Productivity of Core Infrastructure Investment in China: An Input-Output Approach*

Zhifeng Wang, Guiying Laura Wu, Qu Feng

This version: July 2020

Abstract

This paper examines the productivity effect of two categories of core infrastructure investment in China, by matching a panel of manufacturing firm-level production data with province-level infrastructure investment data. Cross-industry variation in infrastructure reliance using input-output table information is employed to address potential endogeneity issues. We find that firms in an industry that relies more heavily on infrastructure in production experience higher productivity growth from more infrastructure investment. On average the annual rate of return to core infrastructure investment in China is about 23% during 1998 to 2007.

JEL Classification: D24, H54, O18, R15, R53

Key Words: Infrastructure, Productivity, Input-Output Table

*We would like to thank Shang-Jin Wei for his inspiring discussion to motivate this research, and two anonymous referees for their helpful suggestions. Financial support from the MOE AcRF Tier 1 Grant M4011642 at Nanyang Technological University is gratefully acknowledged. Wang is from Center for Economics, Finance and Management Studies, Hunan University, Changsha, Hunan Province, 410006, China. Email: WANG0869@e.ntu.edu.sg (Z. Wang). Wu and Feng are from Division of Economics, School of Social Sciences, Nanyang Technological University, Singapore, 639818. Emails: guiying.wu@ntu.edu.sg (G. Wu), qfeng@ntu.edu.sg (Q. Feng).

1 Introduction

Infrastructure investment has often been regarded as a precursor to economic development. During economic downturns, infrastructure investment is also advocated as an important tool to rejuvenate economy. Since later 2013, the Chinese government has proposed the One Road and One Belt initiative. In early 2018, Donald Trump announced that at least 1.5 trillion US dollar would be spent on repairing and upgrading America's infrastructure in the next ten years. These new waves of big infrastructure investment have raised a question concerning both academic researchers and policy makers for a long time: Is infrastructure investment productive?

Starting from Aschauer (1989), a vast literature has investigated this question using an infrastructure-augmented aggregate production function estimation. However, there is no consensus on the productivity effect of infrastructure using this traditional approach. A major identification challenge lies in the reverse causality between output and infrastructure.[1] The more recent literature makes significant progress on this challenge by focusing on specific types of infrastructure and using more disaggregated data. Such combination allows researchers to infer the causal effect of infrastructure using the instrumental variable approach.[2]

In this paper we estimate the productivity effect of core infrastructure investment in China using an alternative identification strategy. The core infrastructure investment includes two big categories of investment: (1) production and supply of electricity, gas, and water (*electricity* hereafter); and (2) transport, storage, and post (*transport* hereafter). Both of them account for a sizable share in the intermediate input of the manufacturing sector. During 1998 to 2007, the investment expenditures on core infrastructure have been growing at about 10% annually and account for more than 8% of GDP.

Similar to the traditional approach, we apply the production function estimation framework to directly link the core infrastructure to productivity and output. Different from the traditional literature, our basic setup is a firm-level production function estimation by matching a panel of Chinese manufacturing firms with the province-level

[1] Gramlich (1994) provides an excellent survey on the early literature and explains the key challenge arising from reverse causality. Wu et al. (2020) discuss other econometric issues in estimating the productivity effect of infrastructure using an aggregate production function.

[2] Among many others, some leading examples include the river gradient as an instrument for dam in Duflo and Pande (2007); the interaction between coal price and share of thermal generation as an instrument for electricity in Abeberese (2017); and the planned route IV, the historical route IV and the straight-line IV surveyed in Redding and Turner (2015) for roads, highways and railways.

core infrastructure. The fact the production decision is made at the firm-level while the infrastructure investment decision is made at the province-level seems to purge the canonical reverse causality in an aggregate production function estimation. However, several identification challenges remain. First, allocation of infrastructure might not be random. Second, there might be some omitted variables which simultaneously affect productivity and infrastructure. Finally, more productive firms tend to self-select into more productive provinces, which in turn tend to have more infrastructure.

To address these concerns, we first adopt an endogenous productivity process. It decomposes the actual firm-level productivity into the expected productivity and productivity shocks, and models the effect of infrastructure on the expected firm-level productivity through a first-order Markov process. The endogenous productivity process allows for a flexible way to model how the lagged firm-level productivity affects the future productivity. It thus mitigates the reverse causality arising from the self-selection mechanism. Second, to further control for the reverse causality arising from endogenous allocation of infrastructure and the omitted variable bias, we decompose the firm-level productivity shocks into a national-wide aggregate shock, a province-specific aggregate shock, an industry-specific aggregate shock and a firm-specific idiosyncratic shock. After proxying the aggregate components in the shocks using the combination of year, province and industry fixed-effects, the firm-specific idiosyncratic shock is assumed to be orthogonal to the province-level infrastructure investment, which provides the key identification condition.

The new challenge imposed in this strategy, however, lies in that the effect of province-level infrastructure will be absorbed into the province-year fixed effect. To resolve this non-identification problem, we interact the province-level infrastructure with the industry-specific infrastructure reliance constructed using the input-output table.[3] This allows us to identify the productivity effect of infrastructure by asking whether firms in an industry that relies more heavily on infrastructure indeed experience higher productivity growth from more infrastructure investment of a province.

We find that the core infrastructure investment significantly promotes the productivity of Chinese manufacturing firms over 1999-2007. On average firm's productivity increases by 0.058% annually due to infrastructure investment in *electricity*, and by

[3] For a long time, economists have explored the input-output table to allocate aggregate productivity growth to its sources at the level of individual industries (Jorgenson et al., 1987). More recently, researchers have documented enormous heterogeneity across producers in their engagement in input-output linkages (Oberfield, 2018).

0.096% annually due to infrastructure investment in *transport*. With some additional assumptions, we can translate the productivity effects into annual rates of return to core infrastructure investment in China: 20.3% and 25.9% for *electricity* and *transport*, respectively. The estimated rates of return provide an important measure of infrastructure investment efficiency.

Our paper is closely related to Fernald (1999), Li and Li (2013) and Li et al. (2017) in terms of methodology. Fernald (1999) is the first paper that explores cross-industry heterogeneity to estimate the productivity effect of road infrastructure. He applies the growth accounting framework to the industry-level data and interacts the national-level road infrastructure with the industry-level vehicle intensities. Li and Li (2013) and Li et al. (2017) extend this approach to the firm-level data and investigate the investment efficiency of road construction in China.

Our study shares the same core rationale in identification. However, we also make two marginal contributions. First, thanks to the input-output table, we are able to extend the cross-industry heterogeneity from road infrastructure only to a broader category of infrastructure, which is particularly useful in evaluating the efficiency of large scale and general infrastructure investment. Second, the outcome variables in both Fernald (1999) and Li et al. (2017) are the Solow residuals from a growth accounting framework. Such productivity measure is contaminated with unobservable unanticipated shocks and measurement errors in output and is obtained under some restrictive assumptions on production technology and productivity process. We use the control function approach proposed by Ackerberg et al. (2015) to estimate the production function and obtain a consistent estimation of productivity. We also model the dynamics of the productivity using a flexible endogenous productivity process to avoid potential model misspecification.

Our paper also complements the recent works that investigate the productivity effect of general infrastructure in China. Different from Shi and Huang (2014) and Feng and Wu (2018), which explore province-level production and infrastructure data only, the firm-level panel data setting in this paper allows us to control for unobserved factors across years, provinces and industries. Our paper is a close companion to Wu et al. (2020). Both papers model an endogenous productivity process with firm-level productivity and province-level infrastructure. However, the key sources of identification are different. In this paper different industries are pooled together so that the identification comes from the cross-industry heterogeneity on infrastructure reliance.

In Wu et al. (2020) the productivity effect is inferred industry by industry so that within each industry a constructed province-year specific aggregate productivity shock is employed in identification design.

The rest of the paper is organized as follows. Section 2 discusses the empirical strategies that we adopt to address the identification challenges. Section 3 introduces the data and the institutional background. We describe how to construct the industry-infrastructure reliance measures and to select a proper definition for core infrastructure investment under our identification strategy. Section 4 reports the empirical findings on the productivity effect of infrastructure investment. Section 5 concludes and discusses the limitations.

2 Empirical Model

2.1 Output and Productivity

To estimate the productivity effect of core infrastructure investment in China, we match the firm-level productivity from a panel of manufacturing firms with the province-level infrastructure investment. Following the large literature on productivity, such as Pavcnik (2002) and Brandt et al. (2017), we employ a two-step approach in this paper.

