MR. ZHEMING LIU (Orcid ID : 0000-0002-8727-2361)

Article type : Regular Article

Corresponding author mail-id: zengsaixing@sjtu.edu.cn

The Spillovers from Transport Infrastructures onto Firm Productivity: An Analytical and Empirical Study

Abstract

Although issues about the economic spillovers from transport infrastructure have been discussed and debated for decades, a great deal of controversy concerning the direction and magnitude of the economic effects of transport infrastructure remains, and the empirical evidence of those effects in emerging economies is still insufficient. In this study, based on a sample of Chinese manufacturing firms during the period from 1998 through 2007[1], we employed a method combining the difference-in-differences approach with the propensity scoring matching technique to research the effects of improvements in transport infrastructure on firm productivity. With highways used as a typical example of transport infrastructure, the results suggested that a connection to highways boosted firm productivity by an average of 0.043, or approximately

[1] The data that support the findings of this study are available from the corresponding author, upon reasonable request.

0.74% of the sample mean. Moreover, this study provided evidence that the effects of improvements in transport infrastructure on firm productivity were stronger in industries producing non-durable goods and driven by an increase in firm innovation. This study's findings contribute to a reconciliation of the controversy concerning the economic effects of transport infrastructure and enrich the empirical evidence of that effect in emerging economies.

Keywords—Transport infrastructure; firm productivity; innovation; National Trunk Highway Development Program

1. INTRODUCTION

Regarded as the "wheels" of economic activities, transport infrastructure plays a crucial role in promoting the economic development of a country (World Bank, 1994). In past decades, several studies have focused on the economic spillovers from transport infrastructure. In his pioneering work in this field, Aschauer (1989) found that a "core" infrastructure, consisting of streets and highways, airports, electrical and gas facilities, mass transit systems, water systems, and sewers, possessed great explanatory power for productivity in the private economy in the United States. That pivotal finding initiated a trend in the search for important effects from transport infrastructure on the economy. Finding it difficult to make a detailed evaluation of the impact of infrastructure by relying on restricted models of firms' technology and behavior, Morrison and Schwartz (1996) constructed a more complete production theory model of firms' production and input decisions and evaluated the contributions that infrastructure made to manufacturing firms' costs and productivity growth, using state-level data from the United States. Their results confirmed that investment in infrastructure provided a significant direct benefit to manufacturing firms and augmented those firms' productivity growth. However, because the direction of the causation between infrastructure and productivity remained unclear, Fernald (1999) provided evidence that vehicle-intensive industries benefited more from road-building than non-vehicle-intensive industries did, thus suggesting that the correlation between infrastructure and productivity reflected causation from changes in the stock of transport infrastructure to changes in productivity.

However, although some previous studies have revealed positive economic effects from transport infrastructure, a great majority of those studies were based on empirical evidence from developed economies, and there is still insufficient micro-level evidence on the causal relationship between transport infrastructure and firm productivity in emerging economies. Furthermore, because a large number of developing countries and regions are still suffering from an extensive deficit in infrastructure, there is a strong demand for infrastructure worldwide, especially in emerging economies. According to the prediction of Duvall et al. (2015) and Asian Development Bank (2017), almost 57 trillion U.S. dollars would need to be spent on building infrastructure around the world, and of that, developing Asia would need to devote 26 trillion U.S. dollars, over the 15 years from 2016 to 2030, to maintain the region's growth momentum, eradicate poverty, and respond locally to climate change. Given the recent resurgence of infrastructure investment, issues about transport infrastructure's economic spillovers in emerging economies have become very important and meaningful research topics. Berg et al. (2019) compared the differences in the impact of public investment in efficient countries and inefficient ones[2] and identified a counter-intuitive result that increases in public investment spending in inefficient countries did not have a lower impact on growth than such spending did in efficient countries, thus confirming promising prospects for infrastructure investment in emerging economies. The impact of transport infrastructure in developing countries has also attracted the attention of many empirical economists. Employing a difference-in-differences estimation strategy, Datta (2012) evaluated the economic effects of India's Golden Quadrilateral (GQ) Project, the most ambitious highway improvement project since India gained independence in 1947, on the country's nonagricultural private firms, and found that the GQ Project decreased transportation obstacles to production and reduced firms' average stock of input inventories by between 6 and 12 days' worth of production. Escribano et al. (2010) provided a systematic assessment of the impact of infrastructure quality on the total factor productivity of manufacturing firms in 26 African countries, and empirical evidence confirmed that losses from transport interruptions had a significant impact on the productivity of manufacturing firms, especially for those in slower-growing countries. From the perspective of enterprise dynamics,

[2] In the study by Berg et al. (2019), "efficient countries" referred to high-income countries and "inefficient countries" referred to low-income countries..

Shiferaw et al. (2015) provided evidence from Ethiopia showing that improved transport infrastructure there led to a favorable impact on the entry patterns and structure of manufacturing industries. Based on the study of Shiferaw et al. (2015), Moller and Wacker (2017) concluded that infrastructure investment made a significant contribution to an acceleration in growth in Ethiopia. Investigating on the impact of transport infrastructure on the development in colonial Ghana, Jedwab and Moradi (2011) found a strong effect of railroad connectivity on cocoa production due to reduced transport costs, which transformed the economic geography of Ghana durably.

China has also become an emerging economy worthy of attention and study. Since the "reform and opening-up policy" was implemented by the Chinese government in the late 1970s, China has paid great attention to transport infrastructure and has invested a large amount of money in it, thus contributing to the country's "infrastructure boom" during the past several decades (see Figure 1 for detailed information on China's investment in transport infrastructure). For example, in order to cope with the 2008 global financial crisis, the Chinese government enacted a program making an additional investment of ¥4 trillion yuan in 2009 and 2010 to stimulate the economy, and of that amount, approximately 53% was invested in infrastructure projects such as highways and railways (Shi and Huang, 2014). Because the construction of transport infrastructure in China has consumed vast natural and social resources, it is crucial to assess the impact of the investment in transport infrastructure on the economy and to estimate the returns on that investment. The implementation of China's National Trunk Highway Development Program (NTHDP) provides an excellent opportunity to investigate the economic impact of transport infrastructure[3].

{Insert Figure 1 here}

[3] The implementation of the NTHDP led to the rapid development of China's highways in the decade from 1998 through 2007. In that decade, the average annual growth rate of China's highways reached as high as 28.63%, which increased their mileage by nearly 15-fold. However, during that period the average annual growth rate of China's traditional railways was only 1.68%, indicating that China's traditional railway network was already relatively well developed at that time (see Figure 2 for details). Because the NTHDP exemplified an accelerated shift from traditional railways to highways as the dominating form of land transportation in China in the late 20th and early 21st centuries (Li and Shum, 2001), we believed that it was suitable and appropriate to take highways as an example for studying the influence that transport infrastructure had on firm productivity in the Chinese context during the period 1998 through 2007.

{Insert Figure 2 here}

To calculate the spillovers from improved highway accessibility onto firm productivity, we used data from the database of the Annual Survey of Industrial Firms that had been collated by the Chinese National Bureau of Statistics and that covered a sample of Chinese manufacturing firms for the period from 1998 through 2007. Because it was intended to connect a large group of target cities in China, the NTHDP improved the highway accessibility of firms located in target cities and counties lying along the routes of the program's highways, while leaving the highway accessibility of firms located in other counties unaffected. That feature of the NTHDP allowed us to perform a difference-in-differences estimation strategy to compare the firms that were connected to highways with those that were unconnected. However, we felt that differences between the characteristics of firms connected to highways and those of unconnected firms might result in the problem of nonrandom sample selection. To address that problem, we combined the difference-in-differences approach with the propensity scoring matching technique and obtained a cleaner estimate of the impact exerted by improvement in highway accessibility on firm productivity. The resulting estimate suggested that connection to a highway network boosted firm productivity by 0.043 on average, or approximately 0.74% of the sample mean. The effects of improvements in transport infrastructure on firm productivity were stronger in industries producing non-durable goods. The channel through which improved accessibility to a highway affected firm productivity was also examined. Additional empirical results provided evidence that that effect was driven by an increase in firm innovation.

The remainder of the paper is organized as follows. In Section 2, we describe the policy background of highway development in China; in Sections 3 and 4 we introduce the data we used in the study and describe our methodology, including our strategy and the variables we declared; in Section 5 we discuss the results of our empirical analysis; in Section 6 and 7 we present additional results to further investigate the heterogeneity of the main effect across industries and the channel at work; and in Section 8, the final section, we set out our conclusions, discuss the limitations of this study, and provide an outlook for future research.

2. BACKGROUND: THE NATIONAL TRUNK HIGHWAY PROGRAM IN CHINA

The “reform and opening-up policy” implemented by the Chinese government in the late 1970s greatly liberated social productive forces and attracted a large amount of foreign direct investment (FDI) into the Chinese market, thereby contributing marvelously to the launching of China’s economy. However, because that economic prosperity resulted in a rapid increase in society’s demand for transportation, it also brought the side effect of widespread traffic congestion, which was particularly serious on most arterial roads and urban entrances and exits in developed coastal regions in the 1980s. The shortages and poor conditions that plagued China’s transport infrastructure limited its carrying capacity, reduced the operating efficiency of economic activities, and severely restricted further development of the national economy.

In order to surmount the constraints that China’s insufficient infrastructure was placing on greater economic growth, at the end of the 1980s the Ministry of Transport of the People’s Republic of China proposed an ambitious program for the construction of the National Trunk Highway System. The program was approved by the State Council of the People’s Republic of China in 1992 and formally implemented in 1993. The National Trunk Highway Development Program (NTHDP), which was also known as the “Five Vertical and Seven Horizontal” National Trunk Highway Development Program, constructed five vertical (i.e., north to south) and seven horizontal (i.e., east to west) highways in China, with a total length of approximately 35,000 km, to provide a network of highways connecting the national capital, all the other municipalities, all provincial capitals, all other cities with an urban population of 1,000,000 or above, and the majority of cities with urban population in excess of 500,000 (Li and Shum, 2001). At first, the NTHDP was earmarked for completion by 2020. However, in part due to the country’s aggressive fiscal policies to stimulate the economy in response to the Asian Financial Crisis, the construction efforts of the NTHDP accelerated beginning in 1998 (Duncan, 2007; Hou and Li, 2011). As a result, the program was completed at the end of 2007, 13 years ahead of the original plan. The construction of the NTHDP mainly comprised two phases: the “kick-off” phase, between 1992 and 1997, and the “rapid development” phase, from 1998 through 2007 (World Bank, 2007). Roughly 10% of the total mileage was completed during the kick-off phase, and the other 90% of the total mileage was completed during the period 1998 to 2007.

