The investment-driven model has played a prominent role in the fast-economic growth in the People's Republic of China (PRC) over the past 30 years, while transport investment has accounted for a huge proportion of the total investment in fixed assets.

Spillover Effects: There are also positive externalities from network expenditures made by neighboring political units (Lall, 2007). That is because some effects induced by transport infrastructure will extend outside the limits of the investment area, generating spillover effects (Munnell, 1992; Boarnet, 1995, 1996, 1998). Spillover effects in the context of this paper refer to costs or benefits imposed on neighboring political or economic units as the result of some economic activity. The indirect effect is apparent in reducing transport costs and time costs, accelerating the integration of the market, facilitating the rapid flow of the labor force and factors to where they are needed, leading to the dissemination of knowledge and technology, and contributing to improving inter-regional technical efficiency and optimizing resource allocation.

1.2 The High-speed rail system in China

1.2.1 HSR Trains

Trains with an average speed of more than 200 kilometers per hour are widely considered to be the definition of high-speed rail (HSR) service. The Tōkaidō Shinkansen railway that began operations in Japan in 1964 is considered to be the first HSR system. In recent years, the development of HSR systems has accelerated rapidly in many countries because of their potential advantages. Such advantages include their abilities, not just to provide users a time-saving, convenient and punctual travel, but also stronger links between cities.

1.2.2 Chinese HSR Grid System

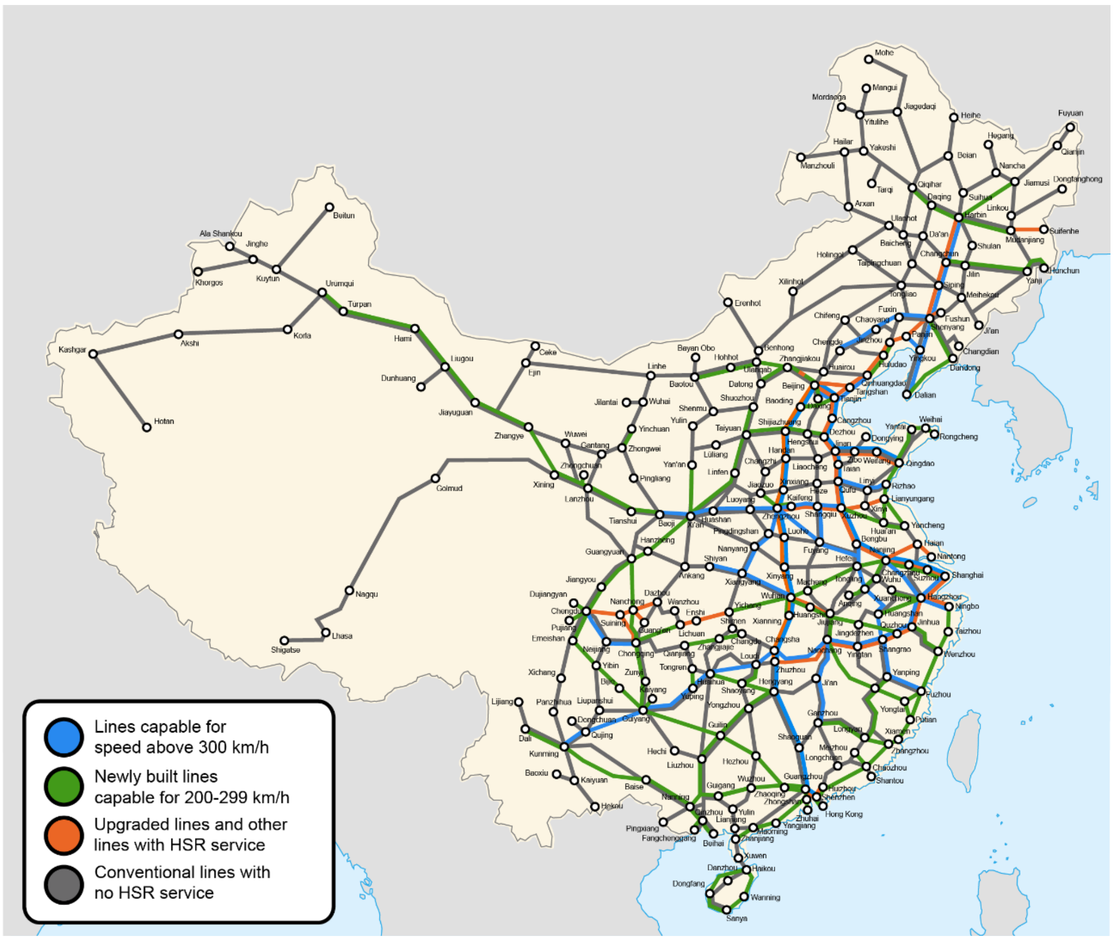
China embarked on its ambitious HSR plan in the early 2000s. The government concentrated on passenger-dedicated HSR lines with rising government budgets. In 2004, the state initiated a "Mid-to-Long-Term Railway Network Plan" to build a national grid with 4+4 high-speed rail corridors[1] of 12,000 km in total length. In 2016, the plan was mostly completed and reorganized as a larger 8+8 HSR grid[2]. The new 8+8 grid was proposed to finish in 2030 and connects most of the country.

[1] The 4+4 corridors contain 8 HSR lines, 4 running north-south and 4 running east-west, with a total length of 12,000km.

[2] In 2016, the Chinese state planner reorganized the national HSR network and doubled the 4+4 grid into 8+8 HSR corridors. Eight Verticals are Coastal corridor, Beijing–Shanghai corridor, Beijing–Hong Kong (Taipei) corridor, Harbin–Hong Kong (Macau) corridor, Hohhot–Nanning corridor, Beijing–Kunming corridor, Baotou (Yinchuan)–Hainan corridor and Lanzhou (Xining)–Guangzhou corridor. Eight Horizontals are Suifenhe–Manzhouli corridor, Beijing–Lanzhou corridor, Qingdao–Yinchuan corridor, Eurasia Continental Bridge corridor, Yangtze River corridor, Shanghai–Kunming corridor, Xiamen–Chongqing corridor and Guangzhou–Kunming corridor.

By the end of 2020, the buildout of HSR in China has provided service over nearly 37,900 km, extended to all provincial-level administrative divisions except Macau and Tibet, and accounts for two-thirds of the world's total high-speed railway networks.

Figure1. Chinese HSR network in 2020



1.3 Study Area—South China

South China is a geographical and cultural region that covers the southernmost part of China. Its meaning varies with context. In this paper, ‘South China’ represents the seven southernmost provinces in China, including Guangdong, Guangxi, Fujian, Yunnan, Guizhou, Hunan and Jiangxi.

This study will focus on South China as its research object for the following reasons. South China is one of the most diverse regions in China, both in geography and economic growth. Its diverse geography has led to poor land connectivity and unbalanced economic development. However, after the initiation of the HSR, the link between cities has been greatly improved.

Figure 2. Map of the study area—South China

1.3.1 Diverse Geography

Fujian is a coastal province whose economy traditionally relies mainly on maritime activities. It is mostly mountainous with Wuyi Mountains forming the border between Fujian and Jiangxi.

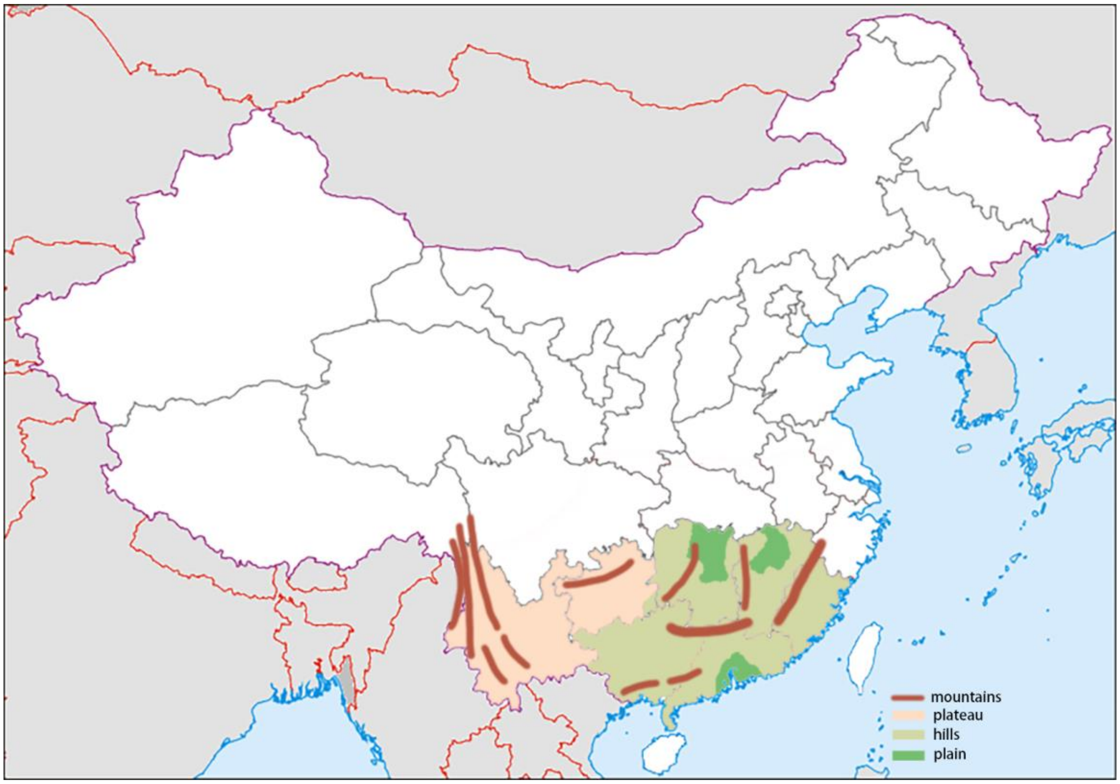
Guangdong province is on the north shore of the South China Sea, part of which lies along the Pearl River Delta. It is separated from the north by the Nan Mountains range.

Guangxi is in the west of Guangdong and bordered by Vietnam in the southwest. It is in mountainous terrain with Nanling Mountains forming the northeast border.

Yunnan and Guizhou are landlocked provinces that lie on the Yunnan-Guizhou plateau. Both of them are noted for a high level of ethnic diversity.

Hunan and Jiangxi are inland provinces in south-central China. Hunan is located on the south bank of the Yangtze River with mountains surround the east, south and west sides. Jiangxi has Gan River and Lake Poyang that finally enter Yangtze River on the north border. It has mountains on the west, east and south.

Figure 3. Illustration of geography condition in South China



1.3.2 Regional development disparity

Because of the diverse geography and mountainous terrain in south China, the land connectivity between these provinces was not strong before the initiation of the HSR. As a result, the economic development of these provinces was to some extent isolated with distinct geographical factors, locational advantages and policies.

Guangdong and Fujian were in the vanguard of the 1978 Deng Xiaoping reform, and took advantage of their access to the ocean. Currently, they are the most prosperous provinces in China, contributing approximately 15% of the total economic output of mainland China. By contrast, Yunnan, Guangxi and Guizhou are the least developed provinces, with GDP far below the national average.

1.3.3 Improvement in land connectivity

Since 2009, with the construction of HSR, land connectivity between south China has been greatly enhanced due to a decrease in travel time and improvement in travel convenience. In 2009, there was no HSR in this region. Travel across the province was costly. For example, travel from Guiyang to Guangzhou required about 2 days. In 2017, there were already 14 lines connecting all provinces in this region. The travel time between Guiyang to Guangzhou decreased to 5 hours, i.e., about 10-fold. Such a dramatic decrease is expected to have a substantial impact on the regional economy.

1.4 Thesis research questions

The overall purpose of this research is to analyze the spatial spillover effects of HSR on economic development in south China. Spillover effects in the context of this paper refer to costs or benefits imposed on neighboring political or economic units as the result of some economic activity.This research will compare the regression results from the OLS models, panel models and spatial panel models to address the following questions.

