NEKN03, Master Programme in

International Economics with a Focus on China

June 2019

Department of Economics

SCHOOL OF ECONOMICS AND MANAGEMENT

Lund University

THE RELATIONSHIP BETWEEN RAIL INFRASTRUCTURE AND REGIONAL DEVELOPMENT IN CHINA

Abstract: The regional disparities in China are increasing, at the same time as the growth moderates. In order to keep successful development in the future, China is rapidly developing the rail infrastructure, as one of the main channels to obtain a balanced regional development. The argument is that regional inequalities will decrease through increased economic interconnection and mobility provided by the rail system. This study aims to investigate how regional development, measured by levels of education, healthcare, employment and environmental pollutants, is affected by the level of rail infrastructure, measured by train travel intensity. A panel dataset of 31 provinces in China during the years 2008-2017 is empirically tested using fixed-effects method and a multivariate approach. The test results indicate that a higher level of rail infrastructure is positively correlated with employment rates, but negatively correlated with population share in higher education. No statistically significant effect was found on level of healthcare or environment. The overall results indicate that the relationship between rail infrastructure and regional development is contradictory. The study concludes that the selection of measurements representing regional development determines the observed correlation with rail infrastructure.

Keywords: Rail infrastructure, Regional development, Panel data, Multivariate approach, China

Author: Sandra Postaroff

Supervisor: Thomas Fischer

OVERVIEW

Title	The relationship between rail infrastructure and regional development in China
Seminar date	2019-06-07
Course	NEKN03, Degree Project in International Economics with a Focus on China, 15 ECTS
Author	Sandra Postaroff
Advisor	Thomas Fischer
Key words	Rail infrastructure, Panel data, Multivariate approach, Regional development, China
Purpose	The purpose of this study is to empirically investigate the relationship between rail infrastructure and regional development on the provincial level in China.
Theoretical perspectives	The analysis is based on theories and earlier research within the field of development and public economics.
Methodolgy	A panel dataset of 31 provinces in China during the years 2008-2017 is empirically tested using fixed-effects method and a multivariate approach.
Results	The main findings indicate that a higher level of rail infrastructure is positively correlated with employment rates, but negatively correlated with population share in higher education. No statistically significant effect was found on level of healthcare or environment.
Conslusion	The overall results indicate that the relationship between rail infrastructure and regional development is contradictory. The study concludes that the selection of measurements representing regional development determines the observed correlation with rail infrastructure.

PREFACE

This thesis is written during the spring of 2019. The writing process has been educative and intriguing. It has contributed to a deeper knowledge and understanding of the subject for me as the author. I hope the reader finds my research to be interesting and enlightening.

I also want to thank my supervisor Thomas Fischer for the guidance and knowledge he has provided throughout the progression of this paper.

Sandra Postaroff

TABLE OF CONTENTS

INTRODUCTION

In the last three decades China has displayed the most remarkable economic transformation that the world has ever seen (The world bank 2012). It has gone from a poor developing country to the fastest growing internationally acknowledged superpower, today being the second largest economy and the greatest exporter in the world (The world bank 2013). However, growth is slowing down, and China is facing issues of increasing disparities between regions (The World Bank 2013). There is a clear divide in China between the rural and urban areas, as well as between inland and coastal areas (Zheng and Kahn 2013). The wealthy big cities in the coastal areas are getting increasingly polluted and overcrowded. While many inland provinces still are underdeveloped and has a hard time improving their living standards through education, employment, healthcare, or by other means (Zheng and Kahn 2013). In addition, traces from the institutional boundaries such as the hukou system still lives on, dividing the people and creating inequalities (Cai 2008).

The regional disparities are a major issue that China needs to confront going forward in order to keep successful development of the country. Therefore, China have started to refocus towards the domestic market to strengthen the domestic development and shift away from being as reliant on investments and exports (Lin 2013; Cline 2010; Dollar 2013). So, with the motive of rebalancing the economy, China has in large switched out the previous export-led growth model to a strategy model focusing on achieving social stability and a sustainable economy (Rodrick 2006; The World Bank 2013). In achieving these goals, it is important for China to improve the economic integration between the provinces (Cai 2008; Zheng and Kahn 2013). One of the main channels the Chinese government uses to attain this objective, is rapid development of the railway system (EESI 2018). The Chinese government has especially done immense investments into developing the rail infrastructure since the implementation of the high-speed rail (HSR) policy in 2008 (Huang et al. 2018). Approximately half of the worldwide investments in HSR is occurring in China (over $300 billion to date) and they are expected to have far more rail lines than the rest of the world combined by 2025, when the HSR system is expected to be completed (The World Bank 2013; EESI 2018; Chen 2019). In terms of deployment speed and service scale, the development of the rail infrastructure in China is enormous and incomparable to similar systems in other countries (Chen 2019).

The tremendous improvements and expansions of the rail infrastructure is believed to improve both the living standard of the population living in the megacities and stimulate economic development of the poorer regions (Zheng and Kahn 2013). The Chinese government's overall agenda is to use the rail transportation system as a mean to obtain a balanced regional development through increased economic interconnection and mobility (Zheng and Kahn 2013). Most theories and literature support the belief that there do exist a positive correlation between regional development and rail infrastructure, but there is no consensus on the topic (Chen 2019). According to the critics, the overall effect of rail infrastructure on regional development is not necessarily positive (See for example Ansar et al. 2016, Chen and Haynes 2017; Button 2017; Wu et al. 2014).

1.1 Research Question

Given that one of the main motives given by the Chinese government for improving the rail transporting system is to stimulate the progression of less developed regions, I find it insightful to investigate the impact of rail infrastructure on regional development. The fundamental questions are if and why regional development is correlated with rail infrastructure. The queries are especially intriguing considering the challenges lying ahead for China caused by regional disparities, and the strategic restructuring from export led growth to instead focus on social stability and sustainable growth that is taking place. The purpose of this study is to empirically investigate the relationship between rail infrastructure and regional development on the provincial level in China by focusing on the well-being of the regional residents. Therefore, this paper, using a multivariate approach, intends to answer the following research question:

How does rail infrastructure affect the level of regional development in China?

This study will contribute to the field of developing economics by answering this question as there currently is no consensus on the regional economic impact of rail infrastructure. China is an interesting case to study because no other country has had the experience of such a rapid development of the rail system before, nor have any other country invested so much in constructing high speed rail as China. Even if there are many other countries in the world that have invested in improved rail systems, the question of how decision makers should understand the rail systems impact on regional development remains a challenge (Chen 2019). Therefore, the research findings of this study can further contribute to the stream of empirical literature on the subject. In addition, there is no previous research (to my knowledge) using the exact same time frame, variables and methods as the present study in investing the relationship between rail infrastructure and regional development in China.

1.2 Data and Methods

In order to answer the research question, data is retrieved from the National Bureau of Statistics of China (NBSC), specifically using China Statistical Yearbooks. The study implements panel data on 31 provinces under a ten-year period stretching between 2008-2017. Four theoretically grounded hypotheses representing subcategories of regional development will be formulated. The used measures of regional development are education, healthcare, employment and environment, and the measure for rail infrastructure is travel intensity. The relationship of interest will then be tested using a linear regression model with a fixed effect estimator.

1.3 Main findings

The main findings of this paper indicate that regional development in the form of employment rates are positively correlated with a higher level of rail infrastructure. However, the case deviates from the common findings in existing research and developing theory, as this study observe that rail infrastructure does not have a statistically significant effect on regional development when it comes to the level of healthcare or environment. In addition, based on developing theory, this paper expected to find a positive relationship between education levels and rail infrastructure. On the contrary this paper observed a negative correlation between the two. Hence, the overall results indicate that the relationship between rail infrastructure and regional development is contradictory. The study concludes that the selection of measurements representing regional development determines the observed correlation with rail infrastructure.

1.4 Disposition

The remainder of this paper is organized as follows. Chapter 2 provides the general theoretical framework for this analysis. Chapter 3 describes China specific characteristics. The methodology is presented in chapter 4. Followed by a presentation of the regression results and robustness checks in chapter 5. Finally, the paper will be end with a concluding discussion in chapter 6.

2. THEORETICAL FRAMEWORK OF TRANSPORTATION INFRASTRUCTURE AND REGIONAL DEVELOPMENT

In this chapter the reader will be provided with basic economic theory and general background on transportation infrastructure and regional development. In revising transportation infrastructure, the main focus will be on rail infrastructure.

2.1 Transportation infrastructure with a focus on the railway system

In economics infrastructure is defined as "the basic systems and services that are needed in order to support an economy" (Cambridge Dictionary 2019). One of the most crucial systems for a country to work effectively is traditionally considered to be the transportation system in developing economy (Popova 2017). Transport infrastructure is an important tool in achieving economic policy goals, as it is a basic condition for an active and vibrant national economy. Governments tend to invest in passenger transportation in order to increase the level of passenger travels. This is done in the believe that it increases development of regions and stimulates economic activity in regions lagging, ergo achieving higher economic growth (Popova 2017).

The overall goal has historically always been economic growth and still is. However, today most countries have started to increasingly put focus on environmental sustainability in transport infrastructure development with the motive of ensuring economic efficiency (UNESCAP 2019). Rail infrastructure is one of the most important areas to develop since it has a smaller ecological footprint than most other transport options such as automobiles, ships and airplanes. There is also evidence of a decrease in domestic flight travels as rail infrastructure improves, especially when implementing high speed rail (HSR) (Chen 2017; Jeng and Su 2013). Furthermore, technology plays a huge role in improving the rail infrastructure. HSR is considered to be one of the most important technological breakthroughs along with intelligent transportation systems (ITS) because they have revolutionized various transportation systems globally (Zheng and Kahn 2013; UNESCAP 2019).

2.2 Defining regional development

Regional development is a broad term that can refer to a number of factors contained in the notion of economic growth in the region and/or in the welfare of its residents. OECD (2019) defines regional development as the supporting of economic activities in regions in a general effort to reduce regional disparities. Historically it has been common to use GDP in the region as a measure of progress. According to OECD this may no longer be a sufficient measure as it cannot tell us anything about the value of increased production in terms of increased living standard in the society. Therefore, they have created the “better life index” to measure material living conditions and quality of life. The index includes the following categories: Housing, employment, education, civic/community engagement, subjective life satisfaction, work-life balance, income, environment, health and safety (OECD Better Life Index 2019).

Within the field of developmental economy, Michalos (2007) likewise claims that it is essential to consider quality of life variables for a country to achieve sustainable economic development and growth. The domains pointed out as main variables are all included in OECD's better life index (employment, income, education, health, housing and civic status) (Michalos 2007). Michalos also points out the distinction between the actual conditions and how the situation is perceived for an individual or region. The distinction between objective and subjective measures is useful since the attitudes towards actual conditions affects the economic behavior, which impacts the living conditions and therefore the development itself (Michalos 2007; Popova 2017). However, the commonly used focus in developing economy is on the objective variables because they can be precisely measured statistically and are more accessible for the researcher.

There are two approaches in the objective school of measuring level of regional development. The first one is the neoclassical approach using only regional per capita income (or GRP per capita), and the second one being the multivariate approach where multiple variables are being used as measures (as is the case in the better life index). According to Lipshitz (1993) four types of measures are especially used in this approach. The four types are public service, consumption, physical or social environment and income-producing variables. Examples of frequently used measures representing the types are healthcare, ownership of sustainable goods, air pollutants and education respectively. Popova (2017) stresses the fact that many researchers consider healthcare to be the main factor, while others particularly emphasizes the sustainability and environmental factors. No matter what factors of regional development being of main interest in individual cases it is evident that they are profoundly interconnected (Popova 2017).

2.3 The relationship between transportation infrastructure and regional development

Sustainable and reliable transportation systems are vital not only for the development of regions, but for the economic growth for entire nations. Transportation infrastructure such as rail transportation is generally considered to significantly affect the level of development because of its impact on the efficiency of the national economy in terms of distribution of goods, production and mobility of people (Popova 2017; UNECE 2019). It also affects regional development through its interconnection with the level of investments, production, human capital, technology and living standards across regions. These factors constitute the main governmental incentives to develop the rail infrastructure. Sustainability can also be added to the list of motives as the environment has become increasingly important in governments decisions on what transportation infrastructures to invest in. Therefore, public administrations tend to pay attention not only to economic and social development but also consider environmental impacts when distributing resources between different transportation projects (Pardo-Bosch and Aguadi 2016). The International union of railways (Union internationale des chemins de fer: UIC) states that one of the main reasons why countries should invest in rail infrastructure over other transportations is because it is environmentally advantageous (UIC 2019).

