Road Infrastructure Investment and Allocative Efficiency: Evidence from China

Mingqin Wu[a], and Linhui Yu[b]

[a]South China Normal University, China

[b]Zhejiang University, China

March 2018

Abstract

This paper empirically investigates whether and how road infrastructure investment improves intra-industry allocative efficiency measured by firm markup dispersion using Chinese manufacturing data. Instrumental variable approach is employed to address the possible endogeneity issues and identify the causal relationship between road infrastructure and markup dispersion. Our baseline empirical results show that improvement in road infrastructure can significantly reduce markup dispersion in industries more reliant on road infrastructure. Further analysis on mechanim indicates that competition effect as reflected by reduced price dispersion dominates cost effect as reflected by reduced marginal cost dispersion in accounting for the flatened markup dispersion due to improved road transportation. Besides, the allocative effect of road infrastructure is more silient in inland regions than coastal regions, and is largely driven by entry and exit of firms.

Keywords: Road infrastructure; Allocative Efficiency; Markup dispersion; China

JEL Codes: O18; O47; R42

1 Introduction

It is widely believed that transport infrastructure invesment and spacial proximity to transport infrastructures are conducive to economic growth. Transport infrastructure is found to impact economy in various ways such as price, trade flows, real income, welfare, agriculture, etc. (Donaldson and Hornbeck, 2016; Donaldson, 2018). The underlying mechanism lies in that better transport infrastructure can reduce trade costs and narrow down interregional price gaps, which lead to increased market integration and reduced regional disparity (Démurger, 2001; Donaldson, 2018); It can also increase firms' access to external markets and expertises, and accelerate interregional factor flow and endowment mobility, which is essential for GDP and real income growth (Banerjee et al., 2012; Qin, 2014; Faber, 2014; Donaldson and Hornbeck, 2016; Baum-Snow et al., 2017a, 2017b). Given such an important role of transport infrastructure in promoting economic growth, few studies so far have formally investigated how transport infrastructure affect allocative efficiency within industries. Actually, this issue is of particular interst to economists because resource misallocation is common around the world especially in developing countries (Hsieh and Klenow, 2014; Banaerjee and Duflo, 2005). Recent studies have exploited the various sources of misallocation across countries and regoins (Restucccia and Rogerson, 2008; Midrigan and Xu, 2014; Lu and Yu, 2015), but it is still empirically unclear whether there is linkage between transport infrastructure and resource misallocation. Moreover, evaluating the effect of road infrastructure on allocative efficiency is also important for better understand the welfare effect and return of transport infrastructure investment.

This paper investigates whether and how construction of road infrastructure influences intraindustry resource allocation. Tot this aim, we empirically examine the effect of road infrastructure investment on markup dispersion of firms within narrowly-defined manufacturing industries. At macro level, dispersion of firm markups matters for welfare or allocative efficiency either in terms of consumer welfare (Arkolakis et al. 2017) or in terms of overall TFP of the economy (Edmond et al. 2015). At micro level, firms with higher markups employ resources at less than optimal levels, and firms with lower markups produces more than optimal level (Lu and Yu, 2015); thus first-best efficiency is achieved when markups converge across firms (products) within the same industry (Robinson, 1934). Improved road infrastructure, on one hand, promotes higher competition and thus condense markup dispersion; on the other hand, it also lowers input costs, in particular of industries relying on highway delivery. Such effects might not be even and may have mixed impacts on firm markup distribution. In terms of resource allocation, transport infrastructure can impact markup dispersion through two dimensions. The first one is the intensive margin, i.e., the adjustment of markups by surviving enterprises. The second is the extensive margin generated by entry and exit of firms. The intensive margin and extensive margin may cancel each other in equilibrium or outweigh each other due to changes in transportation cost caused by transport infrastructure.

Our emprical analysis hinges upon a comprehensive data of road length in China's provinces and Chinese manufacturing enterprises for the period of 1998-2007. To empirically investigate the effect of road investment on markup dispersion, firslty, we recover quantity-based firm-level markups following De Loecker and Warzynski (2012) and De Loecker et al (2016). Secondly, we calculate the dispersion of firm markups (Theil index and several other measures) for each industry and region; Thirdly, we use road length as proxy for road infrastructure investment in a region and construct transport reliance indicators for each manufacturing industry, and rely on their interaction terms to capture the effect of road investment on markup dispersion. To address the possible endogeneity issues and identify the causal effect of road infrastructure on markup dispersion, we use the US transport reliance indicator as the instrument for China in estimation. Fourthly, we explore the mechanism of how markup dispersion respond to road infrastructure; specifically, we regress markups at different quantiles to carefully examine the inner distributional changes of markups; we also examine how road infrastructure affect the distribution of two components of markups, i.e. prices and marginal costs. Lastly, we look into the heterogenous effects of road infrastructure on markup dispersion for different group of firms and in different regions.

Our empirical findings suggest that road infrastructure investment can significantly reduce markup dispersion of firms in industries more reliant on road transportation. These results are robust to alternative indices of markup dispersion and alternative transport reliance rate. Our regressions of markups at different quantiles show that road infrastructure increases markups at the lower quantiles and reduces markups at higher quantiles, clearly show the way how markup dispersion become flattened due to road investment. Our regressions of price dispersion and marginal cost dispersion respectively indicate that both of them are negatively affected by road infrastructure. Given markup dispersion is reduced due to road investment, price changes should outweigh cost changes, which implies that competition effect of road infrastructure dominates its cost effect. Our heterogenous effect tests show that the effect of road infrastructure is greater for inland regions than coastal regions, and stronger for entry/exit firms than surviving firms.

This paper is related to several strands of literature. The first strand of literature is on causal effect of a wide range of transport infrastructure on various economic performance at macro and micro levels. Beginning with Aschauer (1989), a large body of work estimates the effects of road infrastructure investment. Gramlich (1994) provides an excellent survey of the literature. Recent contributions have examined the effects on skill premia in labor market (Michaels, 2008), GDP growth (Banerjee et al., 2012), city growth (Duranton and Turner, 2012), and urban form (Baum-Snow et al., 2017a, 2017b). Shirley and Winston (2004) and Li and Li (2013) find that spending on roads saves firms' inventory costs. Li et al. (2017) show that road infrastructure investment can significantly increase firms' productivity. In contrast, Fleisher and Chen (1997) find that China's transportation infrastructure had no significant effect on its economic growth and firm productivity during 1979-1993. Faber (2014) show that highway network connections had negative growth effects among peripheral counties due to reduced industrial output growth in China.

The second strand of literature is related to the existing studies on transport infrastructure investment in Chinese context. Some of the available evidence supports a positive effect of transport on infrastructure investment. For example, Démurger (2001) employs province-level panel data to show that there is a significantly positive relationship between transportation infrastructure and regional growth. Using data from publicly listed expressway companies, Bai and Qian (2010) find that the average rate of private return to capital in the transport, storage, and postal service sector is around 20%. Faber (2014) support the evidence that reduction in transport cost moves economic activities away from periphery to agglomerated regions. More recent studies provide some microeconometric evidence on the mechanisms by which transport infrastructure influences economic growth in China. These mechanisms include social returns from saving on transport cost (Li and Chen, 2013); efficiency gain via inventory reduction (Li and Li, 2013); and industrial agglomeration or the geographical distribution of industries (Lu and Chen, 2006). Li et al. (2017) find that road infrastructure can increase enterprises' productivities significantly. Our study departs these literatures by examing the effect of road infrastructure on resource allocation.

The Third strand of literature related to our study examines the relationship between trade and average markups, such as the studies of Levinsohn (1993), Krishna and Mitra (1998), Chen et al. (2009), De Loecker et al. (2016), and Lu and Yu (2015). Most of these look at the level of firm markup rather industry markup dispersion. If trade can increase productive efficiency, firm markup may be decreased. Firms within each industry can be affected differently by trade, and thus a firm's efficiency can be increased or decreased within each industry. Therefore the welfare implication differ from the literature when we examine on markup dispersion in each industry.

This paper is the first to examine the microeconomic effect of road infrastructure on resource allocation, especially, there is little microeconometric evidence for the efficiency of transport infrastructure investment in developing countries (Li and Chen, 2013; Li and Li, 2013; Li et al., 2017)[1]. The rest of our paper proceeds as follows. The institutional background of road infrastructure in China is described in Section 2. Section 3 discusses data and variables. Section 4 is empirical strategy. Section 5 shows basic empirical

[1] Better access to transport infrastructure can foster economic growth in developing countries. Donaldson (2014), Banerhee et al. (2014), and Storeygard (2013) find that railways or roads in India, China, and Sub-Saharan Africa improve economic growth. Sotelo (2015) finds positive effects of paving Peruvian roads, with some areas negatively affected.

results; Section 6 further discusses the mechanism; and Section 7 concludes.