Question 1: Did the construction of the HSR system enhance the overall regional economic growth in south China, and if so, by how much? Is the coefficient of the explanatory variable significant? If so, what is the coefficient's value?
Hypothesis: The construction of the HSR grid enhanced connectivity between cities and overall economic growth in this region significantly.

Question 2: Are there interaction effects among geographical units that influence the regression result? Are the coefficients of the OLS and the spatial panel model different? If so, are the spillover effects between cities positive or negative? (Will cities have synergistic effects or competitive effects?)

Hypothesis: The spillover effects between cities are slightly positive. The effects of economic growth in neighboring regions will have positive effects on the economic growth in a given region.

Question 3: Are the effects of HSR on economic growth balanced in developed and backward regions? (Does HSR have a different effect on cities with different levels of economic development?)

Hypothesis: The effect of HSR on economic growth in developed and backward regions will be unbalanced. The developed region will benefit more from the construction of the HSR, and this may exacerbate regional development inequality.

CHAPTER 2. LITERATURE REVIEW

The discussion of the impact of transportation infrastructure on economic development started in the early 1940s. Development economists such as Rodan have argued that underdeveloped countries require large amounts of investments (especially transportation infrastructures) to embark on the path of economic development from their present state of backwardness in his “big push model”. Numerous empirical studies have been carried out on the relationships between transportation capital investment and economic development since the seminal work of Aschauer (1989). Munnell (1992) hypothesized that the highway can cause positive cross-state spillover effects because of the network characteristics. Boarnet (1995) indicated that public capital (highway capital) could influence economic activity by shifting activity from one location to another. Berechman (2005) used a spatial lag model to analyze the construction of the highway on the state, county and municipality levels. The key result is that transportation investments produce strong spillover effects relative to space and time. Unless these factors are properly accounted for many reported empirical results are likely to be overly biased, with important policy implications.

With the rapid development of the Chinese economy, and the enormous investment and construction in transportation infrastructures (highway, high-speed rail), scholars in China have also been eager to conduct empirical research to determine the extent to which there are spillover effects of transportation infrastructures on the regional economy. Xueliang Zhang (2012) conducted an analysis using panel data from 30 Chinese provinces from 1993 to 2009. The explanatory variable was the *total investment in transportation infrastructure.* Zhan concluded that there is a significant impact of transportation infrastructure on economic development. The coefficient of 0.05-0.07

estimated by Zhang using the spatial panel model implied that for every additional unit of the transportation infrastructure investment the regional GDP would increase by 5% to 7%. The estimation results will be overestimated, however, if spatial spillover effects are not considered. Nannan Yu (2013) analyzed the spillover effects of transportation in four different regions. Her results showed that positive spillovers exist due to the connectivity at the national level but vary among 4 regions at the regional level. Shi, Q.Y. (2018) used market potential as the indicator of HSR and took samples of 36 capital cities in China. He demonstrated that high-speed railways' effect on per capita GDP is negative, but positive on the population. For the areas where HSR connects to a network, HSR compresses space and time and promotes cooperation among cities. The effects of HSR on economic growth are unbalanced, the development of high-speed rail is conducive to the further rise of big cities, but not to the development of small and medium-sized cities.

CHAPTER 3 METHODOLOGY

This study intends to employ an empirical framework in investigating the effect of HSR on the regional economy in the south China region. The basic form of the OLS model to be employed will be as follows:

$$lnY_{it} = \beta_0 + \beta_1 lnYT_{it} + \beta_2 lnX_{it} + \varepsilon_{it}$$

Y_i as the dependent variable stands for the real GDP in city i, which is an important indicator for measuring the development level of an economy. YT_{ti} is the explanatory variable market potential, which is purchasing power of neighboring cities weighted by space-time and measures the effect of HSR. While there are many other factors that would affect economic growth, this paper will use the design matrix X_i to account for several control variables to represent other factors, as explained below.

3.1 Choice of Study Unit

Focusing on the impact of HSR on regional economic development in South China, this paper will use 94 prefectural level cities in south China as its data sample.

Many previous empirical studies on this topic have focused on the national level and provincial levels. These levels of focus are suitable to analyze whether or not there is an effect of HSR on economic development from a macro perspective. However, more recent studies on Chinese HSR have begun to use the city as a unit of study with the increased availability of better city-level data. As most of the HSR stations in China were built in prefectural-level cities, it is reasonable to treat cities as separate data points to analyze the exact impact of HSR on a certain region.

3.2 Choice of Panel Data instead of Cross-Sectional Data

The preliminary research of this thesis focused on development between two periods: the year of 2009, during which the HSR was not available, and 2017, during which an extensive HSR network had been established.

However, the construction of the HSR grid in China is a rapid and dynamic process. New lines have opened almost every year, which decreases the travel time and enhances the connectivity between cities along the lines frequently. So, it is better to use panel data with observations from multiple periods to analyze the effect of HSR more accurately. This research will use data of 94 prefectural level cities in 2009, 2011, 2013, 2015 and 2017.

3.3 Choice of the Explanatory Variable

To explore the effects of the HSR system on economic growth, different explanatory variables have been chosen to represent the influence of HSR based on different model settings and questions of analysis. Xueliang Zhang (2012) used asset investment of high-speed rail in different cities as his explanatory variable as he was concerned more about how investment in HSR would affect regional economic growth. Jian Li (2017) used the physical form of transportation (miles of transportation) as an explanatory variable as he tried to measure the effects of roads and highways on economic growth. Shi, Q.Y. (2018) introduced the indicator of Market Potential (purchasing power of neighboring cities weighted by space-time) to quantify the effect of HSR. The market potential is a concept used by economic geographers to measure a specific geographic area's access to the market for inputs and outputs. The author uses the space-time weighted purchasing power of neighboring cities to represent the effect of HSR.

This thesis will use Market Potential (YT) as the main explanatory variable, because time-saving is one of the most important features of the HSR system. It can reduce travel time between cities, thereby facilitating labor flow, market expansion and knowledge spillover. The form of Market Potential (YT) is as follows,

$$YT_{i,t} = \sum_{j=1}^{n} \frac{GDP_{j,t}}{T_{ij,\,t}} \ (i \neq j)$$

Where $GDP_{j,t}$ stands for the gross regional production of city j in the year t. $T_{ij,\,t}$ stands for the travel time between city i and city j in the year of t (minute-based).

Moreover, this study will use the time-lagged explanatory variable YT_{it-2}. As the Chinese train schedules are updated and published every two years. Typically, at the end of a year. So the effect of the HSR will not reflect on the economy immediately at the year. The model used in this analysis will use a time-lagged explanatory variable instead of the regression equation.

$$lnY_{it} = \beta_0 + \beta_1 lnYT_{it-2} + \beta_2 lnX_{it} + \varepsilon_{it},$$

where YT_{it-2} represents the market potential two years earlier.

3.4 Choice of Control Variables

For control variables in the X matrix, there are three major economic growth theories that establish the determinants of regional growth: Neoclassical Growth Theory, Endogenous Growth Theory, and New Economic Geography. The Solow-Swan model in the neoclassical economics framework treats capital and labor as basic inputs of economic activities and contributors to economic growth. In the Endogenous Growth Theory, Romer (1986) and Lucas (1988) treat human capital and knowledge as important endogenous sources for growth and stress on the role of policy in economic development. The New Economic Geography was developed by Paul Krugman and other economists. Spatial concentration and specialization as the persistent sources of regional development disparities.

Papers that analyze the relationship between transportation infrastructure and economic growth often use different groups of control variables selected from these three theories. Berechman (2005) and Nannan Yu (2013) only included Neoclassical Growth Theory factors in their models. Jian Li (2017) considered both Neoclassical and Endogenous Growth Theories variables, while papers that were written by Xueliang Zhang (2012) and Shi, Q.Y. (2018) contain determinants listed by all these three factors. As the drivers of economic growth are complicated, I think it is reasonable to include all factors that have been commonly accepted as growth determinants by economists as control variables.[3] So, the control variable setting of this paper is as follows:

Capital (K), a factor in Neoclassical Growth Theory, this paper will choose data of the *Total Investment in Fixed Assets* to represent capital input for the entire urban economy.

Labor(L), as an indispensable part of Solow's Growth Model, this paper will use *Total Number of Employment* as operational indicators.

Export（E） is another aspect of human capital as Romer thinks firms are learning by exporting. This paper will use *Total volumes of Exports and Imports* in each city as an indicator.

[3] We note, however, that the growth theories alluded to above were developed with structural-equation models comprising systems of simultaneous equations. Hence, endogenous growth cannot be represented with a single-equation reduced-form model.

Human Capital (CS), Romer treats human capital and knowledge as important endogenous sources for growth, this paper will use the *Number of College Students in each city* to represent the human capital level.

Road is an important takeaway from New Economic Geography as economists stressed the importance of transportation costs ignored by earlier economists in trading and forming of core regions. This paper will include *lengths of highways (kilometers)* in each cities' urban and suburban areas as control variables.

Urban, is an important factor in New Economic Geography. This paper will use the *ratio of the non-agricultural population to the total population* at the end of the year to represent the urbanization level.

Urbanization economies of scale (PE), cities are assumed to have the potential to accelerate economic growth because of the effect of urbanization economies of scale. With people in the city, they can share favorable institutions and infrastructure, and gain from diverse knowledge spillovers. Porter argues that knowledge spillovers in specialized, geographically concentrated industries will stimulate growth. Porter's externalities are maximized in cities with geographically specialized, competitive industries. Xueliang Zhang (2012) has defined the variable as follows,
$PEi = \frac{N_i/Y_i}{\sum_i N_i/\sum_i Y_i}$, N stands for the number of firms

Localization economies of scale (LE), is an important argument of New Economic Geography, which indicates that firms will benefit from cluster because of sharing in inputs, labor and knowledge spillovers. Xueliang Zhang (2012) used *LE* to measure this effect.

$\text{LEi} = \frac{g_i/Y_i}{\sum_i g_i/\sum_i Y_i}$, $g_i$stands for value added in industry i.

3.5 Spatial Lag Model

As raised by Munnell and Boarnet (Munnell, 1992; Boarnet, 1995, 1996, 1998), the transportation infrastructure may transfer local economic activities to other regions. Through the diffusion effect, the faster-developed area could drive the economic development of the slower growth area, thus showing a positive spatial spillover effect; at the same time, the transportation infrastructure could produce negative spatial spillover effects through the aggregation effect. The production factors will flow to economically developed regions more easily. In this case, the economic growth of one region may be at the cost of the economic recession of other regions.

Also, Xueliang Zhang (2012) compared the results from the OLS and the spatial lag model, the results show that the feature space model tends to over-evaluate the influence. Jian Li (2017) compared different result from the OLS, Spatial Lag Model and Spatial Tobin Model, which indicated that the spillover effects exist.

As a result, the spillover effects of economic growth in different regions (the effects that the economic growth in neighboring regions will influence the economic growth in a given region) need to be considered in this model.

The basic form of a spatial lag model is as written in equation (1) below. The form of this equation can capture spillover effects of GDP from neighboring cities. It assumes that such dependence in the spatial lag term ρWY of the response variable Y. Where ρ stands for the strength of autoregression of the response variables due to the spillover. W is a matrix representing the spatial structure (i.e. neighbor weights):

$$lnY_i = \rho W lnY_j + \beta_1 lnBT_i + \beta_2 lnX_i + \varepsilon_i \quad (1)$$

W can be defined as comprising basic binary weights

$$W_{ij} = f(x) = \begin{cases} 1, & if\ city\ i\ has\ a\ link\ with\ city\ j \\ 0, & other\ wise \end{cases}$$

CHAPTER 4 DATA PREPARATION

4.1 Data Source

All macro data used in this study (including GDP, total investment in fixed asset, the total number of employment and etc) could be obtained from the “China City Statistical Yearbook” and “China Regional Economic Statistical Yearbook published in 2010, 2012, 2014, 2016 and 2018 (with data on 2009, 2011, 2013, 2015 and 2017).