On a more general note UIC claims that rail transportation plays an important role in mobility policies and serves the regional population. Evidence suggests that a higher level of economic development is achieved in regions with better transport services. If railway lines are made available or improved in less developed regions, it is believed to attract residents, companies and investors increasing the level of development (Popova 2017). Moreover, many researchers have investigated the relationship between rail infrastructure and development in numerous countries by focusing on different aspects using different methods (See for example: Popova 2017; Chen and Haynes 2017; Zheng and Kahn 2013; Zhen 2019; Pardo-Bosch and Aguadi 2016). Several of these researchers have concluded that rail infrastructure is one of the foundations in regional development. Research reports of this kind are of utmost importance for the policy makers due to the great economic, political, social and environmental significance of regional development.

The most straightforward constraint in developing railway transportation is the government's budget constraint. All governmental decision makers need to choose between different public goods, which is usually done using various theories about utility optimization for the nation (Fregert and Jonung 2014). In the processes it is important to put the costs in relation to the estimated benefits. UIC claims that rapidly developing low- and mid-income countries tend to underinvest in rail transportation. Further claiming that it may result in congestions and other transport related challenges creating bottlenecks in the economy (UIC 2019). On the contrary there are those who argue that the large-scale investments in infrastructures such as rail transportation (specifically looking at HSR) is not as effective in increasing economic productivity as often is anticipated (Chen 2019). OECD (2019) further claims that past heavy focus on developing infrastructure (including rail) as the main regional development policy overall has not helped regions lagging to catch up nor significantly reduced regional disparities as previously expected.

In summary, the key governmental incentives in developing railway infrastructure is to reach developmental objectives. On a regional level denoting ensured social well-being, better interconnection between regions and their residents, increased mobility of goods and stimulation of economic activities and growth. Even if there is some debate concerning the effect of rail infrastructure on regional development, the general view is that there do exist a positive correlation between them.

3. REGIONAL DEVELOPMENT AND RAIL INFRASTRUCTURE IN CHINA

After discussing the background and basic theory on transportation infrastructure and regional development, Chapter 3 moves on to specifically examining how rail infrastructure affects the level of regional development in China. The chapter starts of by presenting relevant background information on China's distinctive institutional situation, geographic structure and political strategies for economic growth. The chapter then moves on by describing economic, political and social factors in considering the relationship between rail infrastructure and regional development in China.

3.1 Institutional background

In 1978 China initiated an economic reform that later proved to be the start of one of the most remarkable economic transformations the world has ever seen (The World Bank 2012). The initial goal when China started to open up was to restore vitality in the state owned industrial and commercial economy through a shift to market allocation (Nee and Opper 2012, p. 38). However, it also started a "bottom up" transformation including changes in laws and policies, as well as changes in the informal institutions, norms and behaviors of the Chinese people (Nee and Opper 2014). The government made beneficial changes in micromanagement and macroeconomic policies which slightly opened up a door for the poor rural population to become profit making entrepreneurs through the Household responsibility system. People started to set up private enterprises (at first semi-legally) despite the lack of formal property rights, severe barriers to entry, unfavorable legal restrictions and discriminations against them in terms of access to finance, bureaucracy, corruption and investor protection (Cao 2012; Nee and Opper p.6-10, 38-39). In the light of the Great Leap Forward famine taking place 1959-1961, the incentives to build a better life was strong (Dali et al. 2014). Leading to rapid economic development from below through competition, strong social trust and an increasing command for commodity goods being underprovided by the state-owned firms who focused on the industry sector (Nee and Opper 2012). The living standards for the Chinese population significantly improved as new flows of resources was instigated, especially for the people living in the poorer provinces (The World Bank 2012).

To ensure continued growth the Chinese government answered to the changes in the country. For example, the decentralized fiscal revenues system was implemented mid 1980´s (Han and Kung 2015). But this was just one part in the substantial transformation of the bureaucratic system taking place to provide state actors with incentives and constraint to be pro-reform and not to abuse power (Li 1998). Later, in 1992 the new Socialist Market Economy Chinese style was implemented, and the growth rates in China completely exploded. To achieve this, China used the strategy of an experimental gradual liberalization approach with clear pro-export incentive structures and distorted institutions. They also made sure to be in control by using a heavily state-led scheme and to keep national ownership over key sectors. In addition, they directed outward foreign direct investments and invested a lot in human capital and in other value-added sectors such as technology to stimulate growth (Wade 2003; The World Bank 1993; Rodrick 2006, Wang et al, 2010). These growth strategies were extremely successful in combination with the distinct internal dynamics of China (Lo 2006; Rodrick 2006). In addition, the strategic use of special economic zones (SEZ) and joint ventures has turned out to be profitable for China. SEZ was set up to attract investments and to speed up the technological upgrading by offering preferential treatment to come in and export (Zeng 2011; LO 2006). The foreign companies operating in the SEZ had to firstly operate as joint ventures, a business arrangement where domestic firms were paired up with the foreign firms. This was a way for China to make sure to internalize the positive effects of the foreign company presence in terms of a raise in domestic employment as well as getting spillover effects in terms of knowledge and technology (Nee and Opper 2012).

Furthermore, China's growth rate has been almost 10 percent on average since 1989 until today (Trading Economics 2019). In recent years the growth rate in China has started to slow down implying that the export-oriented growth model is running out of steam (Rodrick 2006). There is also evidence of China starting to refocus towards the domestic market to strengthen the domestic development and shift away from being as reliant on investments and exports (Lin 2013; Cline 2010; Dollar 2013). The World Bank (2019) expects China's growth to slow down to 6.2 percent this year as external and domestic rebalancing continues (The World Bank 2019). In this case indicating China is turning away from the institutional distortions creating the successful growth in the past developmental model. China is adapting to the new situation and is working towards finding a sustainable path forward with the new challenges lying ahead. This shift seems already to have started. For example, domestic household consumption more than doubled during the period 2008-2017 (NBSC 2019). In addition, wages (including minimum wages) are increasing, pointing in the direction of the domestic market expanding (Liu and Shi 2018).

Moreover, one of the most crucial aspects for sustained development in China is the development of the regions that are lagging in order to achieve social stability of the entire Chinese economy (Luo et al. 2014). Realizing this the government of China declared that the core strategy to develop these regions are transportation infrastructure investment allocation policies (Luo et al. 2014). In the end, one of the main challenges in continued success for China lies in the efforts by the government to change the composition of economic activity, spreading it out in the country to the regions needing to catch up. This is of importance due to the division between provinces, with increasing disparities, in terms of economic and population geographics as we shall see in the next section.

3.2 Geographic structure

The current geographical division of regions in China was determined in 1986 (Jiang et al. 2017). Today China constitutes of 31 administrative regions (when excluding Hong Kong, Macau and Taiwan), which can be divided into three economic regions. The eastern regions consist of 11 provinces, Beijing, Tianjin, Hebei, Liaoning, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong and Hainan. The intermediate (central) region includes Shanxi, Jilin, Heilongjiang, Anhui, Jiangxi, Henan, Hubei and Hunan. The western region is covered by Inner Mongolia, Guanxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia and Xinjiang (Jiang et al. 2017; NBSC 2019). A map over the provincial division of China and their respective capitals can be found in Appendix 1.

In terms of economic geographics, there exists great imbalances of economic development between the three economic regions as well as between the urban and rural population within them (Nee and Opper 2012; Jiang et al. 2017). The clear divide between the coastal and inland areas also represents a partition between the traditionally rich (eastern/central) and the poor (western) regions. The three regions each have their distinct characteristics due to diverse resource and natural conditions (Jiang et al. 2017). The eastern coastal region was the first to develop because of the good geographical position, favorable agricultural conditions and abundant resources. It was also here the private enterprise economy grew in scale and the market economy first emerged (Nee and Opper 2012, p. 39). Over time the eastern region established higher quality of the labor force, better technology, more access to finance, and a more solid industrial sector than the other parts, making the coastal area taking on the leading role in the economic development of China (Jiang et al. 2017; Nee and Opper 2012). The intermediate region is characterized by agriculture and heavy industry. Most of China's coal mines is found here (80 percent) along with plentiful of other resources of energy, metals and minerals (Jiang et al. 2017). Most of the intermediate region have climbed out of poverty and followed the eastern regions progression. The western region, on the other hand, covering 57.4 percent of China's landmass has been less fortunate in terms of geographical prerequisites, and has therefore lagged immensely in the economic development compared to the other regions (Jiang et al. 2017). However, the area displays great potential for economic development because of the massive land size and the plentiful mineral resources situated in the area (Jiang et al. 2017).

Moving on to population geographics China is experiencing issues with overpopulation in some areas, and underpopulation in others. Megacities such as for example Beijing and Shanghai, is especially experiencing issues with overpopulation as the dense population mass leads to increased social expenses and living costs, heavy traffic congestions and severe air pollutants having a negative effect on the residents' quality of life (Zheng and Kahn 2013). At the same time some larger cities are believed to be undersized because of the institutional boundaries constituted by the hukou system that still divides the people and creates inequalities (Zheng and Kahn 2013). The Hukou system is a household registration system set up by the government in order to monitor and actively limiting where the people of China is allowed to live based on where they are born (Juneja 2017). It can be especially difficult to move to a more attractive city if the individual is born into a rural hukou, causing distorted incentives and inefficient resource allocation as it prevents the rural workers to migrate to urban sectors (Cai 2008). Dollar (2013) argues that it will be difficult for China to rebalance towards household consumption going forward without reforming the Hukou System. However, the reform needs to be gradual and intelligently constructed to limit the risk of slums appearing in the urban areas as the pressure might increase on already pressured regions that won't be able to absorb large residential inflows (Dollar 2013; Zheng and Kahn 2013). Another important institutional feature that seriously has affected the population geographics in China is the one child policy. China has had a one or two child policy since the beginning of the 1970's until today (CBC 2015). This has resulted in an aging population today which is putting large constraints on the economy. This has lead the Chinese government to urge families to get two children again since 2015, after a long period with one child policy (CBC 2015). However, the young have seen the many benefits of growing up as an only child and therefore usually do not want to get more than one child themselves. The issues regarding China's aging population makes it even more important to develop the regions that are lagging.

3.3 The Chinese government objectives for regional development

The geographic structures of China revealed great differences between the prerequisites and conditions of the regions, affecting their regional welfare and development. Even if China, by their export led growth strategy, has lifted 500 million people out of poverty, become the world's largest exporter and manufacturer and the second largest economy, the social and economic inequalities remains high (The World Bank 2013). The disparities are even increasing in some dimensions such as between regions as well as between the urban and rural residents (The World Bank 2013). Realizing that this is a major issue in achieving sustainable growth, China has gradually switched out the export led growth strategy to give room for a new development strategy better suitable for the new situation. Despite the fact that China's growth moderates, it is likely that China becomes the largest economy and a high-income country within the following decade (The World Bank 2013). In reaching the objectives of being a stable, wellbeing, and high-income society in the next phase of development, China focuses on the social stability in the Chinese economy (Luo et al. 2014; World Bank 2013). This restructuring of the development strategy gradually begun as early as in the late 1990´s, when China started combining export led growth with plans for a more balanced regional development and social stability nationwide. One example being the Western Development Plan launched in 1999 with the incentive to encourage catching-up of the western regions (Luo et al. 2014).

The World Bank and China constructed a joint research team to produce the research report called "China 2030: Building a modern, harmonious and creative society" that was released by The World Bank in 2013. More specifically, the report was constructed by China's Development research center of the state, China´s Ministry of finance and the World Bank. Together they designed a long-term development strategy that extends to 2030 by combining the key elements in China´s 12th Five Year plan. In the report six main objectives for development (both on national and regional level) and how to achieve them is presented.

The first objective presented is to strengthen the foundations for a market-based economy by implementing structural reforms. The structural reforms involving increased competition in all sectors, more intangible public goods and increased efficiency of land use in order to protect farmers and decrease local government dependence on land related revenues. Another crucial structural reform concerns the labor market. In order to increase efficiency, the World bank report (2013) emphasizes measures to increase labor force participation, reforming wage policies and the hukou system so workers can respond to market signals, as well as improving social security instruments including pensions, unemployment and health insurance in all regions. Secondly, they point out increasing cognitive and technical skills of university graduates and investments to build world class universities in order to accelerate the pace of innovation and creation of an open innovation system. So, China keeps putting weight on human capital and education to push the country forward. A third focus is the environment with the motive to address the many environmental challenges, sustain rapid growth and improve the level of wellbeing. The green development and increased efficiency in the use of resources is expected to be achieved through a combination of regulations, public investments, industrial policies, institutional development and market incentives. China is believed to have good conditions in such an ambition due to the large market size (economies of scale), the dynamic private sector and the high investment rates (The World Bank 2013). As the fourth objective of development the World Bank report (2013) points at the promotion of social security and expanded opportunities for all. This means China is aiming towards providing all regions with equal access to public services, social security, finance, and access to jobs. This is considered critical in reversing inequalities steaming from the large rural-urban differences that are closely correlated with the regional disparities, as it will help improve for example healthcare, employment and labor mobility. The fifth objective in the adjusted development strategy is to mobilize more revenues towards the local governments to ensure that they are provided with adequate financing facing rising expenditure responsibilities (The World Bank 2013). The sixth, and final, factor in China's continued path of development pointed out in the World Bank report (2013), is to continue to seek beneficial relations globally. In the past three decades China's integration into the world economy has served it well. In further keeping these international relations, China will continue to benefit from flows of knowledge and ideas going both ways, increased investment opportunities, further specialization, and higher returns to capital (The World Bank 2013).