2 Background

China, one of the most populous countries in the world, is developing at an unprecedented rate.[2] The economic growth of China accelerated at the beginning of the late 1970s. However, transport infrastructure investment in China responded slowly initially, by 1990 the demand for transport services had surged, generating widespread traffic congestion. Before 1990s, the most majority of goods are moved by rail or river in China, and the freight ton-miles moved by road is only less than 5% (Park et al., 2002; Li et al., 2017; Baum-Snow et al., 2016, 2017). Between 1990 and 2010, China constructed an extensive modern highway network. According to China Statistical Yearbooks, road infrastructure investment in China has increased from below 2% of GDP in the mid-1990s to around 6% by the mid-2000s, which is much higher than the 4% average rate for developing countries (World Bank, 2005; Li et al., 2017). The total road length in China in 2015 is 4.57 million kilometers, four times more than the length in 1980. In contrast to road infrastructure, development of railways, the major competitor of roads, has been rather slow in China, at least before the recent high-speed rail boom (Figure 1). The railways grow an annual rate of 6% during 1980-2015, while the annual growth rate of road is 14% during the same period.

[Insert Figure 1 Here]

One thing that we need to note is that the financing structure differs significantly between roads and railroads, in terms of the degree of centralization. Railroads have mainly relied on the central government for financing, as ticket revenue does not accrue to local governments (Li et al., 2017). In contrast, road investment has mostly been financed by local governments through the collection of fees, including road tolls. This financing structure indicates that road infrastructure investment in China is more likely to be determined by local economic conditions, so it is important to understand Chinese local governments' incentives and behaviors on road infrastructure investment in order for conducting this analysis. For another thing, China's political and economic structure is traditionally based on decentralized regional planing, by which provinces are considered to be autonomous economic actors. Actually, studies have concluded that China's domestic market is more fragmented during the 1990s at provincial level (Cheng and Wu, 1995; Young, 2000; Poncet, 2003). For these reasons, we take province as the basic administrative unit for measuring road length and thereafter the analysis of markup dis-

[2] For example, fixed asset investment is 8,200 billion USD in 2015; this increases 600 times compared to the level in 1980, and GDP level is 10,000 billion USD in 2015; this is 149 times the level in 1980.

persion. Besides, we choose province because the jurisdiction's geographic scope should be large enough to cover the spatial spillover effect of roads (Li et al., 2017).

3 Data and variables

3.1 Data

Two main data sets are use in this research. The first is the total road length of China's provinces for the period of 1998-2007. We obtained this data from China Statistical Yearbooks compiled by the National Bureau of Statistics of China. The road length is measured at the end of each year, and does not include roads that may be on trial but not open for general use. Figure 2 shows China's national road length (in log) for the period of 1998-2007. It is obvious that the road length in China experienced a significant jump in early 2000s, which indicated the speed-up of road infrastructure construction since then. In Contrast, the increase of road length in US averaged 4% per year before 1973, but declined to 1% per year afterward (Fernald, 1999), implying that road construction is much faster in China than US in recent several decades.

[Insert Figure 2 Here]

The second is a merged data of the Annual Survey of Industrial Firms (ASIF) database (1998-2007) and a product-level production dataset (2000-2006) provided by the National Bureau of Statistics of China. The ASIF dataset contains detailed information for all state-owned manufacturing firms, as well as for non-stated-owned enterprises that have annual sales of more than RMB 5 million[3]; total output for these accounts for more than 85% of China's industrial output. As an important micro-level dataset in China, the ASIF data has been used in a rapidly growing body of research, including that of Song et al. (2011), Zhu (2012), Holz (2013), Berkowitz et al. (2017), Brandt et al.(2017), etc. A crucial step in obtaining firm markup involves the estimation of production function, which requires the observation of firm-level output in physical terms. As this information is missing in the ASIF data, we use product-level data from the National Bureau of Statistics of China for the period 2000–2006, which contains information on each product (defined at the five-digit product level) produced by the firm, and, in particular, output quantity. As the product-level data and the ASIF data share the same firm identity, we can easily match the two. The key dependent variable, firm markup, which is used to calculate markup dispersion, is estimated using this merged dataset.

[3]This is around 730 thousands USD.

3.2 Variables

3.2.1 Markup dispersion

As our key explained variable is the dispersion of firm markup, we should obtain markups for the first step. Markup is defined as the ratio of price over marginal costs, but firms rarely report price let alone marginal cost of their products, therefore we have to rely on production data to recover markups. We follow the method of De Loecker and Warzynski (2012) to recover firm level markups. Specifically, we assume that $F_{it}(L_{it}, K_{it}, M_{it}, \varpi_{it})$ is the output of firm i at time t where $L_{it}, K_{it}, M_{it}, \varpi_{it}$ represent inputs of labor, capital, intermediate materials, and firm specific productivity shocks.

Consider the following cost minimization problem of firm i at time t:

$$\min \ w_{it}L_{it} + r_{it}K_{it} + P_{it}^{m}M_{it}$$

$$s.t. F_{it}(L_{it}, K_{it}, M_{it}, \varpi_{it}) \geq \overline{Q}_{it}$$

$$L_{it} \geq I[D_{it} = 1]E_{it} \tag{1}$$

where w_{it}, r_{it},and P_{it}^{m} denote wage rate, rental price of capital and price of intermediate materials, respectively; D_{it} is a dummy variable, indicating whether the firm is state-owned enterprise (it takes 1 for SOE, and takes 0 otherwise); and $I[.]$ is an indicator function that takes a value of 1 if the statement in the bracket is true and 0 if not.[4]

Estimation of firm markup requires that choice of an input is free of any adjustment costs and estimation of its output elasticity. Since labor is not freely chosen, especially for SOEs, and capital is often considered to be a dynamic input, we choose intermediate materials as the input to estimate firm markup.

$$L(L_{it}, K_{it}, M_{it}, \lambda_{it}, \pi_{it}) = \ w_{it}L_{it} + r_{it}K_{it} + P_{it}^{m}M_{it}$$

$$+\lambda_{it}[\overline{Q}_{it} - F_{it}(L_{it}, K_{it}, M_{it}, \varpi_{it})]$$

$$+\pi_{it}[I[D_{it} = 1]E_{it} - L_{it}] \tag{2}$$

We take the first order condition to M_{it} :

[4] Since in China SOEs are usually required to employ the minimum level of employment, E_{it}, to fulfill the objective of increasing employment, it is difficult for SOEs to fire workers and quits are rare since SOEs are under political pressure to hire excess labor (Naughton, 1966; Cooper et al., 2015; Berkowitz et al., 2016).

$$\frac{\partial L}{\partial M_{it}} = P_{it}^{m} - \lambda_{it}\frac{\partial F_{it}}{\partial M_{it}} = 0 \tag{3}$$

We arrange the above equation and multiply both sides by $\frac{M_{it}}{Q_{it}}$, and derive

$$\frac{\partial F_{it} M_{it}}{\partial M_{it} Q_{it}} = \frac{P_{it} P_{it}^{m} M_{it}}{\lambda_{it} P_{it} Q_{it}}. \tag{4}$$

by rearranging the above equation, we can obtain the firm markup, which is defined as the ratio of price over marginal cost.

$$markup_{it} = \frac{P_{it}}{\lambda_{it}} = \frac{\partial F_{it} M_{it}}{\partial M_{it} Q_{it}} / \frac{P_{it} M_{it}}{P_{it} Q_{it}} = \frac{\theta_{it}^{m}}{\alpha_{it}^{m}} \tag{5}$$

Note that λ_{it} represents the marginal cost of production at a given level of output, and P_{it} is the price of final good. As the information of the expenditure on intermediate materials and total revenue are both available in the dataset, we can directly obtain this value from the dataset; while $\theta_{it}^{m} = \frac{\partial F_{it} M_{it}}{\partial M_{it} Q_{it}}$ is the output elasticity of intermediate materials, which requires estimation of the production function. According to this equation (5), when there is incomplete competition, markup is the wedge between the input revenue share and output elasticity of this input; $\alpha_{it}^{m} = \frac{P_{it} M_{it}}{P_{it} Q_{it}}$ is the share of expenditure on intermediate materials in total revenue.

