Zhou, Jiangteng (2021) *The effect of transportation infrastructure development on firms' performance: evidence from China.* PhD thesis.

THE EFFECT OF TRANSPORTATION INFRASTRUCTURE DEVELOPMENT ON FIRM PERFORMANCE: EVIDENCE FROM CHINA

Jiangteng Zhou

Submitted in fulfilment of the requirements of the Degree of Economics

Adam Smith Business School, College of Social Science
University of Glasgow

February 17, 2021

Abstract

China's remarkable economic growth during the last four decades was accompanied with rapid infrastructure expansion. Its total length of high-class highways increased from 997,500 to 2,535,400 km from 1997 to 2007, but the following firm-level impacts are still unclear. This dissertation focuses on the impacts of rapid highway expansion on firm performance, composed of three topics. The first topic investigates the impacts of highway construction on firms' productivity, while the second topic focuses on the impacts of highway construction on firm size and size dispersion because the first topic reveals that productivity is strongly related to firms' scales. The baseline empirical studies of these two topics are based on a new constructed geographical highway dataset and the Annual Survey of Chinese Industrial Firms dataset, which enable us to estimate firm-level traffic accessibility. In order to address endogeneity issues, we construct IVs based on historical and counter-factual roads. The results of these two topics confirm that highway construction can promote the increase of productivity, size dispersion, and market concentration; this process is mainly motivated by the scale economies of larger firms, entry and relocation of new and small firms. The third topic focuses on the impacts of highway construction on export activities. It combines Chinese transaction-level customs dataset with the two datasets applied in the previous two topics, which enable us to estimate firms' export transportation costs to the nearest trade posts. In order to address endogeneity issues, new IVs are also constructed according to historical and counter-factual roads in this topic. The results show that highway construction can increase firm-level export value and scope, encourage incumbent or new entry firms to produce more differentiated goods. These impacts are more beneficial for inland regions and motivate Chinese export basket to evolve toward a more diversified structure and increase Chinese overall economic complexity.

Declaration

I declare that, except where explicit reference is made to the contribution of others, that this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Printed Name: Jiangteng Zhou
Signature: Jiangteng Zhou

Acknowledgement

First and foremost, I must say thank you to my PhD supervisors, Dr Sai Ding and Dr Hisayuki Yoshimoto, they provide the initial guidance for me to propose this research project. In the first year of my doctoral studies, I was not a funded student. My supervisors provided essential supports for my scholarship application, including reference letters and application guidance. Without these important supports, I could not become a full-funded student since the second year. During my research experience of the past four years, my supervisors enlightened me through their professional guidance very patiently, which enabled me to complete this research project on time. I strongly believe these constructive guidances will benefit my future development.

Besides my supervisors, I am also very grateful to Dr. Minjoo Kim, Dr Michael Tribe, Prof Jiuli Huang, Prof. Adelheid Holl, and Dr Zhiqiang Zhang respectively. Dr. Minjoo Kim and Dr Michael Tribe knew me well during my master's studies, they provided reference letters for my doctoral application and scholarship application at the very beginning of this project. I appreciate their patient and encouragement during my application. Prof Jiuli Huang seemed like my third unofficial adviser during my doctoral studies, and he brought unique resources to support us to complete this multidisciplinary research project. Prof. Adelheid Holl and Dr Zhiqiang Zhang are exports in economic geography and urban economics, they provided instructive comments for us to find useful instrument variables.

Last, but not least, I must praise the Adam Smith Business School for providing financial support for us to get the initial geographical dataset, which encourages us to initiate this project and bring it to the research frontier. I can feel that our business school and university really care about students' mental and physical happiness, they provide basic academic infrastructure which makes us to carry on this research more smoothly. I must appreciate my family and friends who surround me and encourage me to complete my doctoral studies, their love gives me the courage to overcome all kinds of difficulties.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

China witnessed significant highway expansion after the Asian financial crisis in 1997, its total length of Chinese expressways surpassed 25,000km during the 10th five-year plan (2000 to 2005), and then broke through 65,000km in the 11th five-year plan (2006 to 2010). Although many highway construction plans were proposed during the 1980s, the acceleration of Chinese highway construction was induced by the Asian financial crisis, as part of Keynesian counter-cyclic adjustment to address overproduction issues. These counter-cyclic economic policies were normally considered a great success by Chinese academic world, because they helped China to avoid potential depression and induced its high-speed increase in the next decade (Wen 2013). On the one hand, due to this wave of highway expansion mainly following the logic of Keynesianism, its direct investment return is problematic; on the other hand, the existing economic geography and development economy literature also point out that infrastructure development can generate significant positive externalities, it can reduce transportation costs, increase productivity, and affect the relocation of population and economic activities; during this process, the income disparity across rich and rural regions may also increase (Roller & Waverman 2001, Datta 2012, Faber 2014, Yang & Ng 2015). However, most existing studies focus on highways' aggregate effects, i.e., highways' impacts on region-level output, income, export, or productivity rather than firm-level factors, due to firm-level data constraints; the firm-level impacts of Chinese highway expansion have not been specifically detailed by previous studies.

The first topic (chapter three) is about the impacts of highway construction on firms' productivity, this can be specified from three aspects: first, newly constructed spatial and firm-level proxies to reflect firm-specific traffic accessibility; second, newly constructed instrumental variables, basing on historical and counter-factual road approaches; third, we apply heterogeneity and channel studies to explain the mechanism behind our baseline results. We find that firms closer to newly constructed highways have productivity premiums, a 10% decrease of distance to highways can increase firm-level productivity by 0.2%-0.3%, while this effect is more significant for those firms located in coastal provinces. Highways have lagged impacts on productivity growth, if we replace $Dist$ as one-year-lagged $Dist$, a 10% decrease of lagged $Dist$ is related to a 0.8% or 1.0% increase of firm-level productivity. These results are robust across different types of transportation modes (road, railway, waterway), different productivity measurements (OLS, OP, LP, and ACF productivity), and distance measurements (entrance distances). For channel studies, firms closer to highways have higher inventory level and outsourcing level, while larger firm size and higher productivity are related to higher inventory level and lower outsourcing level, a potential consequence is that firm size dispersion and market concentration will increase during this process. This firm size mechanism is further investigated in the next chapter.

On the basis of the same dataset, the second topic (chapter four) investigates highways' impacts on firm size distribution. This topic was initially inspired by our baseline results, which show that a 10% increase of firm size is related to 1.6% increase of productivity, many times stronger than the impact of highway expansion, implying firm size evolution may play an important role in firm productivity increase. The competition between large and small firms is basically asymmetric; on the positive side, the market selection mechanism will select productive firms to survive, these survival firms tend to be larger, more productive, and produce diversified products (Bernard et al. 2012, Freund & Pierola 2015, Foster et al. 2016); on the other side, higher productivity of some large firms also reflect their market influences and pricing-setting power rather than their production efficiency, sometimes the idiosyncratic shocks of a few giant enterprises can even cause macroeconomic fluctuations (Freund & Pierola 2015, Gabaix 2011, Magerman et al. 2016). The expansion of the road network will change the balance of power across market selection, relocation, industrial agglomeration effects, and reshape competition structure, while there are limited studies that have investigated the spatial impacts of highway development on firm size and market concentration. Our results show that highway expansion tends to increase firm size, but this impact varies from large to small firms, large enterprises tend to grow faster than other firms, while the quantity of small firms gradually increases, then the overall size dispersion tends to increase. This process is motivated by the rapid expansion of large firms, outsourcing activities, and the establishment of new and small firms. Firms with better traffic conditions are more likely to outsource their intermediate inputs and have higher inventory levels, but new entry firms tend to choose addresses around a certain brand of distance from highways, i.e. not very close and not very far from highways. Higher inventory level is correlated with larger firm size, while new entry and outsourcing firms tend to be smaller. These mechanisms can promote the increase of size dispersion and market concentration.