The six governmental objectives for development presented above is in large aimed to decrease inequalities in the population and to reduce regional disparities. This is the case since the Chinese government's main development goals, sustainable growth and social stability, in large relies on a balanced regional development (Luo et al. 2014). In achieving this, employment, healthcare, education, environment, investments and technology is frequently cited.

3.4 The role of the railway system

Over the past 15 years, the railway system in China have developed rapidly due to generous funding from the Chinese government (EESI 2018). China dominates track under construction charts over the world currently, with 37% of heavily rail, 66% of metro and 61% of high speed rail being built there (UIC 2019). The Chinese government has especially done immense investments into developing the rail infrastructure since the implementation of the high-speed rail (HSR) policy in 2008 (Zheng and Kahn. 2013; Huang et al. 2018). The HSR is believed to improve both the living standard of the population living in the megacities and stimulate the economic development of second and third tier cities situated in poorer provinces as it increases intercity travels (Zheng and Kahn 2013). Approximately half of the worldwide investments in HSR is occurring in China (over $300 billion to date) and they are expected to have far more rail lines than the rest of the world combined by 2025, when the HSR system is expected to be completed (The World Bank 2013; EESI 2018; Chen 2019). In terms of deployment speed and service scale, the HSR development in China is enormous and incomparable to similar systems in other countries (Chen 2019). So, the question is why China invests so much money into evolving the railway system?

According to Popova (2017), China's economy largely is defined by the level of infrastructure development. Demerger (2001) adds to this statement with the claim that public transportation such as rail is the most crucial factor in facilitating cross-regional economic integration. The Chinese government themselves argue that the development of the rail system by the construction of HSR is done because of five main reasons. Firstly, it provides a comfortable, reliable and fast way of transporting large numbers of travelers long distance in a densely populated country. In the long run it is also believed to improve economic competitiveness and productivity by linking labor markets together (EESI 2018). Secondly, it supports environmental sustainability as trains uses less energy than other forms of transportation. For example, the growth of HSR has decreased domestic flights significantly, especially flights for shorter distances than 500 km (Chen 2017). Thirdly, it has cultivated an HSR components and technology industry that has emerged from internalizing foreign technologies from joint ventures enabling China to eventually compete on the export market (EESI 2018). The next argument is closely related to the previous one, saying that the development of the rail system stimulated the economy short term as the construction created job opportunities (EESI 2018). The last motivation is that the improvements to the rail infrastructure stimulates regional development, as it facilitates cross-regional economic integration and also enables connections between smaller and larger cities which promotes growth of the smaller ones (EESI 2018). Chen (2019) measured the regional impact of the HSR and found that the rail infrastructure development in China have had a positive regional impact in terms of real GDP growth rates. Chen (2019) further suggests that the impact of the rail system, focusing on HSR, is far reaching. The regional development is positively affected because more and more people do intercity travels, therefore the positive effect is also expected to increase over time (Chen 2019).

On the other hand, there are critics that questions the effectiveness and economic merit of the rail system in China, paying particular attention to the massive investments into HSR construction. For instance, Chen and Haynes (2017) raises the danger of overestimating the effect when deciding to invest in HSR. Button (2017) likewise argue that although policy makers generally consider rail infrastructure investments to be a catalyst for economic development, the estimations that the decisions are grounded on are often over optimistic due to underestimated costs and overestimated anticipated economic growth effects. The Hukou system can also be a bottleneck for the efficiency of the rail system. The improved possibilities of travelling by train may not increase interconnection as people are constrained in their mobility (Cai 2008; Zheng and Kahn 2013). Another skepticism is presented by Wu et al. (2014) who implies that a massive development of the rail infrastructure with high speed links can be problematic because of the high costs. Chen (2019) also lifts the issue that it is more difficult to improve the rail system in poorer provinces. A key issue is probably due to the nature of the payment arrangements, where the local governments typically pays between 40-50 percent to get access to the new bullet train lines and the central government pays the rest (Zheng and Kahn 2013). Ansar et al. (2016) on a similar note raised concerns regarding increased risks for the stability of the financial market, debts and monetary expansion imposed by the large-scale rail infrastructure investment. Hence, according to the critics the overall economic benefit is not necessarily positive, and it is often a challenge to determine the long-term nature of rail infrastructure in order to quantify the regional development effects (Chen 2019).

Despite the skepticism concerning the effect of rail infrastructure, cross-province economic integration and mobility has become increasingly important for regional development in China during the last two decades (The World Bank 2013). According to the Chinese government, rail infrastructure does have a positive effect on regional development, which is also declared to be the case by most theory and research on the subject. This is one of the main motivations for China to invest heavily in expanding and improving the railway system, which they have been doing especially since the HSR policy was put into action in 2008 (EESI 2018).

3.5 Hypotheses formulation

This study investigates how rail infrastructure affects the level of regional development in China, with a specific focus on education, healthcare, employment and environment. In the light of the empirical literature provided above, this study deduces that with all else being equal, higher rail infrastructure levels will result in improved regional development on average in China. To further investigate the relationship of interest the elements of regional development will be tested separately by the following four hypotheses:

H_1: Provinces with a higher level of rail infrastructure have a higher population share enrolled in higher education.

H_2: Provinces with a higher level of rail infrastructure have a higher level of healthcare.

H_3: Provinces with a higher level of rail infrastructure have a higher rate of employment.

H_4: Provinces getting a higher level of rail infrastructure experience a change in the level of environmental pollutions.

4. METHODOLOGY

The general methodological approach in this study is quantitative and deductive. Quantitative in the sense that objective measurements and computational techniques are being used on statistical data (Lundahl and Skärvad, 1999). No qualitative aspects will be included. The analysis approach is deductive as the process started out with a social theory, in this case general theory on the relationship between rail infrastructure and regional development, and then investigated if there exist implications in the data supporting the theory (Lundahl and Skärvad, 1999). The methodology is chosen to provide answers to one basic question in econometrics. Namely, to what extent knowing something about rail infrastructure allows us to predict something about regional development (Angrist and Pischke, 2009). The focus of this study will not be on the absolute values of the investigated relationship, but rather on the direction of the relationship investigated.

In the following sections in this chapter the course of action in conducting the analysis will be presented. Including information about the data, variables and model being used to answer the research question. The performed econometric testing will also be displayed, followed by a discussion of the limitations of this study.

4.1 Data

The empirical analysis in this study relies on panel data of 31 provinces stretching over the time period 2008 - 2017. The data used to compute all variables contained in the model is collected from the National Bureau of Statistics of China (NBSC), specifically using ten China Statistical Regional Yearbooks. The NBSC is the most detailed and comprehensive database of statistical data on China and a China statistical regional yearbook is the collection of statistics on all provinces for the previous year. These include a number of major indicators, quantitatively reflecting the economic and social development in China (NBSC, 2019).

In order to achieve a complete dataset for cross-section and time series data, and to avoid imbalances and inconsistencies in the estimated model, the sample contains 31 provinces and thus excludes Hong-Kong, Macao and Taiwan. The reason for this is limitations in data availability for these regions. In the case of Hong-Kong and Macao this is probably due to the fact that they are special administrative regions (SAR), and in the case of Taiwan it is a very sensitive territorial issue to this date whether or not it is a part of China (Reuters 2018). It should also be noted that, formally, 9 out of the 31 provinces consists of autonomous regions and municipalities (The people's republic of China 2014). A list of the included administrative regions and their formal status is provided in Appendix 2. For simplicity, this study refers to the entire sample as provinces, which is common practice in the literature.

In similar manners, the data excludes the years of observations before 2008 due to limited data access. Another reason to look at the chosen ten-year period is because 2008 was the year when the central government's policies for heavy investments into high speed rail transportation was put into action as the first line opened (Bertie 2019). Since this has resulted in rapid evolvement of the rail infrastructure in China until today, 2008 is considered to be a sufficient and suitable starting point in relation to what this study aims to investigate.

By limiting the number of provinces and years of observations complete information is secured. The data set is strongly balanced, meaning that all the cross-section observations has the same time period. This increases three of the benefits already existing with using panel data: Firstly, greater efficiency of estimations is achieved due to more reliable variability. Secondly, better correction for heterogeneity between observations can be obtained, and thirdly the opportunities to analyze dynamic effects improves (Verbeek, 2012, p.374-376; Angrist and Pischke, 2009, p.222-227).

4.2 Variables

In this section all the contained variables in the analysis will be explained. As an initial useful notification, I want to inform the reader that given the substantial differences between provinces in population size there is a risk that captured effects in the variables is due to unwanted population size effects. To avoid this issue all variables have been converted into per capita measures, which also can be beneficial in interpreting and comparing estimation results.

4.2.1 Dependent variables

The aim of this paper is to investigate the effect of rail infrastructure on regional development by focusing on the well-being of the regional residents in China. This will be done by using a multivariate approach, which means regional development will be measured by multiple variables (Lipshitz 1993). Researchers supporting this method are of the opinion that a country's regional development and welfare is influenced by numerous indicators of development needed to be considered (Milenkovic et al. 2014; Lipshitz 1993). Further stating that regional development is a multidimensional issue and therefore should be treated as such. In using a variety of objective variables there are two commonly used methods; investigation by grouping the variables together as one measure (multivariate statistical analysis) or separate investigation for each variable (Lipshitz 1993). In this study the later of the two will be used. The reason is that this study aims to uncover the regional trends for each variable included in the regional development measure. Furthermore, given what a broad term regional development there are many different variables that may be included in the estimation of it, depending on the aim of the research. In this analysis the dependent variable, regional development, will be categorized into the following four different variables:

1. Education per capita
2. Healthcare per capita
3. Employment per capita
4. Environment per capita.

These variables have been chosen to give an indicator of how rail infrastructure effects regional development on a broad scale. Employment is considered to be one of the most crucial factors in developing regions and is therefore included (OECD 2019). Healthcare, education and environment are also variables commonly used in multiple approach for regional development (Lipshitz 1993). The four variables will be tested separately and are measured by using data retrieved from the China Statistical Yearbooks as described previously. Below follows a deeper explanation and description of each regional development variable individually.

1. Education per capita

Education per capita can be viewed as a "income-producing" factor, but also as a measure of human capital in a region. This analysis assumes that a larger proportion of highly educated people in a province reflects a greater engagement in wealth generating activities and a higher rate of human capital on average. A higher rate of wealth generating economic activities is proven to be an important part of regional development, as is human capital (OECD 2019). Higher education is important both in terms of well-being of the individual and the province. Therefore, a variable for number of people in higher education per 100 000 residents for each province was retrieved from NBSC. This was modified by multiplying the data by one hundred thousand and then dividing it by total population in order to get number of students in higher education in relation to the total number of provincial residents. This measure will be referred to as education per capita from now on and represents the proportion of the provincial population in higher education.

2. Healthcare per capita

Services that are provided mainly by the public sector are also of great significance for regional development. One of the main reasons being that public services to a great extent affects the populations welfare in terms of consumption patterns (The World Bank 2013). A commonly used public service variable used in measuring regional development by multivariate approach is healthcare (Lipshitz 1993). This analysis uses share of available medical technical personnel in healthcare institutions in relation to inhabitants as a measure for the level of healthcare available per person in each province. This study further suggests that in the case of China the level of healthcare is an essential part of regional development since better and more accessible healthcare results in less precautionary savings, more consuming and in general a healthier population on average. These are all important factors to stimulate economic growth (The World Bank 2013).

3. Employment per capita

The employment per capita measure is computed by adding the retrieved data on number of employed persons in urban units, private enterprises and self-employed together and then divide it by the total population in the region. This measure shows the level of employment in the provinces. Employed persons include workers employed in private and township enterprises, rural employment, self-employed, re-employed retired personnel and other employees (as reported by NBSC). The reader should also be informed that the computed rate of employment will get fairly small since it is calculated as the share out of the entire population (including elderlies and children). It would probably be better to use only population of working age (15 to 85), unfortunately this data was not available in the database. However, it is very unlikely that provinces differ in ratio of children in relation to the population since China has had a one or two child policy since the beginning of the 1970's until today (CBC 2015). The variable may capture the factor of some provinces having a larger share of elderlies than others, but this is not considered an issue for this study. Overall, employment is included as one of the measures for regional development as employment reduces regional disparities and stimulates economic activities (The World Bank 2013).