4.2 Calculation of Market Potential

The market potential (YT) is calculated as the time-weighted purchasing power (GDP) of neighboring cities.

$$YT_{i,t} = \sum_{j=1}^{n} \frac{GDP_{j,t}}{T_{ij,\,t}} \ (i \neq j)$$

The travel time of certain HSR lines is available from “Chinese Train Schedules” published every two years. This study has first organized the HSR route network map and illustrated the travel time in 2007, 2009, 2011, 2013 and 2015 in south China. (see appendix 1). Then, the travel time between every two cities could be calculated based on this.

This study will use *Adjacency* and *distance* function in *igraph* package in R to build the network and calculate network travel time. The distance function could calculate the shortest path travel time between every two cities. Then the explanatory variable YT could be obtained. Appendix 2 has shown the results of YT values.

4.3 Statistical Summary of Variables

Table 1. Statistical Summary of Variables

	Description	Unit	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Year			2009	2011	2013	2013	2015	2017
Y	*GDP*	100 million RMB	48.05	593.89	1,040.88	1,734.12	1,835.76	22,490.06
K	*Total Investment in Fixed Assets*	100 million RMB	0.00	412.20	734.20	1,092.30	1,397.00	7,567.80
L	*Total Number of Employment*	10 thousand people	0.00	152.70	219.10	252.10	313.80	1,229.30
E	*Total volumes of Exports and Imports*	10 thousand dollars	0.00	35,481.00	167,111.00	1,266,523.00	439,014.00	53,747,437.00
Road	*Length of Highways*	Kilometers	1,368.00	6,583.00	12,286.00	12,542.00	17,311.00	32,942.00
Urban	*Ratio of non-agricultural population to the total population*	100%	19.00	37.98	46.19	49.16	56.90	100.00
CS	*Number of College Students*	10 thousand people	0.00	1.25	2.40	6.69	4.81	106.73
LE	*Localization economies of scale*		0.00	89.59	127.16	148.38	196.60	495.87
PE	*Urbanization economies of scale*		14.52	58.46	81.83	86.03	109.39	243.68
YT-2	*Market Potential (2 year prior)*		0.00	61.26	198.71	383.34	542.57	3,157.11

CHAPTER 5 MODEL CONSTRUCTION AND ANALYSIS

5.1 Multicollinearity and VIF results

Multicollinearity is the occurrence of high intercorrelations among independent variables in multiple regression models. Although multicollinearity does not reduce the predictive power or reliability of the model, it will affect the calculation of individual coefficients. When independent variables are correlated, changes in one variable are associated with shifts in another variable and an assumption of ordinary least squares regression is violated. The condition of collinearity makes it difficult to estimate a partial regression coefficient cleanly with an appropriate order of magnitude and theoretically correct sign and hence discern the impact on the dependent variable of a unit change in the corresponding explanatory variable, holding constant all other explanatory variables.

The VIF (Variance inflation factor) test is a method to test for the presence of multicollinearity in a set of multiple regression variables. It can measure how much the variance of the estimate of a regression coefficient is inflated due to multicollinearity. If a variable gets a VIF test result higher than 5, it is highly correlated with other variables and may lead to unreliable estimates of regression coefficients.

As a group of independent variables will be included in the model, a variance inflation factors (VIF) test has been done. Table 2 reports the results of the VIF tests. The VIF values of all variables are all between 1 and 3. So there is no need for corrective measures and thus no variable has to be excluded from the set.

Table 2. Results of the VIF tests

Variables	VIF		VIF between two variables	
Market Potential（YT-2)	2.0072			Market Potential（YT-2)
Total Investment in Fixed Assets（K)	1.9151		Market Potential（YT-2)	-
Total Number of Employment （L)	2.5105		Total Investment in Fixed Assets（K)	1.3264
Total volumes of Exports and Imports（E)	2.9927		Total Number of Employment（L)	1.1719
Length of Highways（Road）	1.7276		Total volumes of Exports and Imports(E)	1.3342
Urbanization Rate（Urban）	2.9615		Length of Highways（Road）	1.0043
Number of College Students（NC）	2.0931		Urbanization Rate（Urban）	1.3886
Localization economies of scale（LE）	2.6372		Number of College Students（NC）	1.2155
Urbanization economies of scale（PE）	2.4653		Localization economies of scale（LE）	1.1563
			Urbanization economies of scale（PE）	1.6413

5.2 Model on Total Observations

5.2.1 OLS Modeling

OLS is a basic method to estimate parameters in linear regression models using the least square principle. The first set of regression models are intended to explore the overall relationship between the economic development (GDP) and explanatory variables (YT-2: market potential) using the OLS model.

$$lnY_{it} = \beta_0 + \beta_1 lnYT_{it-2} + \beta_2 lnX_{it} + \varepsilon_{it} \qquad (2)$$

Table 3 summarizes the regression results and each column represents a separate regression. The asterisk indicates whether the t-statistic is significant at the 10% level (one asterisk), 5% level (two asterisks) or 1% level (three asterisks).

In equation (1) in Table 3, no control variables are added, the YT-2 coefficient is significant with a value of 0.2326. With other control variables gradually adding to equation (2)-(9), the coefficient of YT-2 decreases and the adjusted R Squared increases. In equation (9), all independent variables have passed the 10% significance level test. The YT-2 coefficient is 0.0335, indicating that the railway network (economic connectivity with other cities) has a significant positive spillover effect on regional economic growth. It implies that 1 unit increase in market potential leads to a long-term increase in GDP of 0.0335.

Table 3. Results of the OLS model

Variables	1	2	3	4	5	6	7	8	9
Market Potential (YT-2)	0.2326***	0.1256***	0.0931***	0.0570***	0.0595***	0.0409***	0.0374***	0.0290***	0.0335***
Total Fixed Assets(K)		0.4741***	0.3577***	0.2726***	0.2654***	0.2147***	0.1915***	0.1563***	0.1579***
Total Employment(L)			0.5620***	0.5370***	0.4984***	0.4635***	0.4147***	0.6493***	0.6540***
Total Exports(E)				0.1228***	0.1324***	0.0790***	0.0767***	0.0276***	0.0321***
Length of Highways (Road)					0.0785**	0.2040***	0.1842***	0.1694***	0.1724***
Urbanization Rate (Urban)						0.8982***	0.7795***	0.5494***	0.5531***
Number of Students in Colleges (CS)							0.1471***	0.0493***	0.0408**
Localization Economies of Scale (LE)								0.5006***	0.5111***
Porter Externality (PE)									-0.0570*
Constant	5.9258	3.2992	1.2069	0.6187	0.0225	-3.3612	-2.4781	0.7021	0.8629
Adjusted R Squared	0.3909	0.6453	0.7757	0.8391	0.8409	0.8772	0.8895	0.9339	0.9342
p-value	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16
Observations	470	470	470	470	470	470	470	470	470

5.2.2 Heterogeneity and Panel data modeling

As the construction of the HSR grid in China is a rapid and dynamic process, this paper uses panel data instead of cross-sectional data. Panel data includes more information and variability. The dataset used in this study contains observations about 94 cities in south China across time.

For panel data, there might be a heterogeneity effect. Heterogeneity indicates the likely dependence across data observations within the same group. For example, the country's "oriented poverty reduction program" is likely to improve GDP in several cities in Yunan, Guizhou and Jiangxi.

When modeling panel data, if consider a simple least squared linear model(pooling):
$lnY_{it} = \beta_0 + \beta_1 lnYT_{it-2} + \beta_2 lnX_{it} + \varepsilon_{it}$.
This assumes that the model parameters are common across individuals. The estimates are likely to be inconsistent and biased.

Instead, fixed effect and random effect models are commonly used. For fixed effect model: $lnY_{it} = \beta_0 + \beta_1 lnYT_{it-2} + \beta_2 lnX_{it} + \varepsilon_{it} + u_i$, it includes a constant intercept term u_i that varies across individuals but is constant across time. It will estimate the model parameter β_0 and intercept u_i for each of the N groups in the model. For random effect model: $lnY_{it} = \beta_0 + \beta_1 lnYT_{it-2} + \beta_2 lnX_{it} + \varepsilon_{it} + \rho z_i$.The individual-specific effects ρz_i will follow a random distribution instead of a fixed value.

The choice between a pooling model, a fixed-effect model, and random-effect model is typically based on the Hausman test and Breusch-Pagan Lagrange multiplier (LM) test. The Hausman test can be used to decide between fixed and random effect model. The null hypothesis is that the preferred model is random effect, the alternative is the fixed effect. Thus, if the p-value is significant (<0.05), the fixed effect model is more appropriate. The Breusch-Pagan Lagrange multiplier (LM) test is useful to decide between a random effect regression and an OLS model. The null hypothesis is that there is no significant difference across units. If the p-value is significant (<0.05), we can reject the null hypothesis and conclude that the random effect model is more suitable.

As the time frames in this study are small, there is no need for a stationarity test. Table 4 summarizes the regression results and each column represents a separate regression. The asterisk indicates whether the t-statistics is significant at the 10% level (one asterisk), 5% level (two asterisks) or 1% level (three asterisks). The final three rows show the summary statistics for the panel data tests. The test results for 9 separate regressions all imply that the fixed effect models are more appropriate.

In equation (1), no control variables are added, the YT-2 coefficient is significant with a value of 0.1816. With other control variables being added in a step-wise fashion to equations (2)-(9), the coefficient of YT-2 decreases and the adjusted R^2 increases. In equation (9), all independent variables have passed the 10% significance level test. Except for the variable "Localization Economies of Scale (LE), all other variables have positive effect on the economy. The coefficients of YT-2, Length of Highways (Road) and Urbanization Rate (Urban) are 0.0536, 1.4722 and 0.4340.

The results from the fixed effect model and ordinary least squared model (pooling) have shown some differences. Although all variables have passed the 10% significance level test, the coefficients are diverse in value. In comparison with the fixed-effects model, the OLS model has underestimated the coefficients of the lagged YT-2, variable Length of Highways (Road), Number of college students(CS), Localization Economies of Scale(LE) and Porter Externality(PE), and overestimated the coefficients of the others variables. The fixed effect model has captured unobserved variables with some idiosyncrasy per city. When the model gets more complex, the average constant intercept term u_i becomes smaller with less idiosyncrasy values.