The NTHDP vigorously promoted rapid and continuous development of the national trunk highway system. Before the Shanghai-Jiading Highway was opened to traffic in 1988, there were no international-standard dual-carriageway highways in China. However, after the implementation of the NTHDP, China's highway mileage increased extremely rapidly. From 700 km (435 mi) of highway mileage at the end of 1992, the year the NTHDP was approved, to 53,913 km (33,500 mi) at the end of 2007, the year the NTHDP was completed, China's highway mileage increased by nearly 80-fold (see Figure 3 for details). During the 16 years from 1992 through 2007, the average annual growth rate of China's highway mileage was as high as 33.75%, making China's highway mileage greater than that in all other countries in the world except the United States, at the time that the NTHDP was completed.

{Insert Figure 3 here}

3. DATA

The main data set of this study came from the database of the Annual Survey of Industrial Firms (ASIF) collated by the Chinese National Bureau of Statistics (NBS). According to Chinese laws, all qualified firms in China were required to participate in the survey conducted by the NBS. Therefore, the ASIF database covered all industrial firms with annual sales of 5,000,000 RMB (equivalent to approximately 600,000 US dollars at the 2004 exchange rate) or more, in the industries of (1) mining, (2) manufacturing, and (3) production and distribution of electricity, gas and water, from which manufacturing firms accounted for more than 90% of all observations. Because the ASIF database represents a large sample size covering a long time span and contains a large number of detailed firm-specific information, it has been widely used in a growing body of research related to Chinese industrial firms (Qian and Yaşar, 2016; Song et al., 2011). In this study, we focused on manufacturing firms for the period from 1998 through 2007, and we included a total number of 1,651,549 observations in 428 major industries (using four-digit industry codes) in our full sample. We obtained all of the firm-level information needed for this study from the ASIF database. In addition, we collected all of the city-level data used in this study from the China City Statistical Yearbooks.

This study took highways as an example of typical transport infrastructure and focused on their spillovers onto firm productivity. We manually built a highway database that included the specific time that each county-level administrative district became connected to the National Trunk Highway System[4]. In order to obtain highway information at the county level, we collected copies of the *China Road Atlas* published by SinoMaps Press from 1998 through 2007. By comparing the information on highways in the *China Road Atlas* for every pair of consecutive years, we obtained the specific year that each county-level administrative district connected to the National Trunk Highway System[5]. After we had obtained the detailed information of highway openings from the *China Road Atlas*, we then ensured the accuracy of those data by double-checking the information with the news and other reports about planning, construction, and opening of highways available on all levels of official Chinese government websites[6]. The rapid development of China's highways benefited a growing number of county-level administrative districts. As is shown in Figure 4, only 6.69% of the county-level administrative districts in China were connected to the National Trunk Highway System in 1998, but thanks to the implementation of the NTHDP, the proportion increased to 36.60% after just a decade of rapid highway development.

{Insert Figure 4 here}

4. EMPIRICAL STRATEGY AND DEFINITION OF VARIABLES

4.1. Empirical Strategy

The primary purpose of this study was to explore the relationship between transport

[4] The administrative districts of China can generally be divided into four levels: provincial-level administrative districts, prefecture-level administrative districts, county-level administrative districts, and township-level administrative districts. According to the data for 2019, there are totally 34 provincial-level administrative districts, 333 prefecture-level administrative districts, 2,845 county-level provincial-level administrative districts, and 39,945 township-level administrative districts in China. Because China experienced a climax of county-level administrative district mergers in the early 21st century, there were approximately 3,000 county-level administrative districts, whose average population was approximately 480,000, in China at that time. The data set in this study included manufacturing firms located in 2,869 county-level administrative districts, representing more than 95% of all county-level administrative districts in China.

[5] Established in 1954, SinoMaps Press is a large-scale press under the Ministry of Natural Resources and is the most authoritative agency for publishing national legal maps in China. SinoMaps Press publishes the *China Road Atlas* every year to update the conditions of China's roads and highways. Using the highway data of the Ministry of Transport and BeiDou Navigation Satellite System, the highway information in the *China Road Atlas* is very accurate and reliable.

[6] Generally speaking, governments in China are required and willing to disclose the information about the opening of new highways in their territory on official websites through news, reports, interviews, or other forms of publicity.

infrastructure and firm productivity. However, before estimating the impact of transport infrastructure, we realized that regions that had access to better transport infrastructure were probably systematically different from regions that did not. Generally speaking, big cities tended to have better transport infrastructure and firms that were more productive. Therefore, we were likely to overestimate the causal effect of transport infrastructure on firm productivity if we did not properly address the problem of nonrandom placement. Fortunately, the nature of transport infrastructure, such as highway networks, allows them to be regarded as exogenous shocks to regions they pass through (Chandra and Thompson, 2000). In most countries, highway improvement programs usually aim to connect cities with important political status or a developed economy. However, when target cities are connected, regions located between the target cities are passively connected to highway networks. Then, it could be argued that regions located between target cities are connected to highway networks not as a consequence of any political, economic, or any other characteristics they possessed, but merely because of the places where they happen to be located (Datta, 2012). Because the initial plan of the NTHDP aimed to connect the national capital, all the other municipalities, all provincial capitals, all other cities with urban population of 1,000,000 or above, and the majority of cities with urban population in excess of 500,000 (Li and Shum, 2001), we excluded all firms located in those places and only focused on firms that were located in regions between target cities, so that the problem of nonrandom placement could be addressed to the utmost. As a result, our number of observations was reduced from 1,651,549 to 1,102,930. Using the sample that had ruled out firms located in target cities, we adopted an ordinary least squares (OLS) estimation as the first stage of our calculations to examine the relevance of the key variable and the control variables in this study.

However, we reasoned that a systematic difference in total-factor productivity (TFP) could exist between firms that had access to the national trunk highway system and firms that did not. It was likely that regions connected to highways had advantages that made the average TFP of firms located in those regions higher than that of firms located in regions unconnected to highways. Moreover, some unobservable nonrandom factors could exist that were correlated to the explanatory variable, thus leading to a problem of omitted variables and a bias of the OLS estimation. In order to solve those potential problems, we adopted a difference-in-differences approach as the second stage of our calculations to estimate the causal effect of transport infrastructure on firm productivity by comparing firms located in regions unconnected to highways with firms located in regions newly connected to highways during the sample period. That approach ruled out firms that were located in regions that had been connected to highways before 1998 (the first year of the sample period) and reduced the number of observations to 906,595. The difference-in-differences estimation was able to eliminate the influence of all observable and unobservable nonrandom factors that were constant or strongly persistent over time, thereby making the difference-in-differences estimation cleaner and more reliable than the OLS estimation.

In the difference-in-differences estimation, we eliminated the influence of systematic differences in the TFP between the two study groups and also the influence of omitted-variable bias. In order to identify the exact impact of highway improvement on firm productivity and make the estimation even cleaner, we reasoned that it would be better to ensure that the changes of firm productivity during the sample period were driven only by the implementation of the NTHDP and not by other factors. Therefore, we sought to make sure that the firms in the treated group and the control group shared similar characteristics that could affect their productivity. However, it was probable that firms located in counties connected to highways had characteristics that were quite different from those of firms located in counties unconnected to highways in the year prior to their access to highways, thus making the estimation of highways' impact on TFP still likely to be vulnerable to the problem of nonrandom sample selection. In order to solve that problem, we combined the propensity score matching technique with the difference-in-differences approach in the third stage of our estimation procedure. As a result, we obtained a sample of 27,912 pairs of firms (55,824 observations), and that sample was assured to have no significant differences in observable firm characteristics between the two groups.

4.2. Variables and Measurements

4.2.1. Dependent variable: TFP

The main challenge in estimating a firm's TFP is dealing with simultaneity bias. According to the production function (see Equation (1)), output is a function of both production factors and productivity, and productivity is also affected by the inputs of production factors. As a result, it is possible that determinants of production exist, such as productivity shocks, that are unobserved by econometricians. Therefore, in order to obtain a consistent estimate of the production function and assure the accuracy of our TFP estimation, we had to control for unobservable productivity shocks.

Olley and Pakes (1996) and Levinsohn and Petrin (2003) were pioneers in solving the problem of unobservable determinants of production. They used investment and intermediate input, respectively, as proxies for unobservable productivity shocks (in the Olley and Pakes, or OP, method, and the Levinsohn and Petrin, or LP, method), and in so doing they obtained a consistent estimate of the parameters in the production function and eliminated any simultaneity bias in their estimation of TFP (Javorcik, 2004). However, Ackerberg et al. (2015) argued that the techniques of the OP and LP methods suffered from functional dependence problems. Therefore, they suggested an alternative approach (the Ackerberg-Caves-Fraser, or ACF, method) that they based on the OP and LP methods. Because the ACF method further improves the accuracy of estimating TFP, it is widely accepted by economists worldwide (Demmel et al., 2017; Luong, 2013). The ACF method is briefly introduced:

Consider a Cobb–Douglas production function in logs:

$$y_{it} = \beta_k k_{it} + \beta_l l_{it} + \omega_{it}(m_{it}, k_{it}, l_{it}) + \varepsilon_{it} \ , \tag{1}$$

where y_{it} is the log of value added, k_{it} is the log of capital input, l_{it} is the log of labor input, m_{it} is the log of intermediate input, and ω_{it} represents TFP. β_k and β_l represent the coefficients of capital input and labor input, respectively, and ε_{it} is the error term. Here, ω_{it} is a function of m_{it}, k_{it}, and l_{it}. The first stage of the estimation procedure is estimating the conditional expectation of y_{it}, namely $\hat{y}_{it}$, and the approximate estimates of β_k and β_l, namely β_k^0 and β_l^0, by using trans-log function or semiparametric estimation. Then, we can obtain $\omega_{it}(m_{it}, k_{it}, l_{it}) = \hat{y}_{it} - \beta_k^0 k_{it} - \beta_l^0 l_{it}$. Assuming that ω_{it} obeys a first-order Markov process, namely

$\omega_{it} = E(\omega_{it} \mid \omega_{it-1}) + u_{it}$, we can estimate the model and obtain the expectation of ω_{it}, namely $\hat{\omega}_{it}$. After the procedure above, we can obtain

$$\hat{y}_{it} = \beta_k k_{it} + \beta_l l_{it} + \hat{\omega}_{it}(m_{it}, k_{it}, l_{it}) + u_{it}, \qquad (2)$$

by using the moment condition

$$E(u_{it} \left| \begin{matrix} k_t \\ l_{it-1} \end{matrix} \right.) = 0. \qquad (3)$$

Then, β_k and β_l can be estimated, and we can obtain an estimate of TFP (ω_{it})[7].