The third topic (chapter five) focuses on the impacts of highway construction on export activities. Chinese total export value (RMB) increased by 276% from 2000 to 2006, accompanied with a significant trade structure transition. Highway expansion is expected to reduce export transportation costs, especially for those firms located in inland regions, but firm-level studies of how highway expansion affects trade activities are limited. The results in the second topic (chapter four) confirm that firm entry and heterogeneous growth of large and small firms play essential roles during the sample period. Differently from the previous two chapters, a Chinese customs dataset (transaction-level data) is included to estimate firms' travel costs to the nearest trade posts, on the basis of an accumulated travel cost approach. To address the endogeneity issues, new IVs are constructed by the accumulated travel cost approach on the basis of historical roads and counter-factual roads, to reflect firms' travel costs along historical roads or counter-factual roads to the nearest trade posts. Our empirical results show that, from 2000 to 2006, firms closer to seaports (or coastal firms) are more likely to involve in international trade, having higher firm-level export values and higher exports-sales ratios. When we consider the export intensive and extensive margins, coastal firms normally export more types of different products, but focus on a smaller number of destinations. Consequently, firms closer to seaports tend to have lower export intensive margins, i.e., export volume of products, export volume of products in each destination. New exporters tend to emerge in inland regions. When we control the firm-region-industry fixed effects, the results suggest that highway expansion can increase the export value and export intensive margin of those firms not very close to seaports. A possible explanation is that those potential exporters close to seaports have already started exporting. Similarly, new entry processing exporters also tend to choose addresses not very close to seaports, a possible explanation is that those potential processing exporters close to seaports have already started processing exporting because of its lower entry barriers than ordinary exports. When we strictly control firm-industry-region time-invariant heterogeneity, the decrease of travel cost to seaports can reduce firms' processing share in export value, suggesting that highway expansion can induce the processing exporters to transform into ordinary or other exporters.

Chapter 2

Research Background and Highway Data

2.1 Background of Chinese Highway Expansion

In the last four decades, China has witnessed fast economic growth as well as rapid infrastructure expansion, while public investment has been frequently applied as a policy instrument to cope with external shocks and maintain economic growth. Wen (2013) claims that Chinese fixed asset investment including infrastructure investment is closely related to China's cyclical economic crisis. Different from the traditional economic crisis in western countries characterized by overproduction, Chinese economic crisis during the second half of the 20th century was normally accompanied with underproduction and fiscal crisis because China had maintained its highly-centralized economy for decades since 1949. For example, several crises can be identified from China's fixed asset investment history (figure 2.1-1); i.e., 1983, 1989, 1997, 2009, and 2014, characterized by sudden and rapid investment slowdown.[1] The reasons for these crises are various, e.g., from political unrest to external shocks, but the crisis in 1997 induced by the Asian financial crisis was the first overproduction crisis since 1949 (the second was triggered by the global financial crisis in 2008), so the counter-cyclic adjustments like massive fixed asset investment are more efficient to address the issues of insufficient demand. Chinese counter-cyclic economic policies during 1997-2000 were normally considered as successful adjustments by the Chinese academic world, which enables the Chinese economy to skip the potential depression and directly enter the next period of investment expansion (Juglar expansion period).

The whole picture of this dramatic counter-cyclic investment is difficult to elaborate, mostly due to data limitation; however, existing studies tend to support infrastructure investment is

[1] Data source: National Bureau of Statistics of China.

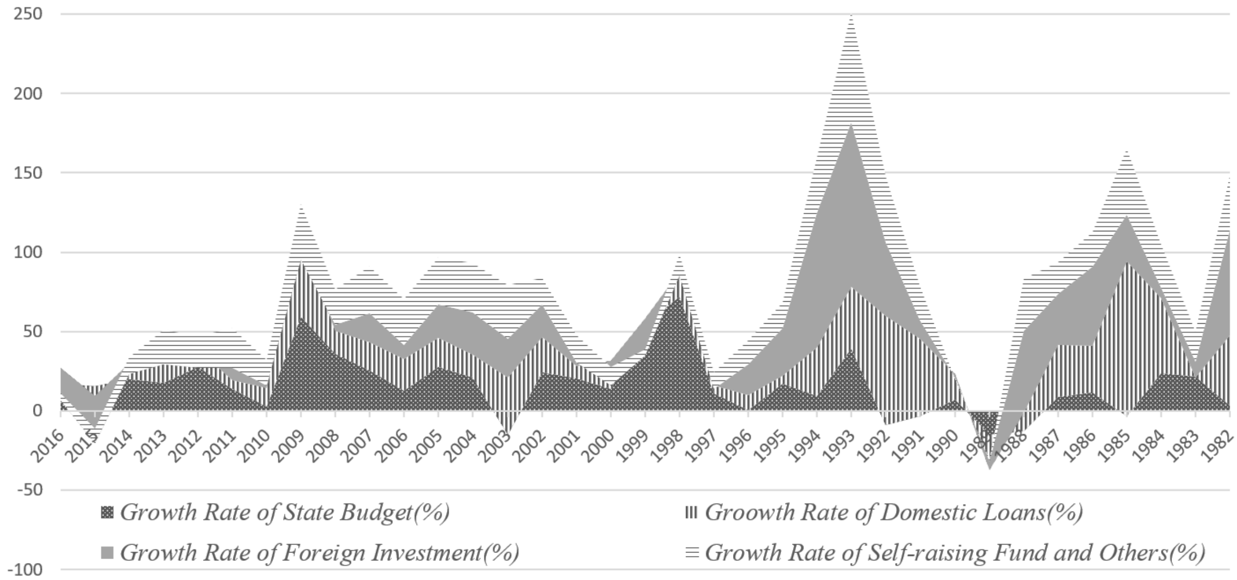


Figure 2.1-1: Growth Rate of China's Fixed Asset Investment by Different Sources of Funds.

an important component of this counter-cyclic adjustment, e.g., China's total highway length increased by 300% from 1997 to 2007, while expressways increased by ten times, from 4,800 to 53,900 kilometers. China's rapid highway expansion is fueled by its National Trunk Highway System Program (NTHS), which started in early 1990s but sped up after 1997 (Faber 2014), consistent with counter-cyclic investment period. This dramatic 'counter-cyclic coincidence' suggests infrastructure expansion could be driven by government's efforts to deal with crises, the logic to motivate these counter-cyclical measures does not have to be business logic (return maximization), but more likely the logic of Keynesianism, so the first priority of policy makers when they approved these infrastructure projects may not have been return maximization but to make up the demand shortfall caused by external shocks.