To estimate production function and acquire θ_{it}^{m}, we follow De Loecker and Warzynski (2012) by using a translog production function, i.e.,

$$\begin{aligned} q_{it} = {} & \beta_l l_{it} + \beta_k k_{it} + \beta_m m_{it} + \beta_{ll} l_{it}^2 + \beta_{kk} k_{it}^2 + \beta_{mm} m_{it}^2 \\ & + \beta_{lk} l_{it} k_{it} + \beta_{km} k_{it} m_{it} + \beta_{lm} l_{it} m_{it} \\ & + \beta_{lkm} l_{it} k_{it} m_{it} + \omega_{it} + \epsilon_{it} \end{aligned} \tag{6}$$

where the lowercase letters represent the logarithm of the uppercase letters; ω_{it} is firm-specific productivity; and ϵ_{it} is an independent and identically distributed error term. There is a large literature on how to solve the estimation bias caused by the unobservable productivity shocks in estimation of production function (Ackerberg et al., 2007). We use the method developed by Ackerberg et al. (2015), thereafter refered to as ACF method, to solve this issue in this paper. We estimate the translog production function separately for each two-digit industry, and calculate the output elasticity of materials as:

$$\theta_{it}^{m} = \widehat{\beta}_m + 2\widehat{\beta}_{mm} m_{it} + \widehat{\beta}_{lm} l_{it} + \widehat{\beta}_{km} k_{it} + \widehat{\beta}_{lmk} l_{it} k_{it}. \tag{7}$$

With the estimated θ_{it}^{m}, we can readily calculate firm markups according to equation

(5). We need to clarify a few practical details when estimating the production function. First, to avoid omitted output price bias in the production function estimation, we use a merged dataset that contains the output in physical terms as described in the data section (Klette and Griliches, 1996). Second, labor input is in physical terms, while capital and intermediate inputs are in value terms. To backout the physical terms of capital and intermediate materials, we first deflate these values using the prices index developed by Brandt et al. (2012), and then use a control function approach developed by De Loecker et al. (2016) to correct the omitted firm-specific input prices.[5] Third, the above estimation assumes that each firm only produces one product, and this might not be the case in reality. So we focus on a group of single-product in the first step to obtain the estimators, and assume that multi-product firms use the same technology as single-product firms in the same industry. This allows us to calculate firm level markups across different products.

In Figure 3, we plot the mean markups for each 2-digit manufacturing industry from 1998 to 2007. The mean markup level is larger than 1 for most industries. The industry with the largest markup is Communication equipment, and the industries with markup levels less than 1 include Garment, footwear and caps, Rubber industry, and Furniture industry.

[Insert Figure 3 Here]

We use several methods to measure the dispersion of markups. The baseline method is a widely used entropy measure, i.e. Theil index. $Theil_{ijt}$ is calculated as follows:

$$Theil_{ijt} = \frac{1}{n_{ijt}} \sum_{f=1}^{n_{ijt}} \frac{y_{fijt}}{\overline{y}_{ijt}} \log(\frac{y_{fijt}}{\overline{y}_{ijt}}) \tag{8}$$

where y_{fijt} is the estimated markup of firm f in industry i, location j, and time t; $\overline{y}_{ijt}$ and n_{ijt} are the average markup and the number of firms in industry i, location j and time t, respectively.

Meanwhile, several alternative measures are employed as robustness checks; the first is the mean log deviation (MLD), which is defined as follows:

$$MLD_{ijt} = \frac{1}{n_{ijt}} \sum_{f=1}^{n_{ijt}} \log(\frac{y_{fijt}}{\overline{y}_{ijt}}). \tag{9}$$

The second dispersion measure is the coefficient of variation (CV), defined as the ratio

[5] To correct this omitted input price bias, we use a control function approach developed by De Loecker et al. (2015). Specifically, the omitted firm-specific input prices are assumed to be a reduced-form function of output prices, market shares, and exporter status and these factors are also interacted with the deflated inputs to construct a flexible control function. For the details of estimation procedure, please refer to Lu and Yu (2015).

of the standard deviation to the mean:

$$CV_{ijt} = \frac{\sqrt{V_{ijt}}}{\overline{y}_{ijt}} \tag{10}$$

where V_{ijt} is the standard deviation of firm markup in industry i, location j, and time t.

The third measure is the relative mean deviation (RMD), defined as the average absolute distance of each unit from the mean and expressed as a proportion of the mean.

$$RMD_{ijt} = \frac{1}{n_{ijt}} \sum_{f=1}^{n_{ijt}} \left| \frac{y_{fijt}}{\overline{y}_{ijt}} - 1 \right| \tag{11}$$

Table 1 reports the average markup dispersions measured by Theil index, MLD, CV and RMD for each two-digit manufactuirng industries. It is obvious that among all two-digit manufacturing industries, Chemical fibers industry has the largest markup dispersion, while the Ferrous metals industry has the smallest markup dispersion.[6]

[Insert Table 1 Here]

3.2.2 Transport reliance rate

Transport reliance rate measures to what extent an industry relies on road infrastructure. Due to the lack of quality data, however, the user cost of vehicle stock is unavailable for Chinese firms. As it is hard to obtain a direct measure of transport reliance rate, we instead construct an indicator to proxy for it. We use the national input-output table published by the National Bureau of Statistics in China in 2002 to construct this indicator. As the input-output table reports each sector's input value from the transport-related sectors including transport equipment manufacturing and transport services, we can calculate the input ratio of a sector from transport-related sectors. Thus, the transport reliance rate of industry i is defined as the ratio of its inputs from transport related industries to its total inputs, specifically,

$$s_i = \frac{\text{Value of inputs from transport-related industries for industry } i}{\text{Total value of inputs for industry } i}$$

s_i varies across industries but is time-invariant during our sample period. Since the sectors in input-output table are more broadly defined than the standard industry classfication, we then match them to each four-digit or three-digit industry according to the concordance table provived by the NBS of China. It worths noting that when counting input values, both the input value from the transport equipment manufacturing industry and the expenditure on road transport services should be counted. The latter informa-

[6]We exclude the tobacco industry from our analysis as (i) there are few observations, and (ii) this is a monopoly industry, protected by the government.

tion, unfortunately, is unavailable in China's input-output table in 2002. To alliviate this concern, we construct an alternative measure of transport reliance rate using China's input-output table in 2012 which reports each industry' expenditure on road transport-related services, and use this measure as a robustness check.[7]

3.2.3 Road length

Road length is commonly used as a proxy for total investment on road infrastructure, especially in developing economies (Donaldson, 2015)[8]. This measure is also applicable to China because Chinese transportation infrastructure investment has mainly been for new facilities rather maintenance (Li et al., 2007). Hence, the road length of China's provinces should be highly correlated with their total investment on road infrastructure. Another advantage of using road length is that it is much easier to measure than investment value, thus leaving little room for measurement errors.

3.2.4 other variables

Except the above key variables, there are some control variables in the regression. The first control variable is the degree of industry agglomeration across geographic regions. Industry agglomeration not only affects product price (competition effect) but also affects production cost (inputs sharing and spillovers), hence both of them are directly linked to the level as well as the distribution of firm markups (Lu and Yu, 2015). We employ the EG index to measure industry agglomeration by following the literature. The higher EG index means the higher degree of geographical concentration[9]. The second are two variables used to control for the size as well as entry barriers of an industry, including the industry average fixed assets and the number of firms in this industry.

The third control variable is the share of SOEs (in terms of output value) in an industry. While SOEs in China benefit from the monoploy power granted by the government, they also need to undertake some social responsibilities (e.g. maintain local employment); thus the proportion of SOEs in an industry (or region) may significantly affect the markup distribution of the industry (or region). The fourth control variable allows

[7] the disadvantages of using the input-output table in 2012 lie in that, year 2012 is beyond our sample period, i.e. 1998-2007, and importantly, industry structures experienced considerable changes with fast technology development in recent decades.

[8] Duranton and Turner (2011) used lane kilometers, which is caculated by multiplying the road length with the number of lanes of this road. This is obviously more accurate to measure road tranport capacity. The information on lane number, however, is unavailable in our data.

[9] Note that the EG index is calculated following Chen and Wu (2014):

$$EG_{ijt} = \frac{G_{ijt}-(1-\sum_j x_{jt}^2)H_{it}}{(1-\sum_r x_{jt}^2)(1-H_{it})}$$

where $G_{ijt} = \sum_j (x_{jt} - s_{jt})^2$ is the spatial Gini coefficient, and x_{jt} is the share of total employment of all industries in location j and year t, $H_{it} = \sum_i z_{it}^2$ is the Herfindahl index of industry i in year t, z_{it} is plant f's share of industry employment at year t.

for the effect of export on firm markup distribution. It is widely acknowledged that exporters and nonexporters are different in terms of productivity and markups within the same industry; and exporter's markup is a weighted average of markups in both domesitic market and foreign market. Thus, we use the output share of exporters in an industry to capture the possible impact of export on industry markup distribution. Lastly, we also use province-level railway length as well as its interaction with transport reliance to control for the effect of railways on markup disperion, because railroad and road are two complementary infrastructures of transportation.