In this study, we followed the ACF method and estimated a separate translog production function for each two-digit industry, and those production function estimates related the log of value added to the log of capital and labor. We adopted intermediate input as a proxy for unobservable productivity shocks. Then, we estimated TFP for each two-digit industry, and we present those descriptive statistics in Table 1.

{Insert Table 1 here}

4.2.2. Explanatory and control variables

We measured *Highway accessibility*, the explanatory variable in this study, as a time-varying county-level dummy variable. If a firm *i* was located in a county that had been connected to the National Highway Network in the year *t*, *Highway accessibility$_{it}$* was coded as 1. Otherwise, it was equal to 0.

In order to avoid the estimation bias caused by omitted variables, to the greatest extent possible, we introduced a series of control variables in our models. The control variables could be

[7] u_{it} is the error term in Equation (2) and Equation (3).

classified into two categories: firm-level control variables and region-level control variables[8]. The firm-level control variables included *Size*, *Age*, *Leverage*, *Fixed assets*, and *Export intensity*. *Size* was defined as the logarithm of the number of full-time employees at a firm. *Age* was defined as the logarithm of the number of years since a firm's founding, and it was assumed to indicate the firm's organizational maturity. *Leverage* was the ratio of a firm's total liabilities to its tangible assets, which indicated the firm's solvency. *Fixed assets* were defined as the logarithm of the fixed assets of a firm. *Export intensity*, which reflected the development strategy of a firm, was defined as the percentage of exports in the total output of a firm. An export-oriented firm (a firm with high *Export intensity*) usually paid more attention to the international trade and overseas market than to the domestic market. Because a firm's size, maturity, solvency, assets condition, and development strategy are all important basic characteristics that affect its productivity (Aw et al., 2008; Chang and Gurbaxani, 2012; Diaz and Sánchez, 2008; Soderbom and Teal, 2001), we controlled those firm-level variables in this study.

Region-level control variables included gross domestic product (*GDP*), *Population density*, foreign direct investment (*FDI*), *Secondary industry ratio, Tertiary industry ratio*, *Railway freight volume*, and *Road freight volume*. For the *GDP*, we used the logarithm of the GDP of a prefecture-level administrative district; *Population density* was the resident population of a prefecture-level administrative district divided by its land area; *FDI* was defined as the logarithm of the foreign direct investment of a prefecture-level administrative district; and *Secondary* (*Tertiary) industry ratio* was the ratio of employees of the secondary (tertiary) industry to all employees in a prefecture-level administrative district. Basically, these five variables reflected the economic condition, demographic characteristics, and industrial structure of a prefecture-level administrative district. In order to rule out the confounding effects and examine the effects of newly built highways, we had to control for the stock of the transport infrastructure. Because road transport and rail transport were the most important modes of transportation for the regions included in our sample, we controlled in our models the transport capacity of roads, which was

[8] As the data of regional characteristics at the county level were unavailable, we controlled prefecture-level regional characteristics in this study. In China, prefecture-level administrative districts are one level higher than county-level administrative districts, and each prefecture-level administrative district includes about 8.54 county-level administrative districts on average. Footnote 3 provided more information about the administrative districts of China.

proxied by *Road freight volume*, and the transport capacity of railways, which was proxied by *Railway freight volume*. In this study, *Railway freight volume* was defined as the logarithm of the amount of freight transported by railways in a year, and the same approach was used for *Road freight volume*.

In order to take into account the time-invariant industry heterogeneity and the time trend, we included four-digit-industry-year fixed effects in our models. Industries were classified on the basis of the "Company industry classification guidelines" enacted by the China Securities Regulatory Commission in 2012, in which 428 separate manufacturing industries were defined (using four-digit industry codes). Moreover, we also controlled for the fixed effects of ownership and region.

Detailed definitions and sources of all variables are given in Table 2, and descriptive statistics, including the means, standard deviations, and Pearson correlation coefficients, for all of the variables in this study, are shown in Table 3[9].

{Insert Table 2 here}

{Insert Table 3 here}

5. EMPIRICAL RESULTS

5.1. Results from the OLS Estimation

First, we performed an OLS estimation to examine the correlation between the dependent variable and the explanatory variable in this study. The estimating equation of the OLS model is:

$$TFP_i = \alpha_i + \beta_{ij} \times Highway\ accessibility_{ij} + \sum \delta_{ik} X_{ik} + \varepsilon_{ijk}\ . \tag{4}$$

Here, *i* denotes a firm, *j* denotes a county in which the firm is located, and *k* denotes a region in which the firm is located. α is the constant term; *X* represents a series of control variables,

[9] Descriptive statistics of variables presented here were based on the sample omitting firms located in target cities.

including the firm-level and region-level control variables, and all fixed effects; β is the coefficient of the explanatory variable; δ represents the coefficients of the control variables; and ε represents random disturbance terms. All regressions introduced below used cluster-robust standard errors.

Table 4 shows the OLS estimation results of this study. The sample used in Panel (A) had omitted firms located in target cities. In Model (1) we included the explanatory variable *Highway accessibility* and all firm-level and region-level control variables, and all fixed effects were included in Model (2) on the basis of Model (1). The empirical results of both of these two models indicated that there was a significant positive relationship between highway improvement and firm productivity. Then, we excluded firms located in regions that had been connected to highways before 1998 (the first year of the sample period) and performed similar regression models, the results of which are listed in Panel (B) of Table 4. The significant positive relationship between highway improvement and firm productivity still held. All in all, the results from the OLS estimations confirmed a positive correlation between highway improvement and firm productivity. Although evidence of a causal relationship was likely to be biased, the OLS results revealed that a potential positive causal relationship might exist between highway improvement and firm productivity, which we then investigated further in the next analysis.

{Insert Table 4 here}

5.2. Results from the Difference-in-Differences Estimation on the Unmatched Sample

In order to eliminate the influences of systematic differences in TFP between groups and an omitted-variable bias on the estimation results, we performed a difference-in-differences estimation. Its regression model is:

$$TFP_{it} = \alpha_{it} + \beta_{ijt} \times Highway\ accessibility_{ijt} + \gamma_{ij} \times Treat_{ij} + \sum \delta_{ikt} X_{ikt} + \varepsilon_{ijkt} \quad . \tag{5}$$

Here, *t* denotes a year. *Treat* is a group dummy variable that is equal to one if a firm belongs to the treated group and equal to zero if a firm belongs to the control group. In this study, we regarded the connection to the highway network as a treatment; firms located in counties that were first connected to the highway network during the sample period (from 1998 through 2007) comprised the treated group, and firms located in counties that were still unconnected to the highway network by 2007 (the last year of the sample period) comprised the control group. Firms located in counties that were connected to highways before 1998 (the first year of the sample period) were excluded from the sample. The coefficient β captures the average treatment effect on the firms in the treated group, and that reflected the effect of highway accessibility on firm productivity[10]. All regressions introduced below used cluster-robust standard errors.

Results from the difference-in-differences estimation are presented in Table 5. For Model (1), only the explanatory variable *Highway accessibility* and the group dummy variable *Treat* were included in the regression model, and the result showed that the improvement of highway accessibility had a positive influence on firm productivity at the 1% significance level. In Model (2), we added all of the firm-level and region-level control variables based on Model (1), and the significant positive influence that improvement in highway accessibility exerted on firm productivity still held. Then, in Model (3), we further added all of the fixed effects, based on Model (2), and again we obtained very similar results. Therefore, the results from the difference-in-differences estimation on the unmatched sample confirmed that improvement in highway accessibility had a significantly positive impact on firm productivity.

{Insert Table 5 here}

5.3. Results from the Difference-in-Differences Estimation on the Matched Sample

5.3.1. Propensity score matching

To address the problem of nonrandom sample selection, we further combined the propensity score matching technique with the difference-in-differences approach. The propensity score matching technique was originally developed for use in the biological and medical research, and later it was introduced to and widely accepted by the fields of economics and management

[10] As a matter of fact, the explanatory variable *Highway accessibility*$_{ijt}$ equals to $Post_t \times Treat_{ij}$. $Post_t$ equals to 0 if time t is in the base period; otherwise, it equals to 1.

(Caliendo and Kopeinig, 2008; Dehejia and Wahba, 2002). Basically, the propensity score matching technique contains two steps. First, the probability (propensity score) of inclusion in the treatment is estimated, according to all samples using characteristic variables as the explanatory variables. Second, for each observation in the treated group, observations in the control group are selected as matched samples on the basis of the closeness of the probability estimated in the first step (Rosenbaum and Rubin, 1983, 1985). By using that procedure, the technique controls for selection bias by restricting the comparison to differences within carefully selected pairs with similar observable characteristics prior to the treatment (Javorcik and Poelhekke, 2017). The matching method of propensity score matching technique mainly includes nearest-neighbor matching, caliper matching, kernel matching, local linear regression matching, spline matching, and the like.

In this study, in the first step, we predicted the propensity scores of all observations with a probit model. The covariables included *Size*, its square and cube, *Age*, its square and cube, *Output*, return on equity (*ROE*), *Leverage*, and *Export intensity*. In order to eliminate the influences of differences in firm performance caused by the heterogeneity of time, industry, geographic location, and corporate ownership, we chose matched pairs of treated observations within the same year-industry cells, following the approach of Javorcik and Poelhekke (2017), and we controlled for dummies for regional and ownership effects in the model as well. In the second step, we adopted the matching method of a caliper-restricted nearest neighbor to build a control group that would be comparable to the treated group. We restricted the difference in the propensity scores between a treated observation and its matched pair in the control group to no more than 0.1%. Moreover, we imposed the nonreplacement method and the restriction of common support. Finally, we had a sample of 55,824 observations after the matching procedure.

Table 6 lists the summary statistics for the full sample and the matched sample. As we can see in Panel (A) of Table 6, in the unmatched sample, for a variety of dimensions, the characteristics of firms located in counties connected to highways were very different from those of firms located in counties that were unconnected to highways in the year prior to their access to highways. According to the results of a *t*-test, almost all of the differences between characteristic variables of firms in the treated group and those of firms in the control group were statistically significant at the 1% level, which convincingly confirmed our concerns about the nonrandom selection issue of the sample in the difference-in-differences approach. However, the propensity score matching procedure solved or at least mitigated the sample selection bias. The results in Panel (B) of Table 6 indicate that there was no longer a statistically significant difference between firms in the treated group and those in the control group in terms of any of the characteristics, after the propensity score matching procedure. Hence, we were assured that firms in the treated group and firms in the control group shared similar characteristics and were comparable prior to the treatment, thus providing an ideal sample for the difference-in-differences estimation.