2.2 The '5-7' Plan

The acceleration of Chinese highway construction after the Asian financial crisis can also be found in China's five-year plan during 2000 to 2010. The 10th five-year plan (2000 to 2005) projected to increase highway length to 1.6 million kilometers, and expressways to 25,000 kilometers. Five years later, the 11th five-year plan (2006 to 2010) proposed to increase highway length to 2.3 million kilometers, and expressways to 65,000 kilometers, some parts of this plan belonged to the part of '5-7' plan in western provinces. The 11th five-year plan also proposed to build or update 1.2 million kilometers roads to connect all township-level administrative regions, connect all village-level administrative regions with asphalt roads or concrete roads in eastern and median provinces, and connect well-qualified village-level regions with asphalt roads or concrete roads in western provinces. The concept of 'villages and towns' comes from the definitions of Chinese administrative division, which divides Chinese administrative regions into five levels, i.e., province, prefecture, county, township, and village. According to the Chinese statistical bureau, Chinese township-level administrative regions include 8,515 urban directs (streets), 20,988 towns, and 9,222 small towns in 2019, each township-level administrative region has about 5,000 to 30,000 population; however, they are not the smallest administrative regions in China. The smallest administrative regions in China are village-level regions, including administrative villages and urban communities, each them normally has a population of thousands.

From these political documents, we can find Chinese road network expansion includes a series of high-quality highway construction projects and ordinary roads construction and updating projects, while our geographical dataset can exactly cover these high-quality highway construction projects. The '5-7' plan was the most important Chinese highway construction project during 1990s and the first decade of 21th century, and was firstly proposed in 1980s, then the technical details were decided by the Ministry of Communications during 1989 to 1990, and approved by the State Council of China in 1993. This road network includes 5 longitudinal highways and 7 latitudinal highways, these highways were mainly composed by Expressway, First Class Road, Second Class Road. The '5-7' plan aims to construct 3,500km highway within 30 years, to connect nationwide important cities, industrial centers, transportation hubs, and trading ports, including metropolis with more than one million people and medium-size cities with more than half million people. However, this process was largely accelerated by external shocks in 1997, just as mentioned above. In 2004, 'The National Expressway Network Plan' was made by Ministry of Communications and approved by State Council, as an important supplement of the '5-7' plan. 'The National Expressway Network Plan' is also called the the '7-9-13' plan, because it aims to construct 7 radial highways from Beijing, 9 longitudinal highways and 18 latitudinal highways; totally 34 roads, 8,500km, covering the 12 roads from the '5-7' plan. The '5-7' plan has been completed by 2007, the '7-9-13' plan was further modified to '7-11-18' plan in 2013, which aims to construct 7 radial, 11 longitudinal, and 18 latitudinal highways, totally 11,800km, projected to be completed by 2030.

The detailed '5-7' construction plan is specified below: the five longitudinal highways,

(1) Tongjiang in Heilongjiang province to Sanya in Hainan province (5,200km), (2) Beijing to Fuzhou in Fujian province (2,500km), (3) Beijing to Zhuhai in Guangdong province (2,400km), (4) Erlianhaote in Inner Mongolia to Hekou in Yunnan province (3,600km), (5) Congqing to Zhanjiang in Guangdong province (1,400km) respectively; the seven latitudinal highways, (1) Suifenhe in Heilongjiang province to Manzhouli in Inner Mongolia (1,300km), (2) Dandong in Liaoning province to Lhasa in Tibet (4,600km), (3) Qingdao in Shandong province to Yinchuan in Ningxia (1,600km), (4) Lianyungang in Liaoning province to Khorgos in Xinjiang (4,400km), (5) Shanghai to Chengdu in Sichuan province (2,500km), (6) Shanghai to Ruili in Yunnan province (4,000km), (7) Hengyang in Hunan province to Kunming in Yunnan province (2,000km).

2.3 Geographical Dataset

Table 2.3-1: Highway Engineering Technique Standard.

1) Top Class: Expressway	2) Second Class: First Class Road
Daily traffic volume: 25,000-55,000 cars for four-lane roads, 45,000-80,000 cars for six-lane roads, 60,000-100,000 cars for eight-lane roads	Daily traffic volume: 15,000-30,000 cars for four-lane roads, 25,000-55,000 cars for six-lane roads
3) Third Class: Second Class Road	**4) Fourth Class: Third Class Road**
Daily traffic volume: 3,000-7,500 trucks	Daily traffic volume: 1,000-4,000 trucks
5) Fifth Class: Fourth Class Road	
Daily traffic volume: less than 1,500 trucks for two-lane roads, less than 200 trucks for single-lane roads	

Note: This table provides the traffic capacity for each type of highway. The ACASIAN dataset mainly contains expressways and first-class highways which are also named high-class highways.

According to Highway Engineering Technique Standard (see Table 2.3-1), highways are classified into five classes, i.e., expressway, first, second, third, and fourth class, while expressways, first and second class highways are normally considered as high-class highways (gaodengji). At the same time, Chinese urban roads are classified into four classes, i.e., urban expressways, main roads, minor roads, and branch roads. Although the name 'urban expressway' (kuaisulu) includes 'expressway', they are normally constructed with the technical standards of first class, some of them are constructed at second or expressway class. Urban expressways are normally independent from other urban roads but connected with highways, they do not have traffic lights and they are not toll roads, while vehicles need interchanges to get access to urban expressways.

The ACASIAN dataset (The Australian Consortium for the Asian Spatial Information and Analysis Network.) is constructed and authorized by Professor Lawrence Crissman from Griffith University. In the existing Chinese infrastructure literature, this dataset is only used by Faber (2014), Liu et al. (2017). This dataset provides the GIS vector data of four types of Chinese transport routes: (1) expressway networks in 1992, 1993, 1998, 2000, 2002, 2003, 2005, 2007, 2010, and 2011, (2) national and provincial road networks in 2007, (3) railroad networks in 1997, 1998, and 2000 (4) navigable waterways, constructed on the basis of existing rivers, lakes, and canals. This thesis only uses the expressway network dataset, while other types of transport routes such as railroads, waterways, national and provincial roads (low-class roads) are controlled by province-level route density. The GIS data of the expressway network is digitalized from published atlases, providing the vector data of 'expressways' and 'first-class roads' defined in Table 2.3-1, distributed in Decimal Degrees (WGS84), with a scale of 1:1M.

However, as our firm-level dataset covers the period from 1998 to 2007, there are only six-year observations that can be matched with the original ACASIAN dataset, so we reconstruct ACASIAN to generate continuous observations from 1998 to 2006. We firstly collect published paper atlases as basic updating information. The map book in 2003 is published by People's Communications Press, the map book in 1998 is published by Xingqiu Press, while other books are published by China Atlas Press. Books in 1998, 1999, and 2000 contain expressways, other highways, and general roads. Books in 2000, 2001, 2002, 2003, 2004, 2005, and 2006 contain high-class highways, other highways, and general roads. Some urban expressways are in province capitals. Book in 2007 contains high-class highways, interchanges of high-class highways, other highways, general roads, and some urban expressways in provincial capitals.

Second, we update the ACASIAN dataset in the missing years (1999, 2001, 2004, and 2006) basing this on published atlases according to a key assumption, i.e., the highway network has tended to expand over time, highways were never destroyed or closed once completed. According to this assumption, the highway network can only expand rather shrink over time, the highway network in 2010 should contain the network in all other years, so the network in 2010 is considered as base-year observation of the largest highway network. The third step is to update highways in 1998, 2000, 2002, 2003, 2005, and 2007, which are provided by the original ACASIAN data. For each year, if a highway was recorded by our published paper atlas but not recorded by the original ACASIAN data, it would be added into the original ACASIAN data; if a highway was recorded by the original ACASIAN data but not recorded by the published paper atlas, no road would be deleted, because the original ACASIAN data is constructed based on more than one atlas for each year, it is expected to contain more information than any single published paper atlas. At the end of this step, we get the updated observations in 1998, 2000, 2002, 2003, 2005, and 2007 (updated ACASIAN data).

