Transport Infrastructure and Productivity: The China Experience[*]

Steven Davis[†]

Meijun Qian[‡]

29/09/2020

Abstract

China has undertaken massive investments in its surface transportation system since the early 1990s, continuing to the present day. We study the impact of these public investments on plant productivity and resource allocation efficiency. We first collect and geocode highly detailed data on China's highway, railway and waterway transport system and its expansion in recent decades. We integrate these transport system data with longitudinal data on the location, inputs and outputs of nearly a half million manufacturing plants. We find that resource allocation is inefficient in China as unproductive units occupy more resources. Not surprisingly perhaps, plants in closer proximity to high-quality transport are more productive than otherwise comparable plants. Using a quasi-experimental approach, we find that productivity rises and its cross-plant dispersion falls with improvements in local transportation infrastructure. While plant entry and exit play an important role in these productivity responses, results suggest that exit and entry in China are inefficient and distorted. Among continuers, plants with relatively high total factor productivity (TFP) before treatment expand relative input usage after treatment. These and other results indicate that better access to transportation intensifies market competition, selects against less productive plants, facilitates the entry of new plants, and promotes the reallocation of factor inputs to relatively productive plants.

Key Words: Transport infrastructure, productivity, dispersion, allocative efficiency, China.
JEL Classification: F10, H54, O18, O40, R10

---

[*] We would like to thank the Booth School of Business for financial support and the University of Chicago Library for research assistant in developing the GIS database for the surfacing transportation system of China. We also thank Australian National University and Australian Research Council (ARC) for continuous support this project through ARC DP190103511.

[†] William H. Abbott Distinguished Service Professor of International Business and Economics, University of Chicago Booth School of Business and NBER, steve.davis@chicagobooth.edu.

[‡] Professor of Finance, College of Business and Economics, Australian National University, meijun.qian@anu.edu.au.

1. Introduction

China has undertaken massive investments in its surface transportation system since 1997, continuing to the present day. In the year 2016 alone, China invested 1.5 trillion CNY into road and railroad construction. Figure 1 in the Appendix summarizes the tremendous expansion in key aspects of China's transportation infrastructure from 1993 to 2013, including motorways, national Highways, railways, high-speed railways, and waterways. Specifically, Panels A and B depict the enormous growth in its National Trunk Highway System (NTHS); Panels C and D show the expansion of its major freight railways and high-speed passenger rail service; and Panels E and F show major navigable waterways. In a highlight, over the period from 1993 to 2013, among the 34,899 zip codes in China, expansion of the NTHS directly impacted 8,418 of them per year and 92% of the total over the 11-year span. These figures are measured in the sense that the locations are served by newly built segments of the NTHS, with a 25-km distance from the zip code enter point.

We investigate how this enormous expansion and improvement in China's surface transportation system affected productivity and resource allocation in its manufacturing sector. In doing so, we bring three contributions to the literature and the academic community.

First, we collect an extensive body of geospatial data on China's surface transportation system from 1993 to 2014 and digitize it into a form that supports detailed quantitative analysis by ourselves and others. We also digitize information about onramps and offramps to the major highways and rail stations, new additions of highways, speed improvement of railways, and new introduction of high-speed railways.

Second, we collect and verify the geolocation of the half million manufacturing plants covered in Chinese census data from 1998 to 2013. This geolocation data allow us to combine our geospatial data with the longitudinal data on manufacturing plants to measure their direct access and distance to the nearest access to any major surface transportation, enabling us to study the effects of China's expanding transport system on productivity and factor allocation in the manufacturing sector.

Third, to the best of our knowledge, this paper is one of the first few to examine the channels through which highways affect firm and aggregate productivity growth and allocative efficiency. Previous research focuses on whether a road or railroad connection affects GDP or population growth. Ghani, Goswami, and Kerr (2016) examine the effects of the Golden Quadrilateral project in India on population, GDP and labor productivity without estimating TFP or exploring the channels of productivity gains. Banerjee, Duflo, and Qian (2012) and Liu, Sheng, and Yu (2017) examine the impact of China's highway expansion on the level and growth of GDP and exports, respectively. Yang (2016) and Huang and Xiong (2018) examine similar questions as ours resulting from the highway expansion. However, their measures of transportation cost and market access are at the prefecture level. This cost measurement intuitively does not capture the transportation price for the plant. Plants that locate within a prefecture with motorways but far away from the exit or entry point face considerable costs before reaching the major transportation route. Our data enable measurement of each plant's distance to the nearest access point to a major surface transportation.

By exploring a quasi-natural experiment –the introduction of China's National Trunk Highway System(NTHS), high-speed railway, and the improvements in the railway and the waterway systems, this paper investigates the role of transport system improvements on industry-level productivity.[1] We conjecture that the improvements work through, first, selection and reallocation responses that shift factor inputs from less- to more-productive plants and, second, productivity gains within plants. We find that TFP increases and dispersion decreases over the sample period, especially for industries for which improved transportation access is vital. We find that, among the plants that are within a close distance to the transportation treatment, those that are situated at a higher position in the industry TFP distribution grow faster after that treatment than those situated at a lower position.

[1] Admittedly, Chinese infrastructure investments are not perfect quasi-experiments. We will apply several methods to address the endogeneity issue: to match treatment and control, to impose restrictions on qualified events, to use instruments in the difference-in-difference (DD) and triple difference (DDD) analyses. In contrast to the straight-line distance as in Banerjee, Duflo, and Qian (2012) and the least-cost path in Faber (2014), we will use historical routes as our instrument as Baum-Snow et al. (2017) did. Donaldson (2018) uses variation from the gradual rollout of the train network over time and shows that the effects hold above and beyond "placebo" rail lines that were planned but not implemented. Placebo comparison is also implemented in Donaldson and Hornbeck (2016). However, given the wide coverage of transportation changes in China and lack of a "placebo", it is challenging for us to implement the same strategy. The matching and instrumental approaches are likely the best we could do for the Chinese setting.

We also investigate exits and new entrants after treatments. We find that, while new entrant rates are higher in the treated sample, exit rates are higher in the matched control sample. This pattern suggests a reduction in economic activities in the controlled sample. We finally decompose the contribution of exits, entrants and reallocation among continuers to TFP improvement over the sample period. The results suggest that reallocation and within-firm improvement contribute to the overall TFP improvement more than exits and new entries.

Overall, the selection and reallocation responses we find alleviate the misallocation-based productivity shortfall in China relative to the United States highlighted by Hsieh and Klenow (2009). They demonstrate that China and India could gain a 30% to 60% improvement in manufacturing TFP if resources could be reallocated at an efficiency level similar to that of the United States.

Our findings are consistent with the work of Syverson (2004a), Raith (2003), and others. The within-plant productivity gains associated with transport system improvements may arise through the following mechanisms: First, an improved transportation system allows for better (or cheaper) access to a wide variety of intermediate inputs. Second, the competitive pressure created by a better transportation network intensifies incentives to improve managerial quality, or "X-efficiency". A quasi-natural experimental setting combined with within-industry difference-in-difference tests will allow us to distinguish these theoretical arguments from other demand- and supply-side causes, such as technology differences and the industrial policy impacts on productivity heterogeneity.

Extant literature shows some evidence of the association between transportation cost and productivity, but there is not a study that has established a direct link between infrastructure and allocation efficiency. Syverson (2004a, 2004b) argues that a lower transportation cost is associated with a higher density of competition, and therefore a lifted lower-bound, higher average productivity levels and less productivity dispersion. The mechanism behind this relation is survivorship, not allocation: lowering transportation costs increases the substitutability of products, and hence increases competition. Faber (2014) documents that, after the introduction of the national highway system, the growth in industrial output decreased in non-targeted peripheral counties because the highways connect metropolitan cities and peripheral large cities. The paper argues for a mechanism that lowering transportation cost between large cities intensifies the conglomeration process by implicitly assuming that firms in the metropolitan cities have either increasing returns to scale or higher marginal profits than others.

Using the Chinese example is desirable not only because China has the world's largest emerging economy, but also because of its vast regional imbalances in terms of both public infrastructure and ubiquitous state interferences that impose serious inefficiencies in the allocation of resources and potentially undermine the sustainability of the country's economic development. Infrastructure changes (e.g., access to new highways that reduce transportation costs) can correct misallocations that are associated with geographic remoteness. However, misallocations on extensive margins (e.g., government credit policies) are likely to persist regardless of geographic convenience or density of demand (Banerjee and Moll (2010)). Therefore, an experiment that improves physical but not economic infrastructure creates an opportunity to decompose misallocations and TFP inefficiency arising from different types of institutional deficiency.

Brandt, Van Biesebroeck, and Zhang (2012) present a comprehensive set of TFP estimates from the same dataset around China's entry into the World Trade Organization (WTO). They show that TFP growth, rather than input accumulation, accounts for the main source of output growth in China. Using reductions in tariffs associated with WTO entry, Brandt, Van Biesebroeck, and Zhang (2012) further show that a procompetitive effect dominates for incumbents' growth while efficiency gain dominates for new entrants. This pattern in export-oriented firms, however, is not necessary robust to all firms.

Infrastructure investment often coincides with economic growth, and at least in part is driven by economic development. Their interaction makes the identification of effects ambiguous: new roads boost growth and faster growth demands for roads. Therefore, we also examine the cross-sectional pattern of improvement associated with regional development, industry policies, and

firm ownership types. Future research could decompose misallocation due to physical as opposed to economic infrastructure. It could give a picture of China's potential economic development if its economic infrastructure were to be improved.

Another strand of literature analyzes the impact of transportation on economic welfare. Banerjee, Duflo, and Qian (2012) study whether access to better transportation enriches or impoverishes the affected regions. As transportation increases factor mobility, new roads could not only draw in new economic activities but also make it easier for human and physical capital to exit. Donaldson and Hornbeck (2016) show that the expansion the railroad system in the United States is fully capitalized into the affected counties' land value, and that the removal of access to that system leads to a 64% fall in land value. While our research is framed similarly, we focus the outcome measurement on and beyond firm-level TFP. Our measurement of aggregate-level gains will focus on productivity and resource allocation efficiency.

The rest of the paper is organized as follows. In Section 2, we review theoretical framework explaining the impact of transportation on productivity and allocative efficiency. Section 3 describes the data: the transportation data in the Geographic Information system (GIS), Manufacture data, and how they are mapped. Section 4 documents the TFP, dispersion and their relation to access to transportation. Section 5 documents the change in TFP, dispersion, input and output in response to the transportation treatment that the sample manufacturing plants experienced during the sample period, as well as plants' heterogeneity responses to the transportation treatment. Section 6 documents exits and new entrants in response to the treatment and decomposes contributions to the TFP through exit, entry and better allocative efficiency in continuous operations. Section 7 concludes the paper.

2. Theoretical Motivation

Lowering transport costs raises the mobility of inputs and outputs including human, material, capital and products. Increased mobility intensifies competition and reduces trade barriers. These changes potentially affect productivity through several mechanisms.

- A selection effect: i.e. the exit of the least productive plants and the entry of more productive ones;
- Improvement in X-efficiency through strengthening managerial incentives;

- Spatial redistribution of economic activities through trade and labor movements;
- Industry factors share adjustment through trade and technology agglomeration.

In Syverson's (2004a) framework, there exists a critical cost c^*, such that only entrants drawing $c<c^*$ will enter to produce. This critical c^* is an increasing function of transportation cost: when transportation cost is low, the critical c^* is low – only more efficient producers enter to produce. Improvement in the transportation system predicts a truncation of the productivity distribution of entry at the left tail. The entry of more efficient producers will also drive less efficient producers to exit. This mechanism works through product substitution. Decreasing transportation costs lower the substitution barrier. Production will be reallocated to select highly productive plants. More productive producers can grab market shares from less productive ones without sacrificing profitability.

Bloom, Sadun, and Van Reenen (2017) estimate that management practices, as a technology input, explains about 30% of the differences in TFP across firms and countries. Lack of competition makes it possible for firms to use inefficient production techniques and still stay in business. Raith (2003) demonstrates that product market competition enforces a positive influence on managerial incentives and hence the productive efficiency of firms. Improvement in the transportation system promotes product market competition, which should be followed by increased total and firm-level productivity.

The above theories are derived based on a single location, customer-driven and closed economy. That is, either a continuum of consumers or a number of firms enter the market and evenly position themselves around a circle of a circumference. Improvement in the transportation system expands economic activities, driving the local economy more open or more integrated with economic activities in nearby regions.

Redding (2016) develops a spatial model that incorporates the rich geography of trade costs and labor mobility with heterogeneous worker preferences across locations. It shows that not only trade shares but also population, and hence production factors and consumption demands may be reallocated. There are both theoretical and empirical studies on the impact of transportation costs on the spatial distribution of economic activities. As Redding and Turner (2015) summarize, many intercity and intracity studies show that adding roads, railroads and transit systems, or removing city barriers causes population to drop in city centers and employment to increase. However, Faber (2014) finds the opposite, but only for the non-targeted area.

Baqaee and Farhi (2019) show that reducing trade barriers, such as transportation costs, allows for trade and redistribution of inputs and outputs across industries and countries. The changes in factor shares improve resource allocation and increase aggregate output.

2.2 Algebra for sector misallocation

A sector's total factor productivity (TFP) is a geometric average of the average marginal revenue product of capital and labor in the sector. The sector productivity depends on each firm's total factor productivity and the distortion of resources allocation within the sector. As derived in Hsieh and Klenow (2009), when total factor productivity of quantity (TFPQ) and total factor productivity of revenue (TFPR) are jointly normally distributed, there is a simple closed-form expression for aggregate TFP:

$$LogTFP_s = \frac{1}{\sigma-1} Log\left(\sum_{i=1}^{Ms} TFPQ_{Si}^{\sigma-1}\right) - \frac{\sigma}{2} Var(\log(TFPR_{si}) \qquad (1)$$

where σ is the constant elasticity of substitution (CES) and M is the number of plants in the industry s. If marginal products were equalized across plants, industry TFP would be a log summation of firm TFP: $LogTFP_s = \frac{1}{\sigma-1} Log(\sum_{i=1}^{Ms} TFPQ_{Si}^{\sigma-1})$ Equation (1) therefore suggests that the negative effect of distortion in the sector can be summarized by the variance of the log TFPR among plants. The larger the dispersion of the marginal products, the worse is the misallocation in the sector.

The existing reduced-form studies on the impact of infrastructure on economic activities do not distinguish changes at the level from reorganization of existing activities in the observed effects. Identifying these two effects is critical for policy implications. Our study proposes an although-imperfect-but-intuitive approach: we use the shift in TFPR distribution to proxy for the change in the level of production, which captures the effects of lowering the input cost or forwarding technology associated with transportation costs, and the dispersion of the TFPR distribution to proxy for the reallocation of production.

3. Data and methodology

A. Highways, railroads and waterways in China

China's National Trunk Highway System (NTHS), also known as the "7918 network", is the world's largest expressway system by length. The name "7918" comes from its composition: 7 radial expressways from Beijing, 9 north-south expressways and 18 east-west expressways. Construction of the system began in 1992 under the National Trunk Highway Development Program, which intended to connect all provincial capital cities and cities with a registered urban population above 500,000. At the end of 2013, the total length of the network was 104,500 km, of which 8,260 km were built in 2013 alone. The system continues to expand today with the further intention to connect all rural areas. Our transportation data also include China's provincial highways developed after 2002, conventional railways, high-speed railways developed after 2004 and waterways.

Geospatial data for the transportation system

We digitalize and generate geospatial data for China's full transportation system, including highways, railways and waterways (but not airways) on a yearly basis from 1993 to 2014. To have consistent geo measures, we use the geospatial mapping in 2014 as the complete universe. The geo-data of the earlier years are therefore mostly a subset of that universe, identified by the actual maps for highways and waterways. For the railways, we double check with a list of changes and additions over the sample period recorded by the Ministry of Railway. For a road, railway, or waterway that existed during the sample period but have been abandoned by 2014, if it is on the physical map, we manually draw and add the data based on the scanned physical map. In our opinion, this approach to constructing the map data incorporates as much information as possible and provides a unified comparable measure. Admittedly, two potential biases might arise with this approach: First, the manually added routes are the best proxy but may not be the exact routes. The deviation is fortunately minimal given the 1:1,000,000 map scale. Second, accesses and exits to highways that existed in 2014 are assumed to exist since the routes existed. Stations and ports of railways and waters, however, are identified by actual maps for each year.

Appendix A details the definition of the transportation and coverage. Figure A1 presents the China motorway maps in the years 1993 and 2013, national highways as they existed by 1993, Railroads in 1993, high-speed railroads in 2013 and waterways in 1994. Year by year examination of the trunk highway shows consistency with the expansion phase described by the Ministry of Communication. The highway system expands dramatically over the sample period in years 1997, 2003 and 2007.

Using 25km or less to the nearest access point of a transportation mode, Appendix Table A1 shows that 12% of China's region, measured by zip codes, are covered by motorways (trunk highway). This ratio increased to 65% in 2013. The average distance to the nearest access to the motorway decreases from 298km in 1995 to 33km in 2012. Access to railway remains mostly the same, consistent with the fact that the improvement in the railway system during this period focuses on speed increases. The high speed railway, however, accounts for a much smaller portion of the accessible transportation network, as only 6% of the region falls within a distance of 25km. Appendix Table A2 shows that 85% of the region has been affected by transportation improvements, using within 50km to the nearest access as the threshold or 53% using 25km as the threshold. The majority of the changes come from the new motorway, new railway and railway speed improvements.

As the new motorway dominates the overall improvement in transportation, we contrast the coverage of trunk highway in 1993 with that in year 2013 in Figure 1a to highlight the drastic changes in surface transportation. We also show the year by year changes of trunk highway in Figure 1b.

[Insert Figure 1 about here]

B. The plant-level manufacturing data

Our plant-level manufacturing database covers all industrial state-owned firms (SOEs) and non-SOEs that have annual sales above five million RMB and some enterprises with smaller sales figures. Three caveats are noted. First, the unit of observation in this database is called a "legal unit". For large enterprises with multiple subsidiaries, each subsidiary enters the sample separately as long as it is registered as a legal unit. Given the dominance of plant/firm coincidence, 88.9% (96.6) of the units report as single-production units in 1998 (2007), we treat the database as plant level. Second, as non-SOEs need to meet a minimum sales amount to enter the database, this criterion affects our identification of a non-SOE's operational exit from decreasing of sales unless it reappears in the sample in a later year. However, there is no such concern for entries because the founding year is observed.

Although the dataset spans 1998-2013, we include 1998-2007 in the analyses and exclude the rest for the following reasons. First, the data in 2008, 2009 and 2010 have poor quality of observation. Second, data after 2010 include no intermediate input observations, making it impossible to compute TFP consistently with those before 2007. We also exclude plants that appear in the database for one year only and those that locate in the bottom 10% of counties by the number of plants hosted (288 counties that each hosts fewer than 13 plants). The final sample covers 457,805 plants and 2,048,671 plants-year observations over the sample period. These plants are located in 2,876 unique counties.

Annual financial and operational information of the firms are filed with the National Bureau of Statistics (NBS). The database also include information about the plants' industry, ownership, legal representative, physical location, contacts and political prefecture. Since China changed its political prefecture over time, we manually verify and align county code over time to be consistent with 2010 standards. Similarly, because the NBS changes industry classifications from the year 2002 to 2003, we adjust the industry codes to make the classifications consistent with those after 2003.