{Insert Table 6 here}

5.3.2. Results of the difference-in-differences estimation

After selecting the control group in which firms shared similar characteristics with firms in the treated group, we performed a difference-in-differences estimation based on the matched sample. Its regression model was:

$$TFP_{it} = \alpha_{it} + \beta_{ijt} \times Highway\ accessibility_{ijt} + \sum \delta_{ikt} X_{ikt} + \varepsilon_{ijkt} \quad . \tag{6}$$

Results from the difference-in-differences estimation on the matched sample are presented in Table 7. We included the explanatory variable *Highway accessibility* in Model (1), and we included the explanatory variable *Highway accessibility* and all of the firm-level and region-level control variables in Model (2). Both of the coefficients of the explanatory variable *Highway accessibility* in these two models were significantly positive. Then, on the basis of Model (2), we further included all fixed effects in Model (3), and the positive relationship between *Highway accessibility* and *TFP* was still significantly positive at the 1% level. Model (1) through Model (3) used cluster-robust standard errors. As robustness checks, we clustered the standard errors of Model (4), Model (5), and Model (6) at the firm, county, and industry level, respectively. The results of Model (4), Model (5), and Model (6) were very similar to the result of Model (3), in which cluster-robust standard errors were adopted. The results from Model (3) through Model (6) indicated that the positive causal relationship between the explanatory variable *Highway accessibility* and the dependent variable *TFP* was robust using standard errors clustered at different levels.

{Insert Table 7 here}

We believed that the difference-in-differences estimation combining the propensity score matching technique was the cleanest estimation of the relationship between improved highway accessibility and firm productivity in this study, so we provided a "back of the envelope" calculation of the impact that improvement in highway accessibility had on promoting firm productivity, based on the estimation of Model (3) that is listed in Table 7, in which the explanatory variable *Highway accessibility*, all firm-level and region-level control variables, and all fixed effects were included. The resulting estimation suggests that a connection to the highway network was able to boost firm productivity by 0.043 on average, or approximately 0.74% of the sample mean[11]. According to the estimation of Fernald (1999), road construction contributed about 1.4% of the productivity growth per year in the United States during the period from 1953 through 1973, and about 0.4% during the period from 1973 through 1989. Comparing with the estimation of Fernald (1999), the resulting estimation of this study was within a reasonable range. Because a county's connection to highway networks benefited all firms located in that county, the positive causal effect of highway accessibility improvement on firm productivity was significant and nonnegligible.

6. HETEROGENEOUS EFFECT OF HIGHWAYS ACROSS INDUSTRIES

As industries have heterogeneous characteristics, transport infrastructure is likely to have a totally differential impact across industries (Duranton et al. 2014; Faggio et al. 2017). From the perspective of products, firms in some industries such as machinery manufacturing industries and transportation equipment industries produce durable goods, while firms in some other industries

[11] According to the empirical results in Table 5, firms in the treated group are less productive before treatment, which is consistent with the findings that the estimation employing the method combining propensity score matching technique with the difference-in-differences approach has the largest estimate in this study. This consistency adds to the credibility of the empirical results in this study to some extent.

such as food & beverage manufacturing industries mainly produce non-durable goods. Therefore, in order to investigate the heterogeneous effect of highway accessibility improvement on firm productivity in different manufacturing industries, we classified manufacturing industries on the basis of the characteristics of their products and divided manufacturing industries into industries producing durable goods and industries producing non-durable goods, following the classification of Cremers et al. (2008). Details of the classification are shown in Table 8. We defined a new dummy variable *Industry_D* to distinguish firms in industries producing durable goods from firms in industries producing non-durable goods[12].

{Insert Table 8 here}

Then, we researched on the heterogeneous effect of highway accessibility improvement on firm productivity across industries, employing the method combining propensity score matching technique with the difference-in-differences approach, on the basis of the empirical analysis presented in Section 5.3. Table 9 presents the empirical results. In Model (1) that is listed in Table 9, we added the interaction term between *Highway accessibility* and *Industry_D* based on Model (3) from Table 7 to examine the industrial heterogeneity of highway accessibility improvement's impact. Moreover, in order to mitigate the problem of omitted variables in the interaction term, we followed the instruction of Balli and Sørensen (2013) and additionally included the interaction term between the explanatory variable *Highway accessibility* and the control variable *GDP*. The result showed that the coefficient of the interaction term was negative but not significant, indicating that industries producing durable goods were likely to benefit less from highway accessibility improvement. In order to further investigate the heterogeneous effect of highway accessibility improvement on firm productivity across industries, we divided the whole sample into two sub-samples according to the classification in Table 8. In Model (2) and Model (3) that is listed in Table 9, we examined the effect of highway accessibility improvement on firm productivity in industries producing durable goods and industries producing non-durable goods, respectively. The results showed that the coefficient of the explanatory variable *Highway accessibility* in Model (2) was 0.010, while the coefficient of *Highway accessibility* in Model (3)

[12] The variable *Industry_D* equals 1 if a firm is in a industry producing durable goods. Otherwise, it equals 0.

was 0.055, around 4.5 times larger than that in Model (2). Moreover, the coefficient of *Highway accessibility* in Model (2) was insignificant, while the coefficient of *Highway accessibility* in Model (3) was significant at the 1% level. The results presented in Table 9 provided some evidence supporting that the effect of highway accessibility improvement on firm productivity was stronger in industries producing non-durable goods.

{Insert Table 9 here}

7. CHANNEL: THE ROLE PLAYED BY INNOVATION

Section 5 presented empirical evidence indicating that an improvement in highway accessibility led to productivity growth in firms connected to the highway network. However, what was the mechanism linking the improvement in highway accessibility and firm productivity? This was a question worth considering and examining.

According to Melitz and Trefler (2012), the enlargement of the market was able to raise the returns on firms' innovation, thereby generating incentives for them to innovate and then increasing their productivity. Because an improvement in the quantity and quality of transport infrastructure could be an effective way to eliminate trade barriers and promote market integration (Faber, 2014), it seemed possible that an improvement in highway accessibility could increase firm productivity by bolstering the innovation performance of firms. In order to examine the role played by firm innovation, we defined firm innovation as the logarithm of a firm's new product output value[13] and made some additional tests employing the method combining propensity score matching technique with the difference-in-differences approach, on the basis of the empirical analysis presented in Section 5.3.

Table 10 lists the results of the additional tests. In the version of Model (1) that is listed in Table 10, we examined the impact of highway accessibility on firms' sales[14]. The result showed that the coefficient of the explanatory variable *Highway accessibility* was positively significant,

[13] In the matched sample, the new product output values of 40,580 observations (about 91% of the full sample) equaled to 0. In order to avoid the loss of observations, we defined the variable *Innovation* as: $Innovation_{it}$=log(new product output$_{it}$+1).

[14] The variable *Sales* was defined as the logarithm of a firm's sales.

indicating that a connection to the highway network increased a firm's sales by 5% on average. The estimation in Model (1) provided some evidence supporting that an improvement in highway accessibility was able to enlarge the size of the market and promote market integration. In Model (2) that is listed in Table 10, on the basis of Model (3) from Table 7, we added the interaction term between *Highway accessibility* and *Innovation* to examine the role played by innovation. The interaction term between *Highway accessibility* and *GDP* was also included in the model to mitigate the problem of omitted variables in the interaction term. The result showed that the coefficient of the explanatory variable *Highway accessibility* was still positive but not significant any more, and the coefficient of the interaction term between *Highway accessibility* and *Innovation* was positively significant at the 20% level, indicating that the positive influence that improvement in highway accessibility exerted on firm productivity was stronger if a firm had a better innovation performance. As a robustness check, we also used a dummy variable for firm innovation to replace the logarithm of the new product output value of a firm in Model (3)[15], and the significance of the explanatory variable and the interaction term were still consistent with that in Model (2). The findings in Table 10 provide some evidence confirming that highway accessibility improvement was able to increase firm productivity by improving the firms' innovation performance.

{Insert Table 10 here}

8. CONCLUSIONS

The literature has had two main research gaps. First, although issues about the economic spillovers from transport infrastructure have been discussed and debated for decades, there is still a great deal of controversy concerning the direction and magnitude of the economic effects of transport infrastructure (Chandra and Thompson, 2000; Lakshmanan, 2011). Second, previous studies in this field have focused mainly on developed economies, and the economic effects of transport infrastructure in emerging economies are still insufficient. In order to help fill those gaps, this study employed a method that combined the difference-in-differences approach with

[15] If the new product output value of a firm was greater than 0, *Innovation* was coded as 1; otherwise, it was equal to 0.

the propensity scoring matching technique and used a sample of Chinese manufacturing firms for the period from 1998 through 2007 to investigate the spillovers from transport infrastructure improvement onto firm productivity. With highways taken as a typical example of transport infrastructure, the results suggested that firms with a connection to the highway network experienced an average boost in firm productivity of 0.043, or approximately 0.74% of the sample mean. The effects of improvements in transport infrastructure on firm productivity were stronger in industries producing non-durable goods. Moreover, this study provided evidence that the effect of improved transport infrastructure on firm productivity was driven by an increase in firm innovation. Thus, the findings of this study contribute to a reconciliation of the controversy concerning the economic effects of transport infrastructure and also enrich the empirical evidence of that effect in emerging economies.

It should be noted that our study had several limitations. First, the study focused only on economic spillovers as a consequence of access to highways. The fact that many other types of transport infrastructure exist (such as railways, airports, long-span bridges, and tunnels), and the probability that the direction and magnitude of different types of transport infrastructures' economic spillovers are likely to be totally different (Melo et al., 2013), should not be ignored. In future research, it would be interesting and meaningful to compare the different economic spillovers from different types of transport infrastructure in a given period and then develop a mechanism to explain why, based on the different characteristics of the various types of transport infrastructure, those effects happened.

Second, this study did not consider the dynamic effects of transport infrastructure on firm productivity. The findings confirmed that an improvement in transport infrastructure had a short-term positive economic effect, but what would be the long-term relationship between an improvement in transport infrastructure and firm productivity? Would the positive economic effects of infrastructure continue to decline as soon as new infrastructure was completed, or would the effect reach its peak several years after completion? When, precisely, would the effect reach its peak? When would the effect disappear? These questions should be answered by future studies.

References

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica, 83*(6), 2411–2451.

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

Asian Development Bank. (2017). *Meeting Asia's infrastructure needs*. Asian Development Bank, Manila.