The final step is to construct the highways that are not included in the updated ACASIAN data, i.e., 1999, 2001, 2004, and 2006. For example, observations in 2006 are constructed based on the updated observations in 2007 and 2005, on the basis of our key assumption that the road network can only expand rather shrink over time. The intuition is that we can delete some roads in 2007 to generate the network in 2006. At the same time, according to the key assumption, we also use the network in 2005 as a comparison to make sure there would not be too many roads to be deleted from the 2007 network. Therefore, there are basically two rules: (1) if a highway is not reported by the published paper atlas in 2006, it would be deleted from the observations in 2007; (2) if a highway is not reported by the atlas in 2006, but reported by ACASIAN data in 2005, it would not be deleted from the observations in 2007. At the end of this step, we get the continuous observations from 1998 to 2007, i.e., Figures 2.3-2 and 2.3-3.

Figure 2.3-2: Chinese Highways Network Expansion from 1998 to 2002

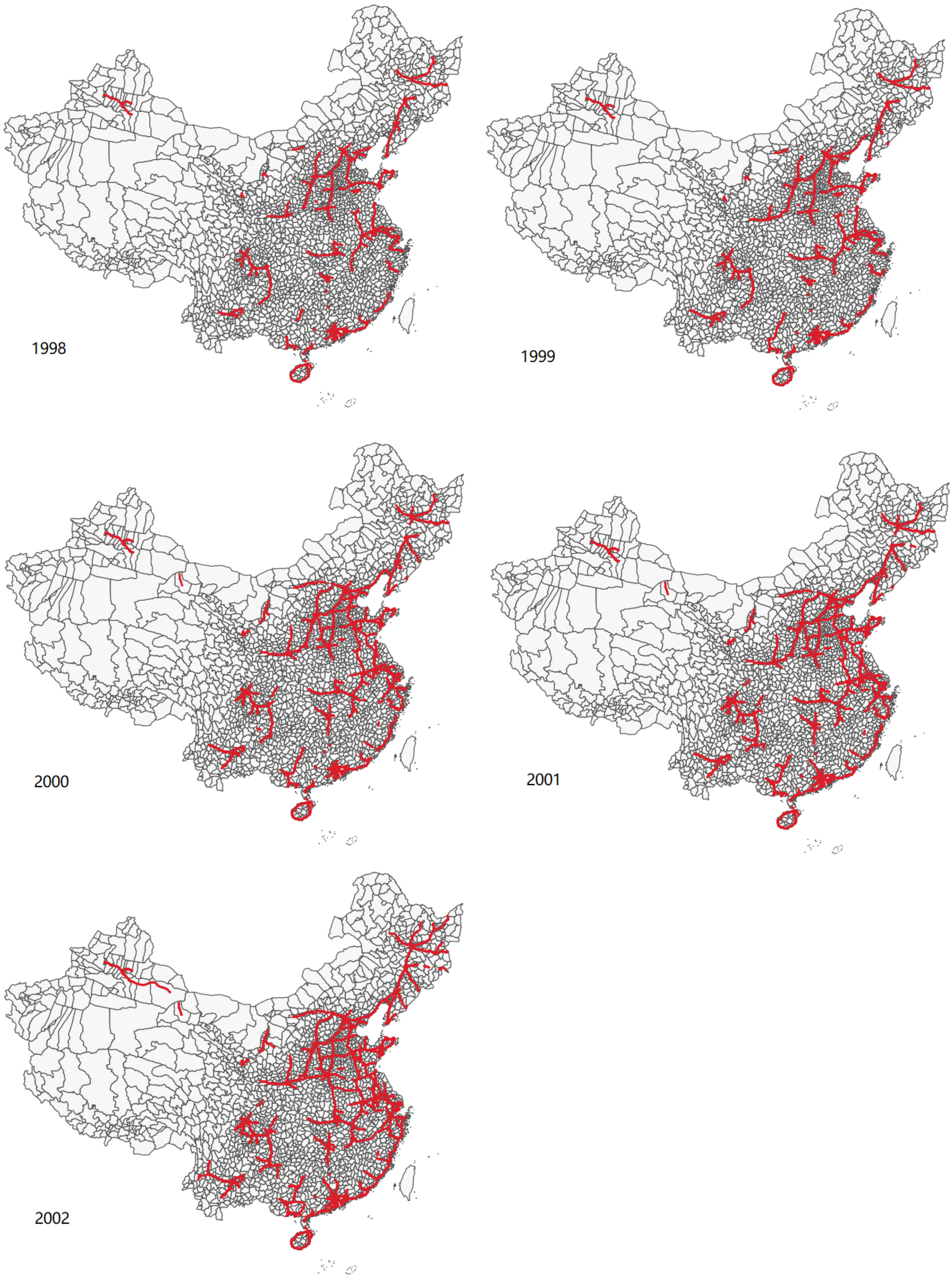
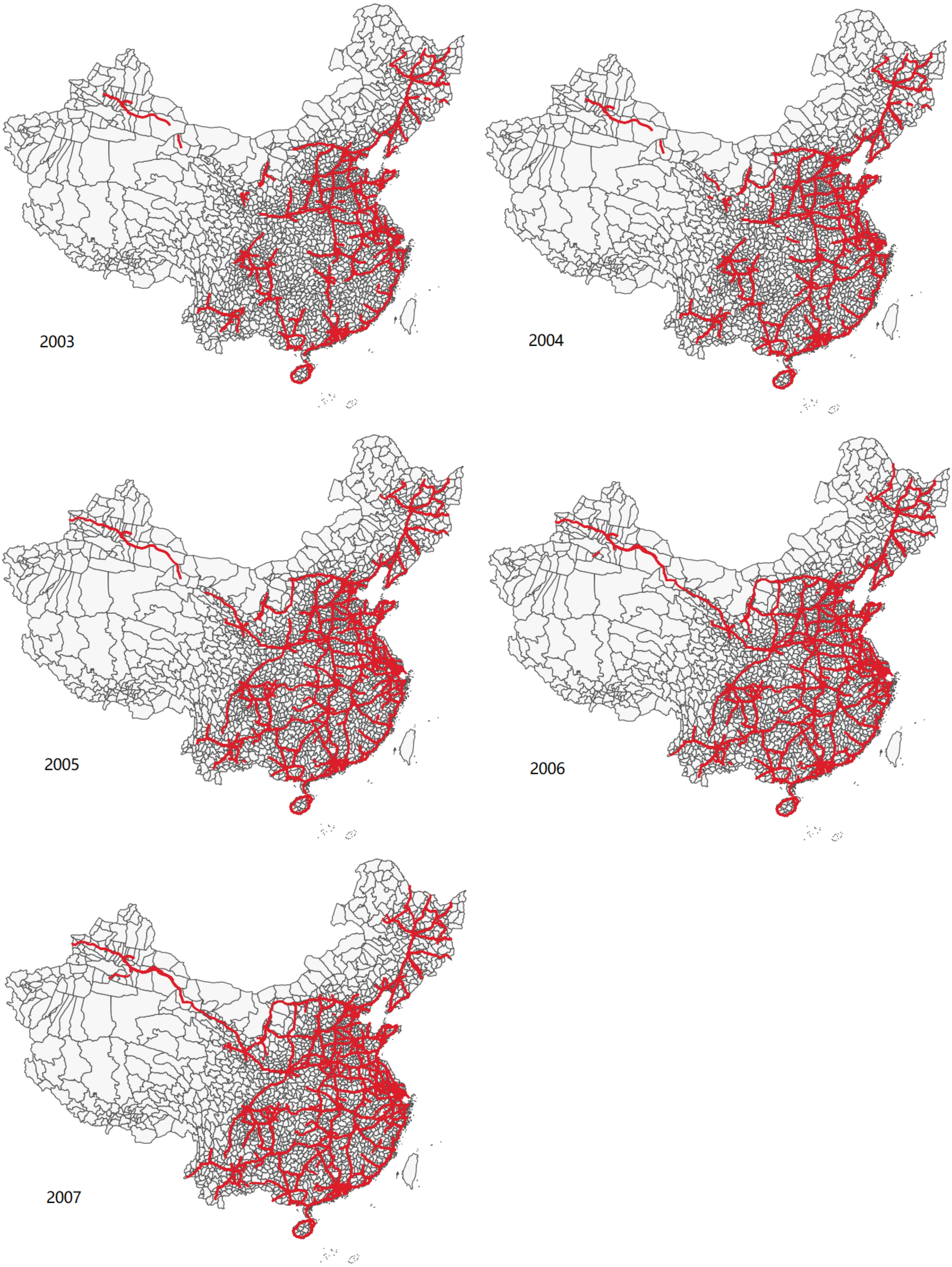