In the first step, consider a two-factor production function in logarithm form:

$$y_{it} = \beta_l l_{it} + \beta_k k_{it} + \omega_{isjt} + \epsilon_{it}, \tag{1}$$

where y_{it}, l_{it}, and k_{it} represent firm i's value added, labor and capital in year t; and ϵ_{it} represents the unobservable shocks to production or measurement errors in the data. Besides i and t, the unobservable firm-level productivity ω_{isjt} has two additional subscripts, which stand for industry s and province j, respectively. It represents productivity shocks that are potentially observed or predictable by firms when they make input decisions, including policy variables such as infrastructure investment. Due to the correlation between the unobserved productivity and factor inputs, the ordinary least squares estimators of elasticities β_l and β_k are inconsistent. We use the control function approach developed in Ackerberg et al. (2015) to estimate this production function. Value of intermediate inputs is used as a proxy for the unobserved productivity in our application.[4]

[4] We apply the Ackerberg et al. (2015) two-stage estimation procedure to estimate the production function industry by industry. In both the first and second stage regressions, we have included province dummies and year dummies to proxy all the potential policy variables, including core infrastructure investment.

In the second step, after obtaining a consistent measure of productivity ω_{isjt} from the first step, we use an endogenous productivity process to model the effect of infrastructure investment on productivity.[5] It explicitly allows infrastructure investment to impact the evolution of productivity through a first-order Markov process:

$$\omega_{isjt} = h_t(\omega_{isjt-1}, g_{jt-1}) + \xi_{isjt}. \tag{2}$$

Equation (2) decomposes the actual productivity ω_{isjt} into the expected productivity $h_t(\omega_{isjt-1}, g_{jt-1})$ and the random shocks ξ_{isjt}. The nonparametric function $h_t(\omega_{isjt-1}, g_{jt-1})$ has two arguments. The first argument ω_{isjt-1} is the lagged or attained productivity of firm i. The second argument g_{jt-1} is the logarithm of the core infrastructure investment flow in province j where firm i is located and in year $t-1$ when the investment is made. The time-to-build assumption implies that it takes time for the infrastructure investment to affect productivity. Similar to Wu et al. (2020), the infrastructure investment g_{jt-1} explicitly appears as a component of the productivity process. Different from Wu et al. (2020), this paper further separates the general infrastructure g_{jt-1} into two different categories of infrastructure g_1 and g_2. That is, $g_{jt-1} = (g_{1,jt-1}, g_{2,jt-1})'$.

As discussed in Wu et al. (2020), the first-order Markov process structure has two important advantages. First, since the contribution of previous infrastructure investment has been reflected in ω_{isjt-1}[6], only the flow measure, instead of the stock, of infrastructure is required. Thus the measurement error issues in constructing capital stock can be avoided. Second, the expected productivity function $h_t(\omega_{isjt-1}, g_{jt-1})$ controls for the lagged productivity ω_{isjt-1}. The productivity effect of infrastructure is thus inferred only from the change between ω_{isjt} and ω_{isjt-1}. This allows us to address the firm's self-selection of location, which is one of the identification challenges.

[5] This modeling strategy follows the spirit of Aw et al. (2011), De Loecker (2013), Doraszelski and Jaumandreu (2013), who model and estimate the productivity effects of R&D investment and learning-by-exporting using an endogenous productivity process.

[6] This point is more evident once we write the productivity process in a recursive way:

$$\begin{aligned}
\omega_{isjt} &= h_t(\omega_{isjt-1}, g_{jt-1}) + \xi_{isjt} \\
&= h_t(h_{t-1}(\omega_{isjt-2}, g_{jt-2}) + \xi_{isjt-1}, g_{jt-1}) + \xi_{isjt} \\
&= h_t(h_{t-1}(h_{t-2}(\omega_{isjt-3}, g_{jt-3}) + \xi_{isjt-2}, g_{jt-2}) + \xi_{isjt-1}, g_{jt-1}) + \xi_{isjt}.
\end{aligned}$$

2.2 Identification Challenges

To establish a causal effect of infrastructure on productivity using equations (1) and (2), we need to address several potential endogeneity issues. First, the allocation of the infrastructure is seldom random. It is most likely dependent on the province-level productivity. If the government officials are able to foresee the future productivity shocks of each province, infrastructure may be placed by the central government into provinces that are expected to have positive shocks to accommodate the higher demand for infrastructure, or into provinces that are expected to have negative shocks to stimulate their economic growth. Meanwhile, a province with better productivity prospects could expect to produce higher output and collect more fiscal revenue in the future, which in turn may allow the province to invest more in current infrastructure via various financing schemes. As the province-level productivity affects all the firms within the province, although equation (2) models the effect of infrastructure using firm-level productivity, it still suffers from the well-known reverse causality between productivity and infrastructure.

Second, the allocation of the infrastructure may also depend on the industry-level productivity. Using the same firm-level dataset, Huang and Xiong (2017) document that firm's geographic locations explain a sizeable dispersion in their productivity, yet the explanatory power of locations varies significantly across industries. Yang (2018) finds industries with larger transportation costs tend to concentrate in locations with better highway access. The geographic concentration of industries implies that an industry-level productivity shock may simultaneously affect firm-level productivity and province-level infrastructure investment. For example, those provinces that are concentrated with energy-intensive industries may investment more in the production of power and electricity in foreseeing a positive shock to those industries. Therefore the reverse causality problem may also arise in equation (2), if the geographic distribution of industries is correlated with the infrastructure investment.

Third, the endogeneity issue may also arise from the omitted variable bias. In addition to infrastructure, many other country-wide, province-wide and industry-wide time-varying factors, such as macroeconomic shocks, local institutional changes, and industrial policy reforms, may have contributed to the firm-level productivity growth, while such factors could also affect the infrastructure investment simultaneously.

Finally, the endogeneity issue may come from the fact that firms choose their location. With spatial sorting, more productive firms tend to select into higher income provinces. Higher income provinces tend to have more infrastructure investment. This implies a potentially positive correlation between firm-level productivity and province-level infrastructure. Therefore it is important to know whether the correlation is due to an underlying process whereby firms with exogenously high productivity locate in provinces with more infrastructure; or whether the correlation is a consequence of infrastructure directly affecting productivity.

2.3 Empirical Specification

To address these identification concerns, we first decompose the firm-level productivity shocks into six components:

$$\xi_{isjt} = u_j + u_s + u_t + u_{jt} + u_{st} + \varepsilon_{it}, \tag{3}$$

where u_j and u_s are the province fixed effect and industry fixed effect which capture the permanent differences in firm-level productivity that are determined by location and industry. u_t is a year fixed effect, which represents those country-wide factors in the productivity that are common across all the firms and might be correlated with the infrastructure investment. u_{jt} and u_{st} represent the province-year fixed effect and industry-year fixed effect, which could control any time-varying province and industry characteristics that might be correlated with the infrastructure investment. Finally, ε_{it} is an idiosyncratic firm-level productivity shock and is orthogonal to the other error components in equation (3) by construction. We then use province dummies, industry dummies, year dummies and the interactions of province and year dummies, and of industry and year dummies to proxy u_j, u_s, u_t, u_{jt}, and u_{st}. This allows us to deal with the endogeneity issues arising from the non-random allocation of infrastructure and the omitted variable bias.

Second, we specify the expected productivity using the following functional form:

$$h_t(\omega_{isjt-1}, g_{jt-1}) = h(\omega_{isjt-1}) + f(\phi'_s g_{jt-1}). \tag{4}$$

The first part $h(\omega_{isjt-1})$ is a non-parametric function of ω_{isjt-1}, for which we experiment with a linear specification, a quadratic specification and an industry-specific linear specification. By controlling the lagged firm-level productivity and allowing it to follow a flexible functional form, we mitigate the reverse causality arising from the self-selection mechanism. This allows us to identify the effect of infrastructure by asking whether conditional on their lagged productivity, firms locating in provinces with more infrastructure investment indeed experience a higher productivity growth.

The second part $f(\phi_s' g_{jt-1})$ is a linear function of the interaction term between ϕ_s and g_{jt-1}, where $\phi_s = (\phi_{1s}, \phi_{2s})'$ is a measure on how the production of industry s has been relying on infrastructure $g = (g_1, g_2)'$. Note that the linear effect of ϕ_s and g_{jt-1} on ω_{isjt} is mixed with u_s and u_{jt}, the industry fixed effect and the province-year fixed effect. Hence, without the interaction term, the productivity effect of infrastructure would not be separated from those fixed effects in estimation. Nevertheless, thanks to the heterogeneity in ϕ_s, the differential effect of infrastructure on productivity can still be identified from firms across different industries. This estimation strategy thus highlights the importance of the industry-specific infrastructure reliance measures.