Aw, B. Y., Roberts, M. J., & Xu, D. Y. (2008). R&D investments, exporting, and the evolution of firm productivity. *American Economic Review*, *98*(2), 451–456.

Balli, H. O., & Sørensen, B. E. (2013). Interaction effects in econometrics. *Empirical Economics*, *45*(1), 583–603.

Berg, A., Buffie, E. F., Pattillo, C., Portillo, R., Presbitero, A. F., & Zanna, L.-F. (2019). Some misconceptions about public investment efficiency and growth. *Economica, 86*(342), 409–430.

Caliendo, M., & Kopeinig, S. (2008), Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22*(1), 31–72.

Chandra, A., & Thompson, E. (2000). Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system. *Regional Science and Urban Economics*, *30*(4), 457–490.

Chang, Y. B., & Gurbaxani, V. (2012). Information technology outsourcing, knowledge transfer, and firm productivity: An empirical analysis. *MIS Quarterly*, *36*(4), 1043–1063.

Cremers, K. M., Nair, V. B., & Peyer, U. (2008). Takeover defenses and competition: the role of stakeholders. *Journal of Empirical Legal Studies*, *5*(4), 791–818.

Datta, S. (2012). The impact of improved highways on Indian firms. *Journal of*

Development Economics, 99(1), 46–57.

Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies, *Review of Economics and Statistics, 84*(1), 151–161.

Demmel, M. C., Máñez, J. A., Rochina-Barrachina, M. E., & Sanchis-Llopis, J. A. (2017). Product and process innovation and total factor productivity: Evidence for manufacturing in four Latin American countries. *Review of Development Economics, 21*(4), 1341–1363.

Diaz, M. A., & Sánchez, R. (2008). Firm size and productivity in Spain: A stochastic frontier analysis. *Small Business Economics*, *30*(3), 315–323.

Duncan, T. (2007). *Retrospective analysis of the road sector, 1997-2005*. Operations Evaluation Department, Asian Development Bank, Manila.

Duranton, G., Morrow, P. M., & Turner, M. A. (2014). Roads and Trade: Evidence from the US. *Review of Economic Studies*, *81*(2), 681–724.

Duvall, T., Green, A., & Kerlin, M. (2015). *Making the most of a wealth of infrastructure finance*. McKinsey & Company, Washington, D.C.

Escribano, Á., Guasch, J. L., & Pena, J. (2010). Assessing the impact of infrastructure quality on firm productivity in Africa: Cross-country comparisons based on investment climate surveys from 1999 to 2005. *World Bank Policy Research Working Paper*, No. 5191. World Bank, Washington, D.C.

Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies, 81*(3), 1046–1070.

Faggio, G., Silva, O., & Strange, W. C. (2017). Heterogeneous agglomeration. *Review of Economics and Statistics*, *99*(1), 80–94.

Fernald, J. G. (1999). Roads to prosperity? Assessing the link between public capital and productivity. *American Economic Review, 89*(3), 619–638.

Hou, Q., & Li, S. M. (2011). Transport infrastructure development and changing spatial accessibility in the Greater Pearl River Delta, China, 1990–2020. *Journal of Transport Geography, 19*(6), 1350–1360.

Javorcik, B. S. (2004). Does foreign direct investment increase the productivity of domestic firms? In search of spillovers through backward linkages. *American Economic Review, 94*(3), 605–627.

Javorcik, B., & Poelhekke, S. (2017). Former foreign affiliates: Cast out and outperformed? *Journal of the European Economic Association, 15*(3), 501–539.

Jedwab, R., & Moradi, A. (2011). Transportation infrastructure and development in Ghana. *PSE Working Paper*, No. 2011–24, Paris School of Economics, Paris.

Lakshmanan, T. R. (2011). The broader economic consequences of transport infrastructure investments. *Journal of Transport Geography*, *19*(1), 1–12.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies, 70*(2), 317–341.

Li, S. M., & Shum, Y. M. (2001). Impacts of the national trunk highway system on accessibility in China. *Journal of Transport Geography, 9*(1), 39–48.

Luong, T. A. (2013). Does learning by exporting happen? Evidence from the automobile industry in China. *Review of Development Economics, 17*(3), 461–473.

Melitz, M. J., & Trefler, D. (2012). Gains from trade when firms matter. *Journal of Economic Perspectives, 26*(2), 91–118.

Melo, P. C., Graham, D. J., & Brage-Ardao, R. (2013). The productivity of transport infrastructure investment: A meta-analysis of empirical evidence. *Regional*

Science and Urban Economics, *43*(5), 695–706.

Moller, L. C., & Wacker, K. M. (2017). Explaining Ethiopia's growth acceleration—The role of infrastructure and macroeconomic policy. *World Development, 96*, 198–215.

Morrison, C. J., & Schwartz, A. E. (1996). State infrastructure and productive performance. *American Economic Review, 86*(5), 1095–1111.

Olley, G. S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica, 64*(6), 1263–1297.

Qian, X., & Yaşar, M. (2016). Export market diversification and firm productivity: Evidence from a large developing country. *World Development*, *82*, 28–47.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician, 39*(1), 33–38.

Shi, H., & Huang, S. (2014). How much infrastructure is too much? A new approach and evidence from China. *World Development*, *56*, 272–286.

Shiferaw, A., Söderbom, M., Siba, E., & Alemu, G. (2015). Road infrastructure and enterprise dynamics in Ethiopia. *Journal of Development Studies, 51*(11), 1541–1558.

Soderbom, M. & Teal, F. (2001). *Firm size and human capital as determinants of productivity and earnings*. United Nations Industrial Development Organization, Geneva.

Song, Z., Storesletten, K., & Zilibotti, F. (2011). Growing like China. *American Economic Review*, *101*(1), 196–233.

Wang, X., Xie, Z., Zhang, X., & Huang, Y. (2018). Roads to innovation: Firm-level evidence from People's Republic of China (PRC). *China Economic Review, 49*, 154–170.

World Bank (1994). The World Bank annual report, 1994. World Bank, Washington, D.C.

World Bank. (2007). China study tour by New Delhi transport office team. World Bank. Washington, D.C.

Table 1. Descriptive statistics of *TFP* for two-digit industries

Manufacturing industries	*N*	Capital	Labor	*TFP*				
				Min	Medium	Max	Mean	S.D.
Food processing	44329	0.143***	0.089***	-1.129	6.050	11.239	6.009	1.293
Food manufacturing	18898	0.169***	0.103***	-1.725	5.767	10.914	5.681	1.448
Beverage manufacturing	13164	0.149***	0.280***	-0.749	5.515	10.343	5.503	1.334
Tobacco processing	1079	0.374***	0.120	1.810	6.940	11.309	7.053	1.711
Textiles	65723	0.123***	0.264***	0.213	5.837	10.937	5.875	0.902
Garments and other fiber products	39314	0.150***	0.174***	0.120	5.981	11.065	6.030	0.878
Leather, furs, and down	17690	0.115***	0.406***	0.690	5.442	9.576	5.503	0.870
Timber processing	13034	0.177***	0.173***	0.062	6.185	10.447	6.180	1.058
Furniture manufacturing	8137	0.157***	0.717	1.340	7.314	11.620	7.360	1.113
Paper making and paper products	24586	0.119***	0.261***	0.686	5.915	11.019	5.960	0.998
Print and record medium reproduction	17345	0.148***	0.239***	-0.606	5.524	9.912	5.367	1.138
Stationery, educational, and sport goods	10635	0.126***	0.106***	0.485	6.180	10.595	6.223	0.991
Petroleum processing and coking product	3195	0.185***	0.142**	-0.137	5.312	11.192	5.312	1.363
Chemicals	55113	0.144***	0.248***	-0.246	5.715	11.387	5.752	1.045
Medical and pharmaceutical products	18034	0.176***	0.236***	-0.036	5.397	10.501	5.403	1.183
Chemical fiber manufacturing	3697	0.115***	0.305	0.810	5.863	10.095	5.900	1.037
Rubber products	9575	0.156***	0.295***	1.169	5.454	9.555	5.510	0.932
Plastic products	36350	0.158***	0.201***	0.562	6.291	10.894	6.327	0.969
Non-metal mineral products	69010	0.112***	0.250***	0.437	6.089	10.635	6.133	1.093
Smelting and pressing of ferrous metals	15147	0.138***	0.156***	1.361	5.948	11.513	6.037	1.275
Smelting and pressing of non-ferrous metals	6567	0.140***	0.345	0.582	5.957	10.948	5.933	1.187
Metal products	40819	0.157***	0.203***	0.826	5.873	10.831	5.939	0.944
Machinery manufacturing	59188	0.155***	0.362***	-2.533	4.946	9.773	4.977	0.952
Special equipment	32304	0.161***	0.267***	-0.306	5.461	10.803	5.449	1.130
Transportation equipment	38018	0.173***	0.223***	-0.953	5.658	12.318	5.691	1.205
Electric equipment	43768	0.193***	0.186***	0.304	5.878	11.677	5.925	1.097
Electronic and telecommunications	24003	0.186***	0.197***	1.061	6.500	12.908	6.612	1.267
Instrument meters and cultural machinery	10831	0.147***	0.387***	0.195	5.481	10.227	5.506	1.073
Crafts and other manufacture	11969	0.120***	0.268***	0.782	5.670	9.889	5.701	0.916

Note:

***, **, and * denote the significance of *p* values at the 1%, 5%, and 10% level, respectively.

Table 2. Definitions and data sources of variables

Variable	Definition	Data source
Dependent variable		
TFP	Estimated using the ACF method. See Section 4.2.1 for details	Database of Annual Survey of Industrial Firms
Explanatory variable		
Highway accessibility	Dummy variable. It equals 1 if a firm *i* is located in a county that has been connected to the National Trunk Highway System in the year *t*. Otherwise, it equals 0.	Manually collected on the China Road Atlas and the official government websites
Control variables		
Firm level		
Size	Logarithm of the number of full-time employees of a firm	Database of Annual Survey of Industrial Firms
Age	Logarithm of the number of years since the founding of a firm	Database of Annual Survey of Industrial Firms
Leverage	Ratio of a firm's total liabilities to its tangible assets	Database of Annual Survey of Industrial Firms
Fixed assets	Logarithm of the fixed assets of a firm	Database of Annual Survey of Industrial Firms
Export intensity	Percentage of export in total output of a firm	Database of Annual Survey of Industrial Firms
Region level		
GDP	Logarithm of the GDP of a prefecture-level administrative district	China City Statistical Yearbooks
Population density	Resident population of a prefecture-level administrative district divided by its land area	China City Statistical Yearbooks
FDI	Logarithm of the foreign direct investment	China City Statistical Yearbooks
Secondary industry	Ratio of employees of secondary industry to all employees in a prefecture-level administrative district	China City Statistical Yearbooks
Tertiary industry	Ratio of employees of tertiary industry to all employees in a prefecture-level administrative district	China City Statistical Yearbooks
Railway freight volume	Logarithm of the amount of freight transported by railways in a year	China City Statistical Yearbooks
Road freight volume	Logarithm of the amount of freight transported by roads in a year	China City Statistical Yearbooks