Table 4. Results of the panel model

Variables	1			2			3			4			5		
	Fixed	Random	Pooling	Fixed	Random	Pooling	Fixed	Random	Pooling	Fixed	Random	Pooling	Fixed	Random	Pooling
Market Potential (YT)	0.1816***	0.1891***	0.2326***	0.1064***	0.1128***	0.1256***	0.1052***	0.1010***	0.0930***	0.1033***	0.0906***	0.0570***	0.0636***	0.0941***	0.0595***
Total Fixed Assets(K)				0.2633***	0.2808***	0.4741***	0.2641***	0.2906***	0.3577***	0.2407***	0.2513***	0.2726***	0.1567***	0.2200***	0.2654***
Total Employment(L)							0.0303	0.2436***	0.5620***	0.0423	0.2990***	0.5370***	0.0597*	0.1889***	0.4984***
Total Exports(E)										0.0482***	0.0925***	0.1228***	0.0273***	0.0933***	0.1324***
Length of Highways (Road)													1.8193***	0.3506***	0.0785**
Urbanization Rate (Urban)															
Number of Students in Colleges (CS)															
Localization Economies of Scale (LE)															
Porter Externality (PE)															
Constant		6.1204	5.9258		4.6223	3.2993		3.2976	1.2069		2.2347	0.6187		-0.2397	0.0225
Adjusted R Squared	0.2537	0.3958	0.3909	0.5831	0.6419	0.6453	0.5826	0.6625	0.7757	0.6796	0.7146	0.8391	0.758	0.7185	0.8409
p-value	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16
Observations	n=94, T=5, N=470			n=94, T=5, N=470			n=94, T=5, N=470			n=94, T=5, N=470			n=94, T=5, N=470		
F test for individual effects	<2.2e-16			<2.2e-16			<2.2e-16			<2.2e-16			<2.2e-16		
Hausman Test	0.0443			0.00012			<2.2e-16			<2.2e-16			<2.2e-16		
Lagrange Multiplier Test	<2.2e-16			<2.2e-16			<2.2e-16			<2.2e-16			<2.2e-16		

Variables	6			7			8			9		
	Fixed	Random	Pooling	Fixed	Random	Pooling	Fixed	Random	Pooling	Fixed	Random	Pooling
Market Potential (YT)	0.0569***	0.0690***	0.0409***	0.0546***	0.0647***	0.0374***	0.0541***	0.0351***	0.0290***	0.0536***	0.0381***	0.0335***
Total Fixed Assets(K)	0.1358***	0.1651***	0.2147***	0.1291***	0.1550***	0.1915***	0.1223***	0.1621***	0.1563***	0.1185***	0.1630***	0.1579***
Total Employment(L)	0.0564*	0.1872***	0.4635***	0.0540*	0.1652***	0.4147***	0.1846***	0.6021***	0.6493***	0.2084***	0.5997***	0.6540***
Total Exports(E)	0.0236**	0.0657***	0.0790***	0.0238**	0.0622***	0.0767***	0.0235**	0.0311***	0.0276***	0.0188**	0.0340***	0.0321***
Length of Highways (Road)	1.5139***	0.3990***	0.2041***	1.4938***	0.37396***	0.1842***	1.5155***	0.1931***	0.1694***	1.4722***	0.1973***	0.1724***
Urbanization Rate (Urban)	0.5094***	0.8699***	0.8982***	0.4800***	0.8019***	0.7795***	0.4513***	0.5974***	0.5495***	0.4340***	0.6050***	0.5531***
Number of Students in Colleges (CS)				0.1008***	0.1543***	0.1471***	0.1020***	0.0602***	0.0493***	0.1080***	0.0573***	0.0408**
Localization Economies of Scale (LE)							-0.1282***	-0.4532***	-0.5006***	-0.1421***	-0.4537***	-0.5111***
Porter Externality (PE)										0.1800***	-0.0323	-0.05700*
Constant		-3.2243	-3.3611		-2.7101	-2.4781		0.1995	0.7021		0.2374	0.8629
Adjusted R Squared	0.7898	0.7909	0.8772	0.7984	0.8082	0.8895	0.8038	0.9201	0.9339	0.8105	0.9183	0.9342
p-value	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16	<2.2e-16
Observations	n=94, T=5, N=470			n=94, T=5, N=470			n=94, T=5, N=470			n=94, T=5, N=470		
F test for individual effects	<2.2e-16			<2.2e-16			<2.2e-16			<2.2e-16		
Hausman Test	<2.2e-16			<2.2e-16			<2.2e-16			<2.2e-16		
Lagrange Multiplier Test	<2.2e-16			<2.2e-16			<2.2e-16			<2.2e-16		

5.2.3 Spatial Panel Data Model

As mentioned earlier, panel data models can model dynamic relations and account for the presence of unobservable heterogeneity. However, for observations of this topic (macro data of 94 prefectural-level cities), there are also interaction effects among geographical units. Typically, the observations in space are not independent. The observations from one point tend to exhibit values similar to those from nearby locations.

A spatial lag model is commonly used to accommodate spatial dependence. It allows observations of the dependent variable Y in area i to depend on observations in neighboring areas j. It assumes that such dependence in the spatial lag term ρWY of the response variable Y. Where ρ stands for the strength of autoregression of the response variables. W is a matrix representing the spatial structure (e.g. neighbor weights). The regression equation is as follows,

$$lnY_i = \rho WlnY_j + \beta_0 + \beta_1 lnYT_i + \beta_2 lnX_i + \varepsilon_i. \tag{3}$$

Recently, spatial econometric methods have been employed intensively with the increased availability of panel datasets. In the fixed effects model, a dummy variable u_i is introduced for each spatial unit and for each time period. This could be estimated using the **splm** package in R,

$$lnY_{it} = \rho WlnY_{jt} + \beta_0 + \beta_1 lnYT_{it-2} + \beta_2 lnX_{it} + \varepsilon_{it} + u_i. \tag{4}$$

The Akaike Information Criterion (AIC) is a test commonly used to compare models with different specifications in terms of how parsimonious their specifications are—i.e., whether additional variables in one or more of the model specifications warrant inclusion in terms of the additional explanatory power they provide. An AIC test can be used to evaluate a model's fit on a set of training data and in which case is computed as:

$$AIC = -2(\log likelihood) + 2k, k\ is\ the\ number\ of\ model\ parameters. \qquad (5)$$

For models using the same datasets, the lower the AIC score, the better the model.

Tables 5 and 6 summarize the results from the fixed-effect panel model and spatial lag fixed effect model. Each column represents a separate regression. As before, the asterisk indicates whether the t-statistics is significant at the 10% level (one asterisk), 5% level (two asterisks) or 1% level (three asterisks).

The final two rows show the summary statistics for the log-likelihood and AIC test values. Compared with the results for the panel model and spatial panel model, the AIC values for the spatial panel models are more negative, which indicates that the spatial panel models have better estimations. For the spatial panel model, with other control variables added to equations (2)-(9), the AIC value decreases. This implies that the model specifications with more regressors are preferred.

For the spatial panel model, , In equation (1), no control variables are added, the YT-2 coefficient is significant with a value of 0.0457. The spatial autoregressive coefficient is significant with a value of 0.1573, indicating that there is a positive spillover of GDP on neighboring cities. With other control variables gradually adding to equation (2)-(9), the coefficient of YT-2 decreases. In equation (9), except for the variable "Total Exports", all other independent variables have passed the 10% significance level test. The coefficient of the market potential (YT-2) is significant with a value of 0.0197. It implies that a one-unit increase in market potential leads to a long-term increase in GDP of 0.0197. The spatial autoregressive coefficient is significant with a value of 0.0969, indicating there are positive spillover effects of GDP on neighboring cities. An increase in GDP in one city will increase the value of GDP in neighboring cities as the synergy effects among each other rather than competition effects take the lead.

Table 5. Results of the fixed effect panel model with AIC values

Variables	1	2	3	4	5	6	7	8	9
Market Potential (YT-2)	0.1816	0.1064***	0.1052***	0.1033***	0.0636***	0.0569***	0.0546***	0.0541***	0.0536***
Total Fixed Assets(K)		0.2633***	0.2641***	0.2407***	0.1567***	0.1358***	0.1291***	0.1223***	0.1185***
Total Employment(L)			0.0303	0.0423	0.0597*	0.0564*	0.0540*	0.1846***	0.2084***
Total Exports(E)				0.0482***	0.0273***	0.0236**	0.0238**	0.0235**	0.0188**
Length of Highways (Road)					1.8193***	1.5139***	1.4938***	1.5155***	1.4722***
Urbanization Rate (Urban)						0.5094***	0.4800***	0.4513***	0.4340***
Number of Students in Colleges (CS)							0.1008***	0.1020***	0.1080***
Localization Economies of Scale (LE)								-0.1282***	-0.1421***
Porter Externality (PE)									0.1800***
log likelihood	-50.44	87.04	87.37	95.73	216.74	250.53	260.97	267.95	276.77
AIC value	104.88	-168.07	-166.74	-181.45	-421.47	-487.06	-505.95	-517.89	-533.55

Table 6. Results of the spatial panel model

Variables	1	2	3	4	5	6	7	8	9
Market Potential (YT-2)	0.0457	0.0316***	0.0311***	0.0314***	0.0232***	0.0204***	0.0203***	0.0197***	0.0197***
Total Fixed Assets(K)		0.1247***	0.1252***	0.1211***	0.0973***	0.0850***	0.0838***	0.0767***	0.0744***
Total Employment(L)			0.0136***	0.0165	0.0314	0.0304	0.0301	0.1607***	0.1790***
Total Exports(E)				0.0112	0.0067	0.0051	0.0055	0.0051	0.0018
Length of Highways (Road)					1.0246***	0.8421***	0.8494***	0.8685***	0.8440***
Urbanization Rate (Urban)						0.3712***	0.3640***	0.3348***	0.3232***
Number of Students in Colleges (CS)							0.0349**	0.0358**	0.0412***
Localization Economies of Scale (LE)								-0.1283***	-0.1388***
Porter Externality (PE)									0.1358*
spatial autoregressive coefficient	0.1573***	0.1323***	0.1322***	0.1309***	0.1052***	0.0999***	0.0978***	0.0982***	0.0969***
log likelihood	212.6154	287.3591	287.5327	288.6595	357.4289	391.3477	393.6129	406.7194	416.2375
AIC value	-231.2308	-378.7181	-377.0655	-377.319	-512.8578	-578.6955	-581.2258	-605.4388	-622.4749

5.2.4 Comparison of the OLS, panel data model and spatial panel data model

Table 7 shows the results of the model on total observations from the OLS model, panel data fixed effect model and spatial lag panel model.

In all models, the estimated coefficient of the explanatory variable, market potential, is statistically discernible from zero at a conventional level of significance, indicating that it has a positive effect on economic development. Estimates of the coefficients range in value from 0.0197 to 0.0536.

However, from the OLS model to the spatial panel model, the models are improved. If we do not consider the heterogeneity effect, our estimation results will be underestimated. If we do not consider the heterogeneity effect and spatial spillover effects, our estimation results will be overestimated. The results from the spatial lag fixed effect model imply that 1 unit increase in market potential leads to a long-term increase in GDP of 0.0197.

Table 7. Comparison of models

OLS model		**Panle Data Fixed Model**		**Spatial Lag Panle Data Fixed Model**	
Market Potential (YT-2)	0.0335***	Market Potential (YT-2)	0.0536***	Market Potential (YT-2)	0.0197***
Total Fixed Assets(K)	0.1579***	Total Fixed Assets(K)	0.1185***	Total Fixed Assets(K)	0.0744***
Total Employment(L)	0.6540***	Total Employment(L)	0.2084***	Total Employment(L)	0.1790***
Total Exports(E)	0.0321***	Total Exports(E)	0.0188**	Total Exports(E)	0.0018
Length of Highways (Road)	0.1724***	Length of Highways (Road)	1.4722***	Length of Highways (Road)	0.8440***
Urbanization Rate (Urban)	0.5531***	Urbanization Rate (Urban)	0.4340***	Urbanization Rate (Urban)	0.3232***
Number of Students in Colleges (CS)	0.0408**	Number of Students in Colleges (CS)	0.1080***	Number of Students in Colleges (CS)	0.0412***
Localization Economies of Scale (LE)	-0.5111***	Localization Economies of Scale (LE)	-0.1421***	Localization Economies of Scale (LE)	-0.1388***
Porter Externality (PE)	-0.0570*	Porter Externality (PE)	0.1800***	Porter Externality (PE)	0.1358*
Constant	0.8629	Constant			
Adjusted R Squared	0.9342	Adjusted R Squared	0.8105		
p-value	<2.2e-16	p-value	<2.2e-16		
Observations	470	Observations	n=94, T=5, N=470		
		F test for individual effects	<2.2e-16		
		Hausman Test	<2.2e-16		
		Lagrange Multiplier Test	<2.2e-16		
		log likelihood	276.77	log likelihood	416.2375
		AIC value	-533.55	AIC value	-622.4749
				spatial autoregressive coefficient	0.0969***

5.3 Model on Group Regression

This study also intends to analyze whether the effects of HSR on economic growth are balanced in developed and backward regions. The observations are divided into developed and backward groups based on GDP values in 2009. The cities with a GDP above the average GDP values are noted as developed groups[4]. The cities with a GDP below the average GDP values are noted as backward groups. The detail information of Group 1 and Group 2 can be found in appendix 3.