Table 1 presents the characteristics and ownership distributions of the sample plants. From the year 1998 to 2007, the number of manufacturing units more than doubles. The sample average age of manufacturing units dramatically drops from 38 to 10. These two patterns suggest enormous new entries, possibly due to both transportation changes and the entry into the WTO, among others.

While the assets grow about 28 % (from 529 to 677 million RMB), firms' sales more than double (from 348 to 817 million RMB). At the same time, there are slight decreases in the financial leverage and the average number of employees hired by each unit. These patterns suggest more efficient utilization of assets and labor inputs, consistent with Brandt et al (2012)'s claims that TFP growth dominates input accumulation as a source of Chinese manufacture output growth.

The ownership structure of these manufacturing units also experienced dramatic changes. Direct state and collective ownership together account for 75% of ownership on average in 1998 but the ratio drops to 14% in 2007. The combined corporation, private and foreign ownership on the opposite increases from 25% in 1998 to 86% in 2007.

[Insert Table 1 about here]

Geo-identification of plants' location

We use the longitude and latitude of each address to identify the plant's geolocation. Specifically, as manually coding millions of address' geo coordinators is unrealistic, we use the center point of each unique zip code (six digits and 34,890 independent observations) to approximate the geolocation of the plant. The zip code geo data are purchased from DataTang, Ltd, a data developing and sharing company in China.[2] We make sure of the data's reliability by first noting that the longitudinal and latitudinal observations have six decimal points and the location is detailed to the village level, and second manually verifying 100 random zip codes' geo data's corresponding village names from Google maps with the actual postal address. The accuracy of the zip geocode validates that our approximation is accurate with a drift less than 1.5 km, which is larger than the radius of most villages.

To allow for combination with other geo data such as population and GDP per capita at the county level, we choose the 2010 county classification as a consistent county boundary measure and base the county geo data on China's National Census 2010 GeoData, available at the China Data Center at the University of Michigan. As the county codes vary during the sample period due to administrative division changes, we manually reclassify them based

[2] We also contact a couple of companies in the U.S., such as Google, that claim to have China's longitudinal and latitudinal observations by zip code. We manually verify their data with Google maps and find that the deviation could often be over 20km.

on 2010[3] category for consistency. Political administrative division change of the county does not affect empirical results because our analyses rely on the physical locations of the plants relative to the locations of transportation networks. However, we will incorporate the political administrative change to address local government policy biases when matching treatment and controlling samples.

In Figure 2a, we plot plants' locations on the Chinese map with old roads as of 1998. In Figure 2b, we plot the NEW plants built during 1998-2007 on the map with the new additions of highways and high-speed railways during this period. As the figures show, all the existing and new entries are located near the highways and railways. For a more detailed view, Figure 2c zooms into the Bohai Bay on the northeast coast in 2007. The zoomed-in picture shows that the majority of the manufacturing plants are located within the 20km buffer of the motorways.

[Insert Figure 2 about here]

C. Estimation of productivity and misallocation

It is a challenge to work with the Chinese NBS firm-level data. We follow Brandt, Van Biesebroeck, and Zhang (2014) to treat the variables used. After preparing the correct input, output, price adjustment and real capital, we use a standard approach of same-year, same-industry cost shares coupled with the assumptions of cost minimization and Cobb-Douglas production to obtain factor demand elasticities.

Measuring plant-level TFP

Following Baily, Hulten, and Campbell (1992) and others, we compute plant-level log TFP as

$$\text{Log}TFP_{\text{et}} = \log Q_{\text{et}} - \alpha_{\text{k}} \log K_{\text{et}} - \alpha_{\text{L}} \log L_{\text{et}} - \alpha_{\text{M}} \log M_{\text{et}}, \qquad (2)$$

[3] The 2010 census shows 2,872 counties including the districts of metro-cities, of which 2,850 can be directly matched with those in the plant sample, covering 92% of the plant observations. For counties in the sample that do not have a direct match with 2010 due to the administrative change, we manually categorize the plant's county location based on its address zip codes' corresponding county in 2010. For any plant whose zip code is missing, we identify the county location based on other plants that have the same county code during the sample years and whose 2010 county code has been identified using the zip code. The last step also allows us to impute a zip code back for the firm. However, as a zip code covers a much smaller area than a county, we avoid using the imputed zip code for empirical analysis. Based on this approach, out of the 2,038,531 full sample, we observe 1,935,707 year*plants' geolocation.

where e and t index plant and year, respectively; Q is real output; K is real capital; L is labor input; M is materials; and α denotes factor elasticity.

Specifically, we measure plant output, Q, as the total value of shipments plus the change in inventories, including both finished products and work-in-progress, then deflated by industry-level price indices. Data on price indices are obtained from the Chinese National Bureau of Statistics in China.

Foster, Grim, and Haltiwanger (2016) and Davis et al. (2014) measure Labor, L, as the total hours of production and nonproduction workers. Our plant level data, however, document the number of employees and total wages paid rather than the hours. Therefore, we use the wage paid per worker.

Foster, Grim, and Haltiwanger (2016) and Davis et al. (2014) use the perpetual inventory method to calculate capital stocks, K, which requires continuous observations from firms' funding year. We apply the similar method, with an adjustment detailed in Brandt, Van Biescebroeck, and Zhang (2012, 2014) to compute real capital stock specifically for the same Chinese database. The procedure estimates the real value of capital stock in the first year that a plant appears in the dataset by inferring the average capital growth rate from the startup year using the age and the nominal capital stock information in year 1993 at the county and industry levels. This procedure assumes a constant growth rate over the years, the same investment deflator (only available at the country level) for all firms for each year, and a constant depreciation rate of 9% for all firms and all years. While this method is closest to the perpetual inventory method, available observations are limited and exclude several industries, including utilities and mining, which accounts for nearly 5% of all firms. Given the large heterogeneity in value freight ratio, the mining industry is particularly interesting for examining the transportation effect.

Foster, Grim, and Haltiwanger (2016) and Davis et al. (2014) are able to detail energy and non-energy material costs: parts, resales, outsourcing and etc. The observations on energy and non-energy in our dataset unfortunately are mostly missing. There are, however, quite complete observations on the aggregated intermediary inputs. Therefore, we use intermediary inputs to measure production materials, M.

We calculate four-digit industry-level cost shares for each input using the Chinese Stock Market & Accounting Research (CSMAR) industry classification system and the Chinese National Statistics Bureau. Measuring factor elasticities at the plant level rather than industry level could be problematic when factor adjustment costs exist (Syverson (2011)), which is common in China.

We use wages to measure labor cost and intermediate input for material cost. Their cost shares, α_L and α_M, take the value of cost over the value of shipments. Given the noise in the real capital stock and the lack of capital income information to infer the capital rental price, we again use two proxies for the capital cost share: the first is to measure it the cost over value of shipment and the second is to set the capital cost share one minus the other two shares, $\alpha_k = 1 - \alpha_L - \alpha_M$. As industry cost shares are noisy and using current year cost shares presumes no adjustment costs (Syverson (2011)), we use time average-cost shares (the current year and the past year) in the TFP calculation (except for the first year in the sample).

Using the above measurements, we compute TFP as in Equation (2) at its log value. Plants with negative values in the input or output are excluded from the computation. In computing the industry average factor shares, we also exclude plants whose own factor share is beyond 0 or 1. As our data have direct observations of labor, wages, inputs, outputs and net assets with unknowns denoted as missing, there is no payroll-industry-imputation issue as discussed in Dunne (1998) and Roberts and Supina (1996). Finally, we winsorize estimated TFP in the bottom and top 1% respectively, and adjust the estimated TFP by the industry-year mean to measure each plant's position in its own industry-year TFP distribution.

Estimate misallocation

Following Syverson (2011), we measure misallocation using productivity dispersion. An improvement in resource allocation results in a more centered distribution. Numerically, we measure dispersion using standard deviation, the value distance between the p99 and p1 distribution points, and the value distance between the p90 and p10 distribution points. We measure both the industry dispersion and locational dispersion.

D. Summary statistics of the estimates

Equal-weighted industry-year adjustment

We report the mean of log(TFP) and the distribution of industry-adjusted log(TFP) year by year in Table 2. Panel A reports the equal-weighted industry-year mean adjustment. We can see that country-wide average log(TFP) increases and the dispersion of industry-adjusted log(TFP) decreases over the sample period. Specifically, the standard deviation decreases from 0.74 in 1998 to 0.46 in 2007. The wedge between the 1% and 99% distribution points narrowed from 3.14 in 1998 to 2.17 in 2007. This reduction in dispersion comes more from the shift of the left tail toward the center than the right tail. While the cut off value at the 1% of distribution shifts rightward for a value of about 0.8, the cutoff values at the 99% shift leftward for about only 0.2 during the sample period. The cutoff values for other distribution points remain mostly the same.

Value-weighted industry-year adjustment

Panel B reports the industry-year adjustment that uses labor input or output value as weights. Compared to Panel A, Panel B shows two interesting patterns. First, the value-weighted industry-adjusted distribution shows the same pattern of dispersion declining over time. However, while the decline is driven by both ends, it mainly comes from the shift of the left tail towards the center. Second, the value-weighted industry-year mean is negative and significantly lower than the equal-weighted industry-year mean. This difference suggests that Chinese manufacturers are dominated by small, relatively more productive plants if measured by the number of units, but by large, relatively less productive plants if measured by market shares. Such an industry distribution suggests that the allocation of resources is inefficient: more resources go to unproductive units.

Dispersion and changes by geographic regions

Panel C report the distribution of industry-adjusted log(TFP) by region. The regions are classified as follows: (1) Eastern, including Beijing, Tianjin, Hebei, Shandong, Shanghai, Jiangsu, Zhejiang, Fujian, Guangdong and Hainan; (2) Central, including Henan, Anhui, Hubei, Hunan, Jiangxi and Shanxi; (3) Western, including Chongqing, Gansu, Ningxia, Qinghai, Inner Mongolia, Tibet, Shannxi, Sichuan, Guizhou, Xinjiang, Yunnan and Guangxi; and (4) NorthEastern, including Heilongjiang, Jilin and Liaoning.

As the panel shows, while productivity in all regions becomes less dispersed, the degree of improvement is larger for regions that were relatively less developed. The wedge between the 1% and 99% cutoff values is reduced by 1.68 in the Northeastern region,1.43 in the Central region, and 1.15 in the Western region, but only 0.97 in the Eastern region, which is the most developed region. Despite larger improvements, other regions remain less efficient than the Eastern region, whose standard deviation of the industry-adjusted log(TFP) and the wedge between 1% and 99% remains the smallest of all.

[Insert Table 2 about here]

Dispersion and changes by ownership type

In Appendix B1, we plot the distribution of industry-adjusted log(TFP) in 1998 and 2007 for the full sample and various subsamples by ownership, region and industry features. In addition to the same pattern in the full sample and cross-regions as illustrated in Table 2, Figure 3 also shows the pattern of productivity changes in SOEs versus that in non-SOEs. In particular, productivity in SOEs is more dispersed than in non-SOEs. However, this gap significantly narrows during 1998-2007. The narrowing primarily results from the striking improvement in the SOEs' productivity. At the same time, although non-SOEs have a right shift in their unadjusted log(TFP), the dispersion has become slightly larger. This makes sense because the business environment has greatly improved over years for the non-SOEs, allowing relatively less productive ones to enter and survive.

4. Empirical evidence: Static cross-section and time series trends

In this section, we focus on the static cross-sectional evidence and time series trends. Specifically, we establish the importance of transportation for productivity level and dispersion.

4.1 Access to transportation and TFP distribution in the cross-section

Transportation improvement facilitates firms' easier and cheaper access to a rich variety of intermediate inputs. On one hand, as predicted by Syverson (2004a) selection mechanism, less productive firms do not survive when situated within a competition circle.

On the other hand, firms in the remote parts of the transportation system have less access to information about technology, management and business practice advancements, hence tend to be at the back of the technology frontier.

We hypothesize that plants near highway exits or entrances should situate higher in the TFP distribution. To test this conjecture, we use GIS to calculate each plant-year's distance to the nearest highway's exit and entrance point, respectively, for all surface transportation methods. We then run the panel regression, where the dependent variable is the industry-year adjusted log(TFP), i.e., the position of the plant in its own industry-year TFP distribution, which is equivalent to controlling for the year*industry fixed effects. The explanatory variables are the distance to the nearest transportation access point. We include plant size, age, ownership and province-year fixed effect in the specification. Including province-year fixed effect is equivalent to controlling for local economic development, population, resources and industry policies.

$$Log\,TFP_{it} = \alpha + \Sigma\beta_{kt}{*}Distance_{ikt} + \Sigma\beta_{jt}{*}Characteristics_{ijt} + Province\text{-}Year\ FE + \varepsilon_{it}\ ; \qquad (3)$$

where, $Log\,TFP_{it}$ is adjusted by industry-year mean and standard deviation, so that the value represents plant *i*'s location on its own industries' TFP distribution in year *t*, in the scale of its own industry-year standard deviation. $Distance_{ikt}$ is plant i's distance to the nearest access point for transportation type *k* in year *t*; $Characteristics_{ijt}$ is plant i's characteristic *j* in year *t*, including all various fixed effects.

The first row in Table 3 reports results from regressions where each transportation type enters the regression, one by one. The second row in Table 3 reports results from regressions where all transportation types enter the regression simultaneously. In Panel A, the dependent variable is industry-year adjust log(TFP) measured as a percentage of standard deviation. All coefficients negative and mostly significant at the 1% level, except for high-speed railways. For example, the industry-year adjusted log(TFP) has a coefficient of -1.85 on the distance to the nearest access to a motorway. In terms of marginal effect, for every 1km the plant is further from the motorway access or exit points, its TFP, controlling for characteristics, ownership and regional effect, moves downward by 1.85% standard deviations in its own sector. This negative relation applies to all other transportation modes, except high-speed railways.

We run the same regression with the labor-weighted or shipment value -weighted industry-year adjusted log(TFP), measured as a percentage of standard deviation, as the dependent variable. As Panels B and C show, the results are qualitatively the same.

[Insert Table 3 about here]

To visually present the relation, we conduct a non-parametric analysis with the distance, then plot the estimated coefficients for each distance range in Figure 3, Panel A. As the plot shows, The TFP is lower when plants locate further away from transportation access points. In Panel B, we plot a heat map by the access to a motorway in 2004. The heat is determined by the average TFP. As the map shows, TFP is higher when the plants are closer to transportation.

[Insert Figure 3 here]

4.2 Locational dispersion

For each location and year, we compute the average deviation of TFP, measured by the absolute value of industry-adjusted TFP, for all plants at that location.

$$DEV_l = \frac{\sum_{i \in l} |\log(TFP_{l,i})|}{N_l} \tag{4}$$

where *l* denotes the location and *i* the plant.

We then run regressions at the location level, with the above deviation as the dependent variable, and the access to surface transportation and local plant density as the explanatory variables. Province-year fixed effects are also controlled. With a similar approach in Figure 3, we run the non-parametric regression and plot heat map for the dispersion of TFP by location.

As Figure 4, Panel A shows, the locational dispersion is larger when the location is further away from transportation. In Panel B, the heat is the locational TFP. The heat is lower when the location is closer to transportation, measured by the 2004 motorway.

[Insert Figure 4 here]

4.3 Time series trend of TFP distribution by industry reliance on transportation

The impact of transportation on productivity dispersion intuitively differs across manufacturing units due to their reliance on a transportation system. For example, the concrete industry has low value-to-weight ratios, hence is confined to local competition, allowing inefficient plants to survive unless a large reduction in transportation cost occurs.

A: The concrete industry as an anecdotal example

In Figure 5a, we plot the distribution of industry-adjusted log(TFP) in the years 1998 and 2007 for the concrete industry. We include plants that produce cement (3121, 3123 and 3124), brick, tile and stone construction materials (3129, 3131 and 3133), but exclude part of the cement industry (3111, 3112, 3122 and 3132) because the number of plants coded in these industries dramatically fluctuates across sample years.

As the figure shows, from 1998 to 2007, the distribution of TFP in the concrete industry becomes much more concentrated in the middle and slightly shifts towards the right. The changes for the concrete industry (Figure 5a) are much more obvious compared to the changes for the full sample (Appendix B).

B: Low value- to-weight versus high value-to-weight

We also contrast TFP distributions in 1998 and 2007 for (e) high value/weight industries and (f) low value/weight industries. The high value over weight industries include gold, silver mining (921, 922), cigarette tobacco products (1620, 1690), silk products (1754, 1763), leather and fur products (1921, 1922, 1923, 1932), cosmetic and fur products (2672, 2674). The low value over weight industries include ore mining (933), tobacco stemming and re-drying (1610), stone quarrying and mining (1011, 1012 and 1013) and wood and bamboo furniture (2110, 2120).

As Figures 5c-5f show, while the dispersion of industry-adjusted log(TFP) become smaller for both groups, the decline in the high value/weight industry is driven by the reduction of markup in the right tail and in the low value/weight industry on both ends. Examining the unadjusted log(TFP) shows that, while both groups radically reduce dispersion over time, industries with low value over weight ratios remain with lower productivity and a larger dispersion.

C: Transportation access at the beginning of the sample period

We also contrast TFP distributions in 1998 and 2007 for (m) plants that have poor transportation access in 1998 but good access in 2007, (o) plants with good transportation access in both 1998 and 2007, and (q) plants with poor transportation access in both 1998 and 2007. Poor access is defined as having no access of any type within 25km, and only one type of access within 50km. Good access is defined as there are at least three types of access within 50km.

As Figures 5m-5q show, among the three groups, only the one with transportation changes from poor to good has an obvious right shift in the unadjusted log(TFP) and the distribution of the industry-adjusted log(TFP) becomes more centered. For plants with always-good or always-poor transportation access, there is no obvious improvement in their productivity and allocation. While plants that situated on the left side seem to disappear from 1998 to 2007, those that situated on the right side appear to enjoy a larger markups in 2007 than in 1998. This result is consistent with Faber (2014) that the new transportation changes further isolate those regions that are skipped in the new network.

[Insert Figure 5 here]

5. The quasi-natural experiment approach

This section examines the treatment effect of transportation infrastructure changes on productivity. As increased competition eliminates the least productive plants and shift resources to the more productive ones, this correction of misallocation induces a more centered distribution. The treatment effect will differ across plants depending on where they are situated on the pre-treatment productivity distribution.

5.1 Transportation treatment and productivity changes: RDD

Define and match treatment and control

For each year, we identify new segments of highways, improvement in the railways and waterways. For the highway and high-speed railway systems, we identify new additions to routes, accesses, exits, or stations. For the conventional railways and water systems, we identify improvement in the speed, transportation capacity and new additions to stations or ports.

The treatment sample starts with all plants that are located within a distance (in km diameter-circle) from any of the new NTHS accesses, exits and railway stations. As such, we have four panels of treatment events based on the distance criteria: 5km, 10km, 25km and

50km. Appendix Table A2 summarizes the treatment events by different criteria. Table A3 summarizes the total treatment sample, using the loosest 50km criteria. Here, we elaborate the approach and present the results, specifically for the 25km criteria.

The treated plants are located within 25km of the access point to any new additions to transportation in that year, conditional on no previous access within 25km. The controls are plants that don't experience any treatment with the 25km criteria, during the five-year window of *[t-2: t+2],* conditional on no previous access within 25km. There are 1,779 treatments and 22,405 controls at the zip code level. We plot all the treated and control locations in Figure 6, Panel A.

Using regression discontinuity design (RDD), we match the treated and control plants based on year, province and neighboring location. The matched sample has 1,440(*2) pairs. We plot the matched treatment and control locations in Figure 6, Panel B. There are 38,091 plants located in the matched treated locations and 12,986 in the matched controls.