4. **Environment per capita**

The environment is one of the most central issues in modern time, not only for China but for the entire world. Using physical environment such as air pollutants as a regional development variable is becoming more and more common in the literature. I have chosen to include it in this analysis since China is facing severe difficulties with air pollutions which greatly affects the welfare of the population in terms of living standard, health and economic situation, and therefore is an important issue in regional development (Mohajan 2014). The variable to represent the environmental situation in the provinces is measured as Sulphur dioxide (SO_2) emissions in waste gas in kg per capita, i.e. how much Sulphur dioxide is emitted (in kg) per person in the provinces. The main reason for this study to use SO_2 instead of carbon dioxide (CO^2) as a measure for pollutions is because SO_2 is a major contributor to the particular matter PM2.5, which is the officially used measure for air pollutions in China (Mohajan 2014). PM2.5 is atmospheric particulate matter that has a diameter smaller than 2.5 micrometers. It is very dangerous to humans and animals when inhaling them, and since they are small and light they tend to stay in the air for a long time (Bliss Air 2019). In addition, nearly all SO_2 in the air comes from human activities with the largest contributor being the process of refining raw materials (Mohajan 2014).

4.2.2 Main independent variable

The main independent variable, rail infrastructure, is computed by collecting data from the China statistical yearbooks. The data used is "Passenger traffic of railways in 10 000 persons". To get this in per capita measures the data is simply multiplied by 10 000 and then divided by the total population. So, in this study the variable for rail infrastructure will be computed as the size of passenger traffic of railways per capita. Passenger traffic of railways refers to the number of passengers transported by railway carriage, calculated as one-way trips no matter the journey length or the ticket fare. Simply put, the variable rail infrastructure in this study refers to how many times one-person travels by train each year, i.e. travelling intensity. This is used as a proxy for how efficient the rail system is in terms of passenger transportation. Efficiency, In the context of this analysis, refers to the level of mobility due to rail infrastructure, e.g. the rail system is assumed to be more efficient if more people choose to travel more frequently by train in China.

4.2.3 Control variables

Grasping the vast size of China in terms of economy, population and geographical magnitude it is important to acknowledge cross-provincial heterogeneity. For example, regions differ considerably in terms of economic and social conditions and demography (See section 3.2). Therefore, control variables are included in the model in order to isolate the effect of rail infrastructure. The control variables account for some of the regional variation that are likely to affect the chosen regional development variables besides the rail infrastructure. Together with the theoretical background offered in chapter 2 and 3 of this paper six control variables from the data available at NBSC is constructed as follows:

1. GRP per capita
2. Local government expenditures on transportation per capita
3. Investment per capita
4. Technology per capita
5. Population
6. Highway infrastructure per capita

These are chosen because they are likely to have effects on the examined regional development variables, and because it is also probable that they will be correlated with the rail infrastructure. A short explanation for each control variable and how they are measured follows.

1. **GRP per capita**

The gross regional product (GRP) is a monetary measure of the market value of all final services and goods produced in a region. GRP per capita is measured as the gross regional product divided by population. This is a commonly used indicator of how well the economy is doing overall on a regional level as mentioned in section 2.1. The variable has a clear correlation with regional development and based on theory this relationship is also true for rail infrastructure. GRP per capita is used as a control variable instead of one of the dependent variables due to three known weakness. The first one being that the measure cannot say anything about whether increases in the value of production leads to higher living standards for the population (Tillväxtverket 2019). Similarly, it cannot tell us if the production increase jeopardizes future generations wellbeing, and income distribution is also not accounted for (Focus Economics 2019). Nevertheless GRP per capita is a good general measure of the regional economic condition and is therefore used as a control variable in this research.

2. **Local government expenditures on transportation per capita**

Local government expenditures on transportation are included as control variables as they are likely to be related to regional development as well as rail infrastructure. They are measured as local government expenditures on transportation in yuan per capita. The study assumes local government expenditures on transportation to be positively correlated with both regional development and the efficiency of the rail system in China.

3. **Investment per capita**

Investments is measured as total investment in fixed assets in yuan per capita. The measure includes investments by state-owned units, collective-owned units, joint ownership units, share-holding units, private units, individuals, foreign investors and others (NBSC 2019). The ratio of investments in fixed assets in relation to the provincial population is considered to be of relevance in this study because it indicates the size of investments made in the provinces. This is of interest as investments are known to be related to regional development. It is also of relevance as the central government in China tends to be more prone to invest in rail infrastructure in regions showing better economic performance, e.g. by attracting a larger share of investments (Nee and Opper 2012).

4. **Technology per capita**

Similarly, to investments, technology is prominently related to both the development of regions and transportation systems (UNESCAP 2019). Therefore, the variable technology per capita is included in this study, measured as number of patent applications granted per capita. This study expects a higher level of technology per capita to be related to a higher level of regional development on average.

5. **Population**

As previously stated, provinces differ substantially in population size. It is included in the analysis as a control variable as it is believed to have an effect on regional development. Larger regions tend to have greater issues with pollutions while smaller regions to a greater extent is experiencing difficulties when it comes to improving their quality of life through education, employment and healthcare (Zheng and Kahn. 2013). The variable is computed as the resident population at year end.

6. **Highway infrastructure per capita**

The last control variable included in the study is highway infrastructure per capita. The data retrieved from the China statistical yearbooks is "Passenger traffic of highways in 10 000 persons". Like in the case of rail infrastructure per capita this is recalculated by multiplying the data by 10 000 and then dividing it by total population in the province. This measures how many times one-person travels by highway per year. This is anticipated to be correlated with both regional development and rail infrastructure.

4.3 Descriptive statistics

In this section descriptive statistics will be presented for all the variables included in this analysis. The sample consists of 310 observations collected on 31 provinces over the ten-year period 2008-2017. The overview consisting of Table 1 below contains the variable name, number of observations, means, standard deviation as well as minimum and maximum values for each variable.

Table 1. Descriptive Statistics

Variables	Obs	Mean	Std.Dev.	Min	Max
Dependent variables					
Education per capita	310	0.0243	0.00910	0.00969	0.0675
Healthcare per capita	310	0.00541	0.00182	0.00220	0.0155
Employment per capita	310	0.287	0.139	0.0990	0.900
Environment per capita	310	16.19	12.55	0.561	64.22
Main independent variable					
Rail infrastructure per capita	310	1.667	1.031	0.0761	6.417
Control variables					
GRP per capita	310	43,936	23,582	9,855	128,994
Local Gvmt expenditures per capita	310	741.8	804.3	31.91	6,359
Investment per capita	310	31,116	15,273	5,185	81,814
Technology per capita	310	0.000713	0.000931	3.18e-05	0.00493
Population	310	4.361e+07	2.750e+07	2.920e+06	1.117e+08
Highway infrastructure per capita	310	17.18	11.56	1.355	66.13

Looking at the descriptive statistics it is clear that there is large variance within the variables, suggesting considerable inter-provincial variances in the sample. This is revealed as the difference between the minimum and maximum values is relatively large for all the variables included in the analysis. This makes China an exemplary case for panel data analysis.

To further investigate the dependent and main independent variables the average value for each province during the investigated period 2008-2017 is computed and illustrated in figures. Figure 1 below illustrates the average education per capita level between 2008-2017. As expected it shows that there is a greater level of highly educated people in the eastern coastal regions than in the western inland areas. Especially Beijing stands out with the most educated residential in China, while the lowest level of highly educated inhabitants can be found in Qinghai and Guizhou.

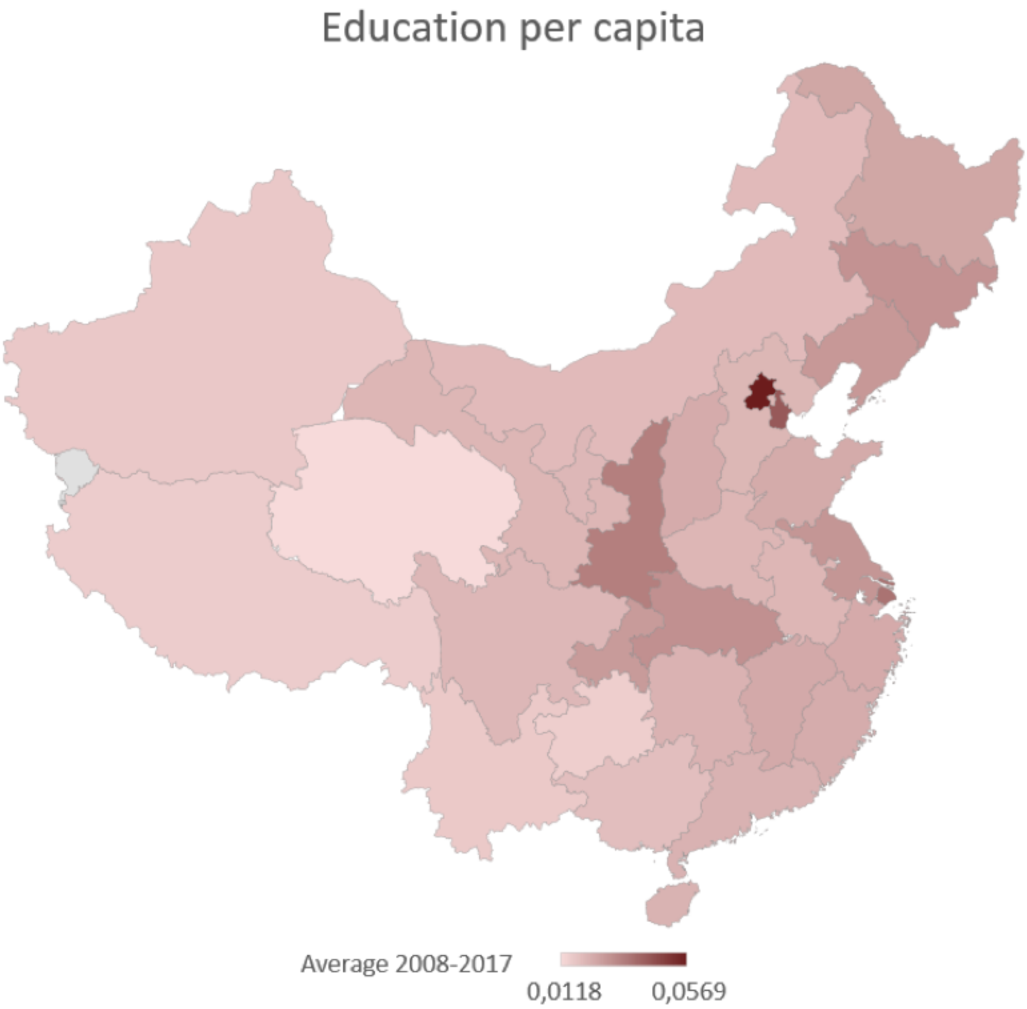

Figure 1. Education per capita on average 2008-2017

When investigating the distribution of healthcare per capita across China, displayed in figure 2, there seems to be more of a north-south diversity rather than between western and eastern regions. Furthermore, it is no surprise that the ratio of number of healthcare personnel per person overall is higher in areas with higher GRP per capita. As can be seen, richer provinces such as Beijing, Shanghai, Guangdong and Inner Mongolia have a higher level of healthcare compared to poorer areas such as Tibet, Jiangxi, Guizhou and Anhui.