Table 8 shows the results of group regression. OLS model, panel model and spatial panel model were all conducted for developed and developing areas. The test results all indicate that spatial panel models are more appropriate. Compare the coefficients of market potential using different models, if we do not consider the heterogeneity effect and spatial spillover effects, our estimation results for the coefficient of the explanatory variable will be overestimated.

The coefficients of market potential using spatial panel model for developed and developing areas are all significant with a value of 0.0293 and 0.0183. This indicates that there are positive effects of HSR on economic development. However, the coefficient for the developed area is larger than the coefficient for the developing area, which indicates that the effects of HSR on economic growth in developed and backward regions are unbalanced. The developed region would benefit more from the construction of the HSR, and this may exacerbate regional development inequality.

[4] The developed group includes 9 (of 9) cities in Fujian Province, 9 (of 14) cities in Hunan Province, 15 (of 21) cities in Guangdong Province, 5 (of 11) cities in Jiangxin Province, 4 (of 14) cities in Guangxi Province, 2 (of 9) cities in Guizhou Province and 3 (of 16) cities in Yunnan Province.

Table 8. Results of group regression

OLS model (Developed Area)		**OLS model** (Developing Area)	
Market Potential (YT-2)	**0.0606*** **	**Market Potential (YT-2)**	**0.0276*** **
Total Fixed Assets(K)	0.1673***	Total Fixed Assets(K)	0.1533***
Total Employment(L)	0.7837***	Total Employment(L)	0.4327***
Total Exports(E)	0.0487***	Total Exports(E)	0.0036
Length of Highways (Road)	0.1082***	Length of Highways (Road)	0.2660***
Urbanization Rate (Urban)	0.2796***	Urbanization Rate (Urban)	0.6959***
Number of Students in Colleges (CS)	-0.0061	Number of Students in Colleges (CS)	0.1087**
Localization Economies of Scale (LE)	-0.7157***	Localization Economies of Scale (LE)	-0.3150***
Porter Externality (PE)	-0.0598	Porter Externality (PE)	0.0915*
Constant	2.4153	Constant	-0.7423
Adjusted R Squared	0.9462	Adjusted R Squared	0.839
p-value	<2.2e-16	p-value	<2.2e-16
Observations	235	Observations	235
Panel Data Fixed Model（Developed Area）		**Panel Data Fixed Model** (Developing Area)	
Market Potential (YT-2)	**0.0347*** **	**Market Potential (YT-2)**	**0.0470*** **
Total Fixed Assets(K)	0.0831***	Total Fixed Assets(K)	0.0893***
Total Employment(L)	1.7915***	Total Employment(L)	0.1077**
Total Exports(E)	0.1159***	Total Exports(E)	0.0090
Length of Highways (Road)	0.9621***	Length of Highways (Road)	1.2975***
Urbanization Rate (Urban)	0.1104*	Urbanization Rate (Urban)	0.6851***
Number of Students in Colleges (CS)	0.0649***	Number of Students in Colleges (CS)	0.3054***
Localization Economies of Scale (LE)	-1.1965***	Localization Economies of Scale (LE)	-0.0663***
Porter Externality (PE)	0.0066	Porter Externality (PE)	0.1609***
Constant		Constant	
Adjusted R Squared		Adjusted R Squared	
p-value	<2.2e-16	p-value	<2.2e-16
Observations	n=47, T=5, N=235	Observations	n=47, T=5, N=235
F test for individual effects	<2.2e-16	F test for individual effects	<2.2e-16
Hausman Test	<2.2e-16	Hausman Test	<2.2e-16
Lagrange Multiplier Test	0.0017	Lagrange Multiplier Test	1.31E-08
log likelihood	242.6976	log likelihood	133.1769
AIC value	-465.3974	AIC value	-246.3558

Spatial Lag Panle Data Fixed Model (Developed Area)		**Spatial Lag Panle Data Fixed Model** (Developed Area)	
Market Potential (YT-2)	**0.0293*** **	**Market Potential (YT-2)**	**0.0183** **
Total Fixed Assets(K)	0.0635***	Total Fixed Assets(K)	0.0805***
Total Employment(L)	1.6754***	Total Employment(L)	0.0621*
Total Exports(E)	0.0754***	Total Exports(E)	-0.0045***
Length of Highways (Road)	0.6161***	Length of Highways (Road)	0.9839***
Urbanization Rate (Urban)	0.0971**	Urbanization Rate (Urban)	0.54711**
Number of Students in Colleges (CS)	0.0465***	Number of Students in Colleges (CS)	0.1346***
Localization Economies of Scale (LE)	-1.2448***	Localization Economies of Scale (LE)	-0.0423
Porter Externality (PE)	-0.00358	Porter Externality (PE)	0.1455***
log likelihood	280.0789	log likelihood	177.7011
AIC value	-444.1579	AIC value	-239.4022
spatial autoregressive coefficient	0.0800***	spatial autoregressive coefficient	0.1156***

CHAPTER 6 RESEARCH CONCLUSIONS AND LIMITATIONS

This thesis has presented an empirical analysis of the effect of the HSR system (market potential as an explanatory variable) on economic development (GDP as a dependent variable) in South China. The data employed in this study comprise observations on 94 prefectural cities in the years 2009, 2011, 2013, 2015 and 2017. A group of control variables[5] has been included in the model. The VIF test values for all variables lie between 1-3, which indicates that multicollinearity between variables is low and that no variables need to be included from model specifications because of collinearity. The OLS model, panel model considering heterogeneity effect, spatial panel model considering heterogeneity and spatial spillover effects have each been estimated using all available observations and group observations.

Considering all observations together, the coefficients of the explanatory variable market potential from the OLS model, panel model and spatial panel model are 0.0335, 0.0536, 0.0197 respectively. The F test, Hausman test, Lagrange Multiplier test and AIC value all imply that the spatial lag model with fixed effect is most appropriate. So, we can infer from these results that: (1) The construction of the HSR system has enhanced the overall regional economic development in this region. A one-unit increase in market potential will lead to a long-term increase in GDP of 0.0197. (2) There are heterogeneity and interaction effects among geographical units. If we do not consider these effects, we may overestimate the impacts of the HSR system development. Moreover, the spatial autoregressive coefficient is 0.0969, which implies that there are positive spillover effects between cities, the cities have synergy effects with their neighbors.

[5] *Total Investment in Fixed Assets, Total Number of Employment, Total volumes of Exports and Imports, Length of Highways, Urbanization Rates, Number of College Students, Localization economies of scale, Urbanization economies of scale*

For group observations, this study has divided observations into two groups based on GDP values in 2009 as developed group and backward group. The test results all indicate that spatial panel models are more appropriate. (3) The coefficients of market potential using the spatial panel model for developed and backward areas are all significant with a value of 0.0293 and 0.0183. This indicates that there are positive effects of HSR on economic development. However, the coefficient for the developed area is larger than the coefficient for the developing area, which indicates that the effects of HSR on economic growth in developed and backward regions are unbalanced. The developed region would benefit more from the construction of the HSR, and this may exacerbate regional development inequality.

Xueliang Zhang (2012) concluded that there is a significant impact of transportation infrastructure on economic development. In his study, using panel data of 30 Chinese provinces from 1993 to 2009, the coefficient of investment in transportation infrastructure was found to be in the range of 0.05-0.07. Nannan Yu (2013) analyzed the spillover effects of transportation in four different regions. Her results showed that positive spillovers exist due to the connectivity at the national level. Compared to related researches, the findings of this study are largely consistent with the others. The spillover effects between cities have been positive and there has been a significant impact of HSR on economic development. The coefficient of market potential on economic development is estimated to be 0.0197, which is lower than 0.05-0.07, but it would appear to be reasonable, as it only represents the effect of HSR and not other types of transportation infrastructure.

In all, the data and methodology used in this thesis are adequate to answer the research questions and hypothesis. The research has proved that there is a positive effect of HSR on regional economic growth in South China and there is no evidence of an unbalanced effect of the HSR on developed and backward regions in South China. There are heterogeneity and interaction effects among geographical units. If these effects are not considered, the impact of the HSR system on economic growth may be overestimated. The results are to a large extent consistent with previous scholars. The methods could be employed and developed to further analyze the effect of the HSR on other regions. The results may differ because of different HSR networks, geography conditions, and other factors, which is useful for the HSR policy decision-making.

For the limitations of this research, this paper has used $YT_{i,t} = \sum_{j=1}^{n} \frac{GDP_{j,t}}{T_{ij,\,t}}\ (i \neq j)$ as its explanatory variable and capture the time-saving characteristics of the HSR. The calculation of $T_{ij,\,t}$ has used route network map and calculates the shortest path travel time between every two cities. However, it is an ideal version of data as it may ignore the time people need to wait and transfer the HSR. Moreover, for the spatial panel model, only the basic binary weights have been employed. A comparison of different neighbour weightings (binary weights, time weights, distance weights…) could be conducted in future studies.

At last, since the coefficient of market potential is roughly 0.02, it represents a proportional change in regional GDP. We can reasonably expect such a growth effect to continue. The construction of the HSR can strengthen links between cities and enhance overall economic growth. In the long run, the investment in the HSR infrastructure would be cost-effective. It can promote the establishment of the domestic market for the specialized labor division and mass product circulation.

However, there is also evidence that the effects of HSR on economic growth in developed and backward regions are unbalanced. The developed region would benefit more from the construction of the HSR. The HSR would boost the flow of labor and migration from backward regions to the developed regions. As a result, the policy should also focus on improving the institutional environment and human capital level and cultivate the new growth center in backward regions.

REFERENCE

Berechman, J., Ozmen, D., & Ozbay, K. (2006). Empirical analysis of transportation investment and economic development at state, county and municipality levels. *Transportation*, *33*(6), 537–551. https://doi.org/10.1007/s11116-006-7472-6

Boarnet, M. G. (1998). Spillovers and the Locational Effects of Public Infrastructure. *Journal of Regional Science*, *38*(3), 381–400. https://doi.org/10.1111/0022-4146.00099

Kim, K. S. (2000). High-speed rail developments and spatial restructuring. *Cities*, *17*(4), 251–262. https://doi.org/10.1016/s0264-2751(00)00021-4

Kim, S., & Lee, Y. H. (2002). Public Sector Capital and the Production Efficiency of U.S. Regional Manufacturing Industries. *The Japanese Economic Review*, *53*(4), 466–477. https://doi.org/10.1111/1468-5876.00020

Kojima, Y., Matsunaga, T., & Yamaguchi, S. (2017). The impact of new Shinkansen lines (Tohoku Shinkansen (Hachinohe – Shin-Aomori) and Kyusyu Shinkansen (Hakata – Shin-Yatsushiro)). *Transportation Research Procedia*, *25*, 344–357. https://doi.org/10.1016/j.trpro.2017.05.412

Li, J., Wen, J., & Jiang, B. (2017). Spatial Spillover Effects of Transport Infrastructure in Chinese New Silk Road Economic Belt. *International Journal of E-Navigation and Maritime Economy*, *6*, 1–8. https://doi.org/10.1016/j.enavi.2017.05.001

Ortega, E., López, E., & Monzón, A. (2012). Territorial cohesion impacts of high-speed rail at different planning levels. *Journal of Transport Geography*, *24*, 130–141. https://doi.org/10.1016/j.jtrangeo.2011.10.008

Rosenstein-Rodan, P. N. (1943). Problems of Industrialisation of Eastern and South-Eastern Europe. *The Economic Journal*, *53*(210/211), 202. https://doi.org/10.2307/2226317