We are less concerned about idiosyncratic events that might have created a bias around a specific event or in a specific region, because they are likely to be only noises in the large panel of events. The province-year fixed effect will also reduce the influence of such noises. As follows, we will examine cross sectional patterns of the TFP changes around the event to establish a link with transportation. In short, our study does not argue that infrastructure is the only cause for TFP changes, but seeks to prove that it is indeed one among many that matter.

Treatment effect and falsification tests

Despite clear theoretical frameworks and evident cross-sectional patterns, several factors may obfuscate the estimated treatment effect. First, in an emerging economy that experiences continuous economic and institutional reforms and is exposed to extensive political interference from the central and local governments, it is inevitable that many other unobservable events could randomly coincide with the infrastructure changes.

Second, due to the Chinese central government's five-year National Development Plans, there is often a five-year cycle during which certain particular industries boom. We hope that matching on industry and year significantly reduces the bias caused by the government's industrial favoritism policy.

Finally, while it is salient that, based on the trend of productivity distribution, resource allocation efficiency improves over time, which factors have contributed to the efficiency improvement is unclear. Government's industrial policies, infrastructure changes and institutional improvements may all have contributed to this efficiency improvement. These factors may move in the opposite direction for reasons irrelevant to transportation.

Nevertheless, we present the treatment effect and falsification tests in Table 4. The dependent variables are industry-year adjusted log(TFP) and the growth rate of input and output, respectively. The treatment effect is estimated by assigning year *t*, the true event year, as the event year in estimation. The falsification tests are conducted by assigning year *t+/-s* as the event year in estimation.

Not surprisingly, the estimated treatment effect is not as clear cut as we usually see in an event study. However, the pattern in Panel A clearly suggests the following: (1) The reversal causality story can be excluded. The transportation changes are not likely driven by the productivity increases, as the falsified *t-1* and *t-2* event years have an insignificant or negative impact on the productivity change indicators. (2) There are some significantly positive treatment effects in the years *t* and *t+1*.

In Panel B, we examine the three-year mid-term effect. The growth of productivity, input and output are measured from *t-1* to *t+2* or from *t* to *t+3*. The estimation shows a mostly a positive and somewhat significant treatment effect.

[Insert Table 4 here]

5.2 Instrumental approach

Another approach to addressing the potential endogeneity issue is to use instruments. Following Baum-Snow et al. (2017), we utilize the historical 1962 road and railway network as the instrument for the transportation treatments that occurred in our sample period. This identification strategy is valid in the sense that the roads and railroads in 1962 affect plants current performance only because of their influence on the current transportation network. To measure the instrument, we calculate the distance between each zip code and the nearest road and railway in 1962.

In the first stage, we use the historical road access to explain the transportation changes that occurred in each treated location. In the case of simultaneously occurring changes, we use the one within the shortest distance. We also include the location's distance to the nearest prefectural-level or county-level city, whichever is closer. In the second stage, we use the predicted distance of transportation changes to define treatment and control, and match them based on year, province and neighbouring locations. The dependent variables in the 2nd stage are the same as those in Table 4. The fixed effects, control variables in Table 4, and those in the 1st stage are all included.

$$\text{Distance to new access} = \alpha + \beta_1 Road1962 + \beta_2 Rail1962 + \gamma X + \varepsilon$$

$$Log(TFP)\ or\ growth\ of\ Q, L, M, K = \alpha + \beta * Predicted\ Treated$$
$$+\beta_1 Road1962 + \beta_2 Rail1962 + \gamma X + FE + \varepsilon \tag{6}$$

We present the results in Table 5. The first column shows the results from the 1st stage. The distance to the historical roads and railroads significantly predict the distance to the new transportation access points. The rest three columns present the regression results for the zip code-level average TFP in year *t+1* after the treatment, where the treatment is defined as the predicted distance from the 1st stage. As the columns show, the equal-weighted, labor-weighted and output-weighted industry-year adjusted log(TFP) at the zip code level is significantly higher in the treated locations than in the matched control locations. These results suggest that transportation improvement positively influences productivity and resource allocation.

[Insert Table 5 here]

5.3 Asymmetric effects

The impact of resource reallocation upon transportation cost reduction depends on firms' current competitiveness or technology (Syverson (2004b)). Firms that are located at the high end of the TFP distribution are likely to gain more resources, when new access to transportation is introduced. This effect should also be larger for industries with high shipment cost over value than for those with low shipment cost over value.

To conduct the test, we divide the treatment plants into four quartiles and examine how productivity, input and output grow three years after the treatment and differ across the quartiles. We run the following regression to detect the asymmetry in difference-in-difference.

$$Growth_{it-1:t+3} = \alpha_i + \Sigma\gamma_k{*}Treat{*}Quartile_k + \gamma_{ij}{*}Control_{ij} + Province\text{-}year\ FE + \varepsilon_{it}\ ; \qquad (6)$$

where $Growth_{it:t+3}$ is the changes in the shipment, employment, real capital and material, respectively. The growth rate is computed as the change in output between year *t* and *t+3*, divided by the simple average of output (or factor input) in *t* and *t+3*. This growth rate measure is symmetric around zero, lies in the closed interval [-2, 2], facilitates an integrated treatment of births and deaths, and is identical to the log difference up to a second-order Taylor series expansion. α_i is industry fixed effect. *Treat* is the dummy indicating treatment and equals one for treatment plants with event occurred in year *t*. $Quartile_k$ denotes the plants' location on the pre-event TFP distribution. Control variables include plant size, age, existing access to transportation, ownership and province-year fixed effect. ε_{it} is the random shock.

As Table 6 shows, after the treatment, plants with pre-event productivity in the bottom and top quartile lose resources and the two middle quartiles grow better. The pattern is consistent with the hypothesis that the distribution of productivity becomes more centered, with the worst plants losing markets and the best ones reducing markups.

[Insert Table 6 about here]

5.4 The change in within-industry dispersion and the change in average distance to transportation

Another DID approach is to examine how within-industry TFP dispersion changes in relation to the plants in this industry that are impacted by the transportation changes. The categorization of industries is based on the two-digit industry classification system in the CSMAR and Chinese NBS, and the plants are grouped into 40 industries. We measure the within-industry TFP dispersion with $Var(\log(TFP_{si})$, where *i* denotes plant in sector *s*. The change in industry dispersion from time *b* to *t* is therefore: $\Delta_{b:t}\ Var(\log(TFP_{si})$.

We measure the average distance (AD) and weighted average distance (WD) to transportation mode *M* for sector *s* in year *t* as follows:

$$AD_t^{s,M} = \frac{\sum_l N_{l,t}^s d_{l,t}^M}{\sum_l N_{l,t}^s} \qquad (7a)$$

$$WD_t^{s,M} = \sum_l S_{l,t}^s d_{l,t}^M \qquad (7b)$$

where $N_{l,t}^{s}$ is the number of plants in industry *s*, at time *t*, that are in location *l*; $S_{l,t}^{s}$ is plants' share of industry *s* activities (measured by output or labor), at time *t*, that are in location *l*. $d_{l,t}^{M}$ is location *l*'s (center point) distance to the nearest access point to mode *M* (transportation type) at time *t*. The change in the average (w/o weights) distance to transportation mode *M* for plants in sector *s* is measured using both the fixed plant (fixed spatial distribution) approach and letting the spatial distribution vary over time.

$$\Delta AD_{b,t}^{s,M}(fixing\ spatial) = \frac{\sum_l N_{l,b}^{s} d_{l,t}^{M}}{\sum_l N_{l,b}^{s}} - \frac{\sum_l N_{l,b}^{s} d_{l,b}^{M}}{\sum_l N_{l,b}^{s}} \tag{7c}$$

$$\Delta WD_{b,t}^{s,M}(fixing\ spatial\ ) = \sum_i S_{i,l,b}^{s} d_{l,t}^{M} - \sum_i S_{i,l,b}^{s} d_{l,b}^{M} \tag{7d}$$

$$\Delta AD_{b,t}^{s,M}(vary\ spatial\ ) = \frac{\sum_l N_{l,t}^{s} d_{l,t}^{M}}{\sum_l N_{l,t}^{s}} - \frac{\sum_l N_{l,b}^{s} d_{l,b}^{M}}{\sum_l N_{l,b}^{s}} \tag{7e}$$

$$\Delta WD_{b,t}^{s,M}(vary\ spatial\ ) = \sum_i S_{i,l,t}^{s} d_{l,t}^{M} - \sum_i S_{i,l,b}^{s} d_{l,b}^{M} \tag{7f}$$

We run regressions with rolling window observations. As Table 7 shows, the shorter the distance, the larger is the reduction in TFP dispersion, indicating more efficiency in the within-industry allocation of resources.

[Insert Table 7 here]

6. Exit and new entry

6.1 Asymmetric exit and entry

Productivity-enhancing resource reallocation involves the exit of inefficient plants and the entry of relatively efficient ones. We investigate how events of interest influence the pace and character of plant exit and entry following the approach in Table 8 of Davis et al. (2014). The first step is to sort plants into quartiles defined by their situation in their own-industry TFP distribution at *t*. The second step fits logistic models for the probability of plant exit by time *t+2*, allowing the probability to differ by the time-*t* TFP quartiles interacted with an indicator for whether the plant was affected by the event of interest at *t*. The estimated model tells us how the probability of exit by *t+2* varies with the plant's prior location in its own-industry TFP distribution and how this probability function shifts after the treatment event. For example, suppose that the events of interest are public investments in transport infrastructure. Then we can immediately discern whether the reduced shipping costs and intensification of competitive pressures associated with these events yield greater exit rates of low-TFP plants.

Table 8 Panel A shows two clear patterns. First, consistent with the hypothesis, the exit rate is the highest in the least competitive units, and declines when plants move up in the TFP distribution. The same pattern exists in the treated and control group. However, the control sample, surprisingly, has higher, although insignificantly differently, exit rates than the treated sample. The latter result is consistent Faber (2014) that regions that are passed by in China's new trunkway system experience deterioration in economic activities.

We also investigate how these events affect the pace of plant entry by quartiles of the plant-level TFP distribution. In Panel B, we compare the probability of plants entering in years *t* to *t+2* after the treatment by their performance in year *t+3*. Two patterns appear. First, plants that locate higher on the productivity distribution in year *t+3* are more likely to be new entries during years *t* to *t+2*. Second, the probability that the high productive plants are new entries is significantly larger in the treated sample than in the control sample.

In Panel C, we compare the exit and entry rates among all treated, the matched controls and the non-matched controls. We find that the exit and enter rates are significantly higher in the treated and the matched controls than in the non-matched-controls. This result suggests that control locations that are not near the treated locations are least affected by transportation changes.

[Insert Table 8 about here]

6.2: Operational status and TFP changes

In Table 9, we regress log(TFP) on plants' operational status with the following three specifications.

(a) $$TFP_{it} = f(plant's\ status\ in\ year\ t+2, controls) + ownership\ FE + Province - Year\ FE + \varepsilon;$$

(b) $$TFP_{it+2} = f(plant's\ status\ in\ year\ t+2, controls) + ownership\ FE + Province - Year\ FE + \varepsilon$$

(c) $$Growth\ of\ TFP\ from\ t\ to\ t+2 = f(plant's\ status\ in\ year\ t+2, controls) + ownership\ FE + Province - Year\ FE \quad (8)$$

Panel A shows that plants that continue operation two years after the treatment have significantly positive (above industry-year average) TFP, while those that exit have significantly negative TFP. This contrast is more striking in the treated than in the controls.

Panel B shows that plants that continue operation from *t* to *t+2* and new entry plants during years *t* to *t+2* both have significantly above industry log(TFP) in year *t+2*. The outperformance is larger in the treated than in the controls. In the treated sample, new entries outperform the continuous ones. The opposite is true in the control sample.

Panel C compares the growth in TFP between the treated and the controls. We find that the continuous plants experienced significant TFP growth in the control sample, but not in the treated sample.

[Insert Table 9 about here]

6.3 Decomposition of TFPs by contributions of new entrants, continuers and exits.

We quantify and decompose the overall TFP effects of the events working through plant entry and exit margins. Following Davis et al. (2014):

$$\Delta TFP_{t:t+2} = (S_{t+2}^{c}\, TFPt_{t+2}^{c} - S_{t}^{c}\, TFPt_{t}^{c}) + (S_{t+2}^{N}\, TFPt_{t+2}^{N} - S_{t}^{x}\, TFPt_{t}^{x}) \qquad (9)$$

where S is employment share, computed for the treatment plants only or for all plants, and C, N, X denote continuers, new entrants and exits. We express the change in TFP as the deviation from the controllers' change in TFP over the same period $\Delta\widetilde{TFP}_{t:t+2}$

$$\Delta TFP_{t:t+2} - \Delta\widetilde{TFP}_{t:t+2} = \left(S_{t+2}^{c}\,(TFPt_{t+2}^{c} - \Delta\widetilde{TFP}_{t+2}^{c}) - S_{t}^{c}\,(TFPt_{t}^{c} - \Delta\widetilde{TFP}_{t}^{c})\right) +$$
$$\Big(S_{t+2}^{N}\,(TFPt_{t+2}^{N} - \Delta\widetilde{TFP}_{t+2}^{c}) - \widetilde{S_{t+2}^{N}}(TFPt_{t+2}^{N} - \Delta\widetilde{TFP}_{t+2}^{c}) - S_{t}^{x}\,(TFPt_{t}^{x} - \Delta\widetilde{TFP}_{t}^{c}) +$$
$$\widetilde{S_{t}^{X}}(TFPt_{t}^{x} - \Delta\widetilde{TFP}_{t}^{c})\Big) \qquad (10)$$

The brackets in the top line isolate the contribution of differences between the treated and the controls among continuing plants; the second line isolates the contribution of plant entrants and exiters (births and deaths).

We decompose changes for the whole sample period as well as for a rolling two-year window. Panel A of Table 10 presents the results for the full sample. The overall TFP in the economy during the sample period increases by 20%, half from the improvement in continuing plants and half from exiters and entrants. However, the two-year rolling window estimation shows that the majority are from the continuing plants, suggesting inefficient exit and entry. This result is consistent Brandt, Kambourov and Storesletten (2019): barriers to entry are the cause of underperformance in new entrants and new entrants' performance deteriorates on average by the end of the second year after entry.

Panel B presents the decomposition of changes in the treated and the control samples, using two-year changes as a measurement. We find that the overall improvement in TFP of 4.5% comes all from the continuing plants. The contribution of exiters and entrants is negative, suggesting not only inefficient exit and entry but also possible crowding out or distortion in existing plants. The changes in TFP in the control locations are overall negative, mainly coming from exiters. This result again is consistent with Faber (2014) that economic activities deteriorate in regions that are passed by in China's new trunkway system.

[Insert Table 10 here]

7. Conclusion

We compile a comprehensive dataset for China's surface transportation from 1995 to 2013 and geocode of longitudinal data on location, inputs and outputs for half million manufacturing plants in the GIS. We show that manufacturing plants located closer to transportation are more productive. Productivity and allocative efficiency increase when access to transportation improves. The improvements come from increased market competition created when unproductive units are forced out, facilitating the entry of new units and reallocating resources among continuously existing units. Plants situated at a higher distribution position in the pre-treatment period gain more resources in the post-treatment period. Exiters and new entrants have contributed to 9% of the increases in TFP.

Our current results suggest an association between transportation and TFP, rather than causality. The difficulty arises from the potential endogeneity issue of transportation construction and TFP. The transportation change could be a response to expected TFP improvement and the misallocation could be negatively associated with TFP. These patterns together generate an endogeneity issue between transportation and misallocation. We reduce this concern to some degree when normalizing the TFP measure within industry and year. More importantly, we address the concern with RDD and instrumental approaches, and falsification test which shows only significant treatment effect only at the actual event time.

References

Alder, Simon, 2014. Chinese Roads in India: The Effect of Transport Infrastructure on Economic Development, Working paper, University of Zurich.

Aschauer D.A., 1989. Is Public Expenditure Productive? *Journal of Monetary Economics* 23(2):177–200.

Baily, Martin Neil, Charles Hulten, and David Campbell. 1992. Productivity Dynamics in Manufacturing Plants. In *Brookings Papers on Economic Activity: Microeconomics*, ed. Clifford Winston and Martin Neil Baily. Washington, DC: Brookings Institution Press.

Banerjee, Abhijit V., and Benjamin Moll. 2010. Why Does Misallocation Persist? *American Economic Journal: Macroeconomics*, 2(1): 189-206.

Banerjee, Abhijit V., Esther Duflo, and Nancy Qian, 2020. On the Road: Access to Transportation Infrastructure and Economic Growth in China. Journal of Development Economics, 145, 442-460.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., & Zhang, Q. 2017. Roads, railroads, and decentralization of Chinese cities. Review of Economics and Statistics, 99(3), 435-448.

Bloom, Nicholas, Raffaella Sadun, John Van Reenen, 2017. Management as a Technology? NBER Working Paper No. 22327

Brandt, Loren, Gueorgui Kambourov, and Kjetil Storesletten, 2019. Barriers to Entry and Regional Economic Growth in China. Working paper, University of Toronto.

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.

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, 339-351.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang. 2014. Challenges of Working with the Chinese NBS Firm-level Data. *China Economic Review* 30, 339-352.

Cosar, A. Kerem and Banu Demir, 2014. Domestic Road Infrastructure and International Trade. Working paper, University of Chicago.

Donaldson, D., 2018. Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, 108, 4-5, 899-934.

Donaldson, Dave and Richard Hornbeck, 2016. Railroads and American Economic Growth: A "Market Access" Approach, *The Quarterly Journal of Economics*, 131(2), 799-858.

Davis, Steven J. and John Haltiwanger, 1999. Handbook of labor economics 3, 2711-2805.

Davis, Steven J., John Haltiwanger, Kyle Handley, Ron Jarmin, Josh Lerner, and Javier Miranda, 2014. "Private Equity, Jobs, and Productivity." American Economic Review, 104(12), 3956-3990.

Dunne, Timothy. 1998. CES Data Issues Memorandum, 98:1. Center for Economic Studies, U.S. Census Bureau.

Ghani, E., A. G. Goswami, and W. R. Kerr, 2016. Highway to Success: The Impact of the Golden Quadrilateral Project for the Location and Performance of Indian Manufacturing," *Economics Journal*, 126, 317–357.

Faber, Benjamin, 2014. Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. *Review of Economic Studies*, forthcoming.

Foster, L.,Grim and J. Haltiwanger 2016. Reallocation in the Great Recession: Cleansing or Not? *Journal of Labor Economics*, 34(1), 293-331.

Helpman, Elhanan, 1998. The Size of Regions. In topics in Public Economics: Theoretical and Applied Analysis. Cambridge: Cambridge University Press, 33-54.

Hsieh, Chang-Tai and Peter J. Klenow, 2009. "Misallocation and Manufacturing TFP in China and India". *Quarterly Journal of Economics*, CXXIV(4), 1403-1448.

Huang, Yuxiao. and Wentao Xiong, 2018. Geographic Distribution of Firm Productivity and Production: A ",Market Access" Approach, Working Paper, Harvard University.

Leibenstein, Harvey, 1966. Allocative Efficiency vs 'X-Efficiency'. *American Economic Review*, 62, 777-795.

Liu, Dan, Liugang Sheng, and Miaojie Yu, 2017. Highways and Firms' Exports: evidence from China. Working Paper, Peking University.

Raith, Michael, 2003. Competition, Risk, and Managerial Incentives. *American Economic Review*, 93(4), 1425-1436.