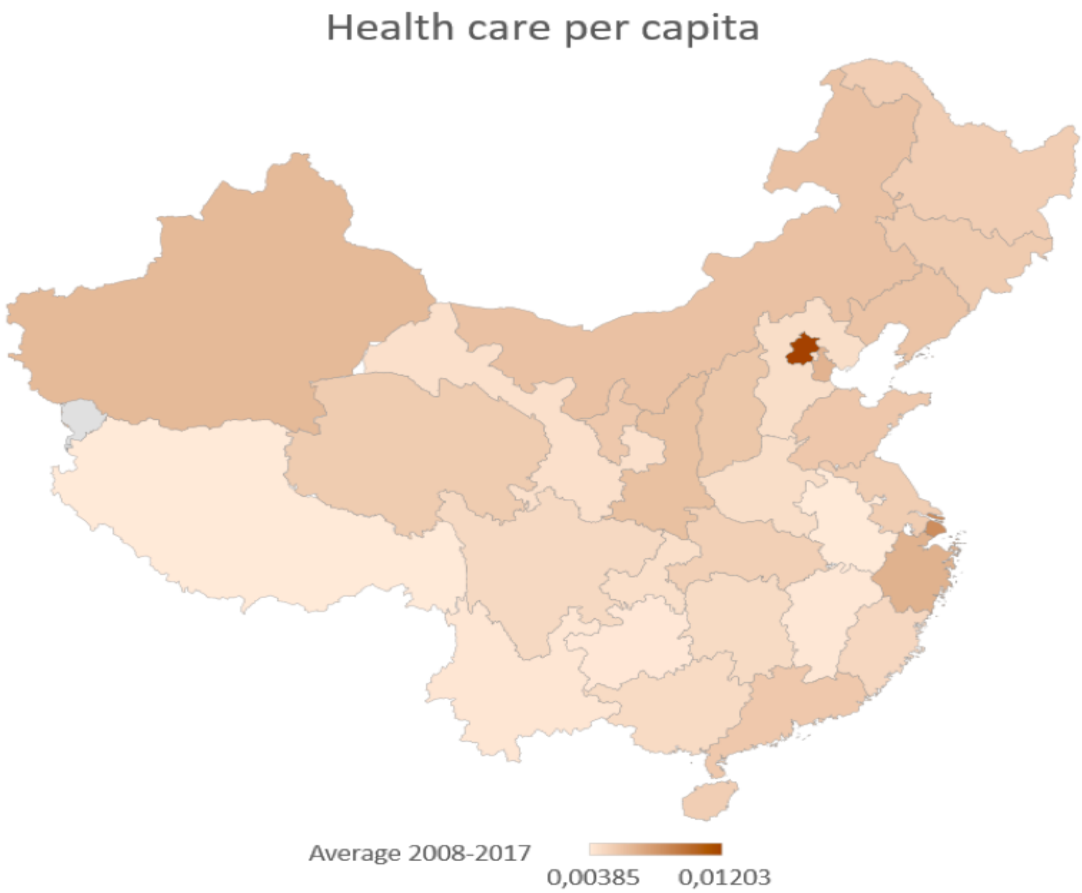

Figure 2. Healthcare per capita on average 2008-2017

Moving on to the level of employment it is clear that the highest worker rates can be found in the eastern coastal region as shown in figure 3. With the highest employment levels seen in Beijing, Shanghai, Zhejiang and Jiangsu. The reason for some regions to show very low employment rates, such as Hebei, Henan and Anhui can be due to a number of factors. One being that a large share of the inhabitants in those regions to a great extent work in surrounding provinces with better job opportunities. For example, it is likely that many people living in Hebei work in Beijing. Another factor can be a larger proportion of residents in some of the poor provinces are non-employed farmers or workers in other unregistered ways. Another plausible reason for the seen inter-provincial differences can be some of the areas having a significantly larger share of elderlies than others on average.

Figure 3. Employment per capita on average 2008-2017

By viewing figure 4 it becomes evident that Inner Mongolia and Ningxia displays the highest levels of Sulphur dioxide per person. While the areas in the south is less related to the direct emissions of SO_2 air pollutants. Some of the provinces showing higher levels of SO_2 emissions is known to hold many of the power industries and manufacturing factories. Examples of such regions are Inner Mongolia, Guizhou and Shandong (Zhang et al. 2018). However, it is often the case that the most developed regions are the drivers of the emissions. Meaning they are the initiators and the main consumers of the outcomes resulting from the emissions in other provinces. Such regions being Guangdong, Shanghai, Zhejiang, Beijing and Tianjin. Zhang et al. (2018) have shown that there exist remarkable inter-regional spillovers of SO_2 emissions within China, contributing to the severe air pollutions in the megacities such as Beijing.

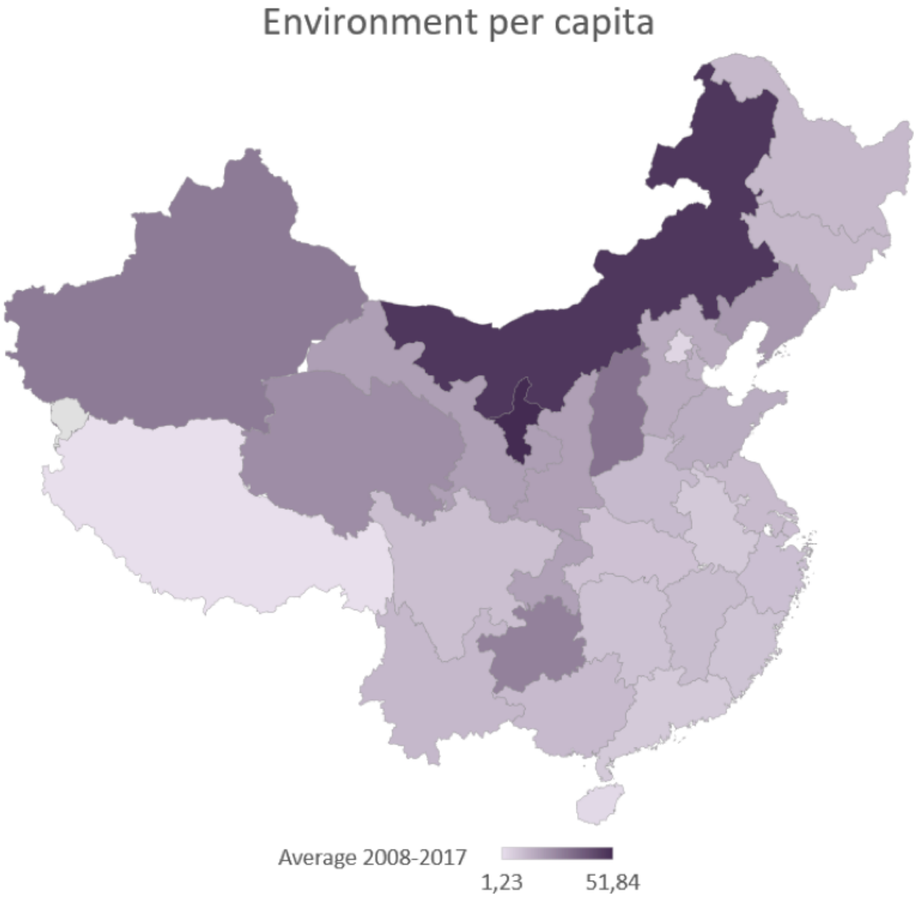

Figure 4. Environment per capita on average 2008-2017

Figure 5 shows the distribution of the main independent variable, rail infrastructure per capita. From the figure we can see the pattern of people living in the eastern and central part of China travels more by train than people living in the western parts. This is likely related to the great investments into developing the rail infrastructure, especially after the implementation of the high-speed rail policy in 2008, focusing on the named regions (Zheng and Kahn 2013). For instance, the first bullet train tracks connected megacities such as Shanghai, Guangzhou and Beijing with nearby cities. China is continuing the improvements on the rail transporting system with the motive of stimulating the progressing of less developed regions.

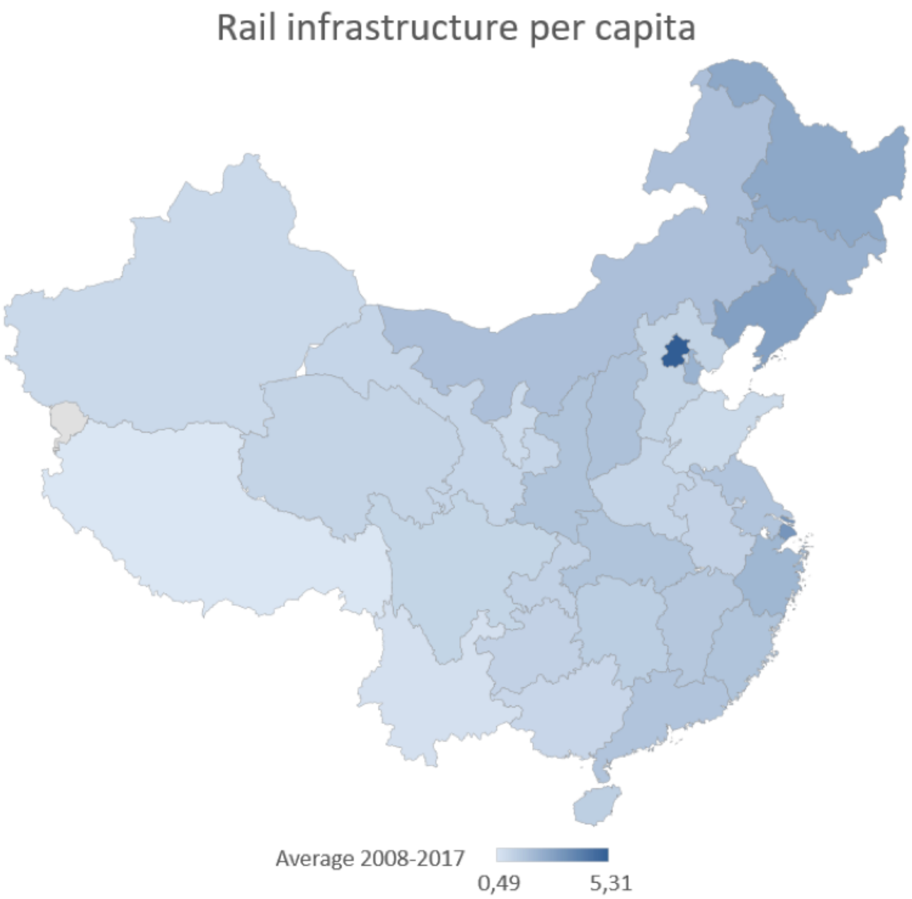

Figure 5. Rail infrastructure per capita on average 2008-2017

4.4 Model Specification

To test the hypothesis formulated in this study ordinary least squares (OLS) estimation method will be used. The purpose with this method is to get a good approximation of the dependent variable by finding a linear combination of the independent variables and a constant (and an error term). The best linear approximation of Y from X´s and a constant is obtained by minimizing the sum of squared residuals (Verbeek 2012, p.7). To test for the relationship between regional development and rail infrastructure the following baseline model is constructed:

Model 1

$$REGIONALDEV'._{it} = \alpha_i + \beta_1 RAILINFRA.pc_{it} + \beta_2 X'_{it} + \beta_3 YEAR_t + \varepsilon_{it}$$

Model 1 estimates the effect of rail infrastructure on regional development. Regional development is separately tested for education per capita, healthcare per capita, employment per capita and environment per capita. “REGIONALDEV’.$_{it}$“ is a matrix containing the four regional variables. The matrix X'_{it} contains the control variables, which are GRP per capita, local government expenditures on transportation per capita, investment per capita, technology per capita, population and highway infrastructure per capita. The constant α_i is the province-specific fixed effects. This accounts for cross-section differences that are fixed over time. In other words, the variable contains differences that are not being investigated but that still may affect the result if not controlled for. For example, there are huge differences regarding geographical conditions (climate, temperature etc.) across China, which affect regional development. Another fixed effect variable can be found in $\beta_3 YEAR_t$ which accounts for changes over time (between years) that affects all provinces equally. For instance, regulations and policies implemented by the central government that apply to the entire nation. Lastly, the error term it captures the unexplained variation in the regional development variables, i.e. it captures the effects that cannot be explained by the included explanatory variables.

4.5 Econometric testing

Tests will be conducted on the data before performing the regression analysis with the purpose of detecting potential biases. Therefore, tests for stationarity, multicollinearity, heteroscedasticity and choice of econometric model follows below.

4.5.1 Stationarity

For the properties of standard estimation and testing procedures to be reliable the assumption about stationarity in the process must be fulfilled (Verbeek, 2012, p. 338-342). A time series process is stationary if the mean, variance and covariance of the ensemble distribution is independent of time (Dougherty 2016, p.481). If not, the regression might be spurious. In this case the results may appear to establish a relationship between two variables that are not related, making the estimators and test statistics misleading (Dougherty 2016, p.490; Verbeek 2012, p.338-342). To check for this, the study performs a Lewis-Lin-Chu (LLC) unit root test for each variable and accounts for trend effects. The null hypothesis is that the series contains a unit root (meaning that variables are non-stationary) (Stata 2019). As the results presented in Appendix 3 shows, the null can be rejected for all the variables except for two. Non-stationarity was detected for environment per capita and investment per capita, meaning that their distribution changes with time (Dougherty 2016, p.484). Therefore first-difference variables are constructed for these two variables which transforms them into being stationary.

4.5.2 Multicollinearity

Being confident that the included variables in the study now are stationary processes, the data will be tested for multicollinearity. The reason is that if there exists multicollinearity among the explanatory variables the regression estimates may be unreliable (Verbeek 2012, p.43-46). Usually it is not an issue if the independent variables are correlated, it only becomes a problem if the correlation is too high since that makes it difficult to identify the individual impact of the variables on the dependent variable (Verbeek 2012, p.43-46). This study performs a variance inflation factor (VIF) test to detect potential multicollinearity. The test measures the linear relationship between the explanatory variables, where a VIF of 10 or more commonly is considered to high (Verbeek 2012, p.44-45). The VIF test results along with a correlogram is displayed in Appendix 4, showing that the correlation between the independent variables are sufficiently low.