Table 3. Means, standard deviations, and Pearson correlation matrix of variables

Variable	Observations	Mean	S.D.	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. *TFP*	1,100,092	5.82	1.150	1													
2. *Highway accessibility*	1,102,930	0.400	0.490	0.131***	1												
3. *Size*	1,102,930	4.810	1.111	0.326***	-0.035***	1											
4. *Age*	1,102,930	2.014	0.824	-0.193***	-0.055***	0.245***	1										
5. *Leverage*	1,102,644	0.700	10.166	-0.013***	-0.001	0.004***	0.004***	1									
6. *Fixed assets*	1,100,092	8.410	1.563	0.380***	0.008***	0.640***	0.148***	-0.015***	1								
7. *Export intensity*	1,102,930	0.161	0.331	0.086***	0.066***	0.189***	-0.021***	0.000	0.004***	1							
8. *GDP*	1,040,678	6.509	0.904	0.232***	0.295***	-0.068***	-0.073***	-0.003***	-0.008***	0.133***	1						
9. *Population density*	1,036,208	580.878	337.058	0.064***	0.159***	-0.015***	0.009***	-0.001	-0.030***	0.103***	0.394***	1					
10. *FDI*	1,022,404	9.850	1.891	0.178***	0.264***	-0.050***	-0.072***	-0.002	-0.018***	0.206***	0.800***	0.353***	1				
11. *Secondary industry*	1,036,208	47.850	12.889	0.109***	0.205***	0.006***	-0.022***	-0.001	0.021***	0.114***	0.546***	0.166***	0.519***	1			
12. *Tertiary industry*	1,036,208	48.838	11.800	-0.066***	-0.143***	-0.023***	0.006***	0.000	-0.030***	-0.103***	-0.404***	-0.079***	-0.441***	-0.835***	1		
13. *Railway freight volume*	885,636	5.621	1.255	0.024***	-0.006***	0.055***	0.016***	0.000	0.076***	-0.117***	-0.091***	-0.223***	-0.169***	0.042***	-0.058***	1	
14. *Road freight volume*	1,035,479	8.302	0.707	0.146***	0.207***	-0.039***	-0.042***	-0.002	0.005***	0.057***	0.706***	0.132***	0.495***	0.381***	-0.271***	0.198***	1

Note:

***, **, and * denote the significance of p values at the 1%, 5%, and 10% level, respectively.

Table 4. OLS estimations

Variable	Panel (A)		Panel (B)	
	Model (1)	Model (2)	Model (3)	Model (4)
	Dependent variable: *TFP*			
Highway accessibility	0.136***	0.008***	0.189***	0.011***
	(62.76)	(3.84)	(76.13)	(4.61)
Size	0.234***	0.311***	0.241***	0.314***
	(167.33)	(237.95)	(156.97)	(216.34)
Age	-0.356***	-0.235***	-0.360***	-0.235***
	(-260.16)	(-192.57)	(-241.67)	(-175.94)
Leverage	-0.007	-0.002	-0.006	-0.001
	(-1.63)	(-1.27)	(-1.61)	(-1.24)
Fixed assets	0.200***	0.162***	0.194***	0.158***
	(209.35)	(183.34)	(184.75)	(160.76)
Export intensity	0.025***	-0.010***	0.009**	-0.000
	(7.59)	(-3.11)	(2.50)	(-0.09)
GDP	0.406***	0.373***	0.415***	0.382***
	(142.55)	(39.05)	(135.07)	(37.24)
Population density	-0.000***	-0.000***	-0.000***	-0.000***
	(-15.67)	(-4.35)	(-13.03)	(-3.65)
FDI	-0.011***	0.009***	0.001	0.007***
	(-10.45)	(6.05)	(1.26)	(4.65)
Secondary industry	-0.001***	-0.005***	-0.002***	-0.005***
	(-5.32)	(-23.66)	(-13.07)	(-25.18)
Tertiary industry	0.006***	0.005***	0.006***	0.005***
	(36.72)	(23.82)	(33.69)	(21.33)
Railway freight volume	0.031***	0.008***	0.036***	0.015***
	(34.39)	(4.06)	(37.12)	(7.19)
Road freight volume	-0.098***	0.010**	-0.106***	0.008*
	(-40.28)	(2.17)	(-41.24)	(1.65)
Constant	1.577***	1.997***	1.539***	1.963***
	(76.41)	(4.16)	(71.30)	(3.96)
Industry-year fixed effects	No	Yes	No	Yes
Ownership fixed effects	No	Yes	No	Yes
Region fixed effects	No	Yes	No	Yes
Sample size	874556	874555	718049	718049
R-squared	0.287	0.541	0.289	0.543

Notes:

1. ***, **, and * denote the significance of *p* values at the 1%, 5%, and 10% level, respectively.

2. *t*-statistics are in parentheses.

3. All regressions in the table use cluster-robust standard errors.

Table 5. Difference-in differences estimation: Unmatched sample

Variable	Model (1)	Model (2)	Model (3)
	Dependent variable: *TFP*		
Highway accessibility	0.398***	0.013***	0.012***
	(99.64)	(2.93)	(2.70)
Treat	0.044***	-0.028***	-0.015
	(5.39)	(-2.86)	(-1.59)
Size		0.156***	0.157***
		(46.19)	(47.24)
Age		0.150***	0.102***
		(26.24)	(17.10)
Leverage		0.000	0.000
		(0.33)	(0.40)
Fixed assets		0.030***	0.027***
		(14.41)	(13.31)
Export intensity		-0.000	0.010
		(-0.02)	(1.57)
GDP		0.626***	0.326***
		(78.15)	(21.83)
Population density		-0.000	-0.000***
		(-1.35)	(-3.81)
FDI		-0.000	0.006***
		(-0.10)	(3.77)
Secondary industry		-0.002***	-0.003***
		(-11.03)	(-12.38)
Tertiary industry		0.001***	0.004***
		(3.09)	(17.01)
Railway freight volume		0.016***	0.019***
		(8.41)	(10.10)
Road freight volume		0.034***	0.012**
		(6.66)	(2.50)
Constant	5.670***	0.238***	2.523***
	(1607.37)	(4.57)	(20.04)
Industry-year fixed effects	No	No	Yes
Ownership fixed effects	No	No	Yes
Region fixed effects	No	No	Yes
Sample size	904190	718049	718049
R-squared	0.033	0.174	0.240

Notes:

1. ***, **, and * denote the significance of *p* values at the 1%, 5%, and 10% level, respectively.

2. *t*-statistics are in parentheses.

3. All regressions in the table use cluster-robust standard errors.

Table 6. Summary statistics for the full sample and the matched sample

Variable	Panel (A) Unmatched sample (N=702,085)				Panel (B) Matched sample (N=55,824)			
	Mean		t-test	p-value	Mean		t-test	p-value
	Treated	Control			Treated	Control		
$Size_{t-1}$	4.853	4.940	12.814	0.000***	4.852	4.855	-0.410	0.681
$Size_{t-1}^2$	24.669	25.667	13.859	0.000***	24.657	24.685	-0.300	0.762
$Size_{t-1}^3$	131.157	139.896	14.163	0.000***	131.100	131.280	-0.220	0.824
Age_{t-1}	1.992	2.072	15.469	0.000***	1.994	1.988	0.760	0.445
Age_{t-1}^2	4.644	5.010	18.151	0.000***	4.657	4.641	0.590	0.556
Age_{t-1}^3	11.789	13.098	18.896	0.000***	11.851	11.812	0.420	0.673
$Output_{t-1}$	10.944	11.044	2.295	0.011**	10.953	10.944	0.180	0.854
ROE_{t-1}	0.088	0.075	-8.152	0.000***	0.088	0.087	0.250	0.806
$Leverage_{t-1}$	0.662	0.701	2.356	0.009***	0.663	0.666	-0.680	0.496
$Export\ intensity_{t-1}$	0.168	0.149	-9.560	0.000***	0.166	0.166	0.010	0.989

Note:

***, **, and * denote the significance of p values at the 1%, 5%, and 10% level, respectively.

Table 7. Difference-in-differences estimation: Matched sample

Variables	Model (1)	Model (2)	Model (3)	Model (4)	Model (5)	Model (6)
	Dependent variable: *TFP*					
Highway accessibility	0.245***	0.048***	0.044***	0.044***	0.043**	0.044*
	(23.59)	(3.30)	(2.70)	(2.70)	(2.49)	(1.86)
Size		0.132***	0.095***	0.095***	0.095***	0.095***
		(5.28)	(3.75)	(3.75)	(3.29)	(2.59)
Age		-0.004	0.081	0.081	0.080	0.081
		(-0.06)	(1.22)	(1.22)	(1.11)	(1.17)
Leverage		-0.008	0.012	0.012	0.010	0.012
		(-0.38)	(0.55)	(0.55)	(0.49)	(0.54)
Fixed assets		0.034**	0.049***	0.049***	0.049***	0.049***
		(2.24)	(3.09)	(3.09)	(3.03)	(2.68)
Export intensity		-0.101**	-0.081*	-0.081*	-0.081*	-0.081
		(-2.15)	(-1.74)	(-1.74)	(-1.67)	(-1.49)
GDP		0.633***	0.270**	0.270**	0.290**	0.270*
		(10.54)	(2.09)	(2.09)	(2.11)	(1.89)
Population density		0.000*	0.000	0.000	0.000	0.000
		(1.79)	(0.08)	(0.08)	(0.11)	(0.08)
FDI		0.013	0.020	0.020	0.019	0.020
		(0.94)	(1.28)	(1.28)	(1.20)	(1.20)
Secondary industry		0.001	-0.005	-0.005	-0.005	-0.005
		(0.29)	(-1.21)	(-1.21)	(-1.26)	(-1.10)
Tertiary industry		0.008***	0.003	0.003	0.003	0.003
		(2.88)	(0.77)	(0.77)	(0.79)	(0.64)
Railway freight volume		-0.001	0.004	0.004	0.004	0.004
		(-0.12)	(0.31)	(0.31)	(0.30)	(0.29)
Road freight volume		0.047	0.085*	0.085*	0.091*	0.085*
		(1.10)	(1.79)	(1.79)	(1.74)	(1.74)
Constant	5.795***	-0.157	2.034*	2.178*	1.858*	2.034*
	(1116.20)	(-0.35)	(1.95)	(1.93)	(1.75)	(1.67)
Industry-year fixed effects	No	No	Yes	Yes	Yes	Yes
Ownership fixed effects	No	No	Yes	Yes	Yes	Yes
Region fixed effects	No	No	Yes	Yes	Yes	Yes
Sample size	55675	47043	47043	47043	46904	47043
R-squared	0.068	0.152	0.496	0.496	0.496	0.496

Notes:

1. ***, **, and * denote the significance of *p* values at the 1%, 5%, and 10% level, respectively.