Redding, Stephen J., 2016. Goods trade, factor mobility and welfare, *Journal of International Economics*, 101, 148-167.

Redding, Stephen J. and Matthew A. Turner, 2015. Transport Costs and the Spatial Organization of Economic Activity, in *handbook of urban and Regional Economics*, ed. by G. Duranton, J. V. Henderson, and W. Strange, vol.5, chap.20, pp. 1339-1398.

Restuccia, Diego, and Richard Rogerson, 2008. Policy Distortions and Aggregate Productivity with Heterogeneous Plants. *Review of Economic Dynamics*, 11, 707–720.

Roberts, Mark J., and Dylan Supina. 1996. Output Price, Markups, and Producer Size. European Economic Review, 40(3), 909-921.

Stigler, George J., 1976. The Xistence of X-efficiency. American Economic Review, 66(1) 213-216.

Syverson, Chad, 2004a. Market Structure and Productivity: A Concrete Example. *Journal of Political Economy*, 112(6) 1181-1222.

Syverson, Chad, 2004b. Product Substitutability and Productivity Dispersion. *Review of Economics and Statistics*, 86(2).

Syverson, Chad, 2001. What Determines Productivity? *Journal of Economic Literature*, 49:2, 326-365.

Yan, Y. 2017. Transportation Infrastructure, City Productivity Growth and Sectoral Reallocation; Evidence from China. Working Paper, UCLA.

Figure 1: Surface Transportation Changes in China: Motorway example

A: The orange lines indicate Motorway map in 1993 and the blue lines, 2013.

B: Year-by-year changes in China's Motorway from 1993 to 2012.


1993


1994


1995


1996


1997


1998


1999


2000


2001


2002


2003


2004


2005


2006


2007


2008


2009


2010


2011


2012

Table 1: The sample description

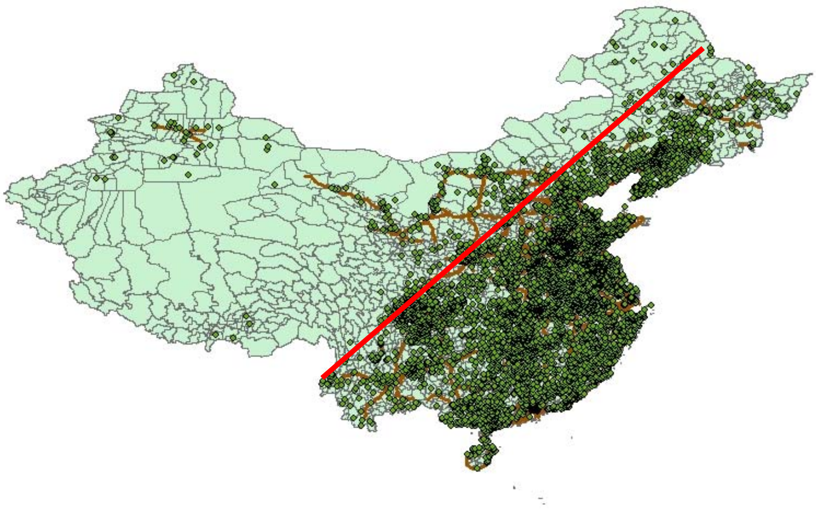
Our empirical analyses use Chinese plant-level data during year 1998-2007. We describe the annual sample annual characteristics in Panel A: firm size (total assets), leverage (long-term + short-term liability/total assets), sales, age, and the distribution of ownership types across years.

		Characteristics					Ownership type (%)				
Year	# of units	Assets (RMB million)	Sales (RMB Million)	Emp-loyee	Leve-Rage (%)	Age	State	Collective	Corporate	Private	Foreign
1998	133,202	529	348	295	65	38	37	38	8	10	7
1999	150,588	530	356	301	65	25	34	37	11	10	7
2000	150,388	555	401	276	64	19	30	35	16	11	8
2001	159,630	542	413	268	61	16	25	30	26	11	8
2002	172,514	553	446	260	60	14	21	27	33	11	8
2003	183,687	584	515	257	58	12	16	23	40	11	9
2004	250,073	517	497	217	59	10	11	17	50	11	11
2005	266,342	560	588	219	56	10	9	17	53	10	11
2006	295,602	585	655	209	55	10	8	8	64	10	11
2007	276,505	677	818	218	54	10	6	7	65	10	11

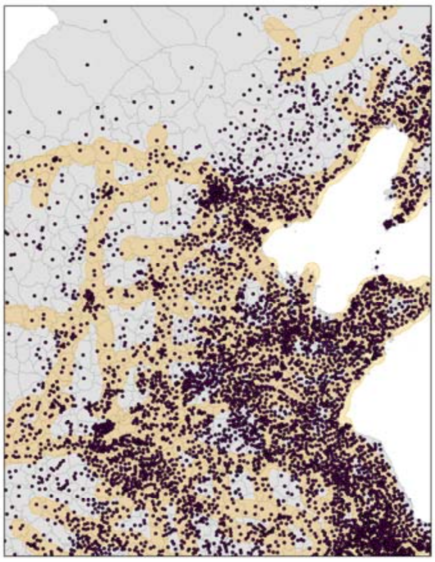
Figure 2: Plant locations in China

The first graph plots all the plants in China in the year 1998 with old railroads. The second graph plots the addition of new plants during years 1998-2007 and the new additions of the highway and high-speed railways during this period.

(a): Plants locations in 1998 relative to the National Highway



(b): Locations of NEW ENTRIES of plants 1998-2007 relative to the Motorway and high-speed railway



(C) Motorway's 20km buffer and plants' locations in 2007, a Northeast coast example

Table 2: Summary statistics of the sample's TFP.

In table we report the distribution of plants' total factor productivity computed as LogTFP_{et} = logQ_{et} – $\alpha_K$log$K$et – $\alpha_L$log$L$et – $\alpha_M$logM_{et}, (1). Panel A reports its annual mean and standard deviation, and its industry-year-adjusted values' percentage cut off values on distribution. Panel B reports the value (labor)-weighted industry-year mean and standard deviation, and its value-weighted industry-year -adjusted distribution. Panel C reports the value (output value)-weighted industry-year mean and standard deviation, and its value-weighted industry-year -adjusted distribution by region.

The regions are classified as follows: Eastern: Beijing, Tianjin, Hebei, Shandong, Shanghai, Jiangsu, Zhejiang, Fujian, Guangdong, Hainan; Central:Henan, Anhui, Hubei, Hunan, Jiangxi, Shanxi; Western: Chongqing, Gansu, Ningxia, Qinghai, Inner Mongolia, Tibet, Shannxi, Sichuan, Guizhou, Xinjiang, Yunnan, Guangxi; NorthEastern: Heilongjiang, jilin, Liaoning.

Panel A: Full sample

	Industry-Year mean and std, Log(TFP)			Distribution of industry-year-adjusted log(TFP)						
Year	obs.	Mean	STD	P1	p10	p50	p90	P99	P90-P10	P99-P1
1998	93,225	1.12	0.46	-1.83	-0.44	0.02	0.45	1.30	0.89	3.14
1999	106,287	1.20	0.48	-1.92	-0.47	0.03	0.48	1.20	0.95	3.13
2000	113,769	1.18	0.49	-1.91	-0.48	0.03	0.48	1.37	0.96	3.27
2001	127,772	1.18	0.47	-1.95	-0.44	0.02	0.47	1.18	0.90	3.13
2002	138,404	1.19	0.47	-1.88	-0.45	0.02	0.47	1.19	0.92	3.07
2003	149,101	1.21	0.42	-1.58	-0.42	0.01	0.46	1.06	0.88	2.64
2004	209,281	1.24	0.40	-1.20	-0.42	-0.01	0.46	1.12	0.88	2.33
2005	214,884	1.32	0.41	-1.18	-0.45	-0.01	0.48	1.14	0.92	2.32
2006	232,433	1.38	0.41	-1.16	-0.45	-0.02	0.48	1.20	0.93	2.36
2007	212,696	1.44	0.40	-1.04	-0.44	-0.02	0.48	1.13	0.92	2.17

Panel B: Value-weighted industry-year adjusted log(TFP)

Value weighted by Labor input (L)										
	Industry-Year mean and std			Distribution of industry-year-adjusted log(TFP)						
Year	Obs	Mean	STD	p1	p10	p50	p90	p99	P90-P10	P99-P1
1998	120,564	-0.14	0.54	-2.38	-0.67	-0.09	0.38	1.12	1.05	3.50
1999	130,872	-0.12	0.56	-2.32	-0.70	-0.06	0.44	1.10	1.13	3.42
2000	129,054	-0.12	0.56	-2.24	-0.69	-0.08	0.43	1.20	1.12	3.44
2001	138,299	-0.10	0.52	-2.16	-0.61	-0.06	0.42	1.10	1.03	3.26
2002	148,201	-0.09	0.51	-1.95	-0.60	-0.05	0.43	1.14	1.04	3.09
2003	158,239	-0.08	0.45	-1.69	-0.56	-0.06	0.42	1.02	0.97	2.71
2004	221,598	-0.07	0.42	-1.29	-0.52	-0.08	0.41	1.07	0.94	2.36
2005	236,582	-0.07	0.42	-1.26	-0.53	-0.07	0.43	1.07	0.96	2.33
2006	262,583	-0.07	0.42	-1.23	-0.53	-0.08	0.42	1.11	0.95	2.34
2007	244,458	-0.06	0.40	-1.09	-0.51	-0.07	0.42	1.04	0.93	2.13

Value weighted by Output (value Q)

	Value (Q)-weighted Industry-Year mean and std			Distribution of industry-year-adjusted log(TFP)						
Year	Obs	Mean	STD	p1	p10	p50	p90	p99	P90-P10	P99-P1
1998	120,564	-0.25	0.55	-2.50	-0.79	-0.20	0.27	1.01	1.06	3.50
1999	130,872	-0.26	0.56	-2.51	-0.85	-0.20	0.30	0.96	1.15	3.47
2000	129,054	-0.25	0.56	-2.37	-0.83	-0.20	0.30	1.11	1.13	3.48
2001	138,299	-0.22	0.52	-2.32	-0.75	-0.18	0.30	0.98	1.05	3.29
2002	148,201	-0.21	0.51	-2.08	-0.73	-0.17	0.31	1.02	1.05	3.11
2003	158,239	-0.18	0.46	-1.78	-0.67	-0.16	0.31	0.90	0.98	2.68
2004	221,598	-0.16	0.42	-1.38	-0.62	-0.16	0.32	0.98	0.94	2.36
2005	236,582	-0.16	0.42	-1.36	-0.63	-0.16	0.33	0.97	0.96	2.32
2006	262,583	-0.16	0.42	-1.33	-0.63	-0.16	0.33	0.99	0.95	2.32
2007	244,458	-0.14	0.40	-1.19	-0.59	-0.15	0.34	0.97	0.93	2.16

Panel C: By region: Industry adjusted TFP

	Eastern					Western				
Year	Obs.	Mean	STD	P90-p10	P99-P1	Obs.	Mean	STD	P90-p10	P99-P1
1998	70,321	0.02	0.45	0.89	3.04	14,908	-0.06	0.59	1.24	3.60
1999	75,669	0.02	0.50	0.99	3.22	16,925	-0.09	0.62	1.32	3.66
2000	75,328	0.03	0.48	0.95	3.21	17,209	-0.11	0.63	1.36	3.82
2001	85,587	0.02	0.46	0.90	2.97	17,240	-0.09	0.59	1.24	3.50
2002	94,281	0.02	0.46	0.93	2.99	17,712	-0.07	0.57	1.23	3.32
2003	104,549	0.01	0.41	0.88	2.54	17,827	-0.08	0.53	1.20	2.99
2004	153,714	0.01	0.40	0.88	2.29	20,868	-0.04	0.48	1.09	2.75
2005	162,097	0.00	0.40	0.90	2.20	23,067	-0.03	0.49	1.11	2.69
2006	178,295	-0.01	0.39	0.88	2.14	25,114	-0.01	0.48	1.10	2.63
2007	167,476	-0.01	0.38	0.87	2.07	22,447	0.01	0.45	1.06	2.45

	Central					North Eastern				
Year	Obs.	Mean	STD	P90-p10	P99-P1	Obs.	Mean	STD	P90-p10	P99-P1
1998	22,787	0.07	0.57	1.10	3.80	8,567	-0.24	0.78	1.68	4.09
1999	25,105	0.03	0.59	1.20	3.71	8,895	-0.09	0.70	1.49	4.01
2000	23,758	0.05	0.61	1.21	3.90	8,715	-0.14	0.71	1.50	4.05
2001	22,951	0.01	0.58	1.14	3.70	8,398	-0.08	0.66	1.33	3.92
2002	23,112	0.02	0.54	1.10	3.34	8,792	-0.09	0.63	1.30	3.72
2003	23,000	0.03	0.50	1.08	2.92	9,251	-0.08	0.53	1.16	3.06
2004	28,575	0.03	0.45	1.01	2.55	12,649	-0.07	0.49	1.11	2.75
2005	31,182	0.05	0.45	1.05	2.56	14,263	-0.07	0.48	1.12	2.59
2006	35,575	0.07	0.49	1.10	2.86	17,457	-0.05	0.47	1.10	2.57
2007	32,877	0.07	0.44	1.03	2.37	16,178	-0.05	0.45	1.05	2.41

Table 3: Access to transportation and TFP distribution

This table include all 1,935,707 plant*year during 1998-2007 that have the GIS location identified. We run the panel regression with the following specification:

$$\frac{logTFP_{i,s,t} - \overline{logTFP_{s,t}}}{\sigma(LogTFP)_{s,t}} = a + \beta Distance_{i,t} + \gamma D(> 100km)_{i,t} + +\delta Z + FE_{p,t} + \varepsilon_{i,s,t}$$

The dependent variable is logTFP adjusted by industry mean and standard deviation for each industry each year. In Panel A, B, and C, the industry-year mean are equal weighted, labor input weighted, and output value weighted.

The explanatory variables are the plants' distance, measured in km, to the nearest access/exist point of each type of surface transportation, an indicator for distance longer than 100km, firm size, age, ownership types, and province-year fixed effect. In the first row, each transportation type enters regression independently. In the second row, the regression specification include all transportation types together. ** and * respectively denote significance at the 1% and 5% level.

Panel A: Industry-year equal-weight -adjusted log(TFP), unit in % of standardized deviation

Distance to	Motorway	Railway	Waterway	National Highway	High Speed Railway
Each transportation alone	-2.00***	-12.50***	-4.00***	-17.77***	-1.77
	[-4.46]	[-14.24]	[-8.76]	[-14.84]	[-1.05]
All transportations together	-1.80*	-15.50***	-1.99***	-7.35***	0.50***
	[-3.30]	[-16.35]	[-3.75]	[-5.15]	[2.80]

Panel B: Industry-year labor-input-weight -adjusted log(TFP), unit in % of standardized deviation

Distance to	Motorway	Railway	Waterway	National Highway	High Speed Railway
Each transportation alone	-2.85***	-18.41***	-3.78***	17.17***	-0.12
	[-7.33]	[-24.04]	[-10.14]	[-17.58]	[-1.28]
All transportations together	-2.36***	-18.69***	-2.33***	-4.92***	0.49***
	[-5.03]	[-20.34]	[-5.26]	[-4.09]	[-7.87]

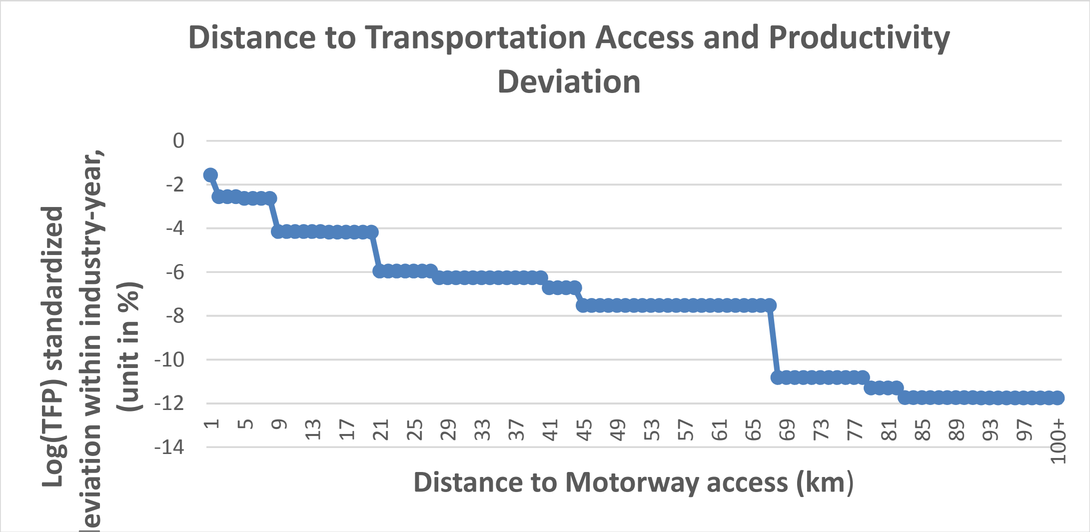
Panel C: Industry-year output-value-weight -adjusted log(TFP), unit in % of standardized deviation

Distance to	Motorway	Railway	Waterway	National Highway	High Speed Railway
Each transportation alone	-3.77***	-21.18***	-4.52***	-21.34***	0.04
	[-6.98]	[-23.01]	[-10.06]	[-18.51]	[-0.4]
All transportations together	-3.05***	-20.97***	-2.55***	-7.25***	0.43***
	[-5.85]	[-20.75]	[-5.21]	[-5.51]	[-6.36]

Figure 3: Non-parametric approach: Distance and productivity

Panel A presents the piece-wise coefficients from the non-parametric regression of productivity level on distance bin. The x-axis is the distance and y-axis is the industry-year adjusted productivity.

$$\frac{logTFP_{i,s,t} - \overline{logTFP_{s,t}}}{\sigma(LogTFP)_{s,t}} = a + \sum_b \beta_b\, Bin_{i,t,b} + \gamma D(> 100km)_{i,t} + +\delta Z + FE_{p,t} + \varepsilon_{i,s,t}$$



Panel B presents the productivity heat map in year 2004. For each distance bin to the nearest motorway access, the map plots the mean of industry-year adjusted log(TFP) for all plants in the bin. The TFP is categorized into 5 classes, with the warmest (coldest) colour denotes the highest (lowest) value. The black line is the motorway network existed in the same year.