Sasaki, K., Ohashi, T., & Ando, A. (1997). High-speed rail transit impact on regional systems: does the Shinkansen contribute to dispersion? *The Annals of Regional Science*, *31*(1), 77–98. https://doi.org/10.1007/s001680050040

Shi, Q.Y. (2018). High-Speed Railway and Regional Economic Growth: An Empirical Study Based on Market Potential. *American Journal of Industrial and Business Management*, 8, 83-102. https://doi.org/10.4236/ajibm.2018.81006

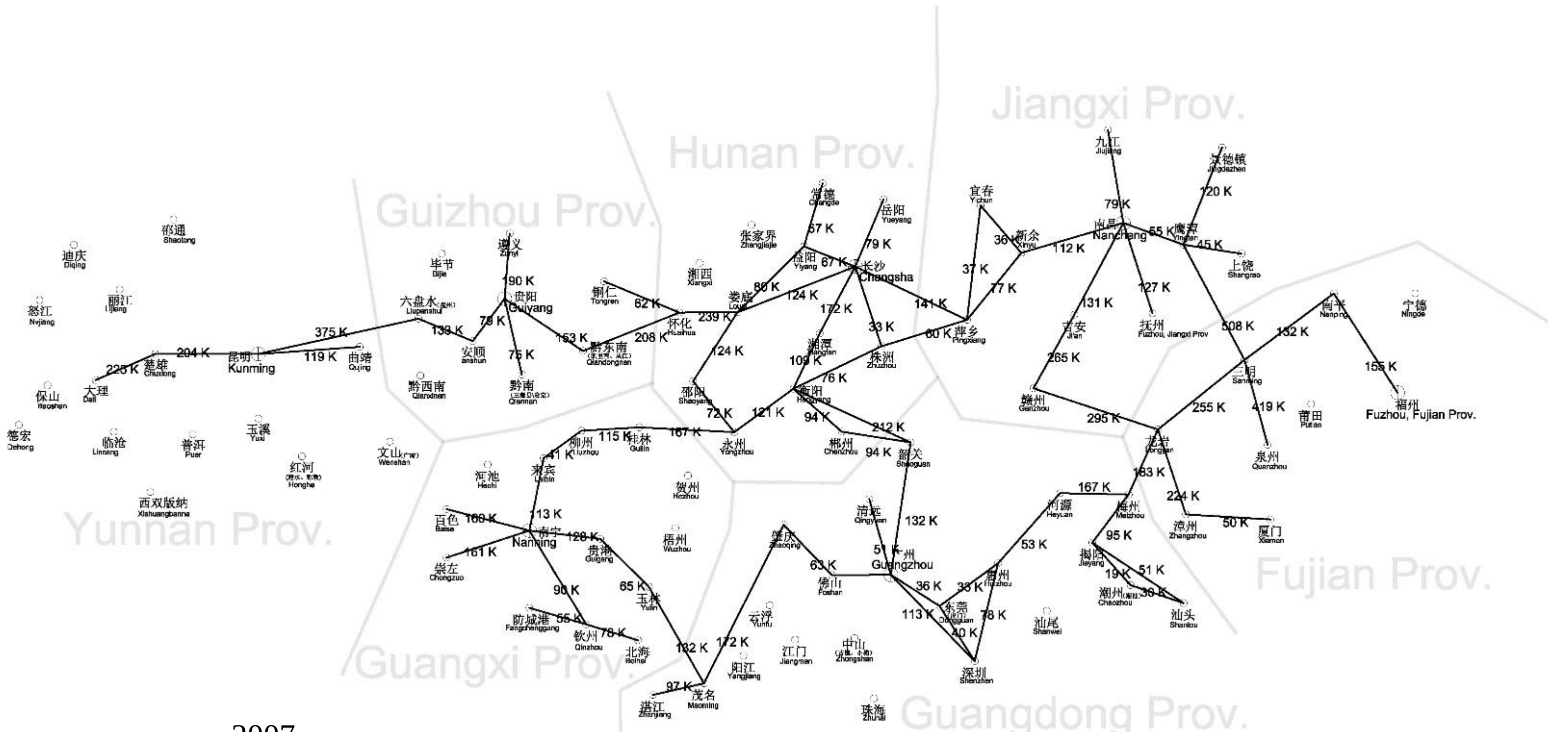
Xueliang, Z. (2013). Has Transport Infrastructure Promoted Regional Economic Growth? — With an Analysis of the Spatial Spillover Effects of Transport Infrastructure. *Social Sciences in China*, *34*(2), 24–47. https://doi.org/10.1080/02529203.2013.787222

Yu, N., de Jong, M., Storm, S., & Mi, J. (2013). Spatial spillover effects of transport infrastructure: evidence from Chinese regions. *Journal of Transport Geography*, *28*, 56–66. https://doi.org/10.1016/j.jtrangeo.2012.10.009

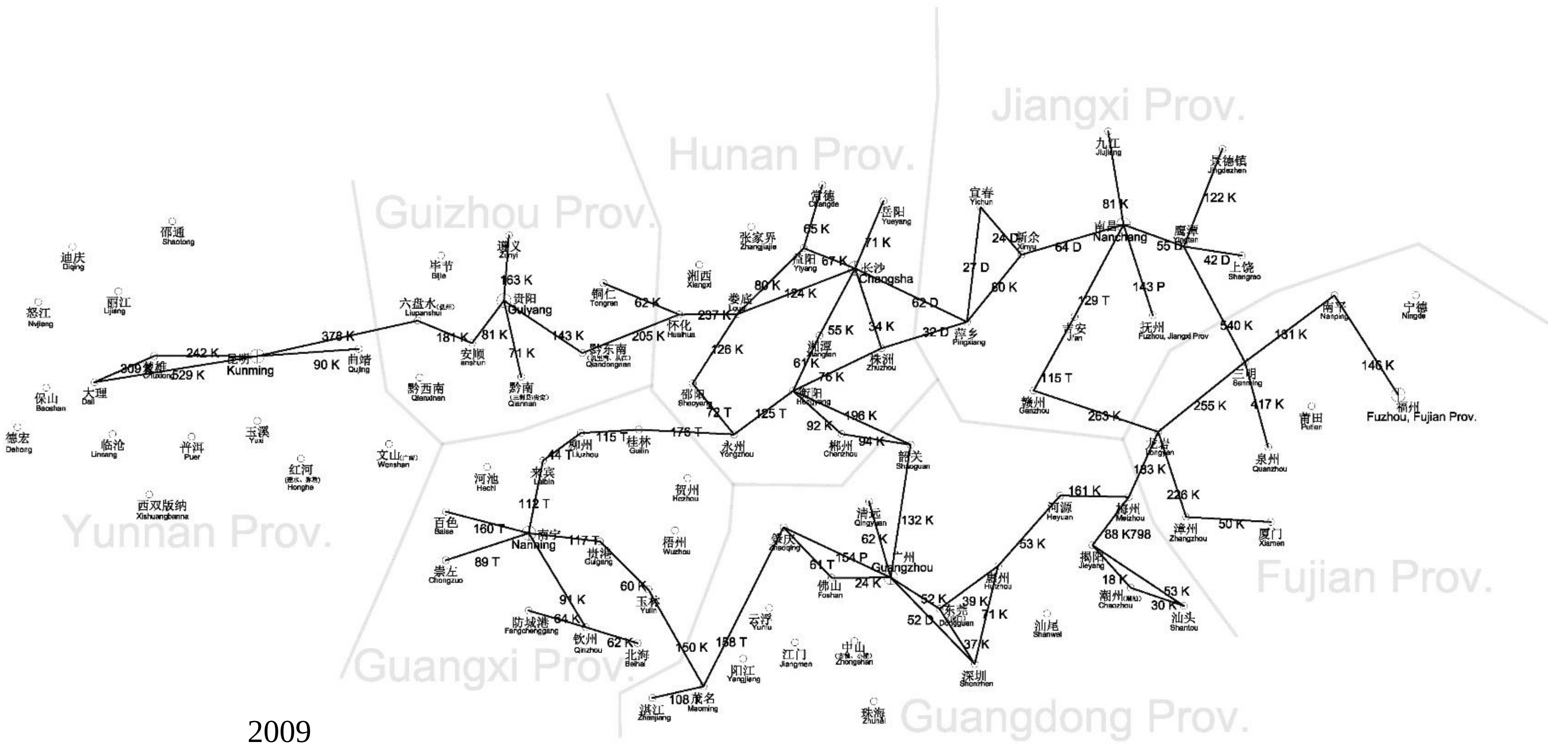
Zou, W., Chen, L., & Xiong, J. (2019). High-speed railway, market access and economic growth. *International Review of Economics & Finance*. https://doi.org/10.1016/j.iref.2019.11.014

APPENDIX

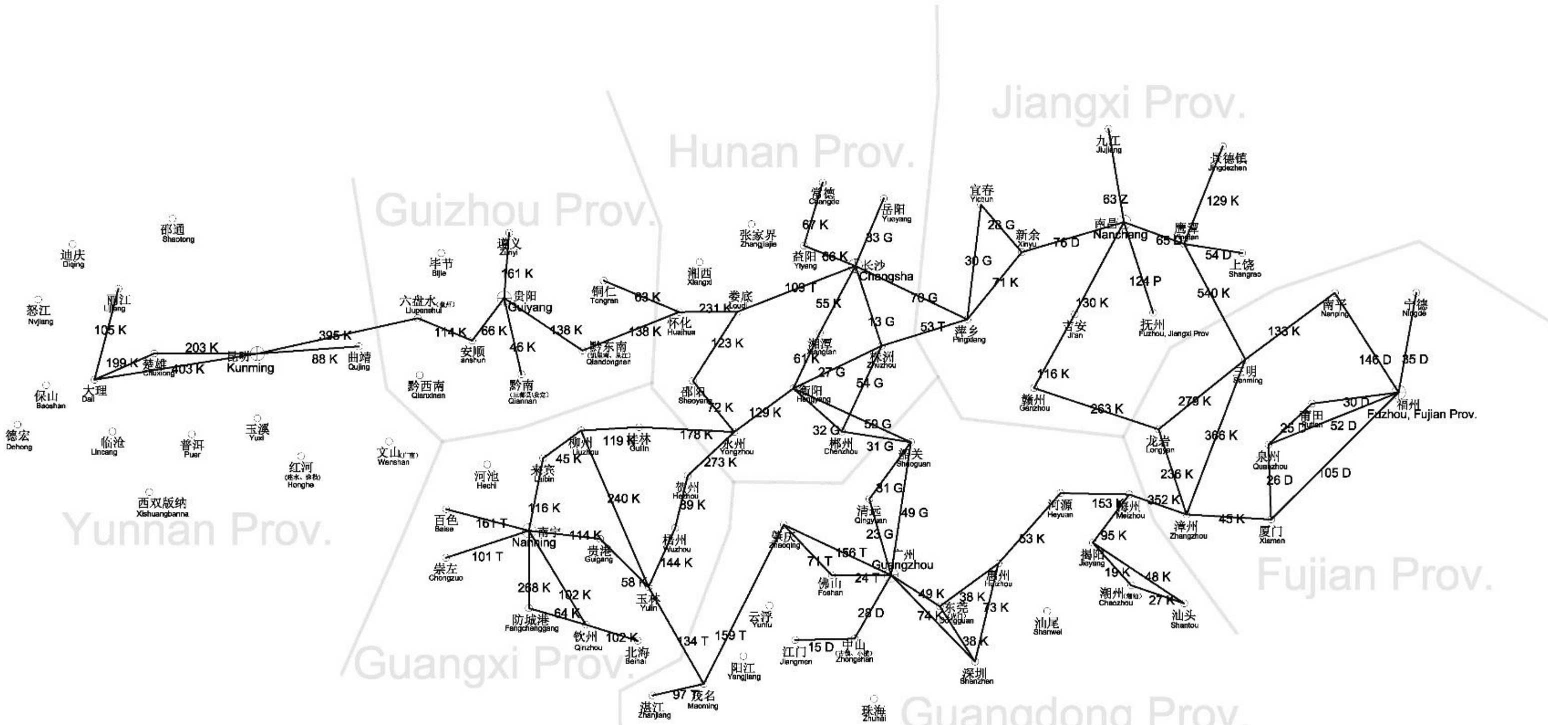
Appendix1. Train network and travel time *(number indicates travel time between two cities, G represents the HSR, K represents regular fast rail)*