2. *t*-statistics are in parentheses.

3. In this table, Model (1) through Model (3) use cluster-robust standard errors. Model (4), Model (5), and Model (6) use standard errors clustered at the firm, county, and industry level,

respectively.

Table 8. Classification of industries with durable and non-durable goods

Industries with durable goods	Industries with non-durable goods
Smelting and pressing of ferrous metals	Food processing
Smelting and pressing of non-ferrous metals	Food manufacturing
Metal products	Beverage manufacturing
Machinery manufacturing	Tobacco processing
Special equipment	Textiles
Transportation equipment	Garments and other fiber products
Electric equipment	Leather, furs, and down
Electronic and telecommunications	Timber processing
Instrument meters and cultural machinery	Furniture manufacturing
	Paper making and paper products
	Print and record medium reproduction
	Stationery, educational, and sport goods
	Petroleum processing and coking product
	Chemicals
	Medical and pharmaceutical products
	Chemical fiber manufacturing
	Rubber products
	Plastic products
	Non-metal mineral products

Table 9. The heterogeneous effect of highway accessibility improvement across industries

Variables	Model (1)	Model (2)	Model (3)
	Dependent variable: *TFP*		
	Whole sample	Industries producing	Industries producing
Highway accessibility	0.066	0.010	0.055***
	(0.52)	(0.30)	(3.04)
*Highway accessibility*Industry_D*	-0.027		
	(-0.72)		
Industry_D	0.000		
	(.)		
Size	0.095***	0.092*	0.091***
	(3.76)	(1.87)	(3.10)
Age	0.078	0.331**	-0.007
	(1.18)	(2.49)	(-0.10)
Leverage	0.012	-0.007	0.028
	(0.53)	(-0.19)	(1.24)
Fixed assets	0.049***	0.073**	0.043**
	(3.07)	(2.23)	(2.39)
Export intensity	-0.082*	-0.003	-0.107**
	(-1.74)	(-0.03)	(-2.08)
GDP	0.273**	0.390	0.262*
	(2.10)	(1.31)	(1.85)
*Highway accessibility*GDP*	-0.002		
	(-0.12)		
Population density	0.000	-0.001	0.000
	(0.06)	(-0.78)	(0.28)
FDI	0.021	0.051	0.007
	(1.29)	(1.48)	(0.42)
Railway freight volume	0.004	-0.008	0.008
	(0.31)	(-0.36)	(0.57)
Road freight volume	0.086*	0.209**	0.043
	(1.79)	(2.21)	(0.86)
Secondary industry	-0.005	-0.017*	-0.002
	(-1.21)	(-1.72)	(-0.50)
Tertiary industry	0.003	-0.010	0.007
	(0.75)	(-1.02)	(1.58)
Constant	1.956*	0.812	2.563**
	(1.87)	(0.35)	(2.28)
Industry-year fixed effects	Yes	Yes	Yes
Ownership fixed effects	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes
Sample size	47043	15558	31485
R-squared	0.496	0.520	0.490

Notes:

1. ***, **, and * denote the significance of *p* values at the 1%, 5%, and 10% level, respectively.

2. *t*-statistics are in parentheses.

3. All regressions in the table use cluster-robust standard errors.

Table 10. Channels: The role played by innovation

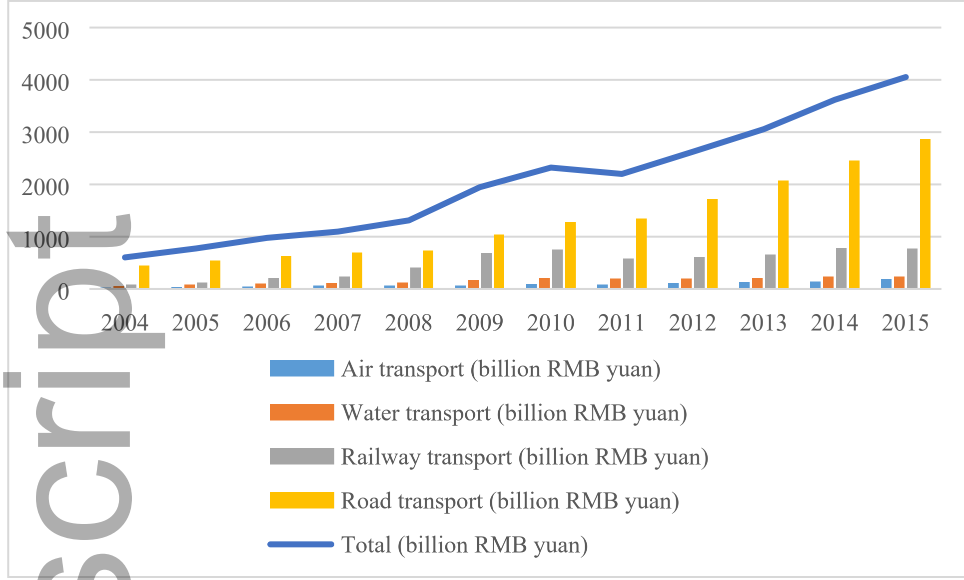
Variables	Model (1)	Model (2)	Model (3)
	Dependent variable: *Sales*	Dependent variable: *TFP*	
Highway accessibility	0.050***	0.038	0.043
	(4.09)	(0.30)	(0.34)
*Highway accessibility*Innovation*		0.007^{+}	0.063^{+}
		(1.34)	(1.30)
Innovation		0.005	0.027
		(1.45)	(0.88)
Size	0.345***	0.094***	0.095***
	(17.11)	(3.71)	(3.73)
Age	0.200***	0.091	0.088
	(3.62)	(1.38)	(1.34)
Leverage	0.020	0.011	0.012
	(1.47)	(0.51)	(0.53)
Fixed assets	0.185***	0.049***	0.049***
	(13.87)	(3.06)	(3.08)
Export intensity	-0.072*	-0.084*	-0.083*
	(-1.91)	(-1.80)	(-1.78)
GDP	0.235**	0.279**	0.275**
	(2.40)	(2.14)	(2.10)
*Highway accessibility*GDP*		0.000	-0.001
		(0.01)	(-0.03)
Population density	0.000	0.000	0.000
	(0.69)	(0.07)	(0.07)
FDI	0.016	0.019	0.019
	(1.28)	(1.17)	(1.19)
Railway freight volume	0.018*	0.004	0.004
	(1.88)	(0.31)	(0.29)
Road freight volume	0.055	0.085*	0.085*
	(1.57)	(1.76)	(1.77)
Secondary industry	-0.007*	-0.005	-0.005
	(-1.91)	(-1.19)	(-1.20)
Tertiary industry	-0.001	0.003	0.003
	(-0.21)	(0.77)	(0.75)
Constant	4.401***	2.108*	2.144*
	(5.62)	(1.85)	(1.88)
Industry-year fixed effects	Yes	Yes	Yes
Ownership fixed effects	Yes	Yes	Yes
Region fixed effects	Yes	Yes	Yes
Sample size	47043	47043	47043
R-squared	0.613	0.496	0.496

Notes:

1. ***, **, *, and $^{+}$ denote the significance of *p* values at the 1%, 5%, 10%, and 20% level, respectively.

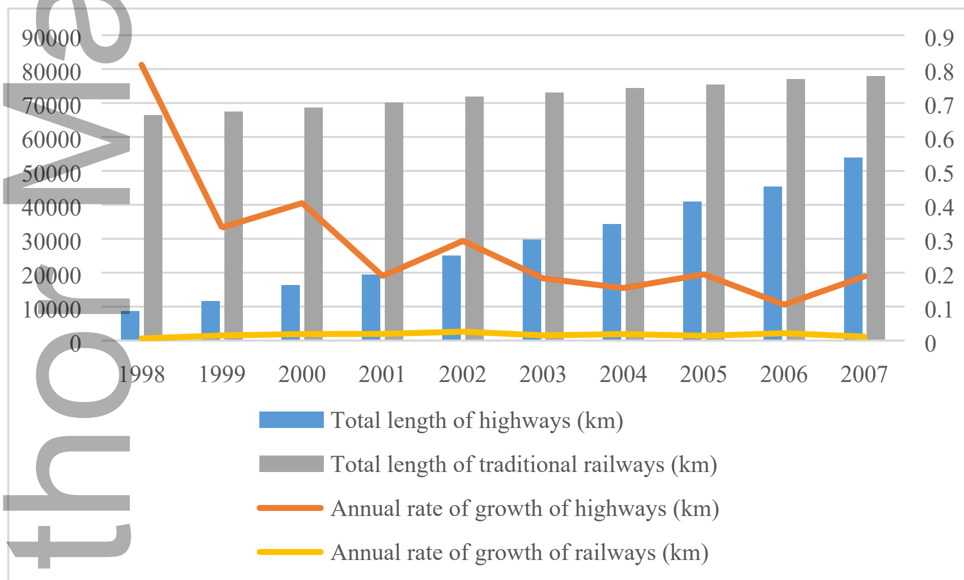
2. *t*-statistics are in parentheses.