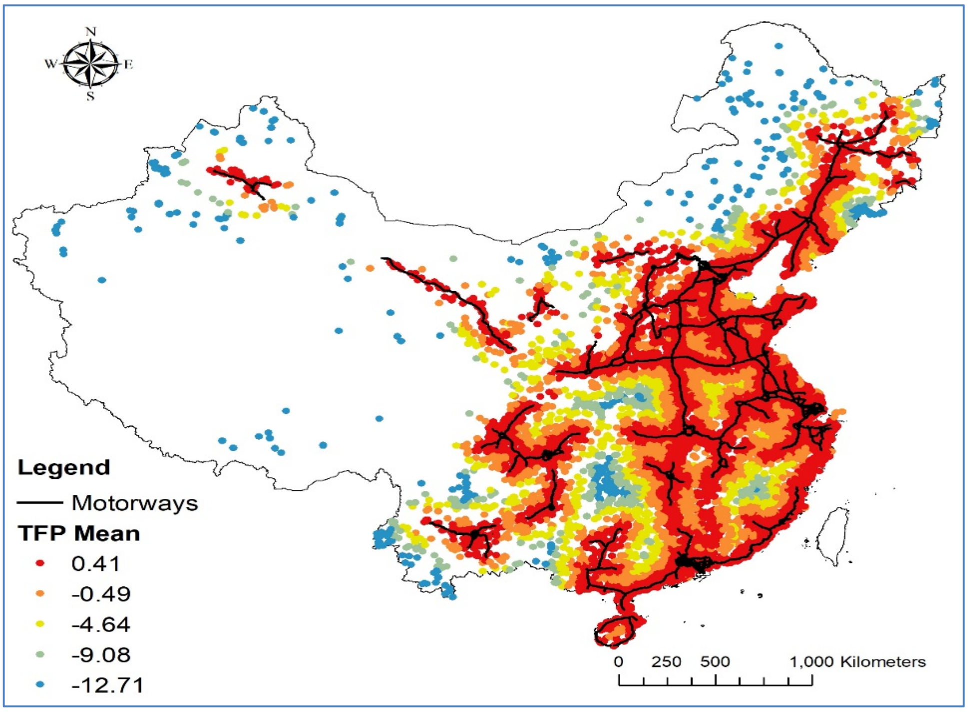
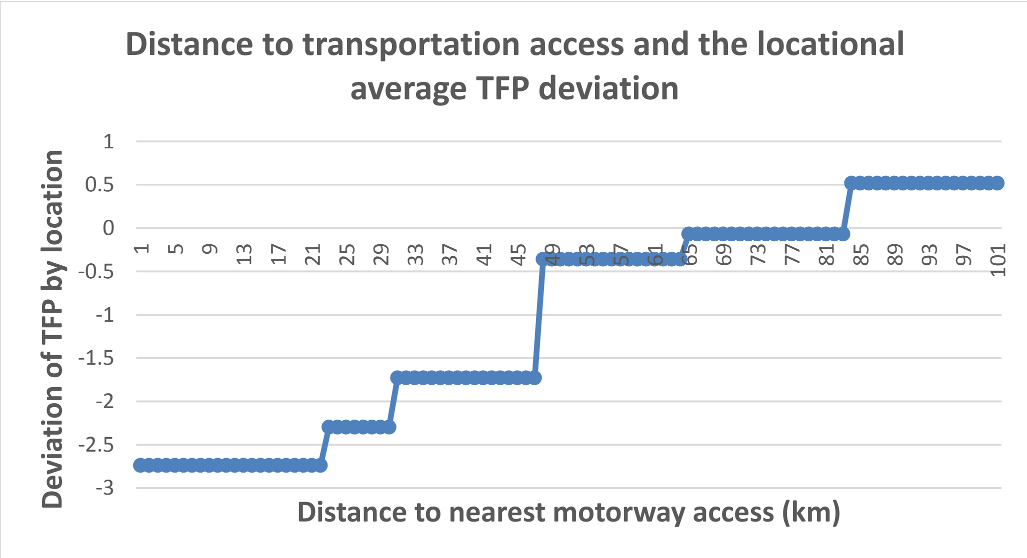


Figure 4: TFP deviation by location and the location's distance to transportation

Panel A presents the piece-coefficients from the non-parametric regression in the location-year panel, in which the dependent variable is the average deviation (from its own industry-year mean) for all plants located in this location, $DEV_l = \frac{\sum_{i\in l}|\log(TFP_{l,i})|}{N_l}$, and the explanatory variable is the distance to the nearest access to motorway. The density of plants in the location and year fixed effects are controlled.

$$\frac{\sum_{i\in l}|\log(TFP_{l,i})|}{N_l} = a + \sum_b \beta_b Bin_{l,t,b} + \gamma D(> 100km)_{l,t} + FE_{p,t} + \varepsilon_{l,t}$$



Panel B presents the dispersion heat map in year 2004. The dispersion is the locational DEV_l categorized into 5 classes. The warmest (coldest) colour denotes the highest (lowest) value. The black line is the motorway network existed in the same year.

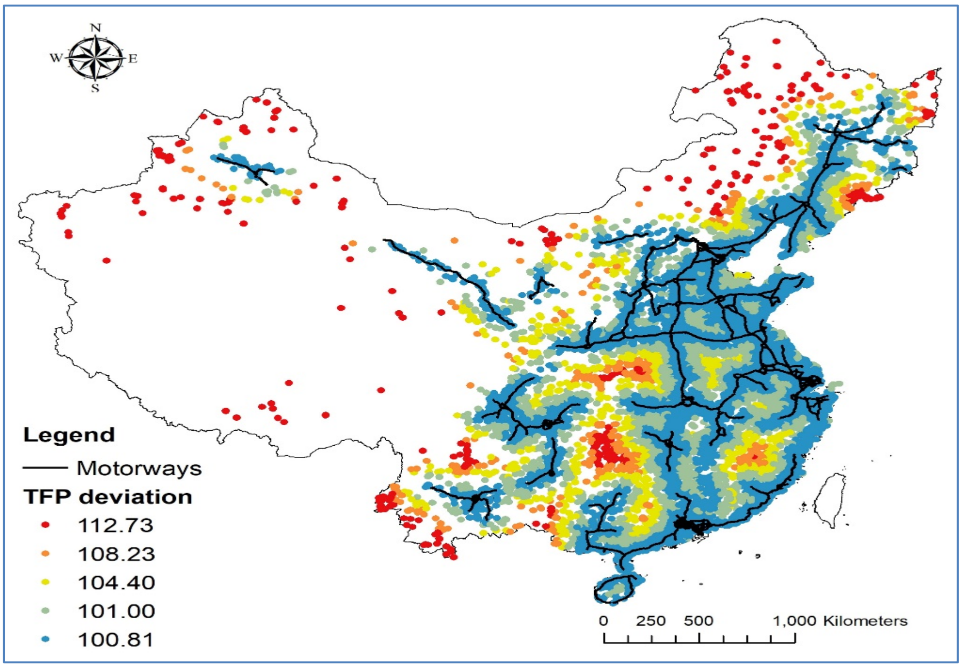
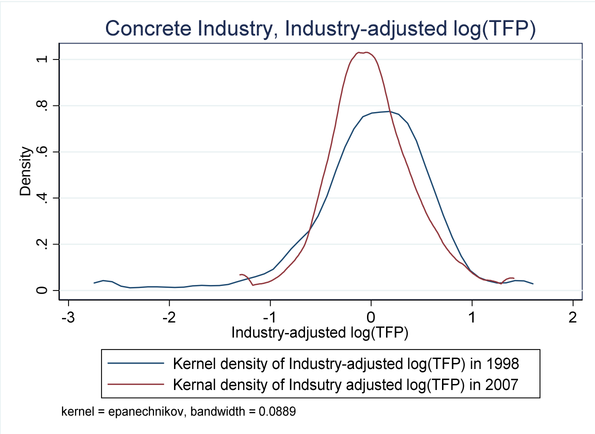
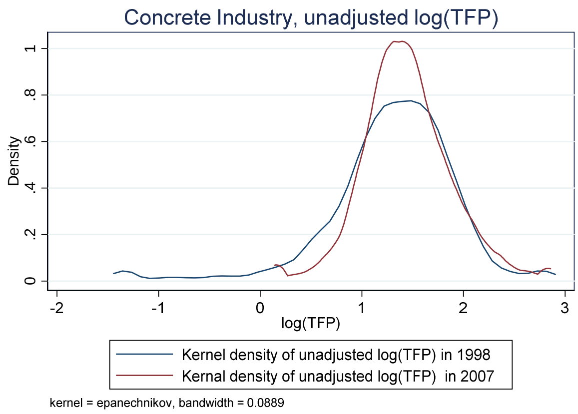


Figure 5: TFP distribution changes that are potentially related with transportation

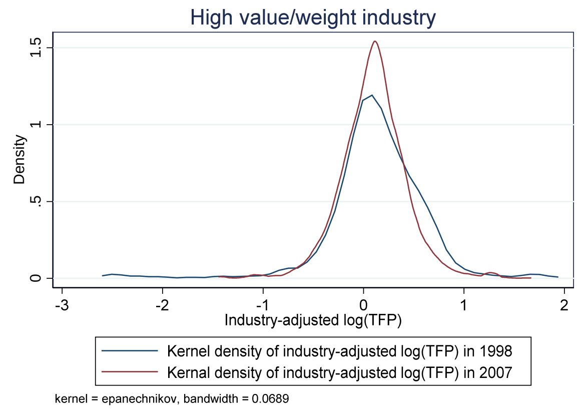
(I) By Industry characteristics



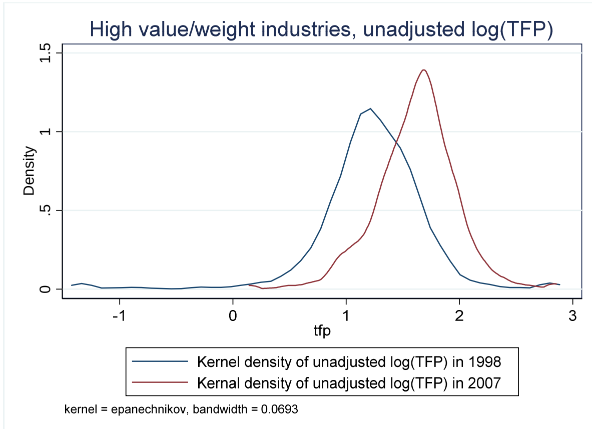
(a)



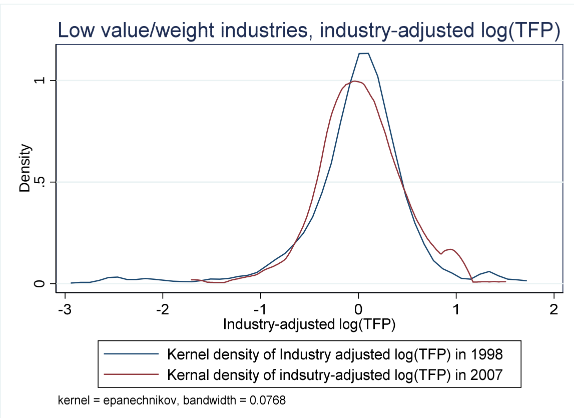
(b)



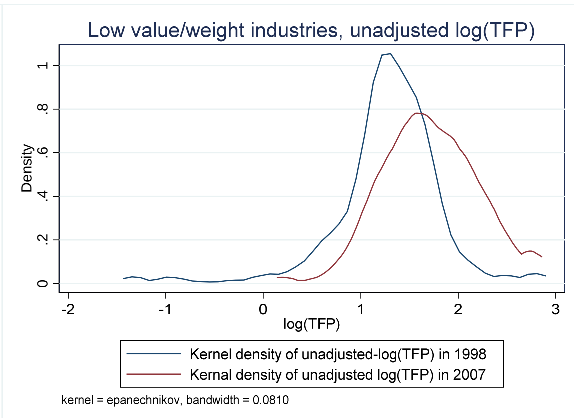
(c)



(d)

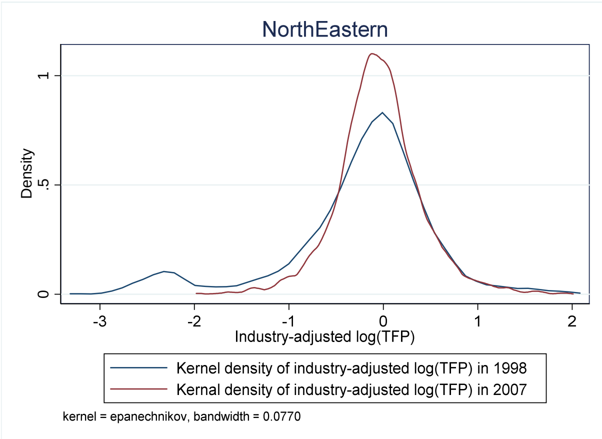


(e)

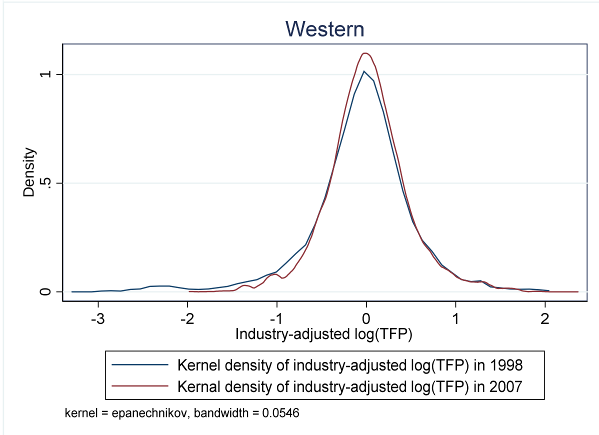


(f)

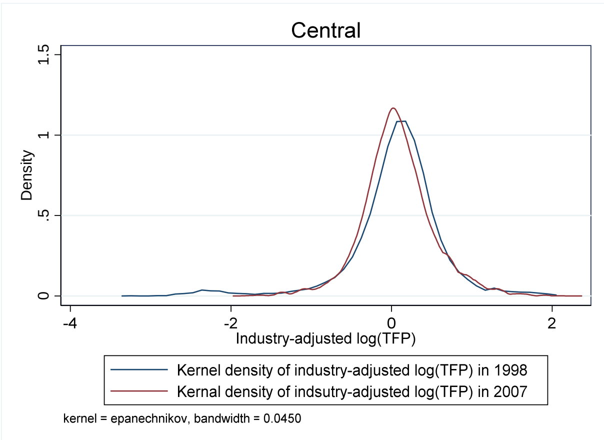
(II) By pre-existing conditions that are potentially related to transportations



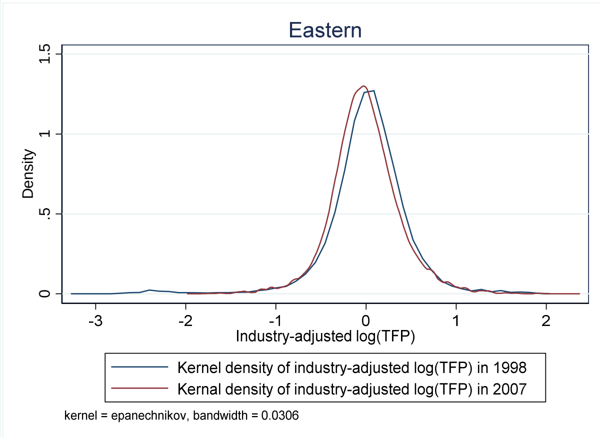
(g)



(h)

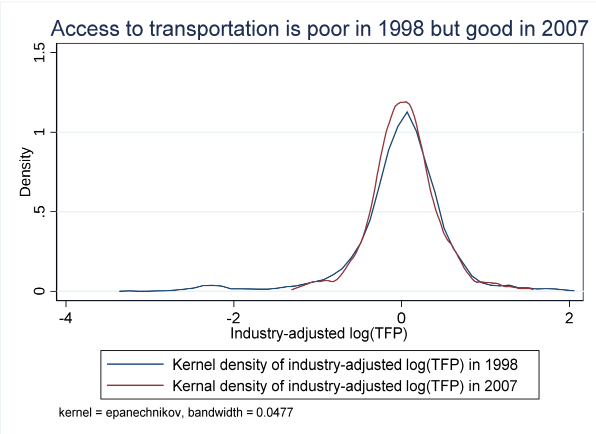


(i)

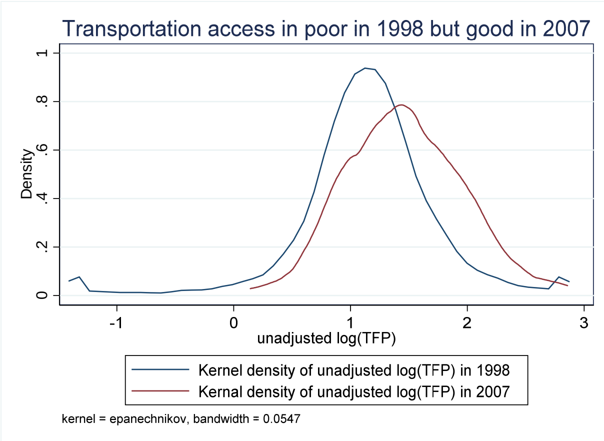


(j)

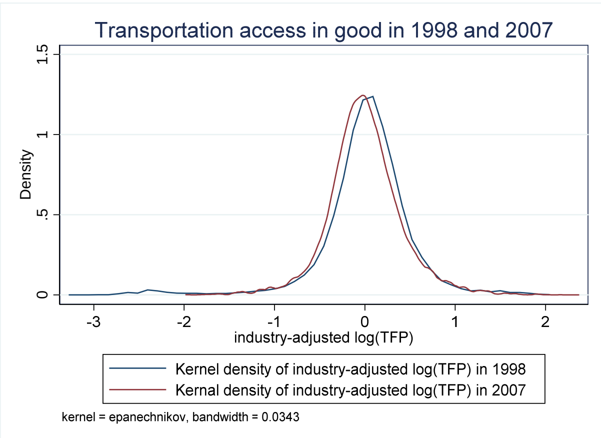
(III) By the change of access from 1998 to 2007



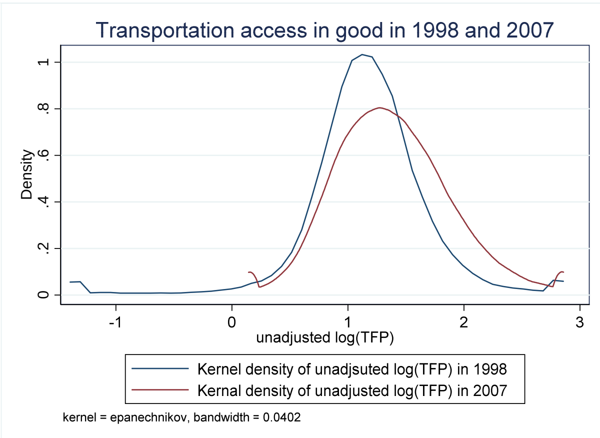
(m)



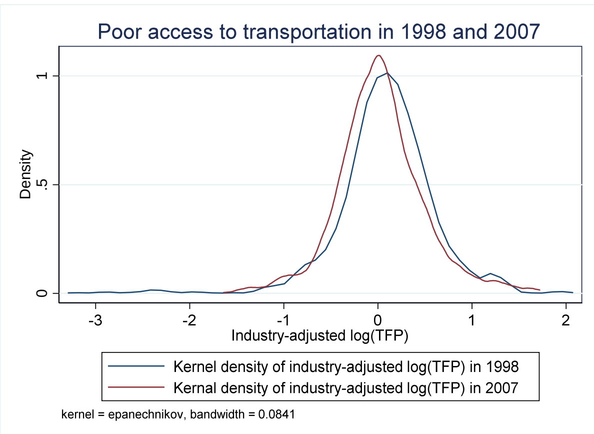
(n)



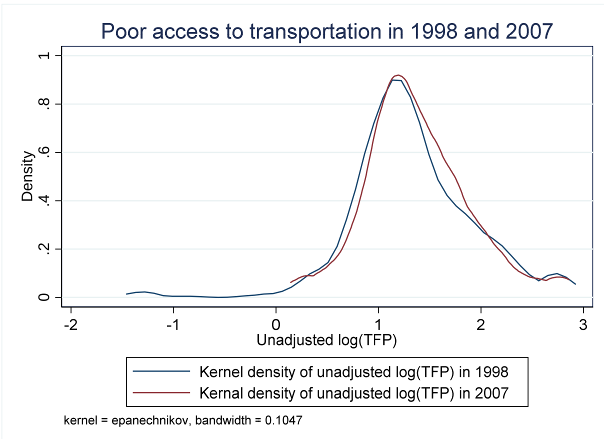
(o)



(p)



(q)



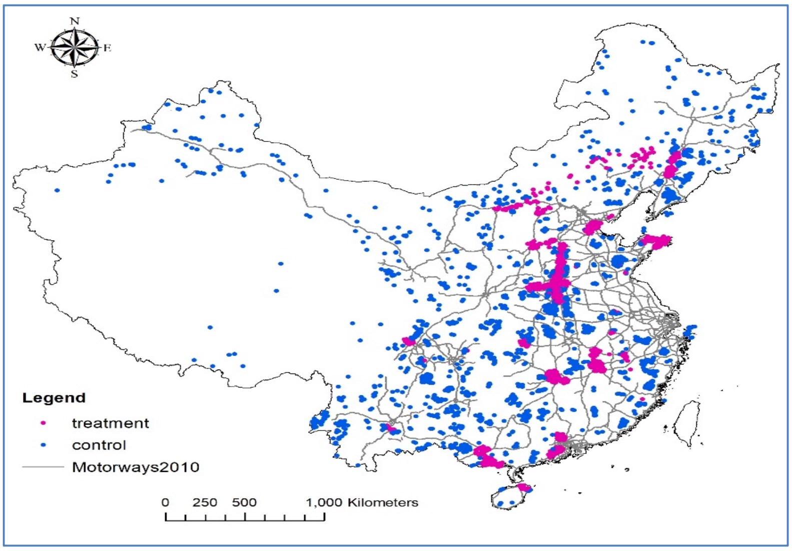
(r)

Figure 5: Treatment and control samples

Treated are plants locates within 25km (or 5, 10,50km when distance threshold changes) to the access point to any new transportation additions in that year, condition on no previous access within 25km. Controls are plants that don't experience any treatment, within 25km criteria, during the five years window of [t-2: t+2], condition on no previous access within 25km.