4.5.3 Heteroscedasticity

Next, the study controls for heteroscedasticity since one of the most important conditions in regression analysis is that the variance of the error terms is constant and that all the independent variables observed have the same dispersion (Dougherty 2016, p. 290-293). This condition is called homoscedasticity. If there do exists heteroscedasticity in the model the unexplained variation in the dependent variable will change as the value of the independent variable changes. To test for this a White test for panel data is executed where the null is panel homoscedasticity. The results show that there are signs of heteroscedasticity in the models. Therefore, I will use the command for robust standard errors when running the regressions to make sure that there is no heteroscedasticity in the model and that the estimator is robust to some misspecifications. (Dougherty 2016)

4.5.4 Choice of econometric model

After testing for stationarity, multicollinearity and heteroscedasticity some further clarifications regarding the regression model are required. As previously stated the regression method used in this analysis to test the hypothesis will be an OLS regression. In the OLS regression it is possible to use different estimators depending on which one that fits the dataset better. Hence tests on whether to use a pooled OLS regression model, a random effect model or a fixed effect model will be executed. Firstly, a Breusch-Pagan Lagrangian Multiplier (BPLM) test is done to establish if random effect or pooled OLS fits best into the data (Dougherty 2016, p.541). The null is that the variance of the random effect is zero, which would mean that I should run a pooled regression since all observations would have the same intercept. Since the resulting p-values from the test are highly significant for all four regressions the pooled OLS option can be ruled out.

As a second step a Hausman test is performed to distinguish between the random effect and fixed effect method. The null is that the individual specific effects are random. It is rejected if the difference between fixed and random effect estimators are significantly different from zero, in which case the fixed effect estimator should be used as it will be more efficient (Verbeek 2012, p.385-386). The results from the Hausman test was unanimously for the four different regressions. For the regressions having healthcare per capita and employment per capita as dependent variables the p-value was highly significant (0.0000), meaning that fixed effect estimator should be used. As for education per capita it was unable to compute positive values and for environment per capita the p-value was insignificant (0.5800). However, as pointed out by Verbeek (2012) these results should not automatically be interpreted as evidence that a fixed effect estimator as a tool is inappropriate. In the case of China there exists great differences between regions as many of the provinces have massive land areas, populations and GDPs (Worldatlas 2019; NBSC 2019). Big differences across provinces over the timeline investigated is also implied due to the rapid growth China has experienced during the last couple of decades. It is therefore assumed in this analysis that heterogeneity exists over time periods as well as on the cross-sectional level. This indicates that fixed effect estimation should be used in all four regression models in order to isolate the relationships of interest, seeing that there are many unobservable that otherwise may affect the results.

4.6 Limitations

It has been debated whether or not NBSC is a credible database or if misreporting exists on macro-level data in cases when there is something to gain from embellishing the true values (Kawasaki 2018; Lau 2018). However, the data is collected by different organizations, institutions and statistical bureaus, which is likely to contribute to the transparency in the compilation of the data at the same time making manipulation of it harder. In addition, it is in China's best interest that the data publications are accurate as it can benefit the country if researcher have access to true values in order to make predictions, give recommendations and explain developing patterns. Another reason for using NBSC data is simply because there is no source on provincial data more detailed and thoroughly provided as in the China statistical yearbooks.

Nevertheless, a second limitation to this study is related to data availability. Data for some main variables was largely missing for the period prior to 2008 thus limiting the time period investigated to ten years as the last year reported was 2017. On similar basis the study excludes three regions from the sample. Even though the sample size of time series is fairly small, panel data makes it possible to use useful cross-sectional information capturing both provincial and time effects which helps overcome this issue as pointed out by Verbeek (2012). Hence the time frame is considered sufficient to investigate the relationship of interest. A further potential limitation to the study is the issue of time variant observations getting thrown out of the sample as fixed effects estimation is used in the model. This may be an issue as it can cause the estimated effect of rail infrastructure on the regional development variables being biased downwards. However, using fixed effect estimation has great benefits outweighing this potential limitation. The main one being reduced likelihood of omitted variable problems, but also because it controls away some permanent differences between provinces that are not being investigated but may affect the outcome.

5. RESULTS

Chapter 5 presents the main findings of the study. The results are discussed, and robustness checks will be conducted to control for the validity of these results.

5.1 Benchmark Results

Table 2 below shows the benchmark results from running model 1 specified in section 4.4. The results are displayed separately for the four defined subcategories in regional development: (1) Education, (2) Healthcare, (3) Employment and (4) Environment.

Table 2. Benchmark results

	(1)	**(2)**	**(3)**	**(4)**
***Dependent variables*:**	**Education**	**Healthcare**	**Employment**	**Environment**
Rail Infrastructure	-0.001***	-0.000	0.034***	0.807
	(0.000)	(0.000)	(0.008)	(0.696)
GRP	-0.000	-0.000	0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Gvmt. Expenditures	-0.000	-0.000**	0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Investment	0.000***	0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Technology	-2.123***	-0.714***	3.336	-250.749
	(0.439)	(0.189)	(7.960)	(658.761)
Population	0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Highway Infrastructure	-0.000	-0.000	-0.001	-0.016
	(0.000)	(0.000)	(0.000)	(0.028)
Year (FE)	YES	YES	YES	YES
Constant	0.021***	0.006**	0.019	-14.174
	(0.007)	(0.003)	(0.122)	(10.103)
R-squared	0.528	0.567	0.847	0.478
Observations	279	279	279	279
Number of Provinces	31	31	31	31

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
All variables are computed in per capita measures.
Fixed effects estimation for time specific and cross-provincial effects.

The presented results in table 2 displays the approximated relationship between rail infrastructure and the chosen regional development variables. The coefficients show the change in the dependent variables as the independent variable increases by one unit (Dougherty 2016, p.98). As shown in table 2, the benchmark results exhibit no significant correlation between rail traveling intensity and healthcare levels in the provinces, which means that H_2 stating that there is a positive relationship should be rejected. The estimated relationship between SO_2 emissions and rail infrastructure is also proving to be insignificant, resulting in the rejection of H_4 stating that a relationship exists. Number of people enrolled in higher education appears as having a negative relationship with the levels of rail infrastructure (-0.001) with a statistical significance at a one percentage level, leading to the rejection of H_1 in which a positive correlation was declared. Furthermore, according to the results in table 2 there does exists a statistically significant positive effect on employment ratios (0.034) at the one percentage level. Meaning that the only hypothesis that is accepted, according to these findings, is H_3 saying that provinces with a higher level of rail infrastructure have a higher rate of employment.

Based on the benchmark results, this study rejects three out of the four hypotheses formulated in section 3.6. Out of the four regional development variables investigated only employment rates is positively affected by a higher level of rail infrastructure.

5.3 Robustness checks

After presenting the benchmark results the robustness will be tested by making a number of changes to the model, and to some of the variable specifications. If the results remain roughly similar after making these changes, the results are considered robust. The tests will be conducted in the same way as previously, where each regional development variable was tested separately. Six robustness checks will be carried out as follows. The regression will be run as a dynamic model, the measure for the main independent variable will be changed, monetary values will be logged, the model will be tested when excluding municipalities and autonomous regions respectively, and lastly GRP per capita will be used as a dependent variable.

5.3.1 Dynamic Model

As a first robustness check the model is transformed into a dynamic one that accounts for lagged effects of the rail infrastructure. This is done when it is reasonable to hypothesize that the dependent variable may depend on both current and lagged values of the explanatory variable of interest (Dougherty 2016, p.413-415). In this case, it would mean that changes in rail infrastructure in the previous year is likely to impact the regional development the following year. For example, large investments into rail infrastructure one year is expected to improve the employment opportunities in the second year through the mechanism of changes in commuting behavior (by providing a shorter travelling time). Therefore, the following dynamic model is created:

Model 2

$$REGIONALDEV'\cdot_{it} = \alpha_i + \beta_1 RAILINFRA.pc_{it-1} + \beta_2 RAILINFRA_{it} + \beta_3 X'_{it} + \beta_4 YEAR_t + \varepsilon_{it}$$

The lagged independent variable, β_2RAILINFRA.pc_{it-1}, in model 2 is interpreted as the effect on regional development this year from a one percentage unit change in last year's rail infrastructure. In all other aspects model 2 has the same structure and interpretation as model 1.

Table 3. Results from dynamic model

Dependent variables:	(5) Education	(6) Healthcare	(7) Employment	(8) Environment
Lag Rail Infrastructure	-0.001	-0.001	0.015	0.448
	(0.001)	(0.000)	(0.016)	(1.310)
Rail Infrastructure	-0.001	0.000	0.023	0.479
	(0.001)	(0.000)	(0.014)	(1.184)
GRP	-0.000	-0.000	0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Gvmt Expenditures	-0.000	-0.000**	0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Investment	0.000***	0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Technology	-2.114***	-0.705***	3.106	-257.674
	(0.440)	(0.188)	(7.965)	(660.323)
Population	0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Highway Infrastructure	-0.000	-0.000	-0.001	-0.015
	(0.000)	(0.000)	(0.000)	(0.029)
Year (FE)	YES	YES	YES	YES
Constant	0.021***	0.007**	0.007	-14.517
	(0.007)	(0.003)	(0.123)	(10.171)
R-squared	0.529	0.571	0.848	0.478
Observations	279	279	279	279
Number of Provinces	31	31	31	31

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
All variables are computed in per capita measures.
Fixed effects estimation for time specific and cross-provincial effects.

The results in table 3 shows that when applying a dynamic model to the data all of the main results becomes insignificant, making it meaningless to extract any more information from the table. Significance usually drops when using lags if the investigated timeframe is relatively short, which is the case in this study (Keele and Kelly 2006). The insignificant results can be due to the variables being subject to more short-term variation. Implying that dependent variables only depend on the current values of the explanatory variable. The results from the dynamic model does not discredit the robustness of the benchmark model if that is the case, since which model to use depends on what assumption that are made about when the effect of rail infrastructure appears. The insignificant results from the dynamic model may also indicate that there is no causational relationship present. Meaning that rail infrastructure only is correlated with regional development but does not have a causal effect on it.

5.3.2 Changing the specification of the main independent variable

As a second step in the robustness checks the main independent variable, rail infrastructure, will be measured in a different way. Instead of using how many times one-person travels by train each year, rail infrastructure will now be measured as how far one-person travels by train each year in kilometers. In other words, traveling distance will replace traveling intensity. The variable for highway infrastructure will be changed in the same way in order to keep consistency. Furthermore, since the values are very large on average the variables are being logged. Giving them a slightly different interpretation since it is now a linear-log model. The coefficient represents the unit change in the regional development variables for a percentage change in rail infrastructure. The results from using how far residents travel as a measure for rail infrastructure is presented in table 4 below.

__Table 4. Results from changing the measure for rail infrastructure__

	(9)	(10)	(11)	(12)
__Dependent variables:__	**Education**	**Healthcare**	**Employment**	**Environment**
Ln(Rail Km)	0.001	0.000	0.024***	0.740
	(0.000)	(0.000)	(0.009)	(0.729)
GRP	-0.000***	-0.000	0.000***	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Gvmt Expenditures	-0.000*	-0.000**	0.000***	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Investment	0.000***	0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Technology	-1.940***	-0.625***	18.999**	66.135
	(0.436)	(0.185)	(8.015)	(650.399)
Population	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Ln(Highway Km)	0.002**	0.000	0.035***	0.090
	(0.001)	(0.000)	(0.013)	(1.019)
Year (FE)	YES	YES	YES	YES
Constant	0.000	0.000	-0.481***	-20.413
	(0.010)	(0.004)	(0.185)	(15.032)
R-squared	0.521	0.568	0.840	0.475
Observations	279	279	279	279
Number of Provinces	31	31	31	31

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
All variables are computed in per capita measures.
Fixed effects estimation for time specific and cross-provincial effects.

Similarly, to the benchmark results a correlation between the rail infrastructure and employment levels can be established with a significance at the one percentage level. The results show that a one percentage increase in travel length per passenger results in a 2.4 unit increase in employment levels. The results show no significant relationship to education per capita as it did in the benchmark results. However, the results from changing the measure of rail infrastructure overall shows similar patterns and therefore confirms the rejection of H_2 and H_4, while confirming H_3.

5.3.3 Logging the monetary values

When the values of variables are very large (as often is the case with monetary value) or if they are suspected to be skewed in some direction it is common practice in econometrics to normalize such variables by using logs. Further reasons for using logged variables is the easier interpretation of the results or that it improves the model fit. The reason for logging the monetary values as a robustness check in this analysis is to investigate if the model fit improves and whether or not the main results stay robust to such an alternation. The results from logging the monetary values GRP per capita, local government expenditures on transportation per capita and investment per capita are presented in table 5.