3. All regressions in the table use cluster-robust standard errors.



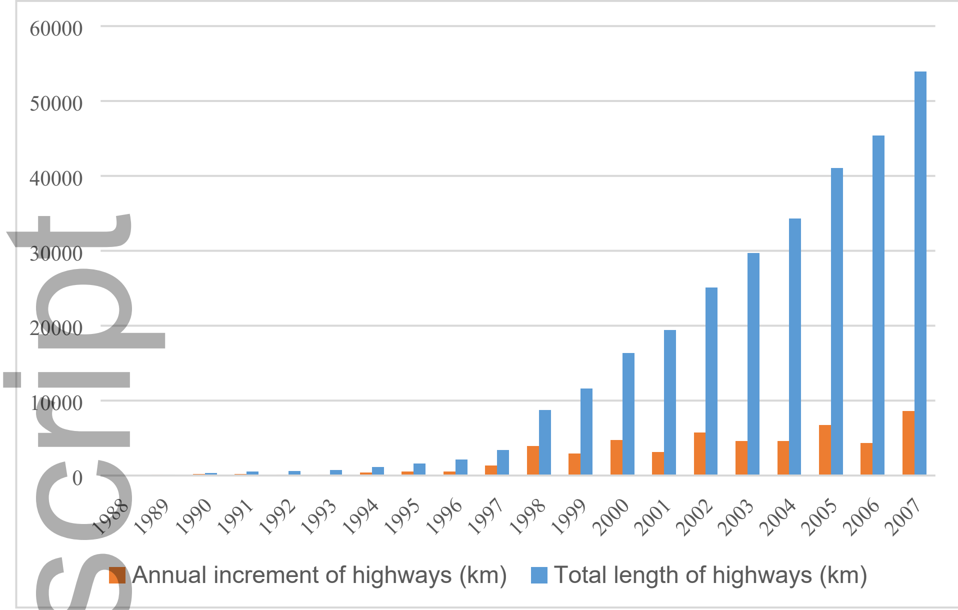
Data Source: CSMAR Database

Figure 1. Fixed-asset investment in transportation in China (2004-2015)



Data Source: China Transportation Yearbooks

Figure 2. Comparison between growth of highways and railways (1998-2007)



Data Source: China Transportation Yearbooks

Figure 3. The development of highways in China (1988-2007)

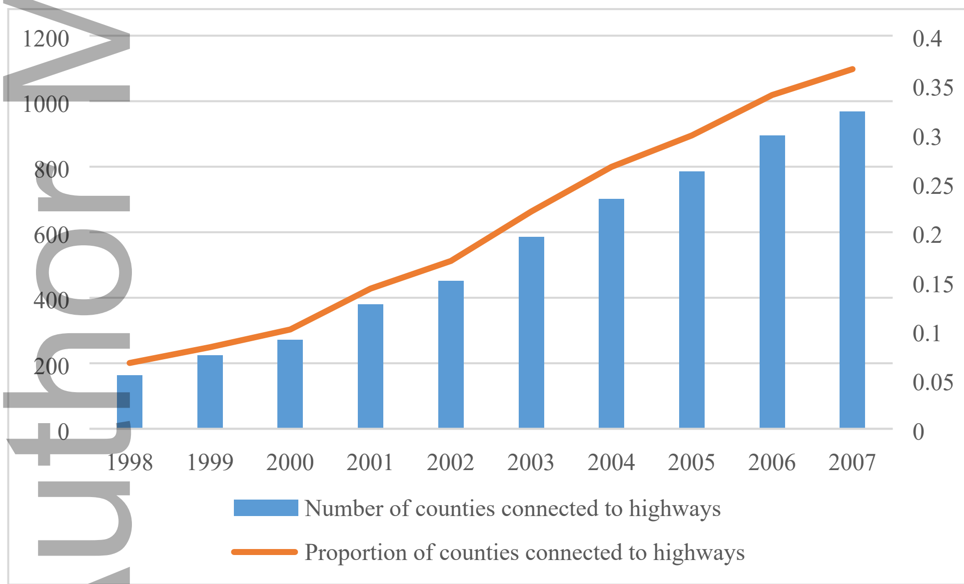
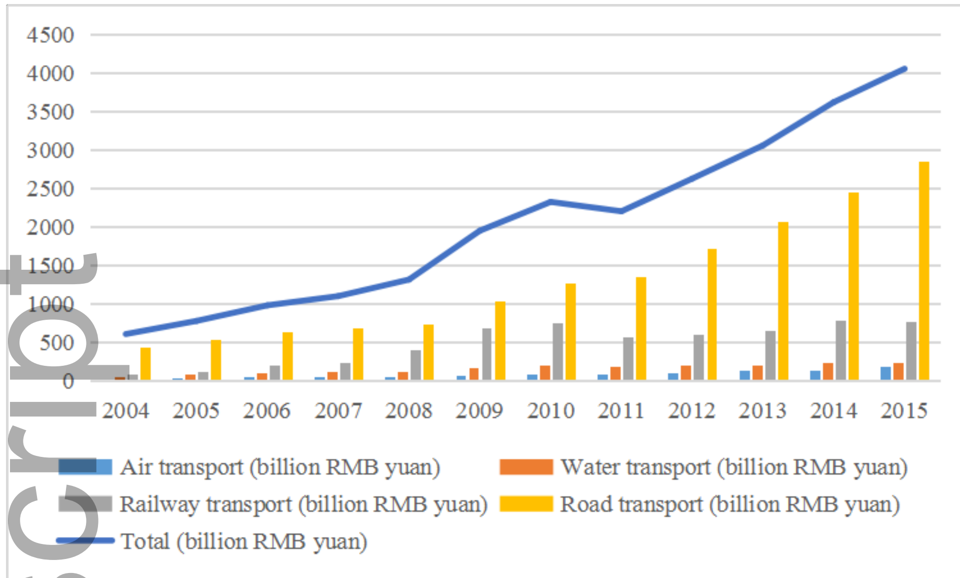
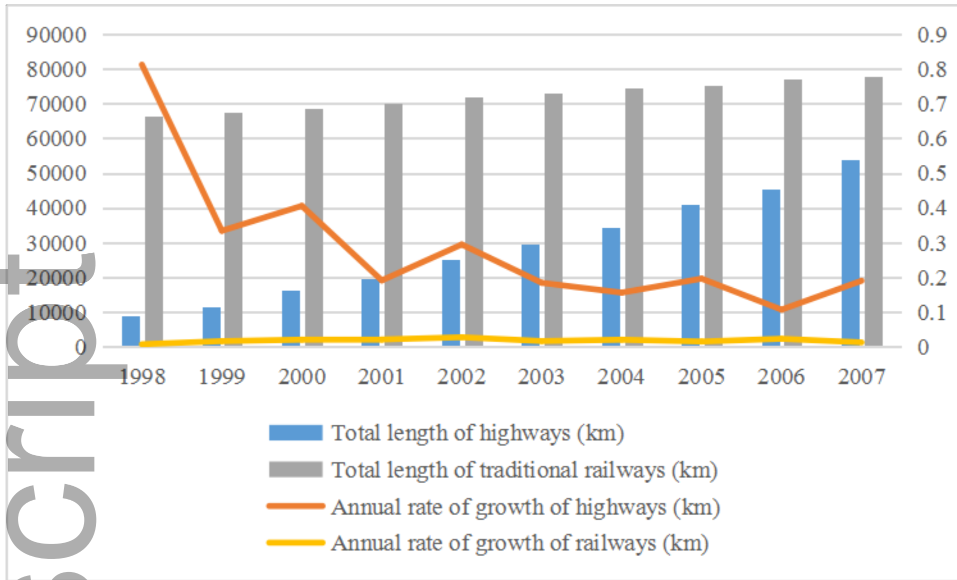


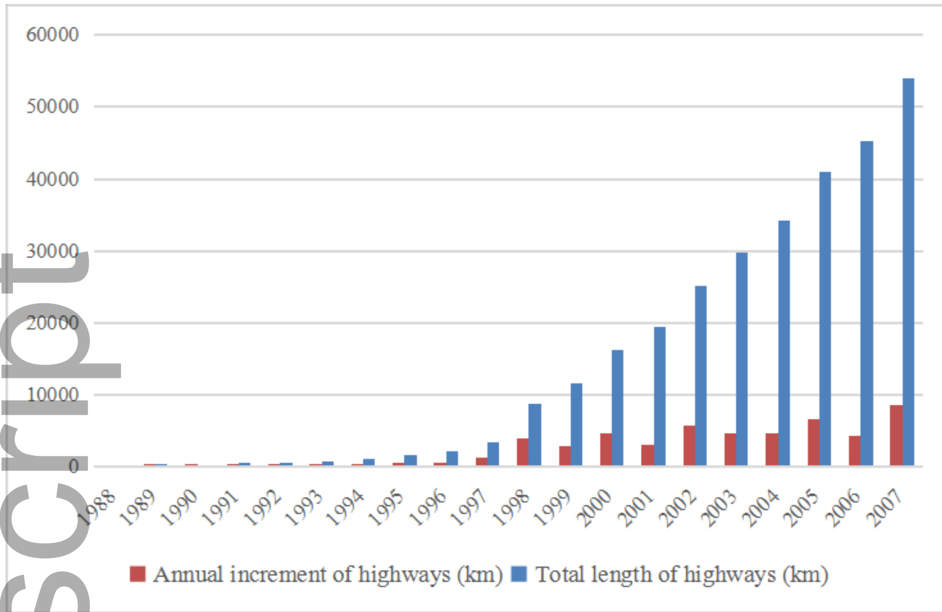
Figure 4. Changes in the proportion of counties connected to highways (1998-2007)



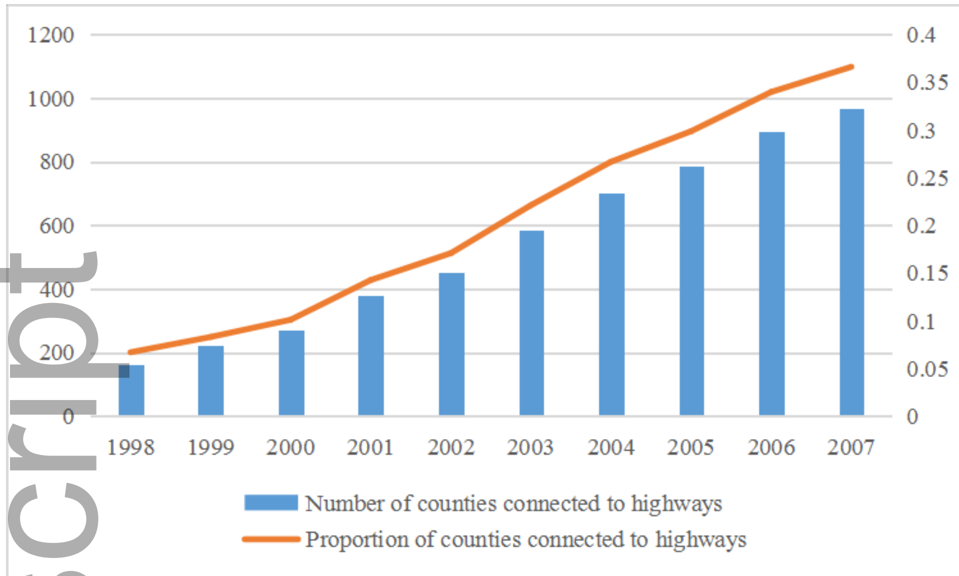
rode_12701_f1.tif



rode_12701_f2.tif



rode_12701_f3.tif



rode_12701_f4.tif