Combining equations (3) and (4), we obtain an empirical specification for the endogenous productivity process (2):

$$\omega_{isjt} = h(\omega_{isjt-1}) + \alpha_1 \phi_{1s} g_{1,jt-1} + \alpha_1 \phi_{2s} g_{2,jt-1} + u_j + u_s + u_t + u_{jt} + u_{st} + \varepsilon_{it}. \quad (5)$$

The key identification assumption in our approach is that the firm-level productivity shock ε_{it} is orthogonal to the province-level infrastructure investment, or formally,

$$E\left(\varepsilon_{it} | \omega_{isjt-1}, g_{jt-1}, \phi_s\right) = 0. \quad (6)$$

The underlying rationale of this assumption is that, the policy makers will not adjust the infrastructure of a province, in light of an idiosyncratic firm-level productivity shock ε_{it}, which is not correlated with any aggregate shock at the national, province and industry level. Under this assumption, conditional on lagged firm-level productivity ω_{isjt-1}, the correlation between current firm-level productivity ω_{isjt} and province-level infrastructure g_{jt-1} reflects causation from changes in infrastructure stock to changes in productivity level, i.e. the causal effect of infrastructure investment on productivity growth.[7]

As the output y_{it}, productivity ω_{isjt} and infrastructure investment g_{jt-1} are all in logarithm form in equation (1) and (5), by definition, the output elasticity of infrastructure investment in category 1 is

$$\frac{\partial y_{it}}{\partial g_{1,jt-1}} = \frac{\partial y_{it}}{\partial \omega_{isjt}} \frac{\partial \omega_{isjt}}{\partial g_{1,jt-1}} = \frac{\partial \omega_{isjt}}{\partial g_{1,jt-1}} = \alpha_1 \phi_{1s}, \quad (7)$$

which depends on the coefficient α_1 and the industry-infrastructure reliance ϕ_{1s}.[8] If α_1 is found to be positive, it implies that firms in those industries that rely more

[7] Notice that g_{jt-1} is the logarithm of the core infrastructure investment flow thus it reflects changes in infrastructure stock. And equation (5) has included the logarithm of lagged productivity ω_{isjt}.

[8] The implicit assumption here is that the effect of core infrastructure on firm-level productivity depends on how the production of different industries relies on the core infrastructure differently.

on infrastructure g_1 benefit disproportionately from investment in g_1. It therefore suggests that this category of infrastructure investment is productive. Similarly, the output elasticity of infrastructure investment in category 2 is $\alpha_2\phi_{2s}$. Since values of ϕ_{1s} and ϕ_{2s} are obtained from data, the parameters of interest in equation (5) are α_1 and α_2.

3 Data and Background

3.1 Firm-level Data

The firm-level data come from the Annual Survey of Industrial Firms conducted by China's National Bureau of Statistics, covering years from 1998 to 2007. All state-owned enterprises and non-state-owned enterprises with sales revenue above 5 million Chinese Yuan are included, containing detailed information about firm characteristics, output and input, and balance sheet variables. It is a dataset that has been widely employed in many empirical studies on various topics through the lens of firm-level productivity, such as Wu (2018) on capital misallocation, Hsu et al. (2020) on international trade and Feng et al. (2019) on productivity dynamics.

We strictly follow Brandt et al. (2012) in matching the annual data into a panel and deflating the output and input data using their price indices. In particular, we have used the price indices of intermediate inputs updated in Brandt et al. (2019).[9] Real capital stock is constructed by the perpetuity inventory method. The value added and intermediate inputs information is reported in the survey. The number of employees is used to proxy the labor input. As explained in Wu et al. (2020), the petroleum processing and coking industry and the smelting and pressing of nonferrous metals industry (coded as 25 and 33 in the Survey) are excluded from our analysis, to rule out the possible contamination from high inflation and large price volatility. Thus we apply our empirical exercises to the remaining 27 two-digit manufacturing industries listed in Table 1.

To ensure that our empirical findings are not driven by extreme values, we first drop those observations whose growth rates of output, capital stock, labor, and intermediate input are greater than the 99-percentile value or smaller than the 1-percentile value of

[9] Brandt et al. (2019) admitted that their price indices for intermediate inputs are wrong due to some programming mistake in Brandt et al. (2012), and they updated the price indices for intermediate inputs on their website. Since the deflated intermediate inputs variable is used as the proxy for firm-level productivity, we find the correction of the price indices does lead to more sensible and more favorable results in this study.

the corresponding distributions. To exclude the possibility that some extremely large firms may drive province-level infrastructure, we also drop the largest five firms in terms of sales revenue by province and year. The upper panel of Table 2 provides the summary statistics for the firm-level variables used in the subsequent analyses.

3.2 Core Infrastructure and Industry-Infrastructure Reliance

In the literature on infrastructure investment in the U.S., such as Aschauer (1989) and Gramlich (1994), the infrastructure usually refers to highways, mass transit, airports, electrical and gas facilities, water and sewers, which are usually known as the core infrastructure. In China, however, detailed information on such specific type of infrastructure investment is not available. Instead, the China Statistics Yearbooks and the China Fixed Investment Statistical Yearbooks have reported total investment in fixed assets in four industries by province and year. This includes the following four categories: (1) production and supply of electricity, gas, and water; (2) transport, storage, and post; (3) information, transmission, computer service, and software; and (4) management of water conservancy, environment, and public facilities. Depending on the nature of research questions and the design of identification strategies, different researches on China have defined infrastructure investment with a combination of different categories.[10] We next discuss how to define a proper infrastructure investment under our context.

We construct the industry-infrastructure reliance measures using the national input-output table published by the China's National Bureau of Statistics in 2002. This table reports the input and output values for 122 industries spanning the whole Chinese economy. The industry classification in the input-output table is more detailed than those in the Annual Survey of Industrial Firms and in the Yearbooks on total investment in the fixed assets. Hence we first combine the five-digit industries in the input-output table into two-digit industries to be consistent with the manufacturing industries in the Survey and the infrastructure industries in the Yearbooks.

We then define the industry-infrastructure reliance ϕ_s as share of the value of infrastructure investment as an intermediate input to total intermediate input value of industry s. Intuitively, ϕ_s measures that among all the intermediate input to the production of industry s, how much input comes from the infrastructure investment.

[10] For example, Shi and Huang (2014) use the four categories all together. Naughton (2007), Feng and Wu (2018) and Wu et al. (2020) measure infrastructure investment as the sum of the first three categories. Bai and Qian (2010) only focus on the first two categories.

Table 1 presents our constructed values of industry-infrastructure reliance, representing 4 categories of infrastructure investment for 27 manufacturing industries. Here, $\phi_s = (\phi_{1s}, \phi_{2s}, \phi_{3s}, \phi_{4s})'$, $s = 1, 2, \cdots, 27$. To see whether the constructed measure of industry-infrastructure reliance is plausible, we cross check the variations in the table by industries and by infrastructure categories. High-energy-consuming industries, such as raw chemical materials and chemical products industry, nonmetal mineral products industry, smelting and pressing of ferrous metals industry, rely on the production and supply of electricity, gas and water industry most. If one industry is highly dependent on the raw materials, it will generally need good transport conditions. Consistent with this intuition, nonmetal mineral products industry, smelting and pressing of ferrous metals industry and timber processing, bamboo, cane, palm fiber industry heavily depend on the transport, storage and post industry.

<insert Table 1 here>

At the bottom of Table 1, we report the mean and standard deviation of the reliance measures. Taken as a whole, across the four categories of infrastructure, manufacturing industries rely on transport, storage and post industry most, which on average accounts for 4.4% of the total intermediate inputs. The contribution of the production and supply of electricity, gas and water industry ranks the second, on average, 3.3%. This is consistent with the economic rationale, as these two categories of infrastructure are crucial to the continued operation of the production units and the distribution of the final products. The weight of information, transmission, computer service, and software industry is much lower, on average, 1.2%. Furthermore, the industry-variation in the reliance on this type of infrastructure is also much smaller than those in the first two categories.[11] The contribution of the management of water conservancy, environment, and public facilities industry is trivial and nearly zero across all industries, which is shown in the last column of Table 1. This is not surprising as investment in this category mainly aims to enhance the urban amenities so that has little direct effect on the manufacturing sector.