Figure 2.3-3: Chinese Highways Network Expansion from 2003 to 2007



Chapter 3

Impacts of Highways on Firms' Productivity

3.1 Introduction

Due to China's rapid infrastructure expansion being closely related to counter-cyclic adjustments, the logic motivating these adjustments may not follow business logic (return maximization), but more likely the logic of Keynesianism. The assessment of economic return includes direct impacts of capital formation and spillover effects over regions with newly constructed road networks, while highways' firm-level impacts have not been investigated in detail by previous studies. This chapter investigates the impacts of high-class highway expansion on firm-level productivity. Existing studies find that the development of transportation infrastructure seems to be a double-edged sword. On the positive side, it can reduce firms' inventory costs, direct logistics costs, and logistics time costs (Fernald 1999, Datta 2012, Yang & Ng 2015), then increase firms' profits and productivity; at the same time, it can also promote the relocation of population and economic activities; while more concentrated firms and population will encourage knowledge sharing and connections between upstream and downstream firms, then generate agglomeration economies and scale economy (Holtz-Eakin & Schwartz 1995, Roller & Waverman 2001, Faber 2014). On the other hand, the improvement of transportation system can strengthen the competitive advantages of rich regions, so the relocation of population and economic activities will deepen the income disparity across rich and rural regions (Faber 2014). In addition, if China's rapid road expansion started from 1997 is essentially counter-cyclic expansion, the high level of public investment may have a crowding-out effect and tighten credit constraints on private sector (Cavallo & Daude 2011).

The contribution of this chapter can be specified from three aspects: (1) Different from previous Chinese research, we calculate firms' distances to expressways to capture firm-specific traffic accessibility based on the manufacturing firm and geographical highway dataset. (2) To address potential endogeneity issues, we construct IVs through historical road and counter-factual road approaches; these new constructed IVs are efficient and provide consistent results across different fixed effects and robustness checks. (3) On the basis of IV and baseline regressions, we investigate two possible channels which may explain these highway construction effects, i.e., inventory channel and learning-by-doing effect in international trade. We also investigate two heterogeneous effects, i.e., heterogeneous impacts across coastal and inland provinces, and heterogeneous impacts across goods with high or low value-weight ratios.

The baseline regressions show that firms closer to newly constructed highways have productivity premiums. The decrease of firms' distance to highways is induced by two mechanisms: on the one side, highway development can reduce $Dist$ of address-unchanged firms; on the other side, highway development can also promote firms' entry and relocation, then reduce the overall $Dist$. a 10% decrease of distance to highways can increase firm-level productivity by 0.2%-0.3%, and this effect in coastal provinces are stronger than in inland provinces. Highways have lagged impacts on productivity growth, if we replace $Dist$ as one-year-lagged $Dist$, a 10% decrease of lagged $Dist$ is related to a 0.8% or 1.0% increase of firm-level productivity. These results are robust across different types of transport modes (road, railway, waterway), different productivity measurements (OLS, OP, LP, and ACF productivity), and distance measurements (entrance distances). Firms closer to highways have higher inventory levels and outsourcing levels, while larger firm size and higher productivity are related to higher inventory levels and lower outsourcing levels, a potential consequence is that firm size dispersion will be increased during this process. This firm size mechanism is further studied in the next chapter.

This chapter is structured as follows. Section 2 reviews the literature about infrastructure construction effects, instrument variable construction, and productivity estimate approaches. Section 3 introduces the data sources and research methodology. Section 4 analyzes the empirical results. Section 5 concludes.

3.2 Literature Review

3.2.1 Infrastructure and Agglomeration Economies

Most economic geography literature supports the view that transportation infrastructure has both aggregate effects and relocation effects. Aggregate effects mean regional economic performance can benefit from well-developed road network, because more convenient traffic environment can increase firms' market accessibility, firms' efficiency to find suitable suppliers, idea sharing, then strengthen connections between upstream and downstream firms (Xu 2017, Banerjee et al. 2012) which are expected to have positive impacts on productivity growth. On the other hand, a more convenient traffic environment attracts more rivals to enter rural regions and increase competition, then promotes entry of high-efficiency producers, exit of low-efficiency producers, and relocation of population and economic activities between rich and rural regions. During this process, migration flow from rural to rich regions will slow the growth in rural areas, because infrastructure improvement increases employment accessibility and strengthen the comparative advantages in rich regions (Faber 2014, Gibbons et al. 2016).

Spatial aggregation of population and economic activities in rich regions normally leads to agglomeration economies. Without agglomeration economies, cities should keep to a small scale, while a metropolis should never form because clustered population and economic activities will lead to higher rent prices, more competition and traffic congestion, and downward force on price and profits (Duranton et al. 2015). However the reality is that more than half population are living in urban areas, and this proportion is projected to reach 70% by 2050[1]; metropolises are growing stronger, attracting intelligent minds to create a booming new economy, which can only happen if agglomeration economies can generate larger positive externalities than agglomeration diseconomies.

Transportation infrastructure is the amplifier of agglomeration economies. Similar to the mechanism of transportation infrastructure, agglomeration economies can also strengthen input-output and producer-consumer linkage, decrease matching costs, and promote knowledge sharing and spillover (Duranton & Puga 2004). When agglomeration economies and diseconomies reach a dynamic balance, the distribution of different sectors will present a concentric zone model (Alonso et al. 1964). According to bid rent theory, different sectors

[1] According to the projection of World Bank.

such as retail and manufacturing industries will compete for land use rights, then tertiary industry will occupy the Central Business District (CBD), surrounded by manufacturing and agricultural industries. This fact implies that even if manufacturing firms closer to traffic lines may benefit from agglomeration economies, it does not mean 'the more recent the better', because if they are too close to the CBD they have to face additional competition from other sectors and get downward pressure on their revenue and profit. However, empirical studies tend to neglect agglomeration diseconomies and focus on a linear relationship between traffic environment and economic performance, which may explain why studies using different measurements provide conflicting results.