Table 5. Results with logged monetary values

	(13)	(14)	(15)	(16)
Dependent variables:	**Education**	**Healthcare**	**Employment**	**Environment**
Rail Infrastructure	-0.001***	-0.000	0.040***	0.733
	(0.000)	(0.000)	(0.009)	(0.675)
Ln(GRP)	0.016***	0.005***	0.103***	8.732***
	(0.002)	(0.001)	(0.036)	(2.871)
Ln(Gvmt Expenditures)	-0.000	-0.001***	0.036***	-0.992
	(0.000)	(0.000)	(0.011)	(0.833)
Ln(Investment)	-0.000	-0.000	0.006	0.811**
	(0.000)	(0.000)	(0.005)	(0.365)
Technology	-2.064***	-0.720***	39.026***	148.065
	(0.309)	(0.146)	(7.058)	(558.645)
Population	0.000***	0.000	0.000	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Highway Infrastructure	-0.000	-0.000	-0.001*	-0.038
	(0.000)	(0.000)	(0.000)	(0.027)
Year (FE)	YES	YES	YES	YES
Constant	-0.150***	-0.047***	-1.325***	-110.081***
	(0.017)	(0.008)	(0.394)	(31.190)
R-squared	0.703	0.662	0.844	0.494
Observations	262	262	262	262
Number of Provinces	31	31	31	31

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
All variables are computed in per capita measures.
Fixed effects estimation for time specific and cross-provincial effects.

Comparing the R-squared values in table 5 to table 2 it can be confirmed that the model fit improves in the cases of education per capita and healthcare per capita, stays nearly unchanged for employment per capita and declines slightly for environment per capita. However, the differences in model fit are not considered sizeable. Looking at the relationship between the regional development variables and rail infrastructure, the results are almost the same as in the benchmark model. The coefficients are insignificant for healthcare and environment, but significant at the one percentage level for education and environment in the case when logged monetary values are being used in the model. The results in table 5 therefore confirms the direction of the relationships established in the benchmark model for the education and employment variables. The robustness check also coincides with the benchmark results confirming H_3, while rejecting H_1, H_2 and H_4.

5.3.4 Excluding municipalities and autonomous regions from the sample

Municipalities and autonomous regions are formally described as having the same economical, jurisdictional and political rights as the provinces (China Today 2008). However, in praxis there do exist differences due to the complex political situation in China. In the case of the municipalities, Beijing, Tianjin, Shanghai and Chongqing, they are under direct administration of the central government, have a smaller land size and are much richer compared to other regions (China Today 2008; NBSC 2019). The autonomous regions, Inner Mongolia, Ningxia, Xinjiang, Guangxi and Tibet, can to some extent also be expected to show some divergence due to their status as ethnic independent areas (china.org.cn 2019). Therefore, a robustness checks where municipalities and autonomous regions are excluded from the sample is performed to investigate how it affects the benchmark results.

Table 6. Results from excluding municipalities

Dependent variables:	(17) Education	(18) Healthcare	(19) Employment	(20) Environment
Rail Infrastructure	-0.000	0.000	0.021***	0.575
	(0.000)	(0.000)	(0.008)	(0.807)
GRP	-0.000	0.000**	0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Gvmt Expenditures	-0.000	-0.000*	0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Investment	0.000***	0.000***	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Technology	-0.801**	-0.059	22.951***	-252.284
	(0.342)	(0.111)	(7.965)	(846.167)
Population	0.000*	-0.000**	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Highway Infrastructure	0.000	0.000**	-0.001	-0.019
	(0.000)	(0.000)	(0.000)	(0.036)
Year (FE)	YES	YES	YES	YES
Constant	0.012**	0.007***	-0.045	-15.539
	(0.005)	(0.002)	(0.113)	(11.968)
R-squared	0.756	0.874	0.873	0.492
Observations	243	243	243	243
Number of Provinces	27	27	27	27

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
All variables are computed in per capita measures.
Fixed effects estimation for time specific and cross-provincial effects.

Table 6 shows the results of using model 1 with the four municipalities excluded from the sample, showing similar trends to the benchmark results. Except that the statistical significance disappears for the relationship between rail infrastructure and education level. Meanwhile, the explanatory power of rail infrastructure on employment level decreases to (0.021) from (0.034) in the benchmark model but remains significant at the one percentage level.

Table 7. Results from excluding autonomous regions

Dependent Variables:	(21) Education	(22) Healthcare	(23) Employment	(24) Environment
Rail Infrastructure	-0.002***	-0.000	0.041***	-0.187
	(0.001)	(0.000)	(0.009)	(0.453)
GRP	-0.000	-0.000	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Gvmt Expenditures	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Investment	0.000***	0.000***	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Technology	-1.993***	-0.747***	2.272	-416.502
	(0.478)	(0.209)	(8.571)	(408.838)
Population	0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Highway Infrastructure	-0.000	-0.000	-0.000	-0.018
	(0.000)	(0.000)	(0.000)	(0.018)
Year (FE)	YES	YES	YES	YES
Constant	0.023***	0.007*	-0.033	-8.534
	(0.008)	(0.004)	(0.148)	(7.038)
R-squared	0.495	0.540	0.847	0.629
Observations	234	234	234	234
Number of Provinces	26	26	26	26

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
All variables are computed in per capita measures.
Fixed effects estimation for time specific and cross-provincial effects.

Table 7 displays the results from excluding the autonomous regions. The statistical significance remains at the one percentage level for the effect of rail infrastructure on both education and employment level. The coefficient of rail infrastructure on education per capita decreases from (-0.001) to (-0.002) when excluding autonomous regions. A plausible explanation for this slight drop may be due to residents in autonomous regions being more inclined to go to school if rail infrastructure increases than in other regions. On the other hand, the coefficient of rail infrastructure on employment per capita increases from (0.0034) to (0.041) implying that the employment rate is more affected by rail infrastructure in other regions than in the autonomous regions.

5.3.5 Using GRP per capita as one of the dependent variables

As a last robustness check the original regressions will be tested excluding GRP per capita as a control variable and instead including it as a dependent variable. GRP per capita is considered an important indicator for economic development in regions, which makes it reasonable to include it on the left-hand side. In addition, GRP per capita may absorb some of the effect on the regional development variables that otherwise would be captured by the other explanatory variables.

Table 8. Results from GRP per capita

VARIABLES	(25) GRP	(26) Education	(27) Healthcare	(28) Employment	(29) Environment
Rail Infrastructure	0.019	-0.002***	-0.000*	0.053***	1.136*
	(0.018)	(0.000)	(0.000)	(0.009)	(0.651)
Gvmt Expenditures	0.000**	-0.000*	-0.000**	0.000***	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Investment	0.000***	0.000***	0.000***	0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Technology	-22.005	-2.425***	-0.786***	30.746***	210.375
	(15.270)	(0.373)	(0.160)	(7.324)	(559.744)
Population	-0.000	0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Highway Infrastructure	0.001	-0.000	-0.000	-0.000	-0.012
	(0.001)	(0.000)	(0.000)	(0.000)	(0.028)
Year (FE)	YES	YES	YES	YES	YES
Constant	10.438***	0.021***	0.006**	0.012	-14.291
	(0.276)	(0.007)	(0.003)	(0.132)	(10.118)
R-squared	0.952	0.525	0.566	0.820	0.474
Observations	279	279	279	279	279
Number of Provinces	31	31	31	31	31

Standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1
All variables are computed in per capita measures.
Fixed effects estimation for time specific and cross-provincial effects.

Table 8 depicts the results from excluding GRP per capita as a control variable from the original four models, as well as the results from the model using GRP per capita as a dependent variable. The relationship between rail infrastructure and GRP per capita reveals to be statistically insignificant in this setup. When excluding GRP per capita as a control variable, in the benchmark model, all four regional development variables show significant relationships with rail infrastructure. This can be due to some effects previously being captured by GRP per capita now is being absorbed in the rail infrastructure variable. One interesting aspect being that rail infrastructure now appears to have a positive correlation with SO_2 emissions, which is significant on a ten-percentage level. Similarly, healthcare per capita turns up to be negatively related to rail infrastructure with a significance level of ten percent. The coefficients and significance level for education and employment per capita is similar to the benchmark results. Even though the effect increases for both variables, especially for the relationship between rail infrastructure and employment per capita.

In summary, the robustness checks overall confirms the results from the benchmark model. The positive relationship between employment level and rail infrastructure seems to be especially robust to changes.

6. CONCLUDING DISCUSSION

In this paper, the impact that rail infrastructure has on the regional development on a provincial level in China was explored. The paper further specified that the variables of interest in regional development are education, healthcare, employment and environment levels, since focus was on the living standards of the regional residents. To test the relationship between these subcategories and the level of rail infrastructure, the study ran OLS regression with fixed effects estimation. The checks conducted in large proved the results to be robust. The finding is that only two of the regional development variables are affected by rail infrastructure. The first one, number of students in higher education, shows results inconsistent with the theory as it displays a negative correlation with rail infrastructure. This study suggests that this may be explained by more people moving to the areas with better rail infrastructure, which makes the ratio of residents in higher education decline (as the measure is in per capita). The second variable, employment rates, demonstrates a highly significant and positive relationship with rail infrastructure, which is in line with the theoretical implications of this paper. The result confirms that interconnection and mobility is an important factor in stimulating economic activities and that rail infrastructure contributes to regional development across China.

As for the level of healthcare and environment no significant effect was found. One explanation could be that there simply is no correlation in the way the data, variables and model was specified in this study. Another explanation for the found results for environment is that a large part of the SO_2 emissions comes from other sources than transportation. If the main pollutants come from industries that are geographically bounded (for example oil refineries and coal mines) and if those industries are unaffected by the rail system, then the emissions should remain unchanged by the improved rail infrastructure. As for the insignificant result for healthcare it is likely that it does not display a relationship because healthcare already is developed in the regions due to dense populations, making it unnecessary to travel to other areas to get good healthcare. A second explanation could be that the once that the majority of the once that would need better healthcare, making the hospitals hiring more healthcare personnel, is the rural residents that cannot afford the train ticket fee or the charge of a hospital visit.

In conclusion, the main result from this analysis is that the relationship between rail infrastructure and regional development is contradictory. The study further concludes that the selection of measurements representing regional development determines the observed correlation with rail infrastructure. The analysis also highlights the important role of rail infrastructure in stimulating interconnections and mobility leading to improved employment rates. However, as long as the Hukou system is not completely abolished, increased interconnection and improved possibilities of traveling by train may not have the desired effect on regional development. This is because it ties the Chinese people to a specific area, depriving them of the possibility to relocate in response to market demand. This in turn leading to economic growth deprived of increased productivity. So, regional disparities seem to be highly affected by the Hukou system as it creates inequality, insecurity and inefficiency in the Chinese economy. This is why I suggest further research being conducted on how a dissolution of the hukou system would affect the regional development variables investigated in this paper. Furthermore, given that the national rail planning strategies were intended to decrease regional disparities and facilitate social stability it would be enlightening to investigate how the impact have varied between regions.

REFERENCES

Angrist. Joshua and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist Companion*. Princeton University Press.

Ansar, A., Flyvbjerg, B., Budzier, A. and Lunn, D. (2016). *Does infrastructure investment lead to economic growth or economic fragility? Evidence from China*. Oxford Review of Economic Policy, Vol.32, Nr.3, pp.360-390.

Berti, Adele (2019). *Timeline: Profiling the evolution of China´s high-speed rail network*. Available online: https://www.railway-technology.com/features/china-high-speed-rail-evolution/ [Accessed 19 May 2019]

Bliss Air (201). *What is PM2.5 and why should you care.* Available at: https://blissair.com/what-is-pm-2-5.htm [Accessed 27 May 2019]

Button, K. (2017). *High speed Railways: Do they produce economic growth?*. Mercatus Center Research Paper. George Mason University.

Cai, F. (2008). *Approaching a Triumphal Span: How far is China towards its Lewisian Turning Point?*. UNU-WIDER Research Paper 2008/09.

Cambridge Dictionary (2019). *Infrastructure*. Available at: https://dictionary.cambridge.org/dictionary/english/infrastructure [Retrieved: 22 May 2019]

Cao, Lan (2012). *Informal Institutions and Property Rights*. Faculty Publications 1427. College of William & Mary Law School.

CBC News (2015). *5 things to know about China´s 1-child policy.* Available online: https://www.cbc.ca/news/world/5-things-to-know-about-china-s-1-child-policy-1.3294335 [Accessed 19 May 2019]

Chen, Zhenhua (2017). *Impacts of high-speed rail on domestic air transportation in China*. Journal of Transport Geography. Nr.62 pp.184-196.

Cline, W. (2010), *Renminbi Undervaluation, China's Surplus, and the US Trade Deficit,*Peterson. Institute for International Economics. Policy Brief 10-20 August 2010.