Given that our empirical exercises only apply to the manufacturing sector and that the industry-infrastructure reliance plays a key role in our identification, in this paper we will only consider infrastructure investment in the first two categories, that

[11] The Yearbooks only started to report investment in the category of (3) information, transmission, computer service, and software since 2003. That is another important reason why we only focus on the first two categories of infrastructure investment in this paper.

is investment in (1) production and supply of electricity, gas, and water; and (2) transport, storage, and post. Following the literature, we name these two categories as the core infrastructure. The lower panel of Table 2 presents the summary statistics for the core infrastructure investment data used in the subsequent analyses.

<insert Table 2 here>

Table 3 provides the real growth rate and investment-to-GDP share of core infrastructure investment aggregated from 30 provinces. The growth rate of infrastructure investment is more significant since 2004. Although a few early years witness a negative growth rate, overall, investment in both types of infrastructure has been growing very fast during our sample period, with an average annual growth rate at 14.3% in the production and supply of electricity, gas and water and at 9.4% the in transport, storage and post. The infrastructure investment to GDP ratios have been relatively stable over the years. On average investment in the production and supply of electricity, gas and water accounts for 3.3% of GDP and investment in transport, storage and post amounts to 5.1% of GDP.

<insert Table 3 here>

Table 4 reports the real growth rate and investment-to-GDP share of core infrastructure investment by province. The last row of Table 4 presents the standard deviations across the provinces. It thus highlights the substantial variation in the core infrastructure investment among different provinces, which offers an important source of identification.

<insert Table 4 here>

Compared with those works that focus on a specific type of infrastructure, which typically use physical measures of infrastructure, for example, the length of the highways, our monetary measure of infrastructure has two advantages. First, it better controls for unobservable variation in the quality of infrastructure, which is usually harder to capture by physical measures. Second, with a monetary measure of infrastructure, the output elasticities estimated from our production function can be easily translated into rates of return of investment, which provide an intuitive benchmark for the investment efficiency. There is, of course, a downside of using a monetary measure. Due to the data limitation, we are not able to further decompose these two big categories of investment into smaller subcategories, whose productivity effects may differ from each other. Thus in this paper we are only able to answer whether as a whole core infrastructure investment is productive in China. It therefore complements many recent works that study the productivity effect of a specific infrastructure.

3.3 Institutional Background

Bai and Qian (2010) and Shi and Huang (2014) discuss the source of funds and different roles played by various jurisdiction levels in infrastructure investment in China. Wu et al. (2020) summarize two stylized facts which explain why we use the infrastructure investment data at the province level. First, most infrastructure investment are made by state-owned or state-controlled enterprises with funds from both the central government and the local governments. Second, among various jurisdiction levels the provincial government plays a key role in infrastructure investment decision.

Why the Chinese governments have a strong incentive in infrastructure investment? According to Démurger (2014), the rationale for the central government is twofold. First, infrastructure development is necessary to support the rapid economic growth of the country that fuels an ever-increasing demand for infrastructure services. Second, infrastructure development is needed to fight worsening regional inequality by promoting the catch-up of lagged inland provinces with coastal provinces. As for the local governments, a leading view, represented by Li and Zhou (2005), Zhang et al. (2007) and Xu (2011), argues that under China's regionally decentralized authoritarian system, infrastructure investment has been adopted as the most effective instrument by the local governments as their response to the GDP yardstick competition. In particular, consistent with the general findings in the literature, such as Bellak et al. (2009) and Vijil and Wagner (2012), infrastructure is regarded as an important determinant in attracting FDI and promoting exporting. On the other hand, Shi and Huang (2014) argue that the Chinese governments tend to use infrastructure investment as a choice for reviving its economy when they expect a large negative productivity shock. Such institutional background highlights the importance of addressing the endogenous allocation of infrastructure and the omitted variable bias in our identification strategy.

4 Empirical Results

4.1 Productivity Effects

Table 5 reports the estimation results for equation (5) in a special case of $\alpha_1 = \alpha_2$. We start with the simplest case with a linear specification for the lagged productivity. As the firm-level productivity is estimated by industry, column (1) reports the results by including the industry, province and year dummies, to control for the differences in firm-level productivity due to industry, geographic locations and the countrywide aggregate productivity shocks. The coefficient of the infrastructure investment is 0.168, which is large and significant. In column (2), we add industry-year and province-year fixed effects into the regression, to further control the endogeneity due to omitted industry-specific and province-specific timing varying factors. The coefficient on infrastructure investment in column (2) further declines to 0.150.

<insert Table 5 here>

The functional form of the productivity process could also affect the productivity effect of infrastructure investment. To shed light on this concern, we first add the quadratic term of the lagged productivity into the productivity process, which brings the coefficient of infrastructure investment to be 0.159, as shown in column (3). However, the quadratic term itself is insignificant. We next add the interaction between lagged productivity and industry dummies into the regression, allowing for an industry-specific first-order Markov process. The joint significance test shows that they are highly relevant in the regression. The coefficient of infrastructure investment changes slightly to 0.156 in column (4). In column (5), after including the quadratic term of lagged productivity and its interaction with industry dummies, the coefficient varies slightly to 0.158, where the quadratic term is insignificant again and the interaction terms are still jointly significant. Thus we take column (4) as our benchmark estimates, with the assuring fact that the results across columns (3) to (5) are very stable.

Our results therefore suggest that there will be an upward bias in the estimation of output elasticity of infrastructure investment if the reverse causality is not fully controlled. This finding is in line with the findings in Feng and Wu (2018) using aggregate data. Nevertheless, after addressing a set of the identification issues using our empirical strategy, our most conservative estimate still suggests that firms in an industry that relies more heavily on infrastructure indeed experience higher productivity growth from more infrastructure investment. This implies that the core infrastructure investment is productive.

Table 6 presents the estimation results for equation (5), and shares the structure of Table 5. As expected, the coefficients on the combined infrastructure investment in Table 5 are in the middle of the coefficients on the individual infrastructure investment, while the significances of the coefficients on the lagged productivity terms remain the same. In the column (4), the coefficient on investment in *electricity* is 0.122, while the coefficient on investment in *transport* is 0.233, both of which are highly significant.

<insert Table 6 here>

The output elasticity of infrastructure investment in *electricity* is $\alpha_1\phi_{1s}$. Given that the estimates of α_1 are significantly positive in Table 6, the productivity effect of infrastructure on firms increases with the industry-infrastructure reliance. That is, firms in industries that rely more on infrastructure gain more from infrastructure investment. For example, infrastructure investment is more beneficial to firms in the industry of nonmetal mineral products than those in industry of Tobacco processing.

To see the average output elasticity of infrastructure investment across the manufacturing sector, we pool all industries together to obtain $\alpha_1\bar{\phi}_1$, where $\bar{\phi}_1$ is the industry average of industry-infrastructure reliance reported in the second last row of Table 1. With an estimate for $\alpha_1 = 0.122$ under the specification of column (4) of Table 6 and a value of $\bar{\phi}_1 = 0.033$, the industry average output elasticity of infrastructure investment in *electricity* is $\alpha_1\bar{\phi}_1 = 0.004$. Given an average annual growth rate of 14.3% in infrastructure investment in *electricity* as reported in the last column of Table 3, on average firm's productivity increases by 0.058% annually, *ceteris paribus*. Similarly, firms' productivity increases annually by 0.096%, due to $\alpha_2 = 0.233$, $\bar{\phi}_2 = 0.044$ and an average annual growth of 9.4% in infrastructure investment in *transport*.

4.2 Robustness Checks

Table 7 presents the results using an alternative measure of infrastructure investment in the two categories. To further mitigate the reverse causality, following Ozyurt (2009) and Feng and Wu (2018), we use the newly increased fixed assets (NIFA hereafter) (*xinzeng guding zichan touzi* in Chinese) reported by the National Bureau of Statistics of China. NIFA measures investment in fixed assets that have been *used* for production after the process of construction and purchase is completed. As argued by Feng and

Wu (2018), NIFA is less likely to be affected by the current level of output. The estimates in Table 7 share a similar pattern presented in Table 6, but with smaller magnitude. For example, in columns (4) of both tables, the estimated α_1 decreases from 0.122 to 0.082 and α_2 decreases from 0.233 to 0.166. One possible explanation is that, investment in infrastructure may increase both firm-level demand in the short-run and firm-level productivity in the long-run. By using the newly increased fixed assets as a measure of infrastructure investment, we further tease out the short-run demand effect and obtain a long-run productivity effect of infrastructure investment in Table 7.