In Table 2, we report the summary statistics of all the dependent variables and independent variables.

[Insert Table 2 Here]

4 Econometric strategy

Inspired by Fernald (1999) and Li et al. (2017), we specify the following baseline empirical model in equation (12)

$$y_{ijt} = \alpha_0 + \alpha_1 s_i dr_{jt} + \alpha_2 X_{ijt} + a_i + a_j + a_t + \varepsilon_{ijt}, \tag{12}$$

where y_{ijt} refers to the markup dispersion in industry i, province j and year t; s_i is the transport reliance rate of industry i; dr_{jt} is road growth rate in province j and year t; the interaction term between transport reliance rate and road length is used to estimate the differential effects of road investment on industry-and-region markup dispersion over time. The estimated coefficient of the interaction term between s_i and dr_{jt} is of our key interest and expected to be negative, meaning that the higher reliance of an industry on road trasport, the smaller markup dispersion of this industry in regions where more roads are built. In other words, improving road infrastructure of a region will reduce the markup dispersion of local industries with high reliance rate on road infrastructure.

Province fixed effects, a_j, account for all region-specific shocks that do not change over time. These shocks include local policy changes and different types of infrastructure investment. Year fixed effects, a_t, are included to account for all time effects that are the same across provinces and industries. Industry fixed effects, a_i, account for all industry fixed effects, such as technological improvements.

X_{ijt} is a vector of control variables that affect markup dispersion, including share of SOEs (in terms of output value) in an industry, industry agglomeration, industry average fixed assets, the number of firms in an industry, output share of exporters in an industry, railway length, etc.

There are two major sources of endogeneity issues in equation 12. First, transport reliance rate is suffered from measurement errors. On one hand, due to lack of quality data on transport reliance at disaggregated industry level, we have to rely on input-output table to obtain a proxy measure of transport reliance rate for each industry. On the other hand, Chinese input-output table may be affected by the globalization-led but industry-specific productivity shocks in China, such as information technology revolution in around 2000s; and these shocks may prevent us from accurately estimating the reliance of each industry on road transportation in gerernal purpose.[10] To address this concern, we construct an instrumental variable by recalculating the ratio s_i with the US input-output table in the same year. The transport reliance rates of US and China are highly correlated (their correlation coefficient is 0.728); we plot the transport reliance rates of manufacturing industries in the US and China in Figure 4.[11] In addition, the transport reliance rates in US should be uncorrelated with the China-specific industry technology shocks or adjustment, it can only affect Chinese industries through world technology diffusion channel; thus avoiding possible endogeneity bias.[12] It worths noting that although transport reliance rates in China and the US are constant during our sample period, their interactions with road length are changing over time, which enables us to identify our panel data estimates of differential effects of road infrastructure across industries.

[Insert Figure 4 Here]

Second, road infrastructure investment is possibly suffered from reverse causation issue. Although road construction decisions are usually made by the central or local governments based on their fisical revenue and budget, which makes these decisions exogenous to industry-specific factors, it is, however, possible that the relatively smaller markup dispersion (implying higher competition) of an industry may lead to larger demand for road investment; local government could therefore build more roads as response to this demand. To deal with the estimation bias caused by this reverse causality issue, we introduce another instrument variable, i.e., the historical road length of Chinese provinces in 1992 interacted with the tranport reliance rates in the US. Additionaly, we also include the one-year lagged road growth rate interacted transport reliance rate in the regression models to allivate this concern.

[10] The measurement errors of transport reliance rate caused by such issues may possibly bias the estimate toward zero.

[11] Transportation equipment manufacturing industry is excluded in the figure because it has much larger transport reliance rate than other industries.

[12] We choose US because it is at the world technology frontier; thus it is US that usually leads the technology improvement and industry upgrading of the rest of the world inlcuding China.

5 Impact of road infrastructure on markup dispersion

In this section, we conduct a series of empirical analysis to investigate the impacts of road investment on markup dispersion in manufacturing industries. The emprical results are reported in the following tables and the findings are also interpreted accordingly. Unless otherwise specified, all regressions control for industry, province, and year fixed effects; robust standard errors are clustered at province-industry level. In addition, the transporation equipment manufacturing industry, which has the exceptionally largest transport reliance rate comparing to other industries, is excluded from the analysis.

5.1 OLS estimation results

Tables 3 summarizes the baseline OLS estimation results of equation (12). In Column (1)-(3), we stepwisely add control variables into regressions; regression results reported in Column (2) control for the impact of agglomeration, entry barriers, output share of SOE and output share of exporters; In Column (3), we further control for the effect of railways by including the interaction term of railway growth rate with the transport reliance rate. The results shown in the above three columns indicate that the estimated coefficents of our key interest variables, i.e. the two interaction terms of (the level and the lagged) road growth rate and transport reliance rate, are both negative (albeit the level one is statistically insignficant). This means that the higher reliance of an industry on road trasportation, the smaller markup dispersion of this industry after more roads are built (or better road infrastructure is provided). Note that the coefficient of the interaction term between road growth rate and transport reliance rate has a structural interpretation: it measures the relative importance of road versus vehicle in providing transport services; multiplying this coefficient with the transport reliance of a four-digit industry gives the elasticity of the markup dispersion in each industry with respect to road investment.

[Insert Table 3 Here]

5.2 Instrumental variable estimation results

Since OLS estimation is biased if our key interest variables are endogenous. The first source of endogeneity comes from the measurement error of the transport reliance rates of Chinese manufacturing industires constructed using China's input-output table. To address this issue, we employ the same measure in the US, available in Fernald (1999), as an instrumental variable. While this US measure is highly correlated with that of China, it should not be affected by the industry technology catch-up or policy changes in China, which makes it qualified for serving as an exogenous instrumental variable of transport reliance rate in China. The IV estimation result reported in Column (1) of Table 4 show that the estimated coefficients of our key independent variables, i.e. the two interaction terms, are both negative, and the coefficent of the interaction term with lagged road growth is statistically significant at 1% level. Regarding the magnitude, the coefficent of lagged road growth*reliance (-0.539) is even larger than that of road growth*reliance (-0.120), suggesting that the effect of road infrastructure on reducing markup dipsersion is intertemporal rather than immediate.

The second source of endogeneity comes from the potential reverse causality of road construction and markup dispersion. We address this issue by further replacing the road growth rate with the historical road length of Chinese provinces in 1992, and use its interaction term with the tranport reliance rate in the US as an alternative instrument variable. The IV estimation result is reported in Column (2) of Table 4, which also supports that road construction can significantly reduce markup dispersion in industries with larger transport reliance rates.

[Insert Table 4 Here]

5.3 Robustness Checks

In this subsection, we conduct a battery of additional robustness checks on our results.

Alternative measures of dispersion. In order to test whether our estimation results are sensitive to different measures of dispersion indices, we replace the Theil index of markup dispersion with three alternative measures, i.e. MLD, CV and RMD, and show the IV estimation results in Column (1)-(3) of Table 5. Obviously, the estimated coefficients of the interactions are consistent with our baseline results; and the magnititude of the coefficients is even larger those estimated using Theil index. This robustness check supports our basic findings that improved road infrastructure reduces markup dispersion in industries with larger tranport reliance rate.

[Insert Table 5 Here]

Alternative industry classification. So far, our empirical analysis has been based on narrowly-defined industries, i.e., at four-digit industry classification. To alleviate any concerns about disaggregation bias, we do a robustness check by re-defining the industry at the three-digit classification level; thus the dispersion indices are constructed for each three-digit manufacturing industry in each province. In this practice, the number of observations in the regression sample becomes smaller comparing with the case of four-digit industry classification as a result of decreased number of industries. The IV regression result in Column (4) of Table 5 show that the estimates of the interaction terms (Lagged road growth*reliance and road growth*reliance) are both significant and negative, which indicates that our baseline results are robust to different classification of industry.

Alternative transport reliance rate. To construct transport reliance rate of an industry, both the input value from the transport equipment manufacturing and the expenditure on road transport services should be counted. As the information of an sector's expenditure on (or input from) tranport services is unavailable in China's input-output table in 2002, we then construct an alternative measure of transport reliance rate using China's input-output table in 2012 which has more comprehensive information on an sector's inputs from other sectors. The new index of tranportation reliance rate includes all inputs from transport-related sectors including road transport services and transport equipment manufacturing, we then use this alternative measure as a robustness check. The IV estimation result (using Theil index) as shown in column (5) of Table 5 is much similar as well as comparable to our previous results in terms of both magnitude and significance, suggesting that our empirical findings are not affected by the missing information problem in transport reliance rate calculation.