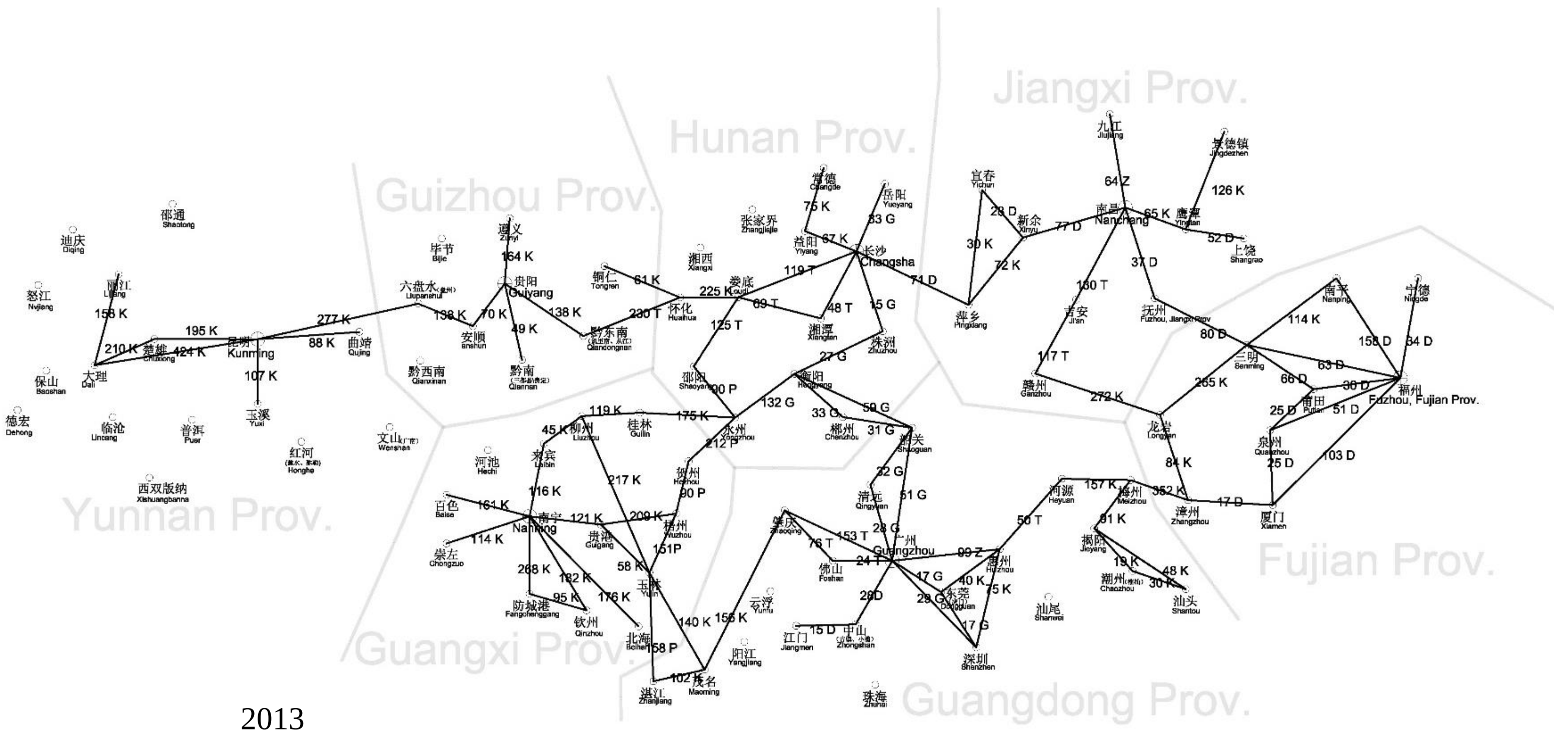
2007


Jiangxi Prov.
Hunan Prov.
Guizhou Prov.
Yunnan Prov.
Guangxi Prov.
Guangdong Prov.
Fujian Prov.
Kunming
Guiyang
Changsha
Nanchang
Nanning
Guangzhou
Fuzhou, Fujian Prov.

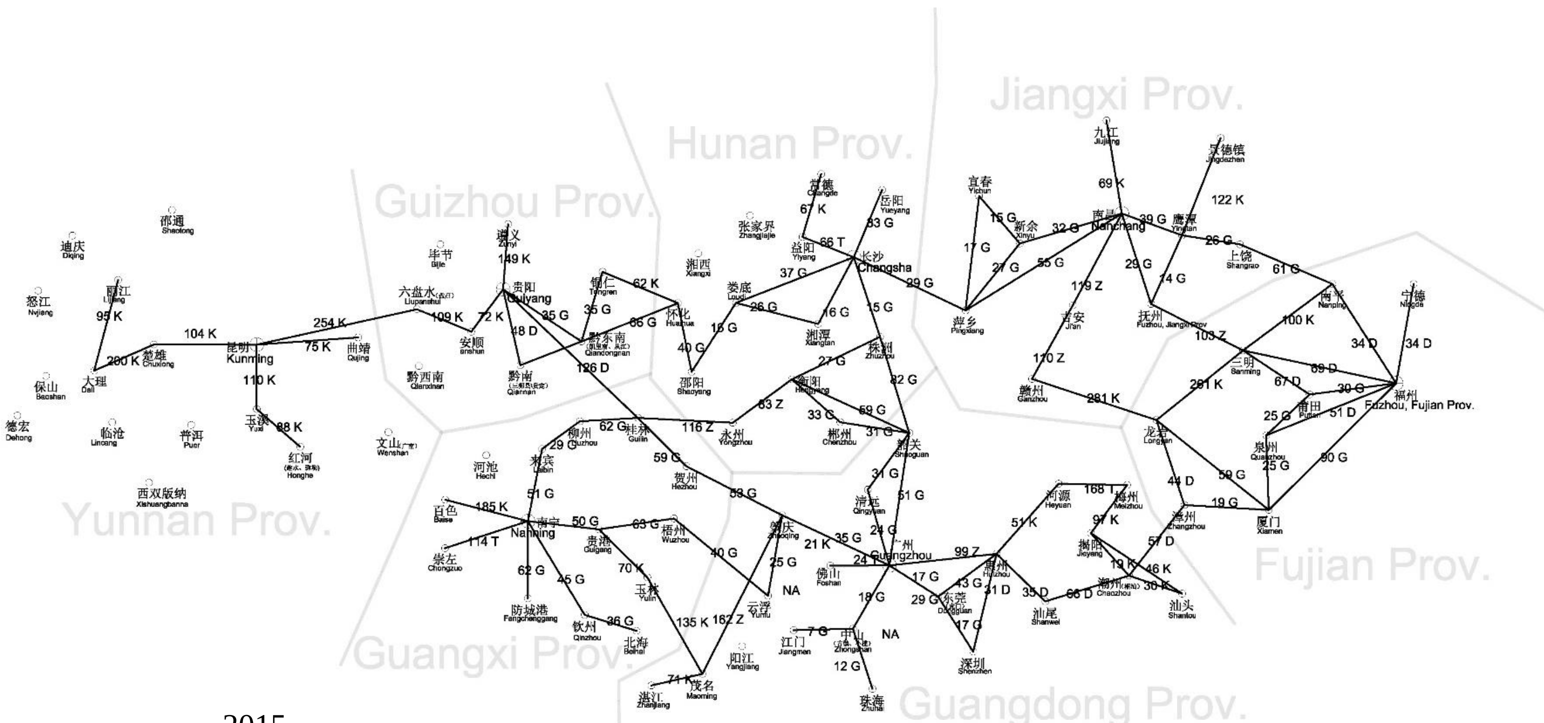
2009


Jiangxi Prov.
Hunan Prov.
Guizhou Prov.
Yunnan Prov.
Guangxi Prov.
Guangdong Prov.
Fujian Prov.
Kunming
Guiyang
Changsha
Nanchang
Nanning
Guangzhou
Fuzhou, Fujian Prov.

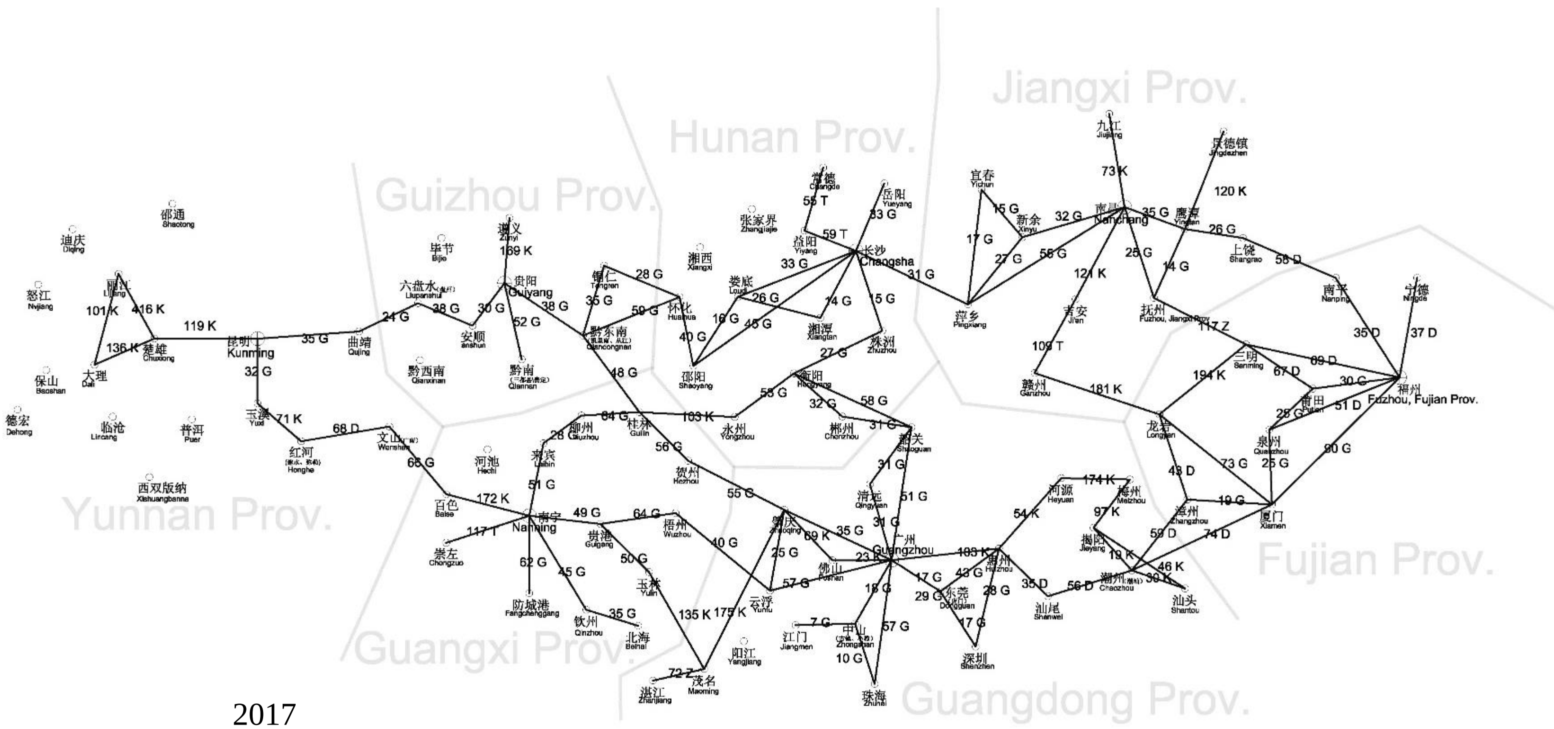
2011



2013


Jiangxi Prov.
Hunan Prov.
Guizhou Prov.
Yunnan Prov.
Guangxi Prov.
Guangdong Prov.
Fujian Prov.
长沙 Changsha
南昌 Nanchang
贵阳 Guiyang
昆明 Kunming
南宁 Nanning
广州 Guangzhou
福州 Fuzhou, Fujian Prov.
厦门 Xiamen
深圳 Shenzhen

2015


Jiangxi Prov.
Hunan Prov.
Guizhou Prov.
Yunnan Prov.
Guangxi Prov.
Guangdong Prov.
Fujian Prov.
Kunming
Guiyang
Changsha
Nanchang
Nanning
Guangzhou
Fuzhou, Fujian Prov.