In Panel A, we plot all treated and controls. In panel B, we plot matched treated and control. Using regression Discontinuity Design (RDD), the treated and control are matched based on Year, province, and neighboring location. There are 1,779 treatments and 22,405 controls at the zip code level, of which 1,440(*2) pairs matched. The motorways in 2010 is also added as a reference.

Panel A: All treated and All controls



Panel B: Matched treated and controls

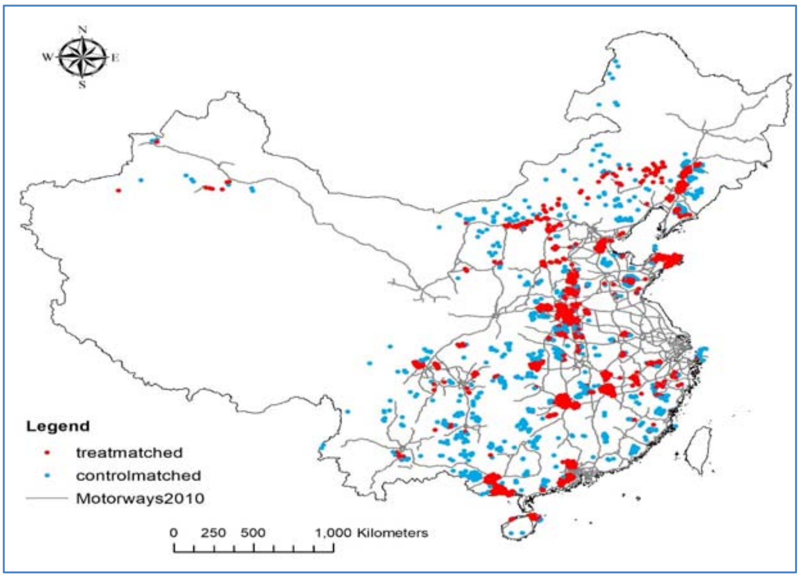


Table 4: Transportation treatment effect on TFP and input, output growth,

We define treatment with plants that gain new transportation access within 25 km in year t, conditional on there is no previous access within 25km. The control plants are those without previous access within 25 km and didn't experience any change in 25km in the past two and future two years [t-2, t+2], the matched for each treatment based on the same year, same province, and neighboring location. There are 1440(*2) pairs of treatment and control at the zipcode level, impacting 38,091 plants as treated, and 12,986 as control plants.

The dependent variables are industry adjusted log(TFP) $logTFP_{i,s,t} - \overline{logTFP_{s,t}}$ or $\frac{logTFP_{i,s,t}-\overline{logTFP_{s,t}}}{\sigma(LogTFP)_{s,t}}$ dispersion of TFP $\frac{\sum_{i\in l}|\log(TFP_{l,i})|}{N_l}$, exiting and entries, and the growth of input and output in year t-2, t-1, t, t+1, and t+2, t+3 respectively.

The outcome variables are observed at the plant level. The control variables include the closest distance to all existing transportation modes, plant size, age, ownership, province-year fixed effect. Panel A presents year to year changes. Panel B presents the 3-year changes around the treatment.

Panel A: Year-by-Year changes in TFP, output, and input Falsification tests of treatment effect

	Ind-adj log(TFP)	Ind-adj log(TFP)/ σ	Growth Q	Growth M	Growth L	Growth K
t-2	0.01	0.01	-0.09*	0.02	-0.03	0.00
	[0.27]	[0.10]	[1.86]	[0.51]	[0.80]	[0.01]
t-1	0.05	0.14	0.01	0.01	-0.07	-0.15**
	[1.64]	[1.49]	[0.22]	[0.16]	[1.56]	[2.10]
t (treat)	0.03	0.05	0.02	0.03	0.12***	0.12**
	[0.85]	[0.53]	[0.73]	[0.86]	[3.71]	[2.31]
t+1	0.09**	0.26*	0.07***	-0.04	-0.03	0.03
	[1.99]	[1.79]	[2.56]	[0.92]	[0.95]	[0.73]
t+2	0.07	0.18	0.01	0.09	-0.03	0.01
	[1.16]	[0.99]	[0.41]	[1.47]	[0.85]	[0.25]
t+3	0.00	0.02	0.01	0.29***	0.05	0.01
	[0.02]	[0.08]	[0.18]	[2.92]	[0.91]	[0.11]

Panel B: Treatment effect measured by 3-years-changes from t-1 to t+2

	Growth of Ind-adj log(TFP)	Growth of Ind-adj log(TFP)/ σ	Growth Q	Growth M	Growth L	Growth K
Changes from t-1 to t+2						
Treatment	7.08*	6.69	0.09*	0.10**	0.03	0.15*
	(1.77)	(0.98)	(1.66)	(1.85)	(0.73)	(1.88)
Changes from t to t+3						
Treatment	0.40	-4.24	0.25***	0.22***	0.18***	0.14
	(0.36)	(0.70)	(2.86)	(2.55)	(2.71)	(1.17)

Table 5: Instrument events using historical routes

This table presents the estimation of treatment effect with instrumental approach. In the first stage, we used the historical road access, *road1962* and *rail1962*, to explain the transportation changes, in the shortest distance, occur in each treated location. We include the location's distance to the nearest prefecture or country level city, whichever is closer. In the second stage, we use predicted distance of transportation changes to define treatment and control, and match them based on year, province, and location-neighbouring. The dependent variable in the 2nd stage takes the same as those in Table 4. The fixed effect, control variables as in Table 4, and those in the 1st stage are all included. Results are at the zip code level.

$$\text{Distance to new access } = \alpha + \beta_1 Road1962 + \beta_2 Rail1962 + \gamma X + \varepsilon$$

$$Log(TFP)\ or\ growth\ of\ Q, L, M, K = \alpha + \ \beta * Predicted\ Treated$$
$$+\beta_1 Road1962 + \beta_2 Rail1962 + \gamma X + FE + \varepsilon \qquad (5)$$

	Stage 1:	Stage 2: TFP in year t+1		
	Y=distance to new transportation access	Industry-year adjusted log(TFP)	Labor-weighted industry-year adjusted log(TFP)	Output-weighted industry-year adjusted log(TFP)
Predicted Treated		0.23***	0.19***	0.22***
		(3.49)	(2.84)	(3.29)
log(Road1962)	0.69***	0.00	0.00	0.00***
	(13.19)	(0.44)	(0.71)	(5.18)
Log(Rail1962)	1.03***	0.00	0.00	0.00
	(21.80)	(0.72)	(0.69)	(0.34)
Log(Distance to city)	1.35***	-0.01	-0.01	-0.01
	(13.86)	(-0.81)	(-0.56)	(-0.61)
Province-Year FE	Yes	Yes	Yes	Yes
R-square	0.16	0.10	0.09	0.10

Table 6: Asymmetric treatment effect at the plant level

Reduction of dispersion comes from the shift of the both tails towards the center, which implies an aymmetic treatment effect depending on plants' pre-event TFP. This table estimates of the piece-wise treatment effects on productivity, by categorizing treatment into four quartiles indicators based on the plants' rank its industry-year adjusted TFP in year t-1. To have sufficient observations in all treatment distance cutoff, this tables uses all treatment, and all others as the redundant group. The dependent variable is the growth of TFP (industry-year adjusted and log value) between year t and t+3. The growth is computed as the difference divided by the simple average of year t and t+3. The standard errors are clustered at the zip code level. * and ** denote the significance level at the 1% and 5% respectively.

	Y= Growth of industry-adjusted TFP from year t to t+3			
Treated	5km	10km	25km	50km
Bottom	0.38	-0.06***	-0.07***	-0.07***
	(0.01)	(3.52)	(5.99)	(7.11)
2nd Quartile	17.54	0.11***	0.11***	0.12***
	(0.63)	(7.16)	(9.94)	(12.75)
3rd Quartile	61.61*	0.07***	0.07***	0.08***
	(2.24)	(4.62)	(5.84)	(8.27)
Top	-7.44	-0.16***	-0.17***	-0.17***
	(0.27)	(9.93)	(13.35)	(17.26)

Table 7: Change of industry dispersion of TFP on change of average distance to transportation

The equal weighted average distance are measured as equation (10a) and (10c). The value weighted average distance are measured as equations (10b) and (10d) and with labor input and shipment output as the weights, respectively. Both the changes of distance and the changes of TFP dispersion are measure over the period [t-2: t]. Coefficients and t-statistics are reported. ** and * denote significance at the 1% and 5% levels respectively.

	Y: $\Delta^s_{t-2:t} Var(\log(TFP_{si})$^0.5 (scaled in %)		
Distance to	$\Delta AD^s_{t-2,t}$	$\Delta lWD^s_{t-2,t}$	$\Delta qWD^s_{t-2,t}$
	Fixed spatial distribution		
Trunkway	0.09	0.08	0.25
	[0.48]	[1.07]	[1.40]
Railway	14.82	-4.54	-8.98
	[0.64]	[-0.20]	[-0.36]
High-speed railway	0.02**	0.01***	0.02**
	[2.18]	[3.44]	[1.97]
	Varying spatial distribution		
Trunkway	0.04	-0.04	0.15
	[1.55]	[-1.23]	[1.00]
Railway	0.95***	1.00***	0.82***
	[15.04]	[4.17]	[22.02]
High-speed railway	-0.01	-0.02	-0.03
	[-0.90]	[-0.92]	[-1.43]

Table 8: Entry and exit of manufacturing plants by location in the distribution of own-industry's total factor productivity, 1998-2007.

The treatment plants are those that located within 25km of any kind of transportation treatment point. The control plants are those that don't experience in any in [t-2,t+2], matched

Panel A: Plant exit probabilities in the first two years post transportation treatment (logistic specification)

Location in own-industry TFP distribution as of the Year t-1	Probability of plant exit by year t+2 (marginal effect)		P-value for difference between the treated and controls
	Treatment	Control	
Bottom Quartile	0.24	0.34	0.16
2nd	0.16	0.21	0.31
3rd	0.18	0.26	0.12
Top	0.17	0.20	0.74

Panel B: Plants in operation in Year t+3

Location in Own-Industry TFP Distribution as of the Year t+3	Probability of Plant in operation t+3 enter the market in year t+2 or year t+1		P-value for Difference Between the Treated and Controls
	Treatment	Control	
Bottom Quartile	0.10	0.08	0.07
2nd	0.13	0.11	0.01
3rd	0.14	0.10	0.09
Top	0.13	0.14	0.00

Panel C: Comparison with matched treated, all control, and non-treated/control

	Treatment	All controls	Non-treated/control
Exit rate by year t+2	0.07	0.09	0.03
p-value of difference	(0.00)	(0.00)	
Enter rate in year t+1 and t+2	0.06	0.07	0.04
p-value of difference	(0.00)	(0.00)	

Table 9: Plant TFP by status, 1998-2007

In this table we analyze plant TFP and the changes on their operational status in year t+2, where t denotes the treatment year. The regressions in panel A, B, C respectively are the following three specifications. TFP are mean-adjusted $\log(TFP_{it})$ within each industry-year cell.

(a) $TFP_{it} = f(plant's\ status\ in\ year\ t + 2, controls) + ownership\ FE + ProvinceYear\ FE + \varepsilon$;
(b) $TFP_{it+2} = f(plant's\ status\ in\ year\ t + 2, controls) + ownership\ FE + ProvinceYear\ FE + \varepsilon$
(c) growth rate of TFP from year t to $t + 2 = f(plant's\ status\ in\ year\ t + 2, controls) + ownership\ FE + ProvinceYear\ FE$

Panel A: TFP in Year t by Plant Status in Year t+2			
Dependent Variable: Plant level Log TFP in Year t	Treatment	Control	P-value for Difference Between the Treated and Controls
Continuous	0.31***	0.24***	0.07
	(0.00)	(0.00)	(0.23)
Exits	-0.26***	-0.10	-0.16
	(0.01)	(0.38)	(0.23)
R-Squares	0.15		

Panel B: TFP In Year t+2, Two Years After treatment, by Plant Status in Year t+2			
Dependent Variable: Plant-level Log TFP in Year t+2	Treatment	Control	P-value for Difference Between the Treated and Controls
Continuous	0.53***	0.37***	0.06**
	(0.00)	(0.00)	(0.03)
New Entrants	0.44***	0. 40***	0.04
	(0.00)	(0.00)	(0.78)
R-Squares	0.17		

Panel C: : Change in Plant-level TFP from Year t to t+2 for continuers			
Dependent variable: growth of Log(TFP) adjusted by mean within each industry-year cell	Treatment	Control	P-value for Difference Between the Treated and Controls
Continuous	0.15	0.45**	0.30*
	(0.46)	(0.04)	(0.08)
R-Squares	0.02		

Table 10: The impact of Transportation improvement on TFP in the manufacturing sector, 1998-2007.

The computation method follows Davis et al (2014). We first compute the average two-year change in TFP for all plants, treated plants, and control plants respectively. TFP is demeaned within each industry-year cell.

$$\Delta TFP_{t:t+2} = (S_{t+2}^{c}\ TFPt_{t+2}^{c} - S_{t}^{c}\ TFPt_{t}^{c}) + (S_{t+2}^{N}\ TFPt_{t+2}^{N} - S_{t}^{x}\ TFPt_{t}^{x})$$

S is the employment share, and C, N, X denote continuers, new entry and exiters.

We then compute the difference-in-difference, the change of TFP in the treated plants as the deviation from the TFP from the controlling plants over the same period $\Delta\widetilde{TFP_{t:t+2}}$

$$\begin{aligned}\Delta TFP_{t:t+2} - \Delta\widetilde{TFP_{t:t+2}} \\ &= \left(S_{t+2}^{c}\ (TFPt_{t+2}^{c} - \Delta\widetilde{TFP_{t+2}^{c}}) - S_{t}^{c}\ (TFPt_{t}^{c} - \Delta\widetilde{TFP_{t}^{c}})\right) \\ &+ \left(S_{t+2}^{N}\ (TFPt_{t+2}^{N} - \Delta\widetilde{TFP_{t+2}^{c}}) - \widetilde{S_{t+2}^{N}}(TFPt_{t+2}^{N} - \Delta\widetilde{TFP_{t+2}^{c}}) - S_{t}^{x}\ (TFPt_{t}^{x} - \Delta\widetilde{TFP_{t}^{c}}) \right. \\ &\left. + \widetilde{S_{t}^{X}}(TFPt_{t}^{x} - \Delta\widetilde{TFP_{t}^{c}})\right)\end{aligned}$$

Panel A: Estimated Average Change in TFP		
	1998-2007	2-year average
TFP Log Change Differential	20.81%	3.06%
Continuing Establishments	9.63%	2.31%
Entry and Exit	11.18%	0.75%

Panel B: Estimated Average Two-Year Post-event Change in TFP: Treated, Control, and Others			
	Treated	Control	Others
TFP Log Change Differential	4.15%	-3.43%	2.10%
Continuing Establishments	5.54%	0.06%	2.29%
Entry and Exit	-1.39%	-3.50%	0.19%

Appendix A: The technical notes for China's transportation network in GIS and summary of coverage.

I. Definitions of transportation system included in the data

(1). National Motorway description

In a strict definition, national trunk way is the newly built part of National highway after 1992 with the highest speed limit among all national highways. By January 2014, the total length is 85,000km. It is longest national trunk way in the world. The speed limit could be as high as 120km/hour. It have at least 2 lanes in each direction of traffic. The data section of the paper includes an overview of its construction timeline. The national and provincial trunk highway network connects administrative regions at and above the county level nationwide.

(2). National highway

There is a total length of 96,000km National Highway (not include National Motorway) in China. In the east and coast region of China, most of the national highway have 2 by 2 lanes or 3 by 3 lanes and a speed limit to 60-100km. In the middle and west China, the national highway often have 2 by 2 or 1 by 1 tertiary highway with speed limit as low as 20km/hour.

(3). Railway

China started to build railway system since Qing Dynasty 1892. It currently has the 2nd longest railway coverage in the world, 121,000km including both passenger and freight, ranked after USA. However, the railway length/population ranks behind 100 in the world. The railway system, aside from the High-speed railway, have two grades: railway and fast-speed railway. The latter refers to trains (both passenger and freight) with average speed between 160km/hour to 250km/hour with current 40,000km in total length.

(4). High speed railway

The high speed railway is developed after 2004. We have data on when each section of the high speed railway is launched up to 2014 and GIS map of all the high speed railway and its stations in operation by January 2014, which covers 19,000km in length. It is the longest high-speed railway coverage in the world. High-speed railway in China defined in China currently include passenger trains only. The average speed is no less than 250km/hour. The highest speed could reach 605km/hour.

(5) Waterway

By the end of 2015, China had 31,300 quay berths for production use. Inland waterways' navigable length was 127,000 km, with graded waterways accounting for 52.2 percent, and the length of high-grade waterways reaching 13,600 km. Our waterway data include all inner land and coastal lines that navigable for ships. The GIS map include these waterways and the ports for coastal lines.

II: Coverage

We have all the National Motorway in GIS map from 1993-2014 and year by year changes. The GIS map year is the year the map is published. The information presented should be treated with 2-3 years of lag to proxy for the actual timing of the transportation existence. By matching the GIS data observations and the highway construction completion data announced by the Ministry of Transportation, we recommend treating the data with 2 years of lag after 2002 and 3 years of lag before 2002.

We have the National highway map as in the year 1993. The actual locations of these roads remain mostly unchanged from 1993-2014. There are major improvements of these roads, including being partially integrated into a part of the National Motorway. Smaller improvement however are not captured by our data.