Larger geographic scope of markup dispersion. One more concern of our dependent variable (i.e. markup dispersion in industry i and region j) is that manufacturing industries may have complex input-output links that often cross provincial borders, and their markets are often broader than a province. To address this concern, we measure markup dispersion at a broader geographic scope, i.e., at national level, and replace the dependent variable y_{ijt} with y_{it}, i.e. markup dispersion of industry i in year t. The IV empirical results using the four measures of markup dispersion are reported in column (1)-(4) of Table 6. It is shown that in all four columns the estimates of both interaction terms are negative and statistically significant, providing further supports to our previous empirical findings.

[Insert Table 6 Here]

6 Further analysis on mechanism

Improved road infrastructure, on one hand, promotes higher competition and thus condense markup dispersion; on the other hand, it also lowers input costs, in particular of industries relying on highway delivery. Such effects might not be even and may have mixed impacts on firm markup distribution. To explicitly understand how road infrastructure investment affect markup dispersion, we disentangle changes of markup dispersion by 1) looking at the response of markup to road infrastructure and transport reliance at different quantiles; by 2) further decomposing markup into price and marginal costs and seperately identifying the differential responses of these two compents of markup; and by 3) heterogeneous effects analysis based on subsamples of firms.

Different quantiles. We first look into how firm markups at different quantiles (i.e. percentiles of 5%, 25%, 50%, 75%, and 95%) as well as at the mean level respond to changes in road infrastructure. The regression results using instrumental variables are shown in Table 7, which indicates that improved road infrastructure significantly enhances firm markups at lower quantiles (P25) while reduces firm markups at higher quantiles (P75 and P95); thus markup dispersion becomes flattened. In addition, we also find that road infrastructure does not impose significant effect on the mean markup level. As firm entry and exit mostly occur at the lower end, these results suggest that firm selection induced by road infrastructure improves markups of firms at lower quantiles; while competition generated by road construction negatively affects markup of firms at higher quantiles.

[Insert Table 7 Here]

Price vs marginal cost. As markup ratio contains the information on both price and marginal costs of a firm, hence the effect of road infrastructure can possibly work through price changes, marginal cost changes, or both. To explicitly understand the mechinism of markup dispersion, we further examine the effect of road infrastructure on price dispersion and marginal cost dispersion, respectively. The regression results reported in Table 8 show that road infrastructure has both negative and statistically significant effects on these two components of markup dispersion; it implies that price and marginal cost both matter as potential channels of reduced markup dispersion due to improved road infrastructure; but the former effect (price) obviously dominates the latter effect (marginal cost), suggesting that reduced markup dispersion is mainly caused by enhanced competition.

[Insert Table 8 Here]

Heterogeneous Effects. So far, we have estimated the average effect of road infrastructure on the dispersion of firm markups in Chinese manufacturing industries. To further shed light on how markup dispersion is affected by road infrastructure, we do the following two subsample analysis.

First, we distinguish between surviving firms and entry/exit firms, because they may respond differently to trade cost reduction due to investment on road infrastructure (Arkolakis, et al. 2017). We conduct sub-sample analysis on these two groups of firms, and report the IV regression results in Column (1)-(2) of Table 9. The results shows that both surviving firms and entry/exit firms are significantly affected by road infrastructure; but the negative effect of road infrastructure on markup dispersion is more salient for the group of entry/exit firms, and this can be observed from the magnitude and significance level of the estimated coefficients of interaction terms.

Second, we allow for the competition degree in different markets. The effect of road infrastructure on markup dispersion should be smaller in more competitive markets than in those monopolized ones (Edmond et al., 2015; Hsu et al., 2014); because markup dispersion in a competitive market is already small, which leaves little room for further reduction of markup dispersion. Following this argument, we divide all firms in our sample into two groups according to their geographic locations, i.e., whether they are located in coastal regions or inland regions. Comparing to China's inland regions, coastal regions are more open to world market due to location advantages and favorable government policies, e.g. special economic zones, coastal open cities, etc. Therefore, markets in coastal regions are more competitive than in inland regions in past decades. Our sub-sample regression results indicate that the negative effect of road infrastructure on markup dispersion is only significant for the group of firms located in inland regions (Column (3)-(4) of Table 9), which is consistent with our expectation.

[Insert Table 9 Here]

7 Conclusion

In this paper, we empirically investigate how road infrastructure affect allocative efficiency within narrowly defined manufacturing industries. The dispersion of firm markups, which usually reflects the degree of product market distortion, is employed to measure the allocative efficiency of resources. We estimate quantity-based firm markups following the method developed by De Loecker and Warzynski (2012), and use several approaches to measure markup dispersion. We employ the transport reliance rate in the US as well as the historical data of Chinese road length as the instrumental variables to address the potential endogeneity issues in estimation. Our basic empirical findings suggest that improvement in road infrastructure can significantly reduce markup dispersion in industries that are more reliant on road transportation. We also explicitly explain how markup dispersion is affected by road infrastructure through a battery of analysis on the underlying mechanism of this effect.

Different from recent literature on the economic impact of infrastructure investment (e.g., Li and Chen, 2013; Li and Li, 2013; Donaldson, 2015; Li et al., 2017), this paper provides a new perspective, i.e., it focuses on the potential benefits of transport infrastructure to allocative efficiency of resources by examining the firm markups distribution within manufacturing industries.

This paper has several policy implications. First, it further sheds light on the belief of the key role of transport infrastructure in promote economic growth and market development. For developing economies, in order to achieve sustainable and fast economic growth and improve market efficiency, it is necessary to further increase the transport infrastructure investment. Second, transport infrastructure investment should be directed to regions where local markets and industries are less competitive as transport infrastructure can improve allocative efficiency and reduce resource misallocation.

References

[1] Ackerberg, Daniel, C. Lanier Benkard, Steven Berry, and Ariel Pakes, 2007. Econometric tools for analyzing market outcomes. In Handbook of Econometrics, Vol. 6A, edited by James J. Heckman and Edward E. Leamer, 4171-4276. Amsterdam: North-Holland.

[2] Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, 2015. Identification properties of recent production function estimators. Econometrica 83(6): 2411-2451

[3] Arkolakis, Costinot, Donaldson, and Rodriguez-Clare, 2017. The Elusive Pro-Competitive Effects of Trade, forthcoming in Review of Economic Studies.

[4] Aschauer, David A., 1989. Is public expenditure productive? Journal of Monetary Economics 23 (2): 177-200.

[5] Bai, Chong'en and Yingyi Qian, 2010. Infrastructure development in China: The cases of electricity, highways, and railways. Journal of Comparative Economics 38 (1): 34-51.

[6] Banerjee, Abhijit, Esther Duflo and Nancy Qian, 2012. On the road: access to transportation infrastructure and economic development, NBER working paper 17897.

[7] Banerjee, Abhijit V., and Esther Duflo, 2005. Growth theory through the lens of development economics. In Handbook of Economic Growth, Vol. 1A, edited by Philippe Aghion and Steven N. Durlauf, 473-552. Amsterdam: North-Holland

[8] Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner and Qinghua Zhang, 2017a. Roads, Railroads and Decentralization of Chinese Cities, Review of Economics and Statistics 99(3): 435-448

[9] Baum-Snow, Nathaniel, Loren Brandt, J. Vernon Henderson, Matthew A. Turner and Qinghua Zhang, 2017b. Transport Infrastructure, Urban growth and Market access in China, working paper.

[10] Berkowitz Daniel, Hong Ma, and Shuichiro Nishioka, 2017. Recasting the iron rice bowl: The reform of China's State Owned Enterprises, Review of Economics and Statistics 99(4): 735-747.

[11] Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang, 2012. Creative accounting or creative destruction? firm-level productivity growth in Chinese manufacturing. Journal of Development Economics 97 (2): 339–51.

[12] Brandt, Loren, Johannes Van Biesebroeck, Luhang Wang, and Yifan Zhang. 2017. WTO Accession and Performance of Chinese Manufacturing Firms. American Economic Review, 107(9): 2784-2820.

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

[14] Chen, Natalie, Jean Imbs, and Andrew Scott, 2009. The dynamics of trade and competition. Journal of International Economics 77 (1): 50-62.

[15] Chen Bin and Mingqin Wu, 2014. Industrial agglomeration and employer compliance with social security contribution: Evidence from China, Journal of Regional Science 2014 54(4): 586-605

[16] Cheng Enjiang and Yanrui Wu, 1995. Market reform and intergration in China in the early 1990s. The case of maize. Chinese Economy Research Unit Working paper, University of Adelaide.