<insert Table 7 here>

An additional robustness check is reported in Table 8. We consider a second-order Markov process, that is the productivity effect of infrastructure investment may lag by two periods, conditional on past productivity:

$$\omega_{isjt} = h_1(\omega_{isjt-1}) + h_2(\omega_{isjt-2}) + \alpha_1\phi_{1s}g_{1,jt-2} + \alpha_1\phi_{2s}g_{2,jt-2} + u_j + u_s + u_t + u_{jt} + u_{st} + \varepsilon_{it}.$$

Compared with the productivity process (5), this specification allows for a longer delay for infrastructure investment to build into firm's productivity. In addition, due to the 2-period lag structure, the reverse causality issue is even less likely to occur. Table 8 shows that the estimated α_1 and α_2 both remain positive and significant in various specifications. Compared with Table 6, the estimated coefficients of α_1 are slightly smaller and α_2 are slightly bigger.

<insert Table 8 here>

4.3 Return to Core Infrastructure Investment

Our empirical exercises find a positive and robust productivity effect of core infrastructure investment. To shed light on the magnitude of the productivity effects we have obtained, with some auxiliary assumptions we then translate the productivity effects into real rates of return from the perspective of firms. To many policy makers, the return rate is key to the decision-making of infrastructure investment,[12] since it provides an intuitive benchmark to evaluate the investment efficiency.

[12] Ismath Bacha and Mirakhor (2018) pointed out that some Muslim countries of the developing world already are indebted as a result of funding infrastructure.

Under our specification for output (1) and productivity (5), the firm-level output elasticity with respect to infrastructure investment in category 1 is $\alpha_1\phi_{1s}$, depending on industry-infrastructure reliance ϕ_{1s}. It is common for all firms in the same industry. When aggregating individual firm's output elasticity to province-level output elasticity, as the distribution of industry could be different across provinces, the province-level output elasticity varies with the weights of firms from different industries:

$$e_{1,jt} = \alpha_1 \sum_s \sum_{i\in(s,j)} \phi_{1s} \times \pi_{isjt},$$

where π_{isjt} is the value-added weight of firm i in industry s in province j in year t. Then under the assumptions that these industries are representative about the whole economy and that the firms in the sample are representative about the population of producers, the economic return, or the marginal product of infrastructure investment in category 1 in each province could be calculated as:

$$r_{1,jt} = e_{1,jt} \times \frac{GDP_{jt}}{G_{jt-1}}.$$

The national-level rates of return are the weighted-averages of those from the provinces:

$$r_{1,t} = \sum_j r_{1,jt} \times \Pi_{jt},$$

where Π_{jt} represents province j's GDP as a share of national GDP in year t.

With these assumptions, the rate of return to infrastructure investment provides another way of interpreting the magnitude of coefficient α_1. Under the specification of column (4) of Table 6, the estimate of α_1 is 0.122. Based on this estimate, Table 9 reports the rates of return to infrastructure investment in *electricity* by province and by year. The last column of Table 9 shows the average return rates of each province over 1999-2007. Across the provinces, the returns vary from 8.2% in Guizhou to 31.0% in Liaoning. The last row of Table 9 shows the national-level return rates averaged across the provinces. Over the years, there is an inverted-U shape in the returns which peaks around 2003. The average rate of return to infrastructure investment in *electricity* during this period is 20.3%. This is similar to the return to capital in electricity and heat water sector calculated in Bai and Qian (2010) using financial accounts.

<insert Table 9 here>

Similarly, we can also translate the coefficient α_2 or the output elasticity of infrastructure investment in category 2 into rates of return. Table 10 gives the return to infrastructure investment in *transport*. Under the specification of column (4) of Table 6, a magnitude of 0.233 for α_2 implies that the average rate of return to infrastructure investment in *transport* in China over 1999-2007 is 25.9%. This is in the similar range to the rate of return to road investment in China, estimated in Li and Chen (2013) and Li et al. (2017) using different approaches. Taking the simple average between the returns of *electricity* and *transport* above, we infer that the average rate of return to core infrastructure investment in China during 1999 to 2007 is around 23%.

<insert Tables 10 here>

Such average rate of return is much lower than those in Aschauer (1989) and Shi and Huang (2014), while it is higher than the discount rate of 8%, the threshold used by Ministry of Housing and Urban-Rural Development of China in evaluating infrastructure projects (Qin, 2016). Considering that the long-run real rate of return to capital investment is about 3 to 4% worldwide, we conclude the general efficiency of core infrastructure investment in China over our sample period.

5 Conclusion

This paper finds that the core infrastructure investment has contributed significantly to the productivity gains in China during the period of 1999 to 2007, by matching the manufacturing firm-level data with the province-level core infrastructure investment data. We use the following empirical strategies to address a set of identification issues. First, we model the productivity of a firm into the expected productivity and a random shock. Second, we decompose the firm-level productivity shock into the aggregate components and an idiosyncratic component. Finally, we model the expected productivity of the firm as a function of its lagged productivity and an interaction of the infrastructure investment of the province where the firm locates and the industry-infrastructure reliance to which the firm belongs. We find that on average firm's productivity increases by 0.057% annually due to infrastructure investment in *electricity*, and by 0.094% annually due to *transport*. Putting together, investment in core infrastructure has an average annual rate of return of 23%.

The combination of a production function estimation framework together with the industry-infrastructure reliance has some unique advantages in addressing the identification issues and obtaining the rates of return to investment. However, there are several methodological issues worth further discussion. First, as the variation in industry-infrastructure reliance comes from how the production of different industries has been relying on the core infrastructure differently, the productivity effect we identified in this paper only captures this specific production channel through which core infrastructure investment affects productivity. Although for manufacturing firms the production channel which varies with industry should capture the majority of the productivity effect of core infrastructure, there could be some other channels which are common across all the firms and do not vary with industries that are not captured in this methodology. The output elasticities and the rates of return in our empirical findings are therefore subject to this caveat.

Second, the fact that the interaction term only varies across industry, province and year implies two limitations. First, our identification strategy only captures the differential effect of infrastructure on productivity. Thus it does not include our empirical strategy would not be immune to the endogeneity issue if there were time-varying industry-province specific productivity shocks, although we believe the majority of the effect from such shocks has been absorbed by the industry-year and the province-year fixed effects in our specification. Second, as the interaction term is not firm-specific, this paper does not investigate the potential mechanisms through which core infrastructure investment enhances the productivity. Such effort usually requires firm-specific instruments, which is more easily available in studying one specific infrastructure investment, for example, Holl (2016) for highways and Fisher-Vanden et al. (2015) for electricity.

Last but not least, as our research framework is a production function estimation using manufacturing firms, it naturally implies that the estimated rates of return in this paper may not fully capture those social returns of infrastructure in other sectors and in other aspects, for example, rural poverty reduction as in Qin and Zhang (2016) and gender equality increase as in Parikh et al. (2015).

References

[1] Abeberese, A. B. (2017). Electricity cost and firm performance: evidence from India. *Review of Economics and Statistics*, 99(5), 839-852.

[2] Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6), 2411-2451.

[3] Aschauer, D. (1989). Is public expenditure productive? *Journal of Monetary Economics*, 23, 177-200.

[4] Aw, B., Roberts, M., and Xu, D. (2011). R&D investment, exporting, and productivity dynamics. *American Economic Review*, 101(4), 1312-1344.

[5] Bai, C., and Qian, Y. (2010). Infrastructure development in China: the cases of electricity, highways, and railways. *Journal of Comparative Economics*, 38(1), 34-51.

[6] Bellak, C., Leibrecht, M., and Damijan, J. P. (2009). Infrastructure endowment and corporate income taxes as determinants of foreign direct investment in Central and Eastern European countries. *World Economy*, 32(2), 267-290.

[7] Brandt, L., Van Biesebroeck, J., and Zhang, Y. (2012). Creative accounting or creative destruction? firm-level productivity growth in Chinese manufacturing. *Journal of Development Economics*, 97(2), 339-351.

[8] Brandt, L., Van Biesebroeck, J., and Zhang, Y. (2017). WTO accession and performance of Chinese manufacturing firms. *American Economic Review*, 107(9), 2784-2820.

[9] Brandt, L., Van Biesebroeck, J., and Zhang, Y. (2019). WTO accession and performance of Chinese manufacturing firms: Corrigendum. *American Economic Review*, 109(4), 1616-1021.

[10] De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics*, 5(3), 1-21.

[11] Démurger, S. (2014). Infrastructure in China. *The Oxford Companion to the Economics of China* (pp. 348-352). Oxford: Oxford University Press.

[12] Doraszelski, U., and Jaumandreu, J. (2013). R&D and productivity: estimating endogenous productivity. *Review of Economic Studies*, 80, 1338-1383.