The length mentioned earlier is the expressway portion. There are also rural highways. The total length was 3.98 million km, connecting 99.9 percent of towns and townships and 99.8 percent of administrative villages as of 2016. Our maps include rural highway as long as they are part of the national highways.

We have data for all the railways and their stations exist in the year 2014 in GIS map, and the data on the timing and location for major railway additions and the speed improvement since 1950. Our data however cannot differentiate passenger versus freight trains by railway sections.

The railway additions and speed improvement are compiled based on the information published by The Ministry of Transportation on the actual data/year of the changes.

The data of high-speed railway introduction are also compiled based on the information published by The Ministry of Transportation on the actual data/year of the changes.

Our waterway data are based on the 1994 map. China has improved the navigation conditions of the Yangtze and Xijiang rivers and the Beijing-Hangzhou Grand Canal, and formed an inland waterway system composed of two horizontal trunk waterways, one vertical trunk waterway, two high-grade waterway networks and 18 high-grade mainstream and tributary waterways. These waterways exist in the database as their condition in 1994, without the improvement as a treatment. As we don't know the low boundary for the waterways to be included in the "navigable" definition, our analyses exclude those with a capacity of 100 tons capacity or below, just to have a clear definition.

We have not incorporated terrain map to cover the difference of sea-level heights between plants and exit/entrance or complication of the terrain. Our surface transportation data don't include airway, oil and gas pipelines.

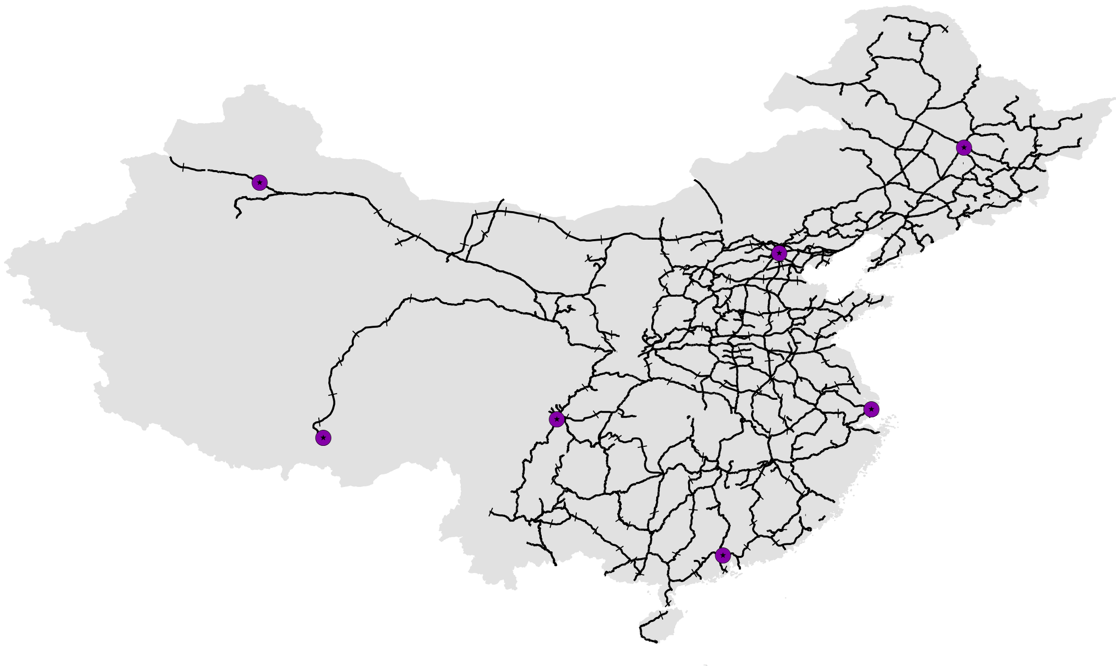
III. Coverage description in maps

Figure A1: The Development of China's Surface Transportation System

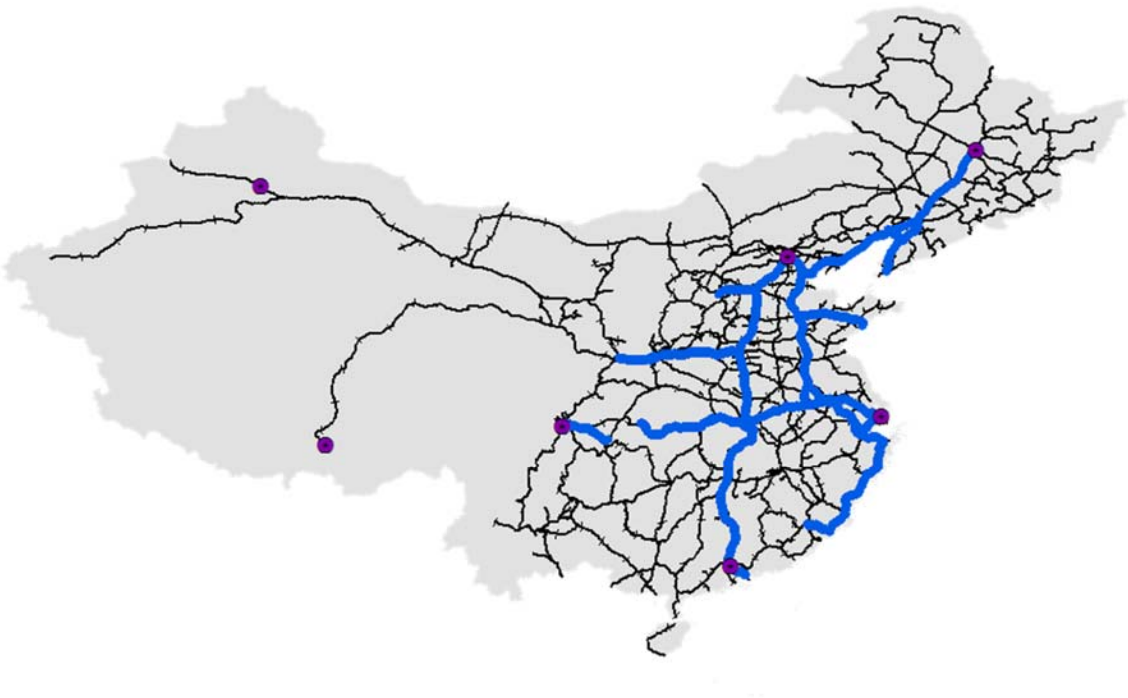
A. National Trunk Highways, 1993

B. National Trunk Highways, 2013

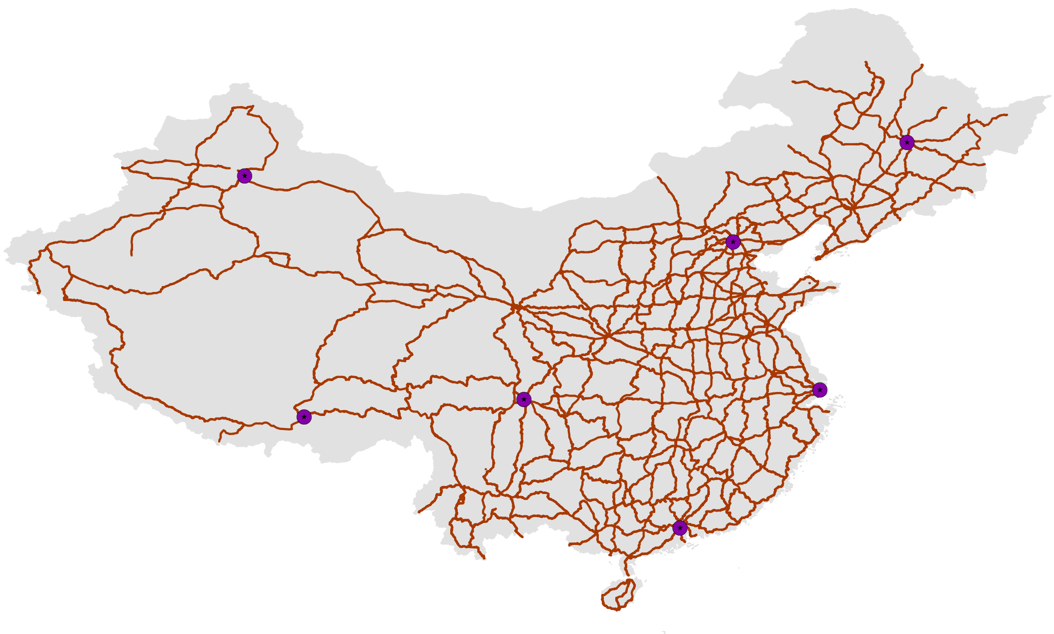
C. High-Capacity Freight and High-Speed Passenger Railway Lines, 1993



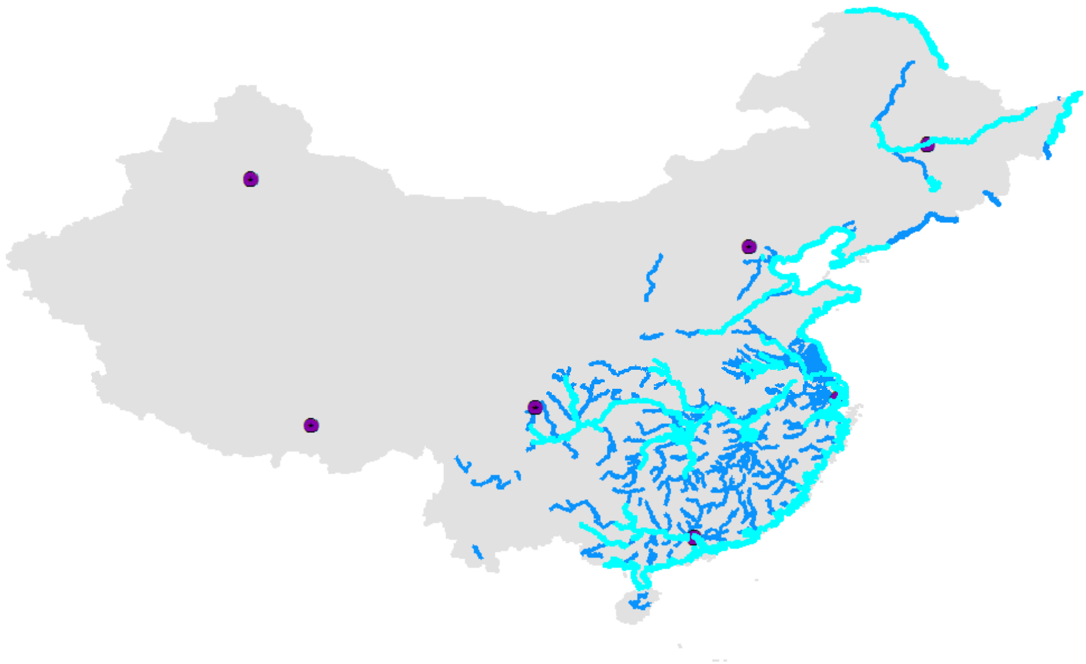
D. High-Capacity Freight and High-Speed Passenger Railway Lines, 2013



E. National highway, 1993



F. Major Navigable Waterways, 1993



Note: The green ones are waterway and coastway with ship capacity 100tons and above. The blue ones has shop capacity less than 100 tons. We exclude the blue ones in the analysis.

IV: Description of the Transportation Network on Coverage and Changes from the perspectives of Geographic coverage measured by zip codes.

Table A1: The coverage of the transportation network summarized by Zip Codes

The sample has a total of 34,889 unique zip codes. Panel A reports the percentage of zip codes whose center point is less than 25km away from the nearest access point to each transportation mode. Panel B reports the average distance from zip code center the nearest access point to each transportation mode.

Panel A: Portion of Zip Codes with nearest access less than 25km					
	Motorway	Railway	High Speed Railway	National highway	Water way
1995	0.12	0.53		0.71	0.52
1996	0.15	0.54		0.71	0.52
1997	0.18	0.54		0.71	0.52
1998	0.20	0.54		0.71	0.52
1999	0.22	0.54		0.71	0.52
2000	0.26	0.54		0.71	0.52
2001	0.31	0.54		0.71	0.52
2002	0.38	0.54		0.71	0.52
2003	0.38	0.54	0.01	0.71	0.52
2004	0.40	0.54	0.01	0.71	0.52
2005	0.44	0.55	0.01	0.71	0.52
2006	0.49	0.55	0.01	0.71	0.52
2007	0.55	0.55	0.01	0.71	0.52
2008	0.58	0.55	0.02	0.71	0.52
2009	0.60	0.55	0.06	0.71	0.52
2010	0.61	0.56	0.10	0.71	0.52
2011	0.63	0.56	0.14	0.71	0.52
2012	0.65	0.56	0.17	0.71	0.52

Pane B: Average Distance from zip center to the nearest access (km)

	Motorway	Railway	High Speed Railway	National highway	Water way
1995	298.89	38.81		19.64	108.87
1996	216.52	38.71		19.64	108.87
1997	179.25	38.61		19.64	108.87
1998	175.09	34.66		19.64	108.87
1999	160.99	34.28		19.64	108.87
2000	147.11	34.28		19.64	108.87
2001	129.17	34.26		19.64	108.87
2002	73.85	34.26		19.64	108.87
2003	73.61	34.24	1299.95	19.64	108.87
2004	70.86	34.23	1299.95	19.64	108.87
2005	61.77	34.01	1299.95	19.64	108.87
2006	55.70	34.01	1299.95	19.64	108.87
2007	50.37	34.01	1299.95	19.64	108.87
2008	46.50	34.01	594.55	19.64	108.87
2009	41.05	33.54	304.31	19.64	108.87
2010	39.02	33.27	268.99	19.64	108.87
2011	34.68	33.12	256.30	19.64	108.87
2012	33.05	33.02	222.39	19.64	108.87

Table A2: The coverage of transportation treatment (using 25km buffer) summarized by Zip Codes

Using 50km as the threshold of access to transportation improvement of any kind, 29,460 out of the total 34,889 zip codes, are affected during 1993-2012. Panel A reports the number of affected zip code by year and the percentage of them whose affected access fall into even a short distance. Panel B reports the number of affected zip codes using the 25km threshold and the percentage of them affected by each transportation mode changes.

	Total	Treat50k	Treat25k	Treat20k	Treat5k
1993	2,307	1	0.47	0.20	0.10
1994	3,282	1	0.45	0.18	0.08
1995	2,428	1	0.44	0.18	0.07
1996	2,993	1	0.52	0.20	0.09
1997	10,,908	1	0.61	0.28	0.16
1998	11,089	1	0.60	0.31	0.18
1999	10,293	1	0.49	0.18	0.07
2000	4,768	1	0.60	0.33	0.20
2001	8,112	1	0.60	0.30	0.17
2002	5,971	1	0.43	0.16	0.06
2003	6,840	1	0.48	0.20	0.09
2004	9,343	1	0.51	0.22	0.12
2005	11,926	1	0.53	0.19	0.08
2006	6,604	1	0.39	0.10	0.03
2007	14,760	1	0.58	0.29	0.17
2008	7,586	1	0.48	0.16	0.07
2009	8,455	1	0.59	0.31	0.19
2010	6,174	1	0.65	0.32	0.18
2011	27,849	1	0.74	0.41	0.22
2012	6,691	1	0.63	0.30	0.16

	Total	NewTrunk25k	NewRail25k	RImpr25k	NewHi25k
1993	2307	0.47	0.00	0.00	0.00
1994	3282	0.45	0.00	0.00	0.00
1995	2428	0.32	0.13	0.00	0.00
1996	2993	0.29	0.24	0.00	0.00
1997	10908	0.30	0.07	0.32	0.00
1998	11089	0.22	0.03	0.40	0.00
1999	10293	0.48	0.00	0.00	0.00
2000	4768	0.06	0.00	0.55	0.00
2001	8112	0.09	0.06	0.48	0.01
2002	5971	0.43	0.00	0.00	0.00
2003	6840	0.37	0.11	0.00	0.03
2004	9343	0.27	0.02	0.24	0.00
2005	11926	0.35	0.21	0.00	0.00
2006	6604	0.39	0.00	0.00	0.00
2007	14760	0.21	0.00	0.44	0.00
2008	7586	0.21	0.27	0.00	0.08
2009	8455	0.13	0.49	0.00	0.17
2010	6174	0.08	0.59	0.00	0.23
2011	27849	0.73	0.14	0.00	0.06
2012	6691	0.00	0.60	0.00	0.27

Table A3: Year distribution of treatment sample (within 50km of new exit/entrance points)

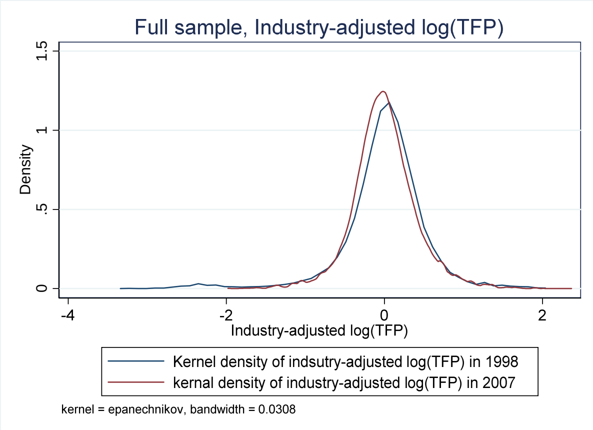
This table presents year-by-year the number of plants that locate within 50km of a new addition to or improvement in transportation system.