[17] Cooper, Russell, Guan Gong, and Ping Yang, 2015. Dynamic labor demand in China: public and private objectives, RAND Journal of Economics 46 (2015): 577-610.

[18] Démurger, Sylvie, 2001. Infrastructure development and economic growth: An explanationfor regional disparities in China? Journal of Comparative Economics 29: 95-117.

[19] De Loecker, Jan, and Frederic Warzynski, 2012. Markups and firm-level export status. American Economic Review 102: 2437-2471.

[20] De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik, 2016. Prices, markups and trade reform. Econometrica 84(2): 445-510

[21] Donaldson, D., 2018. Railroads of the raj: estimating the impact of transportation infrastructure, American Economic Review, forthcoming.

[22] Donaldson, D. and R. Hornbeck, 2016. Railroads and american economic growth: A market access approach. Quartely Journal of Economics 131(2): 799-858

[23] Duranton, G. and M A. Turner, 2011. The Fundamental Law of Road Congestion: Evidence from US Cities. The American Economic Review, 101(6): 2616-2652.

[24] Duranton, G. and M A. Turner, 2012. Urban growth and transportation, Review of Economic Studies 79: 1407-1440.

[25] Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, 2015. Competition, markups, and the gains from international trade. American Economic Review 105(10): 3183-3221.

[26] 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.

[27] Fernald, J. G., 1999. Roads to prosperity? Assessing the link between public capital and productivity. The American Economic Review 89(3): 619-638.

[28] Fleisher, Belton M. and Chen Jian, 1997. The coast-noncoast income gap, productivityand regional economic policy in China. Journal of Comparative Economics 25(2): 220-236.

[29] Gramlich, Edward M., 1994. Infrastructure investment: a review essay. Journal of Economic Literature 32 (3): 1176-1196.

[30] Holz, Carsten A, 2013. Chinese statistics: Classification systems and data sources, Stanford University, SCID Working Paper 471

[31] Hsieh Chang-Tai and Peter Klenow, 2009. Misallocation and manufacturing TFP in China and India, Quaterly Journal of Economics, CXXIV(4): 1403-1448.

[32] Hsu, Wen-tai, Yi Lu, and Guiying Wu, 2014. An anatomy of welfare effects of trade liberalization: The case of China's entry to WTO. Unpublished.

[33] Klette, Tor Jakob, and Zvi Griliches, 1996. The inconsistency of common scale estimators when output prices are unobserved and endogenous. Journal of Applied Econometrics 11(4): 343–61.

[34] Krishna, Pravin, and Devashish Mitra, 1998. Trade liberalization, market discipline and productivity growth: New evidence from India. Journal of Development Economics 56 (2): 447-462.

[35] Levinsohn, James, 1993. Testing the imports-as-market discipline hypothesis. Journal of International Economics, 35 (1-3): 1–22

[36] Li, Han and Zhigang Li, 2013. Road investment and inventory reduction: Evidence from Chinese manufacturers. Journal of Urban Economics (76):43–52.

[37] Li, Zhigang and Yu Chen, 2013. Estimating the social return to transport infrastructure: A price-difference approach applied to a quasi-experiment. Journal of Comparative Economics 41(3): 669-683.

[38] Li, Zhigang, Mingqin Wu, Bin R Chen, 2017. Is road infrastructure investment in China excessive? Evidence from productivity of firms. Regional Science and Urban Economics, forthcoming.

[39] Lu M., Chen Z., 2006. Market integration and industry agglomeration in regional economic development in China. SDX Joint Publishing Company and Shanghai People's Publishing House.

[40] Lu, Yi and Linhui Yu, 2015. Trade liberalization and markup dispersion: Evidence from China's WTO accession. American Economic Journal: Applied Economics 7(4): 1-34.

[41] Michaels, G., 2008. The effect of trade on the demand for skill Evidence from the Interstate Highway System, Review of Economics and Statistics, 90(4): 683-701.

[42] Midrigan, Virgiliu, and Daniel Yi Xu, 2014. Finance and misallocation: Evidence from plant-level data. American Economic Review 104 (2): 422-458.

[43] Naughton, Barry, Growing out of the plan: Chinese economic reform, 1978-1993. Cambridge, UK: Cambridge University Press, 1996.

[44] Park, A., Jin H., Rozelle S., Huang J., 2002. Market emergence and transition: arbitrage, transaction costs, and autarky in China's grain markets. American Journal of Agricultural Economics 84(1): 67-82

[45] Poncet Sandra, 2003. Measuring Chinese domestic and international integration. China Economic Review, 14(1): 1-21.

[46] Qin, Yu, 2017. No County Left Behind? The Distributional Impact of High-speed Rail Upgrades in China. Journal of Economic Geography, 17(3): 489-520.

[47] Restuccia, Diego and Richard Rogerson, 2008. Policy distortions and aggregate productivity with heterogeneous establishments." Review of Economic Dynamics 11 (4): 707-720.

[48] Robinson, Joan, 1934. The economics of imperfect competition. London: Macmillan.

[49] Shirley, Chad and Winston Clifford, 2004. Firm inventory behavior and the returns fromhighway infrastructure investments. Journal of Urban Economics 55 (2): 398-415.

[50] Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti, 2011. Growing like China. American Economic Review. 101(1), 196-233.

[51] Sotelo, Sebastian, 2015. Domestic Trade Frictions and Agriculture, manuscript.

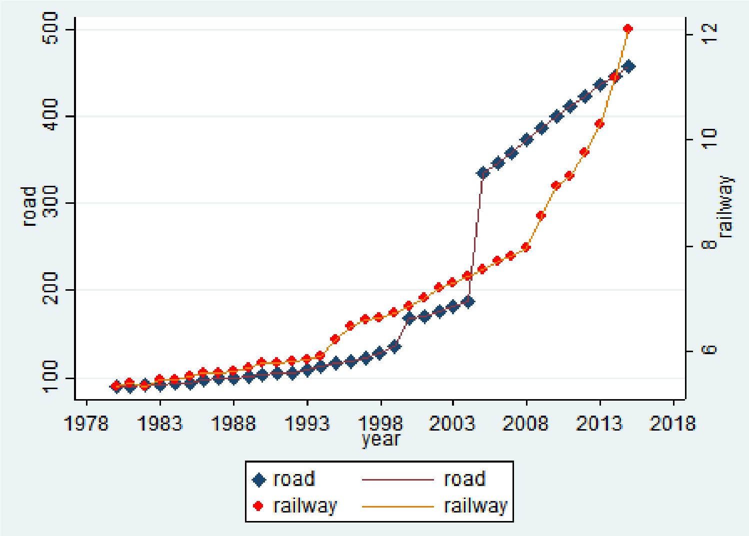
[52] Storeygard, A., 2013. Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa, World Bank Policy Research Working Paper 6444.

[53] World Bank, 2005. Integration of national product and factor markets-economic benefits and policy recommendations, Washington, World Bank.

[54] Young Alwyn, 2000. The razor's edge: distortions and incremental reforms in the People's Republic of China. Quarterly Journal of Economics, 115(4): 1091-1136.

[55] Zhu Xiaodong., 2012. Understanding China's growth: past, present, and future. Journal of Economic Perspectives, 26(4): 103-24

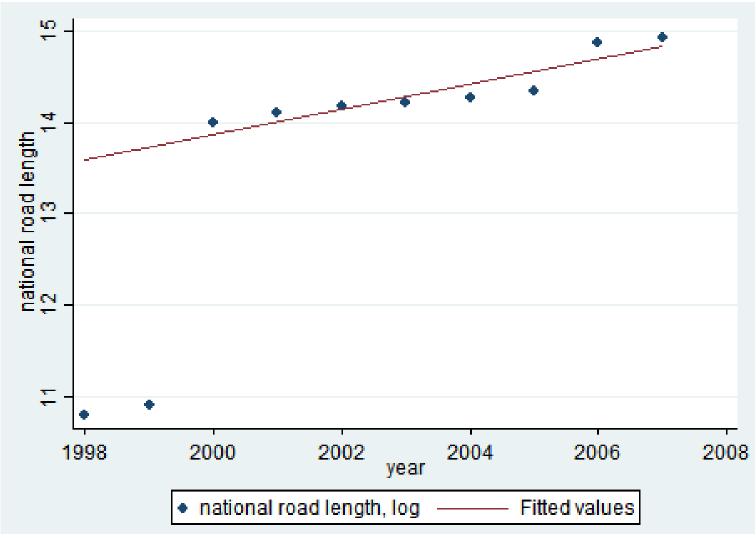
Figure 1 Road length and railway length in China (1980-2015)



Source: China Statistical Yearbook, 1980-2015

Note: The unit of road and railway length is 10,000 kilometers.