[13] Duflo, E., and Pande, R. (2007). Dams. *Quarterly Journal of Economics*, 122(2), 601–646.

[14] Feng, Q., and Wu, G. L. (2018). On the reverse causality between output and infrastructure: the case of China. *Economic Modelling*, 74, 97-104.

[15] Feng, Q., Wang, Z., and Wu, G. L. (2019). Productivity dynamics of Chinese manufacturing firms. *Singapore Economic Review*, 64, 899-919.

[16] Fernald, J. (1999). Roads to prosperity? assessing the link between public capital and productivity. *American Economic Review*, 89, 619-638.

[17] Fisher-Vanden, K., Mansur, E. T., and Wang, Q. (2015). Electricity shortages and firm productivity: evidence from China's industrial firms. *Journal of Development Economics*, 114, 172-188.

[18] Gramlich, E. (1994). Infrastructure investment: a review essay. *Journal of Economic Literature*, 32, 1176-1196.

[19] Holl, A. (2016). Highways and productivity in manufacturing firms. *Journal of Urban Economics*, 93, 131-151.

[20] Hsu, W., Lu, Y., and Wu, G. L. (2020). Competition, markups, and gains from trade: a quantitative analysis of China between 1995 and 2004. *Journal of International Economics*, forthcoming.

[21] Huang, Y., and Xiong, W. (2017). Geographic distribution of firm productivity and production: a "market access" approach, Working Paper.

[22] Ismath Bacha, O., and Mirakhor, A. (2018). Funding development infrastructure without leverage: a risk-sharing alternative using innovative sukuk structures. *World Economy*, 41(3), 752-762.

[23] Jorgenson, D. W., Gollop, F. M., and Fraumeni, B. M. (1987). *Productivity and U.S. economic growth.* Cambridge, MA: Harvard University Press, 1987.

[24] Li, Z., and Chen, Y. (2013). Estimating the social return to transport infrastructure: a price-difference approach applied to a quasi-experiment. *Journal of Comparative Economics*, 41, 669-683.

[25] Li, H., and Li, Z. (2013). Road investments and inventory reduction: firm level evidence from China. *Journal of Urban Economics*, 76, 43-52.

[26] Li, Z., Wu, M., and Chen, B. (2017). Is road infrastructure investment in China excessive? evidence from productivity of firms. *Regional Science and Urban Economics*, 65, 116-126.

[27] Li, H., and Zhou, L. (2005). Political turnover and economic performance: the incentive role of personnel control in China. *Journal of Public Economics*, 89(9), 1743-1762.

[28] Naughton, B. (2007). *The Chinese Economy: Transitions and Growth.* MIT Press.

[29] Oberfield, E. (2018). A Theory of input-output architecture. *Econometrica*, 86(2), 559-589.

[30] Ozyurt, S. (2009). Total factor productivity growth in Chinese industry: 1952-2005. *Oxford Development Studies*, 37(1), 1-17.

[31] Parikh, P., Fu, K., Parikh, H., McRobie, A., and George, G. (2015). Infrastructure provision, gender, and poverty in Indian slums. *World Development*, 66, 468-486.

[32] Pavcnik, N., (2002). Trade liberalization, exit, and productivity improvements: Evidence from Chilean plants. *Review of Economic Studies*, 69(1), 245-276.

[33] Qin, Y. (2016). China's transport infrastructure investment: past, present, and future. *Asian Economic Policy Review*, 11(2), 199-217.

[34] Qin, Y., and Zhang, X. (2016). The road to specialization in argricultural production: evidence from rural China. *World Development*, 77, 1-16.

[35] Redding, S. J., and Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. *Handbook of Regional and Urban Economics*, 5, 1339-1398.

[36] Shi, H., and Huang, S. (2014). How much infrastructure is too much? a new approach and evidence from China. *World Development*, 56, 272-286.

[37] Vijil, M., and Wagner, L. (2012). Does aid for trade enhance export performance? investigating the infrastructure channel. *World Economy*, 35(7), 838-868.

[38] Wu, G. L. (2018). Capital misallocation in China: financial frictions or policy distortions?. *Journal of Development Economics*, 130, 203-223.

[39] Wu, G. L., Feng, Q., and Wang, Z. (2020). A structural estimation on the return to infrastructure investment in China. Working Paper, Nanyang Technological University.

[40] Xu, C. (2011). The fundamental institutions of China's reforms and development. *Journal of Economic Literature*, 49(4), 1076-1151.

[41] Yang, Y. (2018). Transport infrastructure, city productivity growth and sectoral reallocation: evidence from China, Working Paper.

[42] Zhang, J., Gao, Y., Fu, Y., and Zhang, H. (2007). Why could China build up its infrastructure?. *Economic Research Journal* (in Chinese), 42, 4-19.

Table 1 Industry Infrastructure Reliance

Code	Industry (s)	ϕ_{1s}	ϕ_{2s}	ϕ_{3s}	ϕ_{4s}
13	Food processing	0.013	0.026	0.002	0.000
14	Food manufacturing	0.022	0.058	0.010	0.000
15	Beverage manufacturing	0.021	0.063	0.017	0.000
16	Tobacco processing	0.009	0.022	0.002	0.000
17	Textile industry	0.031	0.026	0.007	0.000
18	Garments & other fiber products	0.012	0.034	0.013	0.000
19	Leather, furs, down & related products	0.008	0.028	0.019	0.000
20	Timber processing, bamboo, cane, palm fiber	0.037	0.064	0.009	0.000
21	Furniture manufacturing	0.027	0.064	0.021	0.000
22	Papermaking & paper products	0.047	0.060	0.005	0.000
23	Printing industry	0.020	0.048	0.006	0.000
24	Cultural, educational & sports goods	0.022	0.040	0.009	0.000
26	Raw chemical materials & chemical products	0.096	0.052	0.010	0.001
27	Medical & pharmaceutical products	0.040	0.047	0.018	0.001
28	Chemical fiber	0.038	0.028	0.002	0.000
29	Rubber products	0.035	0.044	0.005	0.000
30	Plastic products	0.035	0.037	0.010	0.000
31	Nonmetal mineral products	0.094	0.087	0.014	0.000
32	Smelting & pressing of ferrous metals	0.072	0.064	0.005	0.000
34	Metal products	0.053	0.047	0.029	0.000
35	Ordinary machinery	0.034	0.045	0.016	0.000
36	Special purpose equipment	0.032	0.045	0.020	0.000
37	Transport equipment	0.019	0.031	0.015	0.000
39	Electric equipment & machinery	0.020	0.039	0.022	0.000
40	Electronic & telecommunications equipment	0.012	0.023	0.008	0.000
41	Instruments, meters, cultural & office equipment	0.015	0.037	0.015	0.000
42	Other manufacturing	0.026	0.038	0.018	0.000
	Industry average	0.033	0.044	0.012	0.000
	Standard deviation	0.023	0.016	0.007	0.000

Notes: 1. The industry-infrastructure reliance is the share of the value of infrastructure investment as intermediate input in an industry to the total intermediate input value of that industry.

2. ϕ_{1s} represents the infrastructure reliance on investment in production and supply of electricity, gas and water.

3. ϕ_{2s} represents the infrastructure reliance on investment in transport, storage and post.

4. ϕ_{3s} represents the infrastructure reliance on investment in information transmission, computer services and software.

5. ϕ_{4s} represents the infrastructure reliance on investment in management of water conservancy, environment, and public facilities.

6. The numbers are calculated using the Input-Output table 2002 by China NBS.

Table 2 Summary Statistics of Variables

Symbol	Definition	Unit(CNY)	Mean	Std. D.	Data Source
y	fiirm value added	1,000	16,252	93,758	China's NBS
k	firm capital stock, defined as in Brandt et al. (2012)	1,000	21,943	135,891	China's NBS
m	firm intermediate inputs	1,000	43,421	225,298	China's NBS
l	firm's number of workers		251	617	China's NBS
G_1	total investment in fixed assets in production and supply of electricity, gas and water	100 million	148	124	Yearbooks
G_2	total investment in fixed assets in transport, storage and post	100 million	221	150	Yearbooks
G_{1new}	newly increased in fixed assets investment in production and supply of electricity, gas and v	100 million	92	89	Yearbooks
G_{2new}	newly increased in fixed assets investment in transport, storage and post	100 million	127	91	Yearbooks

Notes:

1. All the data in monetary value has been deflated by year1998 price indexes.
2. NBS refers to National Bureau of Statistics and Yearbooks refer to China Statistics Yearbooks and China Fixed Investment Yearbooks.
3. All the variables are in logarithm form in regressions.