Chen, Zhenhua (2019). *Measuring the regional economic impacts of high-speed rail using a dynamic SCGE model: the case of China*. European Planning Studies. 27:3, pp. 483-512.

Chen, Z., and Haynes, K. E. (2017). *Transportation infrastructure and economic growth in China: A meta-analysis*. In H. Shibusawa, K. Sakurai, T. Mizunoya, & S. Uchida (Eds.), Socioeconomic environmental policies and evaluations in regional science (pp. 339–357). Springer: Singapore.

China.org.cn (2019). *Administrative Division System*. Available online: http://french.china.org.cn/english/feature/38436.htm [Accessed 20 May 2019]

China Today (2008). *China´s Municipalities*. Available online: http://www.chinatoday.com/city/china_municipalities.htm [Accessed 20 May 2019]

Dali, L. Yang, Huayu Xu and Ran Tao (2014). *A tragedy of the Nomenklatura? Career incentives, political loyalty and political radicalism during China´s Great LeapForward*. Journal of Contemporary China. Vol.23, Nr.89, pp. 864-883.

Demurger, S. (2001). *Infrastructure and economic growth: an explanation for regional disparities in China?* J. Comp. Econ. Nr.29, pp.95-117.

Dollar, D. (2013). *China's Rebalancing: Lessons from East Asian Economic History*. John L. Thornton China Center. Working Paper Series, pp.1-27.
Dougherty, C. (2016). *Introduction to econometrics*. 5th edition, Oxford Press.

EESI (2018). *Fact sheet: High speed rail development worldwide.* Available at: https://www.eesi.org/papers/view/fact-sheet-high-speed-rail-development-worldwide#3 [Accessed 28 May 2019]

Focus Economics (2019). *What is GDP per capita?* Available online: https://www.focus-economics.com/economic-indicator/gdp-per-capita [Accessed 18 May 2019]

Fregert, Klas and Lars Jonung (2014). *Makroekonomi - Teori, politik och institutioner*. 4th edition. Studentlitteratur AB.

Han, Li and James Kai-Sing Kung (2015). *Fiscal Incentives and Policy Choices of Local Governments: Evidence from China*. Journal of Development Economics, nr 116. p.89-104.

Huang, Yaping, Shiwei Lu, Xiping Yang and Zhiyuan Zhao (2018). *Exploring railway network dynamics in China from 2008-2017*. International Journal of Geo-Information. Vol.7, Nr.320. pp.1-19.

Jeng, Chi-Ruey and Chao-Hsu Su (2013). *The Predicament of Domestic Airline Services after the Introduction of Taiwan High-Speed Rail*. Penn State University Press. Vol.52, Nr.1, pp.134-143.

Jiang, Xiushan, Xiang He, Lei Zhang, Huanhuan Qin and Fengru Shao (2017). *Multimodal transportation infrastructure investment and regional economic development: A structural equation modeling empirical analysis in China from 1986 to 2011.* Transport Policy. Nr. 54, pp.43-52.

Juneja, Priyanka (2017). *China's Hukou System: An interview with Fei-Long Wang.* Available at: https://thediplomat.com/2017/07/chinas-hukou-system/ [Accessed 26 May]

Kawasaki, Masumi (2018). *How trustworthy are Chinese statistics? China inflates GDP figures by $478 billion in 2018*. Available online: https://japan-forward.com/how-trustworthy-are-chinese-statistics-china-inflates-gdp-figures-by-478-billion-in-2017/ [Accessed 21 May 2019]

Keely, Luke and Nathan J. Kelly (2006). *Dynamic models for dynamic theories: The ins and outs of lagged dependent variables*. Political Analysis. Vol. 14, Nr.2, pp.186-205.

Lau, Lawrence (2018). *Are Chinese economic statistics reliable?* Available online: https://www.chinausfocus.com/finance-economy/are-chinese-economic-statistics-reliable [Accessed 21 May 2019]

Li, David D (1998). *Changing Incentives of the Chinese Bureaucracy*. American Economic Review. Vol. 88, Nr 2. pp.393-398.

Lin, Justin Yifu (2013). *Demystifying the Chinese Economy*. The Australian Economic Review, vol. 46, no 3, pp. 259-269.

Lipshitz, Gabriel (1993). *The Main Approaches to Measuring Regional Development and Welfare*. Social Indicators Research. Vol.29, No.2, pp.163-181.

Liu, H. and Daqian Shi (2018). *China's role in the new round of restructuring of regional division of labor in East Asia*. Asian Education and Development Studies, Vol. 7, Issue: 4, pp.343-363.

Luo, Xubei, Nong Zhu and Heng-fu Zou (2014). *China's lagging region development and targeted transportation infrastructure investments*. Annals of Economics and Finance. Vol. 15, Nr.1, pp.157-200.

Lo, D. (2006). *Making sense of China's economic transformation*. SOAS Working Papers, No.148.

Lundahl, U. and Skärvad, P. (1999). *Utredningsmetodik för samhällsvetare och ekonomer*. 3.uppl., Lund: Studentlitteratur.

Michalos, Alex (2007). *Development of Quality of Life Theory and Its Instruments*. The selected Works of Alex C. Michalos. Springer, Cham.

Milenkovic, Nemanja, Jovanka Vukmirovic, Milica Bulajie, Zoran Radojicic (2014). *A multivariate approach in measuring socio-economic development of MENA countries.* Economic Modelling. Nr.38, pp.604-608.

Mohajan, Haradhan (2014). *Chinese Sulphur Dioxide Emissions and Local Environment Pollution.* International Journal of Scientific Research in Knowledge. Vol.2, Nr.6, pp.265-276.

NBSC, National Bureau of Statistics of China (2019). Available online: http://data.stats.gov.cn/english/easyquery.htm?cn=C01 [Accessed 10 April 2019]

Nee, Victor and Sonja Opper (2012). *Capitalism from Below: Markets and Institutional Change in China.* Cambridge (MA), Harvard University Press.

Nee, Victor and Sonja Opper (2014). *Markets and Institutional Change in China. The Oxford Companion to the Economics of China*. Oxford: Oxford University Press. p.1-10.

OECD (2019). *Regional development policy*. Available online: http://www.oecd.org/cfe/regional-policy/regionaldevelopment.htm [Accessed 15 May 2019]

OECD Better Life Index (2019). *Better Life Index*. Available at: http://www.oecdbetterlifeindex.org/#/11111111111 [Retrieved: 23 May 2019]

Pardo-Bosch, Francesc and Antonio Aguadi (2016). *Sustainability as the key to prioritize investments in public infrastructures*. Environmental Impact Assessment Review. Vol.60, pp.40-51.

Popova, Yelena (2017). *Relations between Wellbeing and Transport Infrastructure of the Country*. Procedia Engineering. Vol. 178, pp.579-588.

Reuters (2018). *China cracks down on foreign companies calling Taiwan, other regions countries.* Available online: https://www.reuters.com/article/us-china-delta/china-cracks-down-on-foreign-companies-calling-taiwan-other-regions-countries-idUSKBN1F10RC [Accessed 8 May 2019]

Rodrik, D. (2006). *What's so Special About China's Exports?* Harvard University.
Stata (2019). *Panel-data unit root tests*. Available online: https://www.stata.com/features/overview/panel-data-unit-root-tests/ [Accessed 11 May 2019]

The people's republic of China (2014). *Administrative Division*. Available online: http://english.gov.cn/archive/china_abc/2014/08/27/content_281474983873401.htm [Accessed 18 May 2019]

The World Bank (2013). *Report, China 2030: Building a modern, harmonious and creative society.* Development Research Center of the state council, the people's Republic of China.

The World Bank 2019. *Darkening prospects: Global economy to slow to 2.9 percent in 2019 as trade, investment weaken.* Available at: https://www.worldbank.org/en/news/press-release/2019/01/08/darkening-prospects-global-economy-to-slow-to-29-percent-in-2019-as-trade-investment-weaken [Accessed 26 May]

Tillväxtverket (2019). *Breddat mått på regional utveckling*. Available at: https://tillvaxtverket.se/statistik/regional-utveckling/brp--breddat-matt-pa-utveckling.html [Retrieved 22 May 2019].

Trading Economics (2019). *China GDP annual growth rate.* Available at: https://tradingeconomics.com/china/gdp-growth-annual [Retrieved 25 May 2019]

UIC (2019). *Environment.* Available at:
https://uic.org/environment [Accessed 23 May 2019]

UIC (2019). *Sustainable development*. Available at:
https://uic.org/environmental-indicators [Accessed 23 May 2019]

UNESCAP (2019). *Rail transport.* Available at:
https://www.unescap.org/ru/our-work/transport [Accessed 22 May 2019]

UNECE (2019). Transport infrastructure development. Available at:
https://www.unece.org/trans/theme_infrastructure.html [Accessed 22 May 2019]
Verbeek, M. (2012). *A guide to modern econometrics*. 4th edition, John Wiley & Sons Ltd.

Wang, Zhi, Shang-Jin Wei and Anna Wong (2010). *Does a Leapfrogging Growth Strategy Raise Growth Rate? Some International Evidence*. National Bureau of Economic Research Working Papers.

Wade, R. (2003). *Governing the Market: Economic Theory and the Role of the Government in East Asian Industrialization.* Summary to 2003 Edition.

Wu, J., Nash, C. and Wang, D. (2014). *Is high speed rail an appropriate solution to China's rail capacity problems?*. Journal of Transport Geography, Nr.40, pp.100-111.

Zhang, Qian, Yuli Shan, Jun Nakatani, Yuichi Moriguchi (2018). *Inter-regional spillover of China´s sulfur dioxide (SO_2) pollution across the supply chains*. Journal of Cleaner Production, pp.1-15.

Zeng, Douglas Zhihua (2011). *How do Special Economic Zones and Industrial Clusters Drive China's Rapid Development?* Policy Research Working Paper 5583.

Zheng, Siqi and Matthew E. Kahn (2013). *The National Academy of Sciences of the United States of America.* Vol.110, Nr.14, pp.5288-5288.

APPENDIX 1

Map showing provincial names and capitals

APPENDIX 2

List of the 31 administrative regions included in the study

Province	Formal regional status
Anhui	Province
Beijing	Municipality
Chongqing	Municipality
Fujian	Province
Gansu	Province
Guangdong	Province
Guangxi	Autonomous region
Guizhou	Province
Hainan	Province
Hebei	Province
Heilongjiang	Province
Henan	Province
Hubei	Province
Hunan	Province
Inner Mongolia	Autonomous region
Jiangsu	Province
Jiangxi	Province
Jilin	Province
Liaoning	Province
Ningxia	Autonomous region
Qinghai	Province
Shaanxi	Province
Shandong	Province
Shanghai	Municipality
Shanxi	Province
Sichuan	Province
Tianjin	Municipality
Tibet	Autonomous region
Xinjiang	Autonomous region
Yunnan	Province
Zheijiang	Province

Source: The people´s republic of China, 2014.

APPENDIX 3

LLC unit root test and First-Difference for Model 1

	Level		First Difference	
Variable	**t-statistic**	**p-value**	**t-statistic**	**p-value**
Education per capita	-4.0845	0.0000	-	-
Healthcare per capita	-6.3534	0.0000	-	-
Employment per capita	-8.5299	0.0000	-	-
Environment per capita	2.1139	0.9827	-15.3980	0.0000
Rail infrastructure per capita	-6.9481	0.0000	-	-
GRP per capita	-3.7591	0.0001	-	-
Local gvrmt expend. per capita	-5.1750	0.0000	-	-
Investment per capita	1.6699	0.9525	-8.9067	0.0000
Technology per capita	-5.5044	0.0000	-	-
Population	-67.2393	0.0000	-	-
Highway infrastructure per capita	-9.0840	0.0000	-	-

H_0: Panels contain unit roots.

APPENDIX 4

Variance Inflation Factor (VIF) for Model 1

Variable	VIF	1/VIF
Rail infra*	2.68	0.37
GRP*	4.70	0.21
LocGvmExp*	1.35	0.74
Invest*	1.08	0.92
Tech*	3.89	0.26
Population	1.48	0.68
Highway infra*	1.20	0.83
Mean VIF	2.34	

Correlation matrix of independent variables in Model 1

	Rail infra*	GRP*	LocGvm Exp*	Invest*	Tech*	Population	Highway infra*
Rail infra*	1.0000						
GRP*	0.7510	1.0000					
LocGvExp*	0.0303	0.1132	1.0000				
Invest*	0.3487	0.6388	0.3121	1.0000			
Tech*	0.6481	0.8232	0.0220	0.3533	1.0000		
Population	-0.1393	-0.0075	-0.4061	-0.1021	0.1528	1.0000	
Highway infra*	0.0163	-0.0833	-0.1956	-0.2956	0.1074	0.0526	1.0000

*** All variables are reported in per capita measures.**