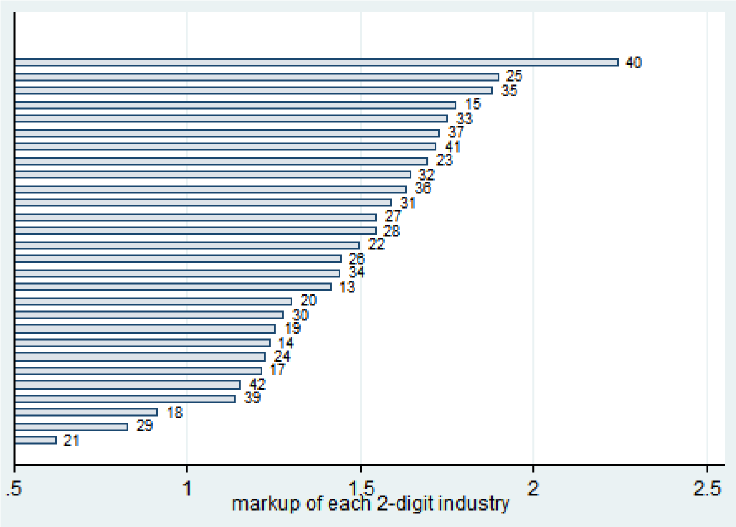
Figure 2 National road length (in log) during 1998-2007



Source: China Statistics Yearbook 1998~2007

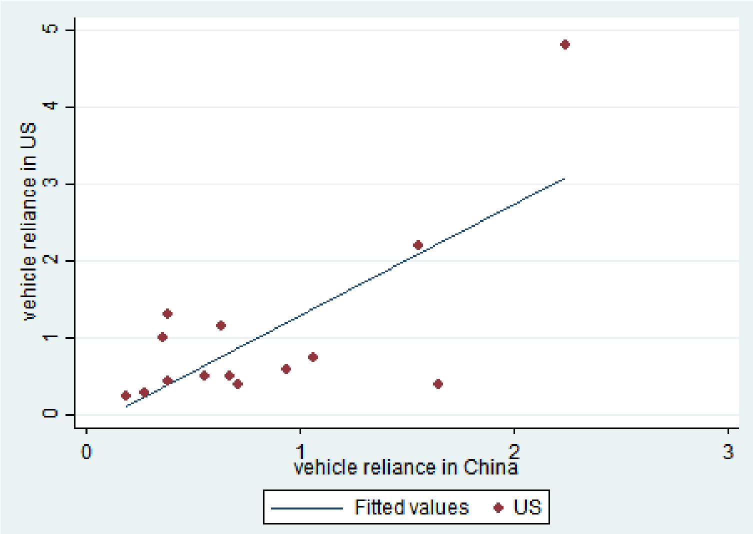
Note: The unit of road and railway length is 10,000 kilometers.

Figure 3 Estimated markup for each two-digit manufacturing industry



Note: 13-food processing; 14-food manufacturing; 15-beverage manufacturing; 16-tobacco manufacturing; 17-texitle; 18-Garment, footwear and caps; 19-leather, fur, and feathers products; 20-timber and wood; 21-Furniture; 22-paper products; 23-printing; 24-culture, education and sports articles; 25-petroleum and coking; 26-raw chemical materials; 27-medicine; 28- chemical fibers;29-rubber; 30-plastics; 31-nonmetallic products; 32-ferrous metals; 33-nonferrous metals; 34-metal products; 35-general machinery; 36-special machinery; 37-transport equipment; 39-electrical equipment; 40-communication equipment; 41-measuring instruments and machinery; 42-artwork and others.

Figure 4 Correlation of transport reliance rate between Chinese and the US manufacturing industries



Note: (1) The correlation coefficient of transport reliance rate between the US and China is 0.728. (2) Manufacturing industries include Food and kindred products, Textile mill products, Apparel & textiles, leather products, Wood and Furniture, Paper products, printing & publishing, Petroleum products, Chemicals, rubber & plastics, Primary metals, Fabricated metals, Miscellaneous manufacturing, Electronic equipment, Telecommunication, computers, Instruments and related, Electric utilities, and Gas utilities.

Table 1 Average markup dispersions for two-digit manufacturing industries.

Industry	Theil	MLD	CV	RMD
Food processing	0.014	0.014	0.170	0.122
Food manufacturing	0.021	0.019	0.213	0.150
Textile	0.011	0.011	0.146	0.110
Garment, footwear and caps	0.019	0.017	0.206	0.138
Leather, fur, and feathers products	0.012	0.012	0.152	0.116
Timber and wood	0.015	0.014	0.175	0.119
Furniture	0.019	0.019	0.192	0.151
Paper products	0.010	0.010	0.143	0.110
Printing	0.014	0.014	0.165	0.124
Petroleum and coking	0.012	0.012	0.154	0.116
Raw chemical materials	0.015	0.015	0.170	0.129
Medicine	0.018	0.018	0.190	0.147
Chemical fibers	0.026	0.026	0.223	0.182
Rubber	0.014	0.014	0.164	0.127
Plastics	0.012	0.013	0.156	0.121
Nonmetallic products	0.012	0.012	0.157	0.117
Ferrous metals	0.010	0.010	0.143	0.106
Nonferrous metals	0.011	0.011	0.145	0.110
Metal products	0.011	0.011	0.151	0.114
General machinery	0.012	0.011	0.153	0.112
Special machinery	0.016	0.015	0.181	0.132
Transport equipment	0.013	0.013	0.163	0.123
Electrical equipment	0.012	0.012	0.155	0.116
Communication equipment	0.014	0.013	0.166	0.124
Measuring instruments and machinery	0.015	0.015	0.175	0.132

Note: (1) markup dispersion is calculated for each four-digit manufacturing industry and then aggregated at two-digit industry level. (2) Theil, MLD, CV, and RMD are four measures of dispersion degree.

Table 2 Summary statistics of variables

Variables	Definition	Mean	S.D.
Theil	Theil index of markups	0.01	0.01
MLD	Mean log deviation of markups	0.01	0.01
CV	Coefficient of variation of markups	0.11	0.09
RMD	Relative mean deviation of markups	0.09	0.07
Theil_price	Theil index of average price	0.29	0.50
Theil_mc	Theil index of average marginal cost	0.29	0.50
Thiel_3 digit	Theil index of markups (3-digit industry)	0.01	0.01
Road growth	Growth rate of road length (%)	0.11	0.21
Reliance	Transport reliance rate (4-digit industry)	0.01	0.01
Reliance2	Transport reliance rate allowing for transport services (%)	0.01	0.01
Agglomeration	EG index (3-digit or 4-digit industry)	14.15	92.11
Number	Number of firms (in log)	1.66	1.36
Fixed asset	Average fixed assets (in log)	9.21	1.47
SOE ratio	share of output from SOEs (%)	0.13	0.31
Export ratio	share of export in total output (%)	0.12	0.22
Rail growth	Growth rate of railroad length (%)	0.04	0.13
Road_1992	Historical road length in 1992 (in log)	10.31	0.78

Note: data source (1) ASIF data of Chinese NBS 1998-2007; (2) China Statistical Yearbook 1992-2007; variables are constructed at 4-digit industry and province level unless otherwise specified.

Table 3 OLS estimation results

	(1)	(2)	(3)
Dep. variables: markup dispersion	Theil	Theil	Theil
Road growth*reliance	-0.066	-0.056	-0.048
	(0.045)	(0.043)	(0.043)
Road growth (t-1)*reliance	-0.020*	-0.026**	-0.023**
	(0.011)	(0.011)	(0.011)
Agglomeration		0.000	-0.000
		(0.000)	(0.000)
Average fixed asset		0.001***	0.001***
		(0.000)	(0.000)
Number of firms		0.003***	0.003***
		(0.000)	(0.000)
Output share of SOE		-0.001***	-0.001***
		(0.000)	(0.000)
Exported share of output		-0.001**	-0.001**
		(0.000)	(0.000)
Railway growth*reliance			0.007
			(0.042)
Railway growth (t-1)*reliance			-0.020**
			(0.010)
Road growth	0.002***	0.001**	0.001***
	(0.001)	(0.001)	(0.001)
Industry fixed effect	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes
Province fixed effect	Yes	Yes	Yes
Observations	43,508	43,415	43,045
R^2	0.142	0.214	0.214

Note: variables are constructed at four-digit manufacturing industry and province level unless otherwise specified; robust standard errors, clustered at industry-province level, are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1;

Table 4 IV regression results

	(1)	(2)
Dep. variables: markup dispersion	Theil	Theil
Road growth*reliance	-0.120	-10.564***
	(0.119)	(2.980)
Road growth(t-1)*reliance	-0.539***	2.817
	(0.206)	(2.999)
Control variables	Yes	Yes
Industry fixed effect	Yes	Yes
Year fixed effect	Yes	Yes
Province fixed effect	Yes	Yes
Observations	41,260	38,976
R^2	0.189	0.191
Underidentification test	[478]***	[9.33]***
Weak identification test	[180]***	[4.7]

Note: (1) variables are constructed at four-digit manufacturing industry and province level unless otherwise specified; robust standard errors, clustered at province-industry level, are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1

(2) Instrument variable is the transportation reliance of each industry in the US in column (1). Instrument variable is the interaction of road length in 1992 and the transportation reliance of each industry in US in column (2).