3.2.2 Evidence of Transportation Infrastructure Effects

Holl (2016) finds that firms' greater proximity to highways has a negative influence on firms' productivity, under the control of agglomeration effects, location effects, historical agglomeration effects and firm specific effects.[2] At the same time, reverse causality could be a problem because the location of new highways could be endogenous as pre-existing factors, such as population and administrative level; to solve potential endogeneity problems, Holl assumes that distribution of ancient roads affects current roads but is independent from the distribution of current population and economic activity, then uses distances to historical roads (Roman Roads and 1760s Postal Roads) as the instruments of the distance to current roads. In addition, her results also show that companies in suburban areas benefit more from highway improvement, manufacturing industry are more sensitive to highway expansion than other industries; suggesting regional and industrial heterogeneity should be considered in Chinese case studies.

There is also different evidence, Garcia-Mila et al. (1996) investigate the influences of public capital on private outputs. Public capital stocks are divided into three main types: transport infrastructure (highways), water and sewers infrastructure, and the rest. They find highways have significant positive effects on private output, slightly smaller than the effects of private capital. However, when they take the difference of these public capital variables, the influence become insignificant, suggesting the possibility of spurious regressions; then they take the difference for every two years and run the same regressions but still get insignificant re-

[2]Agglomerate effects are indicated by population and employment density; location effects include factors such as longitude, latitude, terrain ruggedness; historical controls include historical population; firms specific effects include variables such as firms' age, size and ownership

sults. However, these results cannot deny the importance of highway development, because public capital may have lagged effects on economic output increases, while firms also need time to adjust their production and locations.

Gibbons et al. (2016) apply two approaches to investigate the impact of transport infrastructure construction on firm productivity and employment in U.K. from 1997 to 2008. The first approach focuses on 10,300 electoral wards, investigates highways' aggregate and cross-industry effects on ward-level employment, and plant quantity. Road improvement is indicated accessibility index, instrumented by its lagged variable. Results show road improvements have strong impacts on employment and plant counts, 10% increase of accessibility associates with 3% increase of employment and plant counts for each electoral ward. The second approach focuses firm-level effects, showing the impacts of highway improvement on plant-level employment and output close to zero, suggesting road improvements influence employment mainly through firm entry and exit rather than firm-level employment increases.

Faber (2014) also focuses on highways' aggregate impacts, basing on Chinese region-level dataset and ACASIAN geographical dataset. Similar to Holl's arguments, Faber claims endogeneity issues may rise from some unobservable factors like population distribution and administrative level. Different from existing studies, Faber uses counter-factual roads to construct IVs, and finds that rural regions connected by China's National Expressway Network witnessed slower industries output growth than no-connected counties, counties with more population or closer to province capitals/metropolis are growing faster than others. Faber's results suggest agglomeration effects can promote economic growth when counties near big cities, but nationwide regressions show there are some 'mysterious forces' slowing economic growth for connected cities, they could be agglomeration diseconomies on nationwide perspective, or competition effects. Hence, firm-level studies are needed to confirm whether firms follow similar model when roads are constructed.

Consistent with Faber's study, Baum-Snow et al. (2012) investigates relocation effects on population concentration and economic output, but use different measurements. They divide both highways and railways into radial and ring lines, then count the quantity of radial and ring lines. To solve potential endogeneity problems, they construct instruments based on the 'no initial highways' assumption, i.e., highways before 1990 are not suitable for domestic trade because of their bad quality, then historical roads in 1962 should be exogenous of current shocks like GDP and productivity distribution, but they can also affect the distribution of current radial roads and ring roads, satisfying the requirements of efficient instruments. Their results show infrastructure improvement promotes the decentralization of population and economic activities, but the effects of radial ring lines are much larger than radial lines: both ring roads and radial roads have significant negative impacts on population agglomeration in the city center, but the absolute values of the coefficients of ring roads are much bigger than radial roads (-0.04 versus -0.20); radial railroads and ring roads also have significant effects on city center GDP, but the coefficient of ring roads is still larger than radial railroads (-0.20 versus -0.50). At the same time, heterogeneity studies show infrastructure construction effects become larger after excluding western regions, implying infrastructure investments are more efficient in eastern regions; the coefficients of traffic lines are negative between 1990 to 2000, consistent with nationwide regressions, but the absolute values of these coefficients tend to become smaller or even positive after 2000, suggesting there could be a structural breakpoint around 2000, perhaps because of China joining into WTO in 2001. The existence of structural breakpoint and strong relocation effects are consistent with Faber's research, however, population decentralization may not be the unique consequence of infrastructure investment, slower population growth could also be affected by income level, which is supported by the differences between western and eastern regions.

Existing research do not always find positive effects of infrastructure improvement on economic growth in China, but more significant relocation effects; relocation of population and economic activities make rich regions get more benefit than rural regions, whether for road construction (Faber 2014), railway construction (Baum-Snow et al. 2012), or high-speed railway construction (Xu 2017). At the same time, existing studies tend to use region-level datasets, combined with various spatial indicators, e.g., distance, accessibility index, or road quantity. This study aims to provide firm-level evidence, to explain whether firms' responses to infrastructure improvement are consistent with aggregate-level studies.

3.2.3 Review of IV Construction

When policy makers decide where to construct new highways, their decisions might be affected by economic activities distribution, productivity distribution, or even political factors, which may cause endogeneity problem. To solve this problem, existing research tend to use various instrument variables.

Previous studies have mainly developed five types of instruments to address endogeneity issues, i.e., time-lagged term, historical road, geographical factors, construction plan, and counter-factual road approaches. First, the most straightforward and direct approach is to use time-lagged terms Coşar & Demir (2016), but this approach will shrink sample size. What we prefer to do is to construct geographical-factor-based IVs.

Second, some studies use ancient roads to estimate traffic accessibility index as the IVs of current traffic accessibility (Michaels 2008, Martincus & Blyde 2013, Holl 2016, Duranton & Turner 2012). This approach assumes that, as historical roads were constructed earlier than current roads, their spatial distribution should be independent of current roads; at the same time, the distribution of population and economic activities are initially decided by geographical conditions, this correlation should keep stable over time, so the distribution of ancient roads can be used to predict the distribution of current roads. However, historical indicators are time-invariant, therefore they might be eliminated by fixed effect term in regressions.

Third, Geological factors, such as terrain ruggedness and the distribution of nature resources, can also be used to construct IVs, because the initial geology conditions can influence population aggregation, economic activity, and road construction in the very long history, while these conditions are obviously independent of economic output nor productivity distribution (Holl 2011). However, similar as historical IVs, geology factors can only generate cross-section data.

Fourth, some authors use construction plan to generate counter-factual IVs (Baum-Snow 2007, Michaels 2008, Duranton & Turner 2012, Donaldson 2018). When initial construction plans are proposed, firms and consumers do not have time to change their construction plan or budgets. At the same time, these proposed projects are not unchanged, on the contrary, they may be modified, relocated or canceled during political wrangles, suggesting that those proposed lines can be applied as a kind of counterfactual indicators because they tend to be independent of current productivity shocks. The construction could be time-variant but the data of construction plan is not usually available in practice.