Table 3 Growth and GDP Share of Infrastructure Investment: National Level

	1999	2000	2001	2002	2003	2004	2005	2006	2007	Year average
$\Delta G_1/G_1$	0.103	0.105	-0.115	0.133	0.190	0.410	0.289	0.115	0.058	0.143
$\Delta G_2/G_2$	0.048	0.078	0.136	-0.046	-0.085	0.161	0.220	0.249	0.084	0.094
G_1/GDP	0.030	0.031	0.025	0.026	0.028	0.036	0.042	0.041	0.038	0.033
G_2/GDP	0.055	0.055	0.057	0.050	0.042	0.044	0.048	0.053	0.051	0.051

Notes:

1. Please refer to Table 2 for the definition of G_1 and G_2.
2. G_1 and G_2 are deflated by the fixed asset investment price index. GDP is deflated by the GDP deflator.
3. $\Delta G_1/G_1$ and $\Delta G_2/G_2$ are real growth rates of G_1 and G_2.
4. G_1/GDP and G_2/GDP are the ratios between G_1 and G_2 and GDP.

Table 4 Growth of Infrastructure Investment and its GDP Share: by Province

Province	$\Delta G_1/G_1$	$\Delta G_2/G_2$	G_1/GDP	G_2/GDP
Beijing	0.079	0.116	0.022	0.048
Tianjin	0.124	0.143	0.024	0.047
Hebei	0.099	0.091	0.030	0.049
Shanxi	0.164	0.069	0.058	0.061
Inner mongolia	0.339	0.235	0.076	0.072
Liaoning	0.148	0.097	0.019	0.042
Jilin	0.143	0.116	0.027	0.043
Helongjiang	0.109	0.038	0.020	0.042
Shanghai	0.057	0.160	0.021	0.045
Jiangsu	0.080	0.058	0.025	0.036
Zhejiang	0.119	0.096	0.035	0.056
Anhui	0.192	0.118	0.026	0.047
Fujian	0.070	0.098	0.032	0.052
Jiangxi	0.182	0.079	0.029	0.067
Shandong	0.095	0.056	0.020	0.032
Henan	0.131	0.107	0.028	0.049
Hubei	0.044	0.111	0.052	0.052
Hunan	0.194	0.056	0.027	0.046
Guangdong	0.115	0.061	0.024	0.042
Guangxi	0.216	0.082	0.042	0.056
Hainan	0.178	0.058	0.021	0.079
Chongqing	0.148	0.120	0.033	0.058
Sichuan	0.141	0.020	0.040	0.055
Guizhou	0.250	0.093	0.093	0.089
Yunan	0.256	0.083	0.057	0.080
Shaanxi	0.157	0.088	0.040	0.076
Gansu	0.136	0.057	0.050	0.070
Qinghai	0.109	0.202	0.100	0.124
Ningxia	0.262	0.038	0.080	0.096
Xinjiang	0.123	0.029	0.034	0.073
Province average	0.149	0.093	0.039	0.059
Standard deviation	0.067	0.048	0.022	0.020

Notes:

1. Please refer to Table 2 for the definition of G_1 and G_2.
2. $\Delta G_1/G_1$ and $\Delta G_2/G_2$ are the annual growth rates of G_1 and G_2 over 1998-2007.
3. G_1/GDP and G_2/GDP are the ten-year means of G_1/GDP and G_2/GDP over 1998-2007.

Table 5 Productivity Effects of Infrastructure Investment

Dependent variable: ω_{isjt}

Independent variables:	(1)	(2)	(3)	(4)	(5)
$\phi_{1s}\cdot g_{1t-1}+\phi_{2s}\cdot g_{2t-1}$	0.168***	0.150***	0.159***	0.156***	0.158***
	(0.012)	(0.015)	(0.017)	(0.015)	(0.016)
ω_{isjt-1}	0.875***	0.871***	0.906***		
	(0.001)	(0.002)	(0.024)		
ω^2_{isjt-1}			-0.005		-0.004
			(0.004)		(0.005)
$\omega_{isjt-1}\times$Industry				Y	Y
Joint significancec of $\omega_{isjt-1}\times$Industry (p value)				0.000	0.000
Industry dummy	Y	Y	Y	Y	Y
Province dummy	Y	Y	Y	Y	Y
Year dummy	Y	Y	Y	Y	Y
Industry×Year		Y	Y	Y	Y
Province×Year		Y	Y	Y	Y
Observations	1,216,785	1,216,785	1,216,785	1,216,785	1,216,785
R-squared	0.860	0.861	0.861	0.862	0.862

Notes:

1. The stars, *, ** and *** indicate the significance level at 10%, 5% and 1%, respectively.

2. ω is the firm-level productivity; ϕ_{1s} and ϕ_{2s} are from Table 1; $g_{1t-1} = \ln(G_{1t-1})$; and $g_{2t-1} = \ln(G_{2t-1})$.

Table 6 Productivity Effects of Infrastructure Investment: by category

Dependent variable: ω_{isjt}

Independent variables:	(1)	(2)	(3)	(4)	(5)
$\phi_{1s} \cdot g_{1t-1}$	0.204***	0.133***	0.147***	0.122***	0.123***
	(0.017)	(0.025)	(0.028)	(0.025)	(0.026)
$\phi_{2s} \cdot g_{2t-1}$	0.087***	0.190***	0.184***	0.233***	0.238***
	(0.031)	(0.050)	(0.050)	(0.051)	(0.052)
ω_{isjt-1}	0.875***	0.871***	0.906***		
	(0.001)	(0.002)	(0.024)		
ω^2_{isjt-1}			-0.005		-0.004
			(0.004)		(0.005)
ω_{isjt-1}×Industry				Y	Y
Joint significancec of ω_{isjt-1}×Industry (p value)				0.000	0.000
Industry dummy	Y	Y	Y	Y	Y
Province dummy	Y	Y	Y	Y	Y
Year dummy	Y	Y	Y	Y	Y
Industry×Year		Y	Y	Y	Y
Province×Year		Y	Y	Y	Y
Observations	1,216,785	1,216,785	1,216,785	1,216,785	1,216,785
R-squared	0.860	0.861	0.861	0.862	0.862

Note: Please refer to the notes of Table 5.

Table 7 Productivity Effects of Infrastructure Investment: Newly Increased Fixed Assets

Dependent variable: ω_{isjt}

Independent variables:	1	2	3	4	5
$\phi_{1s}\cdot g_{1t-1,\ new}$	0.147***	0.096***	0.109***	0.082***	0.083***
	(0.013)	(0.020)	(0.022)	(0.020)	(0.020)
$\phi_{2s}\cdot g_{2t-1,\ new}$	-0.238***	0.129***	0.121***	0.166***	0.167***
	(0.017)	(0.035)	(0.035)	(0.036)	(0.036)
ω_{isjt-1}	0.875***	0.871***	0.906***		
	(0.001)	(0.002)	(0.024)		
ω^{2}_{isjt-1}			-0.005		-0.004
			(0.004)		(0.005)
$\omega_{isjt-1}\times$Industry				Y	Y
Joint significancec of $\omega_{isjt-1}\times$Industry (p value)				0.000	0.000
Industry dummy	Y	Y	Y	Y	Y
Province dummy	Y	Y	Y	Y	Y
Year dummy	Y	Y	Y	Y	Y
Industry×Year		Y	Y	Y	Y
Province×Year		Y	Y	Y	Y
Observations	1,216,785	1,216,785	1,216,785	1,216,785	1,216,785
R-squared	0.86	0.861	0.861	0.862	0.862

Note: Please refer to the notes of Table 5.

Table 8 Productivity Effects of Infrastructure Investment: by category

Dependent variable: ω_{isjt}

Independent variables:	(1)	(2)	(3)	(4)	(5)
$\phi_{1s} \cdot g_{1t-2}$	0.164***	0.075**	0.114**	0.067**	0.072**
	(0.023)	(0.033)	(0.047)	(0.034)	(0.034)
$\phi_{2s} \cdot g_{2t-2}$	0.107**	0.266***	0.249***	0.254***	0.277***
	(0.045)	(0.066)	(0.066)	(0.068)	(0.072)
ω_{isjt-1}	0.785***	0.783***	0.649***		
	(0.003)	(0.003)	(0.044)		
ω^2_{isjt-1}			0.021***		0.025***
			(0.007)		(0.006)
ω_{isjt-1}×Industry				Y	Y
Joint significancec of ω_{isjt-1}×Industry (p value)				0.000	0.000
ω_{isjt-2}	0.138***	0.139***	0.357***		
	(0.005)	(0.005)	(0.114)		
ω^2_{isjt-2}			-0.035*		-0.047**
			(0.019)		(0.023)
ω_{isjt-2}×Industry				Y	Y
Joint significancec of ω_{isjt-2}×Industry (p value)				0.000	0.000
Industry dummy	Y	Y	Y	Y	Y
Province dummy	Y	Y	Y	Y	Y
Year dummy	Y	Y	Y	Y	Y
Industry×Year		Y	Y	Y	Y
Province×Year		Y	Y	Y	Y
Observations	547,167	547,167	547,167	547,167	547,167
R-squared	0.881	0.882	0.883	0.883	0.884

Note: Please check the notes of Table 5.