(3) The control variables include Road growth, EG index, Average fixed asset, Number of firms, Output share of SOE, Export share, Railway* reliance, and lagged railway*reliance.

(4) Underidentification test is Kleibergen-Paap rk LM statistics, and weak identification test is Kleibergen-Paap rk Wald F statistics.

Table 5 Robust checks I (markup dispersion measured at industry-province level, IV estimation)

	(1)	(2)	(3)	(4)	(5)
Dep. variables: markup dispersion	MLD	CV	RMD	3-digit ind. Theil	Theil
Road growth*reliance	-0.110	-0.827	-0.543	-0.373*	
	(0.120)	(0.803)	(0.708)	(0.206)	
Road growth(t-1)*reliance	-0.519***	-4.818***	-4.282***	-0.413*	
	(0.199)	(1.370)	(1.207)	(0.212)	
Road growth* reliance2					-0.049
					(0.128)
Road growth(t-1)* reliance2					-0.305***
					(0.113)
Control variables	Yes	Yes	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes
Province fixed effect	Yes	Yes	Yes	Yes	Yes
Observations	40,959	40,959	40,959	20,587	40135
R^2	0.186	0.374	0.278	0.183	0.20
Underidentification test	[478]***	[478]***	[478]***	[53]***	[951]***
Weak identification test	[180]***	[180]***	[180]***	[78]***	[406]***

Note: (1) variables are constructed at four-digit manufacturing industry and province level unless otherwise specified; robust standard errors, clustered at province-industry (4-digit), are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1

(2) Instrument variable is the transportation reliance of each industry in the US.

(3) The control variables include Road growth, EG index, Average fixed asset, Number of firms, Output share of SOE, Export share, Railway* reliance, and lagged railway*reliance.

(4) Underidentification test is Kleibergen-Paap rk LM statistics, and weak identification test is Kleibergen-Paap rk Wald F statistics.

Table 6 Robust checks II (markup dispersion measured at industry level, IV estimation)

	(1)	(2)	(3)	(4)
Dep. variables: markup dispersion	Theil	MLD	CV	RMD
Road growth*reliance	-0.125***	-0.126***	-0.803***	-0.773***
	(0.033)	(0.032)	(0.216)	(0.176)
Road growth(t-1)*reliance	-0.126**	-0.124**	-0.962***	-0.991***
	(0.056)	(0.052)	(0.363)	(0.296)
Control variables	Yes	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
Province fixed effect	Yes	Yes	Yes	Yes
Observations	686,793	686,793	686,793	686,793
R^2	0.732	0.727	0.745	0.690
Underidentification test	[3692]***	[3692]***	[3692]***	[3692]***
Weak identification test	[176]***	[180]***	[182]***	[172]***

Note: (1) variables are constructed at four-digit manufacturing industry or province level unless otherwise specified; Robust standard errors, clustered at province-industry level, are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

(2) Instrument variable is the transportation reliance of each industry in the US.

(3) The control variables include Road growth, EG index, Average fixed asset, Number of firms, Output share of SOE, Export share, Railway* reliance, and lagged railway*reliance.

(4) Underidentification test is Kleibergen-Paap rk LM statistics, and weak identification test is Kleibergen-Paap rk Wald F statistics.

Table 7 Markup regressions at different quantiles (IV estimation)

Dep. variables: markup	(1) Markup P5	(2) Markup P25	(3) Markup P50	(4) Markup P75	(5) Markup P95	(6) Markup Mean
Road growth*reliance	3.135	2.767	0.740	-1.229	-2.276	0.744
	(2.001)	(1.788)	(1.618)	(2.132)	(2.872)	(1.629)
Road growth(t-1)*reliance	4.641	5.856**	-1.075	-7.317**	-11.960***	-1.289
	(2.976)	(2.754)	(2.745)	(3.488)	(4.551)	(2.705)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Province fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	39,694	39,694	39,694	39,694	39,694	39,694
R^2	0.434	0.434	0.505	0.441	0.480	0.502
Underidentification test	[314]***	[314]***	[314]***	[314]***	[314]***	[314]***
Weak identification test	[123]***	[123]***	[123]***	[123]***	[123]***	[123]***

Note: (1) variables are constructed at four-digit manufacturing industry and province level unless otherwise specified; robust standard errors, clustered at province-industry level, are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1

(2) Instrument variable is the transportation reliance of each industry in the US

(3) The control variables include Road growth, EG index, Average fixed asset, Number of firms, Output share of SOE, Export share, Railway* reliance, and lagged railway*reliance.

(4) Underidentification test is Kleibergen-Paap rk LM statistics, and weak identification test is Kleibergen-Paap rk Wald F statistics.

Table 8 Effects of road infrastructure on price dispersion and marginal cost dispersion (IV estimation)

	(1)	(2)
Dep. variables: price/MC dispersion	Price Theil	Marginal cost Theil
Road growth*reliance	-25.599***	-25.202***
	(9.406)	(9.412)
Road growth (t-1)*reliance	1.681	1.751
	(8.055)	(7.994)
Industry fixed effect	Yes	Yes
Year fixed effect	Yes	Yes
Province fixed effect	Yes	Yes
Observations	23,821	23,821
R^2	0.381	0.379
Underidentification test	[138]***	[138]***
Weak identification test	[58]***	[58]***

Note: (1) variables are constructed at four-digit manufacturing industry and province level unless otherwise specified; robust standard errors, clustered at province-industry (4-digit), are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1

(2) Instrument variable is the transportation reliance of each industry in the US.

(3) The control variables include Road growth, EG index, Average fixed asset, Number of firms, Output share of SOE, Export share, Railway* reliance, and lagged railway*reliance.

(4) Underidentification test is Kleibergen-Paap rk LM statistics, and weak identification test is Kleibergen-Paap rk Wald F statistics.

Table 9 Heterogeneous effects (IV estimation)

	(1)	(2)	(3	(4)
Dep. variables: markup dispersion	Surviving Theil	Entry and exit Theil	Inland Theil	Coastal Theil
Road growth*reliance	-0.117	-0.305**	-0.227	0.117
	(0.138)	(0.131)	(0.148)	(0.330)
Road growth(t-1)*reliance	-0.596***	-0.614***	-0.899***	0.436
	(0.186)	(0.208)	(0.192)	(0.769)
Industry fixed effect	Yes	Yes	Yes	Yes
Year fixed effect	Yes	Yes	Yes	Yes
Province fixed effect	Yes	Yes	Yes	Yes
Observations	21,580	39,467	21,678	25,253
R^2	0.168	0.131	0.077	0.205
Underidentification test	[395]***	[396]***	[87]***	[23]***
Weak identification test	[126]***	[163]***	[39]***	[10]***

Note: (1) variables are constructed at four-digit manufacturing industry and province level unless otherwise specified; robust standard errors, clustered at province-industry (4-digit), are reported in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

(2) Instrument variable is the transportation reliance of each industry in the US.

(3) The control variables include Road growth, EG index, Average fixed asset, Number of firms, Output share of SOE, Export share, Railway* reliance, and lagged railway*reliance.

(4) Coastal provinces include Liaoning, Hebei, Tianjin, Shandong, Jiangsu, Zhejiang, Shanghai, Fujian, Guangdong, Guangxi, and Hainan. Inland provinces include Heilongjiang, Jilin, Inner Mongolia, Tibet, Ningxia, Henan, Gansu, Shanxi, Shaanxi, Sichuan, Qinghai, Xinjiang, Yunnan, Anhui, Jiangxi, Hunan, Guizhou, Hubei, Chongqing, and Xinjiang.

(5) Underidentification test is Kleibergen-Paap rk LM statistics, and weak identification test is Kleibergen-Paap rk Wald F statistics.