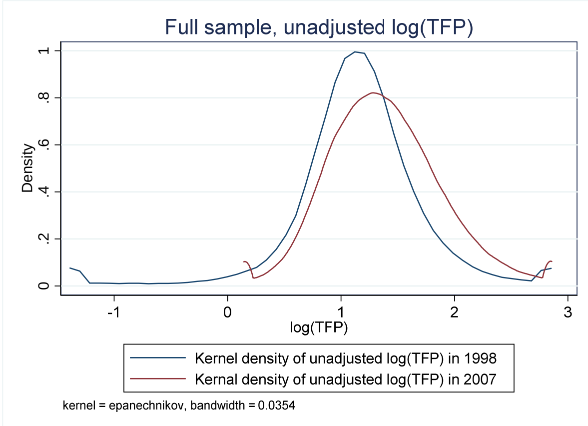
Year	Treatment of any one of the four kinds	Motorway additions	High-speed Railway Additions	Railway Additions	Railway Speed/improvement
1998	66,515	24,926		4,855	48,918
1999	75,609	75,567		42	
2000	30,066	3,927			26,570
2001	55,477	9,299	1,676	4,970	43,159
2002	49,260	49,260			
2003	58,020	53,628	2,375	5,263	
2004	93,113	50,401		623	51,139
2005	92,128	91,631	1,832		
2006	58,581	58,581			
2007	177,818	103,633			141,120
Total	756,587	520,853	5,883	15,753	310,906

Appendix B: Description of TFP of Chinese manufacture firms of their change over the sample period

Figure B1: TFP Distribution overview

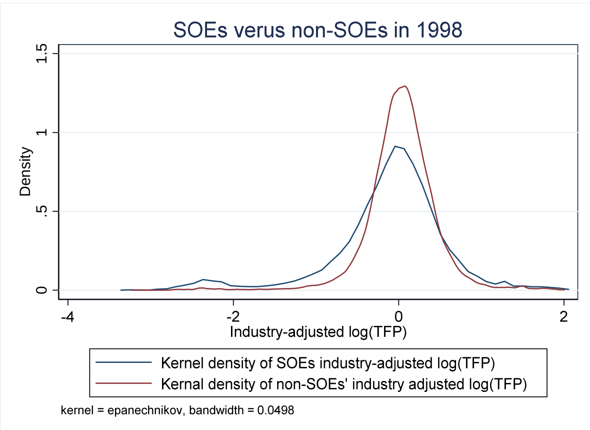
(I) Full sample

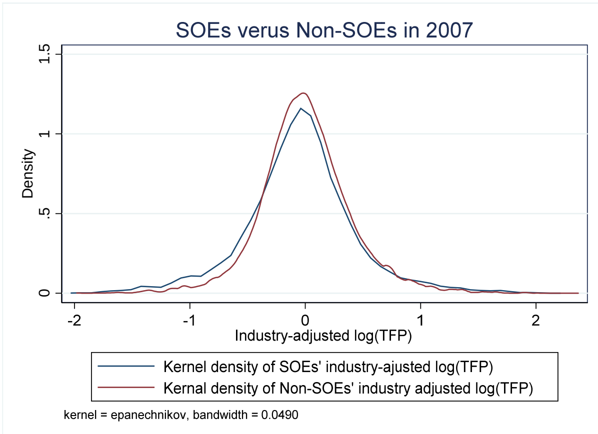




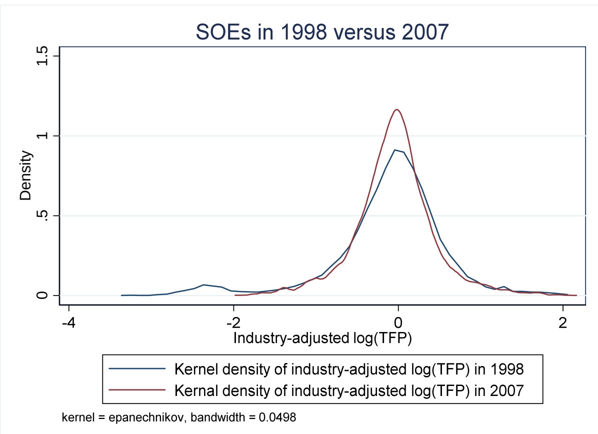
(a) **(b)**

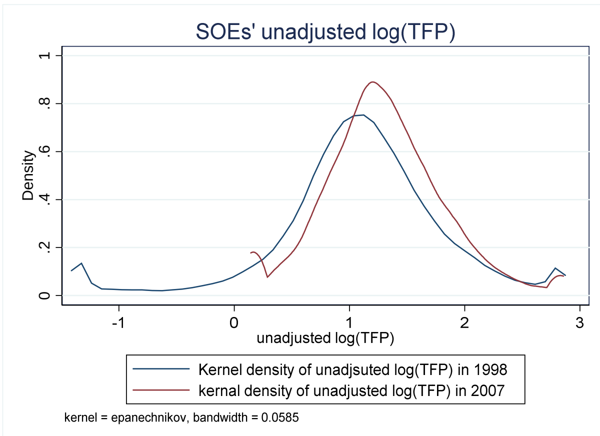
(II) By ownership



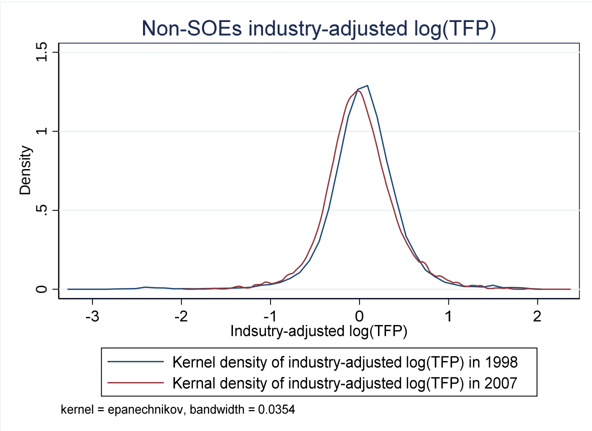


(c) **(d)**

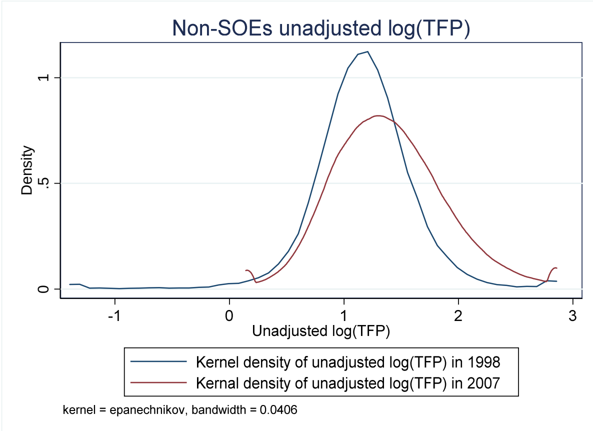




(e) **(f)**

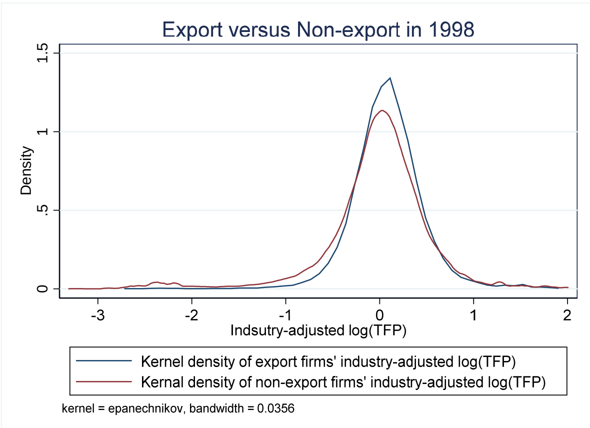


(g)

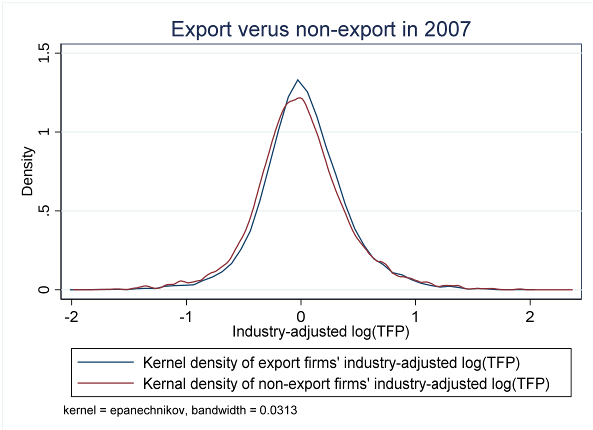


(h)

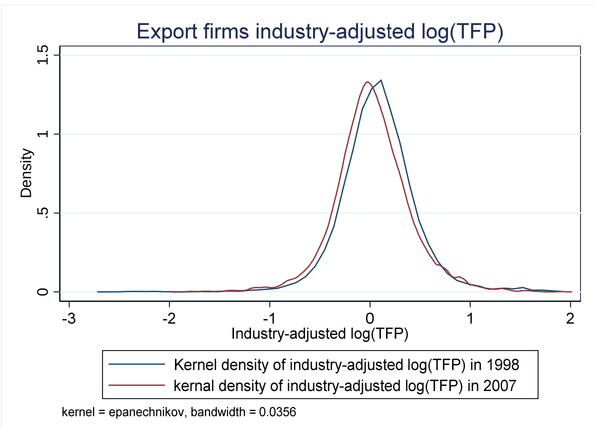
(V): By export



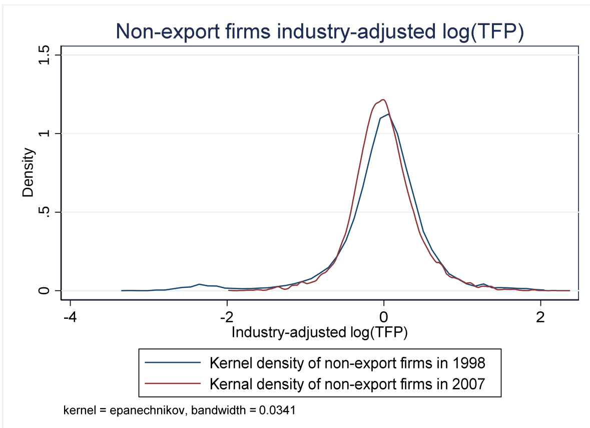
(i)



(j)



(k)



(l)

Table B2: Comparison of TFP distribution by access to and reliance on the transportation system

In this table, we contrast the trend of productivity distribution over time between groups of plants that are categorized by their access to or reliance on the transportation system. As we compare cross industries, we first adjust log(tfp) by industry mean for each industry and each year.

In panel A, we contrast two groups. One that has access to well integrated transportation system at both the beginning and the end of the sample period, but the other only at the end of the sample. We define a plant has access to a well-integrated transportation system, if it is within at least one of the four geographic buffers of 25km of highway or railway roads, 10km buffers for the old national roads or waterways. In panel B, we contract two groups: one for industries have high value over weight ratio and the other low value over weight ratio. We choose Gold, silver mining (921, 922), cigarette tobacco product (1620, 1690), silk product (1754, 1763), leather and fur product (1921, 1922, 1923, 1932), cosmetic and fur product (2672, 2674) industries as the former and Ore mining (933), tobacco stemming and redrying (1610), stone quarrying and mining (1011, 1012, 1013), wood and bamboo furniture (2110, 2120) as the latter.

Panel A: Plants with integrated transportation at the beginning of the sample period versus none

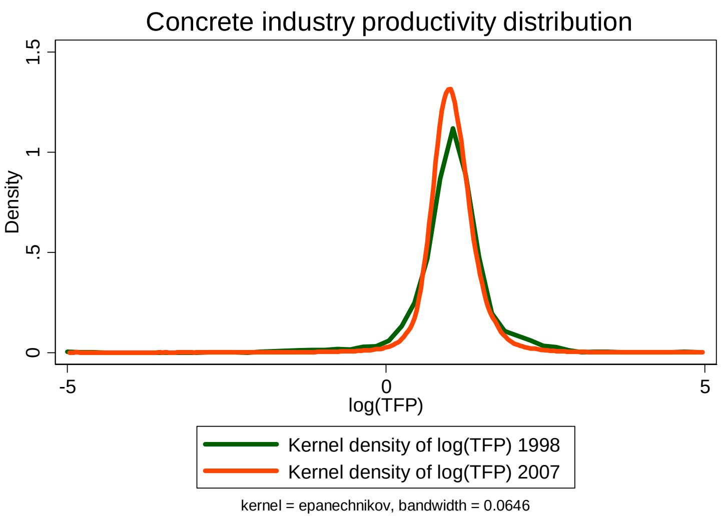
	Plants with integrated transportation at the beginning of the sample period (64%)				Plants with no integrated transportation at the beginning of the sample period (36%)			
	Unadjusted log(TFP)	Dispersion of log(TFP) adjusted by industry mean			Unadjusted log(TFP)	Dispersion of log(TFP) adjusted by industry mean		
Year	Mean	STD	P99-P1	P90-p10	Mean	STD	P99-P1	P90-p10
1998	0.92	0.46	2.84	0.97	0.93	0.48	2.84	1.07
1999	0.88	0.46	2.70	1.01	0.90	0.44	2.65	0.95
2000	0.90	0.49	2.95	1.05	0.95	0.47	2.83	1.00
2001	0.88	0.44	2.60	0.97	0.92	0.39	2.29	0.84
2002	0.91	0.44	2.51	0.96	0.95	0.39	2.34	0.85
2003	0.91	0.40	2.26	0.92	0.95	0.36	2.02	0.82
2004	0.90	0.38	2.06	0.89	0.93	0.34	1.83	0.80
2005	0.94	0.37	1.99	0.87	0.99	0.35	1.85	0.82
2006	0.97	0.36	1.95	0.85	1.02	0.35	1.84	0.82
2007	0.98	0.34	1.78	0.83	1.05	0.34	1.75	0.82
2007-1998	0.06	-0.12	-1.06	-0.14	0.12	-0.14	-1.05	-0.25

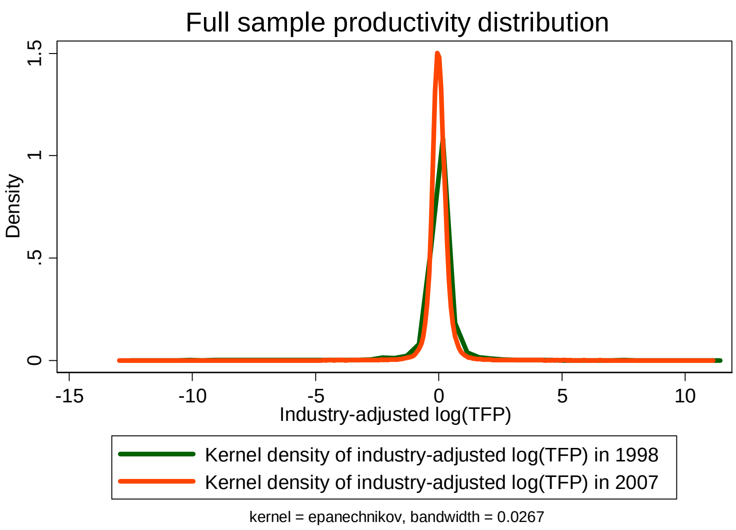
Panel B: Industries with High ratio of value/weight in transportation cost versus Low

	High value/weight industries				Low value/weight industries			
	Industry	Dispersion of log(TFP) adjusted by industry mean			Industry	Dispersion of log(TFP) adjusted by industry mean		
Year	Mean	STD	P99-P1	P90-p10	Mean	STD	P99-P1	P90-p10
1998	0.97	0.66	3.36	0.89	1.03	0.73	4.14	1.03
1999	0.87	0.63	3.29	0.83	0.99	0.57	3.46	0.96
2000	0.96	0.63	4.14	0.92	1.00	0.65	4.04	1.13
2001	0.92	0.62	3.32	0.76	0.97	0.55	3.27	0.87
2002	0.95	0.61	3.44	0.72	0.96	0.67	3.45	0.92
2003	1.00	0.51	2.94	0.69	0.99	0.55	2.96	0.87
2004	1.03	0.46	2.39	0.64	0.99	0.43	2.45	0.76
2005	1.09	0.41	1.98	0.65	1.06	0.52	2.65	0.78
2006	1.13	0.60	2.19	0.69	1.07	0.43	2.12	0.77
2007	1.17	0.34	1.78	0.64	1.15	0.48	2.27	0.79
2007-1998	0.20	-0.32	-1.58	-0.25	0.12	-0.25	-1.88	-0.24

Figure B3: Contrast of TFP distribution from 1998 to 2007

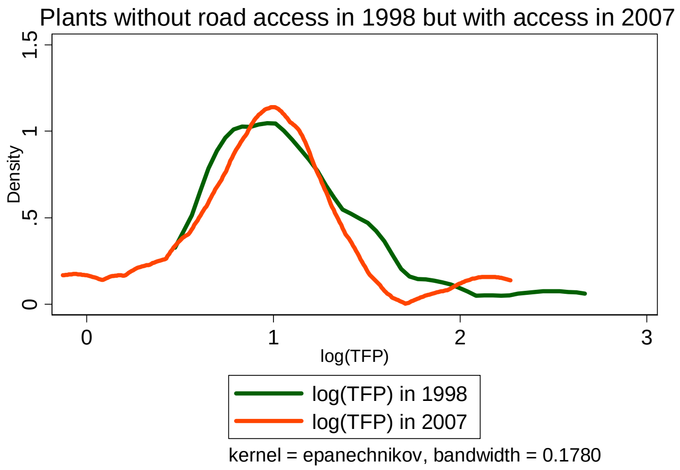
Graph (a) uses concrete industry as an example, include plants in cement (3121, 3123, 3124), brick, tile and stone construction materials (3129, 3131, and 3133), but exclude part of cement industry (3111, 3112, 3122, and 3132) because of dramatic fluctuation in these industry codes' sample size. Graph (b) contrast the distribution of industry-adjusted log(TFP) for the full sample; (c) Plants without access within 20km in 1998; (d) Plants with road access within 20km in 1998; (e) High value/weight industries; (f) Low value/weight industries.

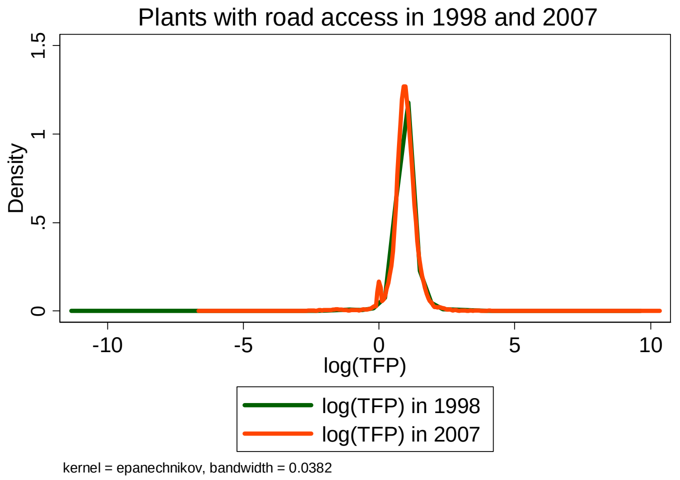




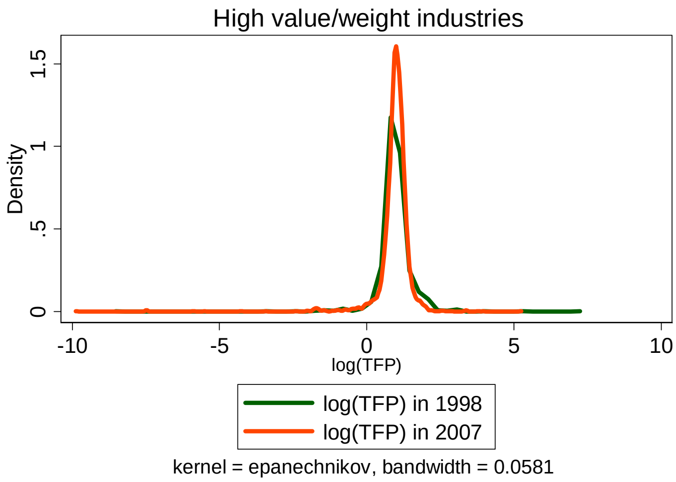
(a): Concrete example

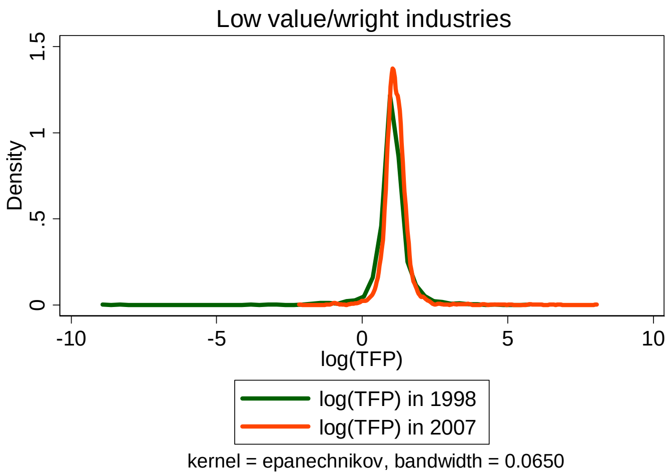
(b): Full sample, industry-adjusted





(c): Plants without access within 20km in 1998 (d): Plants with road access within 20km in 1998





(e) High value/weight industries

(f) Low value/weight industries

Figure B4: TFP and dispersion over time by region