Fifth, Faber (2014), Liu et al. (2017), Tang et al. (2019) use terrain data to construct counterfactual roads as IVs, based on the assumption that policy makers want to minimize their budgets and select the routes with the least construction costs. This approach mainly contains two steps: the first step is to use remote sensing data, which describe land cover and slope information, to generate least cost routes between each couple of node cities (capitals/metropolis); the second step uses a minimum spanning tree approach, to select the cost routes to construct a road network with minimum construction costs in a single continuous network. These counter-factual roads should be independent from unobservable shocks because they are generated only based in a cost minimization rule. At the same time, each policy maker may consider construction costs when they determine the location of new roads, so the distribution of cost-minimized roads should be correlated with current roads. However, there are still some limitations: (1) Similar to historical instruments, least cost paths is also a time-invariant network. (2) This approach replies on 'node cities', while the selection of 'node cities' has to be subjective. In practice, with the economic growth of China, there might be more and more cities become 'node cities'. Consequently, the selection of node cities might be endogenous to some unobservable factors.

According to these studies, IVs based on historical roads, nature resources, and least cost approach are time-invariant; IVs based on infrastructure construction plans could vary over time, which depends on whether the initial plans have detailed specifications for their construction schedule. Time-invariant IVs have some problems in panel regressions, because they could be identified as fixed effects. By linking the node cities with straight lines and generating buffer areas around these lines, Hornung (2015) constructs a time-varying binary instrument, which would be denoted '1' if a point happens to be in the buffer corridors, or '0' when outside the corridors. Similarly, Holl (2016) uses current roads rather than straight lines between nodes cities, to generate buffer areas to filter ancient roads (2000 years ago) as historical time-variant IVs.

Both these studies assume that when policy makers decide where to construct new traffic lines, they are basically connecting some nodes (important cities and sites) on the map, and some unimportant nodes (small cities or towns) are linked in the final project just because they are close to projected lines rather than because of their importance. At the same time, if policy makers decide to link an unimportant node into any potential project, they must modify the project then the length of the projected routes and budgets may increase. Considering the budget constraints, a policy maker will not connect an unimportant node if the distance from an unimportant node is larger than a set upper limit. This upper limit is defined as the efficient distance for which the impact of traffic lines will decay to zero for any node further than this distance. In the study of Holl (2016), the efficient distance of roads in Spain is about 10km, so the radius buffer area is 10km, which means that those IVs outside this 10km distance have little impact on current roads. In this study, we try six different efficient distances, i.e., 10km, 20km, 30km, 40km, 50km, 60km, to generate time-variant IVs.

3.2.4 Review of Different Productivity Estimate Methods

Productivity and Production Function

$$Y_{it} = A_{it} K_{it}^{\alpha_k} L_{it}^{\alpha_l} M_{it}^{\alpha_m}. \tag{1}$$

$$TFP_{it} = A_{it} = \frac{Y_{it}}{K_{it}^{\alpha_k} L_{it}^{\alpha_l} M_{it}^{\alpha_m}}. \tag{2}$$

Total factor productivity is estimated based on Cobb-Douglas production function. Y_{it} represents economic outputs of firm i at time t; K_{it}, L_{it} and M_{it} represent capital inputs, labor input, and material inputs. A_{it} is unobservable Hicks-neutral productivity, indicating those components in output that cannot be explained by input differences, representing the efficiency of economic activities to convert inputs into outputs.

$$ln(A_{it}) = y_{it} - \alpha_l l_{it} - \alpha_k k_{it} - \alpha_m m_{it}. \tag{3}$$

$$= \alpha_0 + \epsilon_{it}. \tag{4}$$

Converting production function into log-transformed form, the productivity can be expressed by equation (3), while ϵ_{it} represents random shocks; α_0 indicates all productivity differences across firms over time.

Productivity Estimate Methods

The fixed effect approach is the basic method to estimate productivity, and requires unobservable plant-specific productivity ω_{it} is consistent over time (ω_i), inputs are exogenous from unobservable shocks ϵ_{it} (Van Beveren 2012). However, these assumptions seem to be too strict to satisfy in practice, which may cause endogenous problems, due to input decisions being highly time-correlated (Ackerberg et al. 2007). In addition, firms' entry and exit are very common in practice, while firms' entry and exit may also be affected by firm-specific factors. If the firm-specific factors can change over time, the regression results for balanced and unbalanced panel (containing or not containing firms' entry and exit) should be different; this is supported by Olley & Pakes (1992). In most cases, thefixed effect method tends to provide biased results.

The Instrumental Variable approach (or GMM) is considered an effective tool to solve the endogeneity problem, as instruments should be exogenous of production function but correlated with endogenous variables. Some studies use first-order lagged input as instruments, based on the assumption that productivity is not self-correlated (Van Beveren 2012). At the same time, higher order of lagged inputs (2nd or 3rd order) might be better to solve the endogeneity problem, but it may lose more observations as well. By contrast, Blundell & Bond (2000) claim that lagged outputs, e.g., double or triple lagged differences, are exogenous of productivity shocks but are strongly enough correlated with level production function, so the lagged differences can be used as instruments in GMM estimation. However, this approach may remove lots of observations when calculating the difference, then the correlation between instruments and endogenous variables might be too weak and lead to downward biased results on input coefficients (Van Beveren 2012). Another candidate instrument is price index; it assumes that firms are living in perfect competitive market, so no firm can influence market price to increase their sales (Ackerberg et al. 2007). Price indices could be efficient instruments in certain industries, but price indices are used to deflate input and output in our study.

Semi-parametric approaches can also solve the simultaneity problem and self-correlation issue efficiently. Olley & Pakes (1992) assume capital input is quasi-fixed input, using investment decision at period $t-1$ to predict productivity changes at t, while the productivity variation satisfies the first order Markovian process. Their estimate contain two steps: firstly they use a non-parametric estimate (polynomial regression) for the unbiased coefficients of flexible inputs, i.e., labor input for two-factor production function, labor and intermediate inputs for three-factor production function. This step actually splits output into two parts, i.e., the part orthogonal to productivity shocks (flexible inputs), and the part correlated with productivity shocks. Although the flexible inputs should be initially correlated with productivity shocks, the polynomial regression can exhaust the information of productivity shocks, then provide unbiased coefficients of flexible inputs. For the second step, the coefficients of capital and productivity are estimated by the GMM approach, based on the assumption that productivity shocks satisfy the first-order Markovian process. However, the OP approach might be constrained by data availability because the real investment could be negative in some years, while negative investment would be identified as missed observations in log-transformed production function, which may cause selection biases.

Similar to the OP approach, Levinsohn & Petrin (2000) use intermediate input as the proxy to indicate productivity. However, intermediate input as a proxy is more likely to suffer from reverse causality issues because labor and intermediate inputs are both flexible input, so they tend to respond simultaneously to productivity shocks. If intermediate input is not strictly exogenous of productive shocks, labor input might fail to be estimated in the first step, which is also called the collinearity problem (Van Beveren 2012). To address this collinearity issue, Ackerberg et al. (2015) assume all inputs are endogenous of productivity shocks, the coefficients of flexible inputs are estimated in the second step, while the first step is applied to remove productivity shocks captured by high-order polynomial terms, which is considered more efficient than the LP approach (Van Beveren 2012).

In general, the semi-parametric approach can largely solve endogeneity problems; price changes are eliminated by price deflators; firms' heterogeneity is partly solved by industry fixed effects and firm fixed effects. We use several semi-parametric approaches to estimate productivity (OP, LP, ACF), and also use OLS productivity as a robustness check.

3.3 Methodology

3.3.1 Data Source

This chapter uses three datasets, i.e., a Chinese manufacture firm dataset, a regional dataset which provides region-level economic indicators, and a Chinese geographical highway dataset. The introduction of the manufacture firm dataset and the regional dataset are specified below.