Table 9 Return Rate to G_1

Province \Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	Year average
Beijing	0.120	0.105	0.224	0.360	0.310	0.499	0.294	0.197	0.224	0.259
Tianjin	0.139	0.141	0.234	0.292	0.311	0.301	0.343	0.336	0.259	0.262
Hebei	0.168	0.191	0.170	0.273	0.332	0.301	0.280	0.195	0.183	0.232
Shanxi	0.130	0.130	0.118	0.166	0.206	0.161	0.122	0.117	0.113	0.140
Inner mongolia	0.202	0.340	0.235	0.174	0.120	0.083	0.051	0.041	0.055	0.145
Liaoning	0.253	0.324	0.250	0.303	0.330	0.347	0.369	0.348	0.266	0.310
Jilin		0.216	0.147	0.224	0.255	0.284	0.267	0.216	0.182	0.224
Helongjiang	0.189	0.262	0.212	0.239	0.337	0.299	0.257	0.222	0.192	0.245
Shanghai	0.149	0.123	0.206	0.292	0.351	0.314	0.284	0.230	0.255	0.245
Jiangsu	0.233	0.224	0.199	0.294	0.271	0.171	0.124	0.151	0.228	0.211
Zhejiang	0.137	0.138	0.124	0.156	0.186	0.178	0.107	0.101	0.128	0.139
Anhui	0.222	0.163	0.211	0.415	0.253	0.405	0.256	0.152	0.154	0.248
Fujian	0.125	0.131	0.117	0.192	0.219	0.244	0.142	0.111	0.118	0.155
Jiangxi	0.310	0.224	0.170	0.165	0.229	0.225	0.177	0.160	0.166	0.203
Shandong	0.291	0.257	0.200	0.263	0.280	0.288	0.290	0.243	0.242	0.261
Henan	0.215	0.201	0.223	0.260	0.299	0.226	0.202	0.186	0.175	0.221
Hubei	0.084	0.083	0.082	0.087	0.082	0.106	0.118	0.131	0.119	0.099
Hunan	0.328	0.246	0.228	0.261	0.232	0.239	0.248	0.180	0.183	0.238
Guangdong	0.183	0.164	0.211	0.222	0.222	0.231	0.153	0.132	0.157	0.186
Guangxi	0.216	0.154	0.110	0.180	0.146	0.136	0.119	0.109	0.088	0.140
Hainan	0.335	0.140	0.194	0.315	0.295	0.271	0.207	0.195	0.230	0.242
Chongqing	0.181	0.263	0.222	0.257	0.244	0.163	0.147	0.114	0.091	0.187
Sichuan	0.134	0.166	0.161	0.207	0.197	0.190	0.135	0.115	0.100	0.156
Guizhou	0.175	0.143	0.076	0.066	0.063	0.052	0.054	0.051	0.055	0.082
Yunan	0.187	0.167	0.155	0.165	0.126	0.139	0.095	0.071	0.060	0.129
Shaanxi	0.112	0.102	0.100	0.149	0.162	0.172	0.128	0.125	0.119	0.130
Gansu		0.109	0.122	0.149	0.167	0.149	0.133	0.145	0.140	0.139
Qinghai	0.069	0.101	0.168	0.101	0.081	0.101	0.075	0.081	0.063	0.093
Ningxia	0.227	0.189	0.121	0.115	0.080	0.087	0.074	0.060	0.060	0.112
Xinjiang	0.194	0.143	0.159	0.171	0.140	0.198	0.251	0.235	0.236	0.192
Province average	0.195	0.188	0.183	0.237	0.244	0.238	0.196	0.170	0.174	0.203

Notes:

1. Please refer to Table 2 for the definition of G_1.
2. Numbers for Jilin and Gansu in 1999 are missing as number of workers is missing in 1998 for the two provinces.
3. Returns are calculated based on α_1 = 0.122 in column (4) of Table 6.

Table 10 Return rate to G_2

Province	1999	2000	2001	2002	2003	2004	2005	2006	2007	Year average
Beijing	0.152	0.286	0.286	0.311	0.211	0.339	0.351	0.252	0.170	0.262
Tianjin	0.183	0.227	0.301	0.229	0.266	0.281	0.378	0.270	0.264	0.267
Hebei	0.205	0.188	0.203	0.236	0.335	0.456	0.414	0.304	0.272	0.290
Shanxi	0.188	0.184	0.212	0.150	0.207	0.295	0.273	0.265	0.308	0.231
Inner mongolia	0.260	0.195	0.187	0.199	0.186	0.155	0.156	0.142	0.164	0.183
Liaoning	0.209	0.215	0.307	0.259	0.373	0.481	0.382	0.358	0.221	0.312
Jilin		0.223	0.266	0.286	0.335	0.337	0.283	0.249	0.274	0.282
Helongjiang	0.177	0.281	0.241	0.220	0.281	0.438	0.459	0.406	0.312	0.313
Shanghai	0.208	0.250	0.371	0.300	0.354	0.276	0.269	0.219	0.174	0.269
Jiangsu	0.233	0.245	0.267	0.237	0.300	0.306	0.320	0.344	0.372	0.292
Zhejiang	0.188	0.169	0.153	0.156	0.217	0.285	0.220	0.184	0.180	0.195
Anhui	0.265	0.254	0.193	0.171	0.251	0.307	0.293	0.267	0.245	0.249
Fujian	0.189	0.168	0.205	0.178	0.208	0.302	0.296	0.267	0.230	0.227
Jiangxi	0.172	0.230	0.248	0.187	0.140	0.162	0.152	0.154	0.175	0.180
Shandong	0.270	0.237	0.243	0.325	0.338	0.503	0.523	0.552	0.477	0.385
Henan	0.310	0.302	0.226	0.212	0.210	0.244	0.275	0.254	0.255	0.254
Hubei	0.214	0.197	0.201	0.195	0.215	0.266	0.225	0.266	0.198	0.220
Hunan	0.218	0.203	0.181	0.210	0.303	0.289	0.413	0.402	0.368	0.287
Guangdong	0.195	0.189	0.187	0.192	0.251	0.335	0.316	0.341	0.314	0.258
Guangxi	0.155	0.167	0.185	0.177	0.217	0.305	0.262	0.242	0.251	0.218
Hainan	0.105	0.105	0.129	0.115	0.105	0.207	0.262	0.325	0.160	0.168
Chongqing	0.181	0.178	0.234	0.194	0.184	0.317	0.212	0.177	0.178	0.206
Sichuan	0.147	0.172	0.176	0.183	0.220	0.273	0.317	0.354	0.313	0.239
Guizhou	0.137	0.147	0.140	0.088	0.097	0.148	0.155	0.162	0.154	0.137
Yunan	0.127	0.134	0.123	0.122	0.160	0.174	0.184	0.140	0.136	0.144
Shaanxi	0.112	0.153	0.130	0.118	0.158	0.192	0.181	0.183	0.148	0.153
Gansu		0.155	0.129	0.134	0.167	0.246	0.201	0.195	0.240	0.184
Qinghai	0.224	0.205	0.083	0.083	0.086	0.136	0.130	0.131	0.126	0.134
Ningxia	0.115	0.119	0.108	0.108	0.126	0.187	0.169	0.175	0.174	0.143
Xinjiang	0.130	0.195	0.164	0.174	0.161	0.206	0.177	0.147	0.193	0.172
Province average	0.206	0.212	0.221	0.216	0.258	0.323	0.317	0.304	0.276	0.259

Notes:

1. Please refer to Table 2 for the definition of G_2.
2. Numbers for Jilin and Gansu in 1999 are missing as number of workers is missing in 1998 for the two provinces.
3. Returns are calculated based on $\alpha_2 = 0.233$ in column (4) of Table 6.