2017

Appendix2. The results of YT values

	City	YT2007	YT2009	YT2011	YT2013	YT2015	YT2017
1	**Fuzhou.Fujian**	47	83	323	540	901	1,101
2	**Xiamen**	56	119	423	702	1,156	1,427
3	**Putian**	0	0	513	769	1,153	1,431
4	**Sanming**	67	121	166	526	722	933
5	**Quanzhou**	40	70	386	608	1,002	1,240
6	**Zhangzhou**	69	137	315	640	1,108	1,353
7	**Nanping**	66	114	195	334	910	1,119
8	**Longyan**	57	154	179	345	799	973
9	**Ningde**	0	0	332	528	831	996
10	**Nanchang**	111	207	351	517	980	1,199
11	**Jingdezhen**	83	138	219	301	523	646
12	**Pingxiang**	161	322	576	691	1,416	1,689
13	**Jiujiang**	101	172	309	430	692	826
14	**Xinyu**	140	274	465	604	1,232	1,495
15	**Yingtan**	127	213	331	458	1,018	1,270
16	**Ganzhou**	41	135	212	276	431	581
17	**Ji'an**	88	160	264	354	576	701
18	**Yichun**	144	284	501	624	1,222	1,467
19	**Fuzhou.Jiangxi**	89	145	253	554	1,017	1,283
20	**Shangrao**	100	170	263	367	853	1,059
21	**Changsha**	164	258	698	879	1,385	1,669
22	**Zhuzhou**	198	350	1,041	1,253	1,740	2,114
23	**Xiangtan**	0	286	561	701	1,616	2,072
24	**Hengyang**	161	264	789	1,009	1,374	1,669
25	**Shaoyang**	118	172	283	403	1,089	1,413
26	**Yueyang**	136	222	622	803	1,145	1,395
27	**Changde**	124	182	346	436	628	829
28	**Zhangjiajie**	0	0	0	0	0	0
29	**Yiyang**	158	252	492	631	897	1,149
30	**Chenzhou**	167	254	797	1,006	1,352	1,625
31	**Yongzhou**	129	193	249	486	804	1,159
32	**Huaihua**	83	122	223	316	846	1,133
33	**Loudi**	133	198	393	621	1,249	1,588
34	**Xiangxi**	0	0	0	0	0	0
35	**Guangzhou**	423	709	1,104	1,718	2,395	2,763
36	**Shenzhen**	285	528	794	1,445	1,972	2,334
37	**Shaoguan**	207	309	942	1,207	1,617	1,905
38	**Zhuhai**	0	648	974	1,262	2,141	1,520

	City	YT2007	YT2009	YT2011	YT2013	YT2015	YT2017
39	**Shantou**	111	170	264	334	878	1,061
40	**Foshan**	504	736	1,125	1,521	2,157	2,465
41	**Jiangmen**	0	666	1,000	1,302	2,509	0
42	**Zhanjiang**	109	141	251	319	506	599
43	**Maoming**	136	175	320	399	629	741
44	**Zhaoqing**	274	258	580	715	1,789	2,230
45	**Huizhou**	398	525	795	1,107	2,186	2,243
46	**Meizhou**	131	205	320	413	593	719
47	**Shanwei**	0	0	0	0	1,415	1,597
48	**Heyuan**	251	352	543	735	1,143	1,259
49	**Yangjiang**	0	0	0	0	0	0
50	**Qingyuan**	344	450	1,309	1,563	2,244	2,303
51	**Dongguan**	530	812	1,219	2,455	3,157	3,798
52	**Zhongshan**	0	848	1,254	1,638	2,881	2,925
53	**Chaozhou**	151	226	338	429	1,136	1,376
54	**Jieyang**	133	201	299	380	950	1,157
55	**Yunfu**	0	0	0	0	1,380	1,726
56	**Nanning**	97	136	208	259	682	898
57	**Liuzhou**	96	136	205	291	708	923
58	**Guilin**	99	144	199	325	852	1,157
59	**Wuzhou**	0	0	198	269	1,003	1,247
60	**Beihai**	80	114	156	200	545	706
61	**Fangchenggang**	74	114	171	180	571	737
62	**Qinzhou**	92	131	192	202	637	825
63	**Guigang**	109	150	250	310	812	1,032
64	**Yulin**	113	149	263	330	596	809
65	**Baise**	76	106	162	205	388	721
66	**Hezhou**	0	0	187	301	1,094	1,342
67	**Hechi**	0	0	0	0	0	0
68	**Laibin**	110	153	223	311	739	963
69	**Chongzuo**	76	126	185	227	471	598
70	**Guiyang**	63	91	168	213	673	981
71	**Liupanshui**	53	73	136	174	414	920
72	**Zunyi**	50	75	132	168	423	535
73	**Anshun**	65	92	169	216	555	1,003
74	**Bijie**	0	0	0	0	0	0
75	**Tongren**	80	116	206	289	630	1,105
76	**Qianxinan**	0	0	0	0	0	0
77	**Qiandongnan**	69	101	190	235	727	1,172

	City	YT2007	YT2009	YT2011	YT2013	YT2015	YT2017
78	**Qiannan**	63	92	176	226	600	811
79	**Kunming**	42	61	103	150	296	770
80	**Qujing**	45	68	111	156	293	905
81	**Yuxi**	0	0	0	150	277	778
82	**Baoshan**	0	0	0	0	0	0
83	**Zhaotong**	0	0	0	0	0	0
84	**Lijiang**	0	0	75	95	186	330
85	**Puer**	0	0	0	0	0	0
86	**Lincang**	0	0	0	0	0	0
87	**Chuxiong**	41	54	94	130	276	523
88	**Honghe**	0	0	0	0	236	582
89	**Wenshan**	0	0	0	0	0	816
90	**Xishuangbanna**	0	0	0	0	0	0
91	**Dali**	33	42	78	104	200	385
92	**Dehong**	0	0	0	0	0	0
93	**Nujiang**	0	0	0	0	0	0
94	**Diqing**	0	0	0	0	0	0

Appendix 3. Details of group observations

Group 1					Group 2			
City	Year	GDP	Province		City	Year	GDP	Province
Guangzhou	2009	9,138.21	Guangdong		Ji'an	2009	584.11	jiangxi
Shenzhen	2009	8,201.32	Guangdong		Shaoguan	2009	578.75	Guangdong
Foshan	2009	4,820.90	Guangdong		Loudi	2009	568.31	Hunan
Dongguan	2009	3,763.91	Guangdong		Honghe	2009	560.88	Yunnan
Changsha	2009	3,744.76	Hunan		Huaihua	2009	559.48	Hunan
Quanzhou	2009	3,069.50	Fujian		Yangjiang	2009	527.27	Guangdong
Fuzhou.Fujian	2009	2,604.04	Fujian		Meizhou	2009	519.29	Guangdong
Nanchang	2009	1,837.50	Jiangxi		Fuzhou.Jiangxi	2009	502.91	jiangxi
Kunming	2009	1,837.46	Yunnan		Bijie	2009	500.01	Guizhou
Xiamen	2009	1,737.23	Fujian		Xinyu	2009	484.17	Jiangxi
Zhongshan	2009	1,566.41	Guangdong		Chaozhou	2009	480.18	Guangdong
Nanning	2009	1,524.71	Guangxi		Wuzhou	2009	453.65	Guangxi
Huizhou	2009	1,414.70	Guangdong		Baise	2009	452.86	Guangxi
Jiangmen	2009	1,340.88	Guangdong		Guigang	2009	437.73	Guangxi
Yueyang	2009	1,272.15	Hunan		Liupanshui	2009	430.16	Guizhou
Changde	2009	1,239.23	Hunan		Pingxiang	2009	421.49	Jiangxi
Maoming	2009	1,231.25	Guangdong		Heyuan	2009	405.50	Guangdong
Zhangzhou	2009	1,178.01	Fujian		Dali	2009	404.50	Yunnan
Hengyang	2009	1,168.01	Hunan		Qinzhou	2009	396.18	Guangxi
Zhanjiang	2009	1,156.67	Guangdong		Shanwei	2009	390.04	Guangdong
Liuzhou	2009	1,046.05	Guangxi		Hechi	2009	382.77	Guangxi
Zhuhai	2009	1,038.66	Guangdong		Jingdezhen	2009	364.03	Jiangxi

Group 1					Group 2			
Shantou	2009	1,035.87	Guangdong		Yunfu	2009	344.51	Guangdong
Zhuzhou	2009	1,024.89	Hunan		Chuxiong	2009	343.95	Yunnan
Guiyang	2009	971.94	Guizhou		Beihai	2009	321.06	Guangxi
Guilin	2009	948.23	Guangxi		Zhaotong	2009	320.45	Yunnan
Ganzhou	2009	940.63	Jiangxi		Chongzuo	2009	304.36	Guangxi
Qujing	2009	870.94	Yunnan		Laibin	2009	303.14	Guangxi
Zhaoqing	2009	862.00	Guangdong		Qiannan	2009	297.97	Guizhou
Qingyuan	2009	861.59	Guangdong		Wenshan	2009	284.90	Yunnan
Chenzhou	2009	843.23	Hunan		Qianxinan	2009	274.83	Guizhou
Jiujiang	2009	831.36	Jiangxi		Qiandongnan	2009	269.73	Guizhou
Longyan	2009	824.88	Fujian		Xiangxi	2009	268.97	Hunan
Jieyang	2009	816.09	Guangdong		Yingtan	2009	256.80	Jiangxi
Sanming	2009	800.24	Fujian		Tongren	2009	251.74	Guizhou
Zunyi	2009	777.64	Guizhou		Fangchenggang	2009	251.04	Guangxi
Xiangtan	2009	739.38	Hunan		Hezhou	2009	249.22	Guangxi
Shangrao	2009	728.50	Jiangxi		Baoshan	2009	221.65	Yunnan
Yichun	2009	700.24	Jiangxi		Puer	2009	211.70	Yunnan
Putian	2009	691.42	Fujian		Zhangjiajie	2009	203.10	Hunan
Yulin	2009	683.49	Guangxi		Anshun	2009	195.70	Guizhou
Yuxi	2009	644.40	Yunnan		Lincang	2009	181.33	Yunnan
Yongzhou	2009	640.04	Hunan		Xishuangbanna	2009	138.64	Yunnan
Nanping	2009	621.65	Fujian		Lijiang	2009	120.67	Yunnan
Ningde	2009	612.28	Fujian		Dehong	2009	115.71	Yunnan
Shaoyang	2009	600.69	Guangdong		Diqing	2009	77.10	Yunnan
Yiyang	2009	591.62	Hunan		Nujiang	2009	48.05	Yunnan