The firm-level dataset is called the Annual Survey of Chinese Industrial Firms, carried out by China's National Bureau of Statistics (NBS). This dataset contains all state-owned enterprises (SOEs) and firms of other types of ownership with annual sales of revenue more than five million RMB (Chinese Yuan), providing their address, financial statements, and employees over the period of 1998 to 2007. These firms are classified into 30 two-digit, 301 three-digit, and 3,010 four-digit manufacturing industries. These industries are defined according to the 2002 classification system, the observations of 1998-2002 are classified by the 1994 classification system, which are updated to the 2002 version in this study.

In order to identify firms' geographical locations, firms' addresses must be converted into longitude and latitude and imported into GIS. We use Stata module 'CHINAGCODE' to identify firms' longitude and latitude on the basis of Application Program Interface (API) provided by Baidu Map Open Platform. However, some regions' names are changed over time, while the Stata module can only identify current address names. Therefore, firms' old addresses are replaced by their most recent address before using Stata module 'CHINAGCODE'. Finally, more than 99% firms' locations are identified in GIS workspace (see Table A1).

Table 3.3-1: Firms Identified by GIS.

Year	Total Obs	Successfully Identified	Fail to be Identified	Identified Rate
1998	165118	164513	605	99.63%
1999	162033	161039	994	99.39%
2000	162885	160269	2616	98.39%
2001	171256	169144	2212	98.77%
2002	181557	176484	5073	97.21%
2003	196222	193781	2441	98.76%
2004	279092	276671	2421	99.13%
2005	271835	270116	1719	99.37%
2006	301961	300067	1894	99.37%
2007	336768	333911	2857	99.15%

Note: This table illustrates how many firms' addresses are converted into longitude and latitude through the Baidu Map Open Platform.

The region-level dataset is used to control population effect and the impacts of other transport modes[3], We collect region-level population, employment, and transportation infrastructure

[3]Based on the information provided by China's statistical bureau, China's transport infrastructure includes road system, railway system, inland waterway, and aviation routes, while the road system carries 75% freight volume.

data from 'China Statistical Yearbook' and 'China City Statistical Yearbook', and merge them into the NBS dataset according to firms' locations. China's administrative regions have three types of administrative rankings, i.e., province level, prefecture level, and county level; 'China Statistical Yearbook' provides provincial-level aggregate statistic indices, such as GDP and road density; while 'China City Statistical Yearbook' provides prefecture-level indices,such as population and employment density.

3.3.2 First Stage: TFP Estimate

In order to investigate the impacts of infrastructure construction on firm-level productivity, the first stage is to estimate firm-level productivity. Productivity is estimated based on three-factor and two-factor production function. The three-factor approach describes the impacts of three production factors, i.e., labor, deflected capital and intermediate value, on the total output. In many cases, industrial output is indicated by sales revenue. This approach may contain some biases because some products might be stored as inventory, while some sold out value may also come from the inventory produced in the previous period. An alternative choice is to add sales revenue and inventory variation up to get real output value, but inventory is hard to price for the given dataset because inventory is the sum of all products produced in different periods, some firms use average sold out price while some others may use production cost, so this study uses total sales revenue to indicate output even if there are some potential biases.

Three-factor production function and productivity are estimated by the following equations (5) and (6). l_{it} indicates logged annual average employees, k_{it} indicates logged tangible fixed asset, while m_{it} represents intermediate input. y_{it} is the log-transformed sales revenue at constant prices. All these inputs and outputs are deflated by the two-digit annual industry deflators to eliminate price variation; these deflators are constructed by Brandt et al. (2012), based on the data published by NBS dataset and China Statistical Yearbook.

$$y_{it} = \alpha_0 + \alpha_l l_{it} + \alpha_k k_{it} + \alpha_m m_{it} + \epsilon_{it}. \tag{5}$$

$$ln(A_{it}) = y_{it} - \alpha_l l_{it} - \alpha_k k_{it} - \alpha_m m_{it}. \tag{6}$$

The two-factor approach uses value added to indicate output, while inputs only include labor and capital. However, value added is not available in the accounting statement and needs to be calculated separately. According to the Statistics of Industrial Value Added (Trial Scheme, published by NBS), value added mainly includes four parts, i.e., (1) profit, (2) tax expenditure such as added-value tax, sales tax, (3) labor expenditure such as wage and welfare expenses, (4) capital replacement and purchase expenditure such as depreciation and interest payment, while all this information is available in our NBS dataset. The two-factor production function is described by equations (7) and (8). y_{it} is the log-transformed value added at constant prices for firm 'i' at year 't', l_{it} is logged annual average employees, while k_{it} indicates logged tangible fixed asset. Similar to three-factor production function, all these inputs and output are deflated by the two-digit annual industry deflators.

$$y_{it} = \alpha_0 + \alpha_l l_{it} + \alpha_k k_{it} + \epsilon_{it}. \tag{7}$$

$$ln(A_{it}) = y_{it} - \alpha_l l_{it} - \alpha_k k_{it}. \tag{8}$$

This study uses several semi-parametric approaches, i.e., OP, LP, and ACF (Olley & Pakes 1992, Levinsohn & Petrin 2000, Ackerberg et al. 2015). For two-factor production function, ACF results are largely consistent with OP results but different from LP and Wooldridge GMM results. The consistent results of the Wooldridge GMM and LP approach and their small return to scales suggest downward biases of input coefficients. Compared with the LP approach, the ACF approach can more successfully solve collinearity issue than the LP approach; compared with the OP approach, the ACF approach does not suffer from selection issue; so the ACF approach is the best choice for the given dataset.

At the same time, productivity estimated by three-factor production function arrives at the same conclusion as two-factor production function, but there are more industries that fail to report statistical significant coefficients. One possible explanation is that intermediate input is closely related to productivity shocks because it is more adjustable than capital input and labor input, thus this three-factor approach may overestimated the coefficients of intermediate input and induce low productivity estimates. Consequently, two-factor productivity is the better choice than three-factor productivity for the given dataset (see Appendix B.3).

3.3.3 Second Stage: Road Construction Effects

Equations (9a) regress firm-level log productivity a_{it} on explanatory and control variables, $Dist_{it}$ represents traffic accessibility. However, $Dist_{it}$ cannot capture road network complexity, i.e., a firm located in a region with low road density will have smaller traffic accessibility than a firm located in a region with intensive traffic lines. So it is necessary to use both province-level traffic route density $Density_{it}$ and firm-level road distances $Dist_{it}$ to capture traffic accessibility; different from $Dist_{it}$ the coefficients of $Density_{it}$ are expected to be positive. These key explanatory variables are controlled by a control vector X_{it}, e.g., population aggregation effects, firm-specific effects, and location effects (see Appendix B). At the same time, four types of fixed effects are also controlled, i.e., firm fixed effects u_i, 4-digit industry fixed effects u_j, region-level fixed effects u_r, and year fixed effects u_t. They can capture time-invariant fixed effects in each county r, each 4-digit industries j, and each firm i respectively; while year fixed effects ϵ_t is controlled separately to capture macroeconomic shocks on all industry.

$$a_{it} = \alpha_0 + \alpha_1 Dist_{it} + \alpha_2 Density_{it} + \mu X_{it} + u_t + u_r + u_j + u_i + \epsilon_{it}. \tag{9a}$$

Similarly, equation (9b) is designed to investigate the channels of how infrastructure development affects firm size distribution and competition structure. The same as baseline regressions, the variable $Dist_{it}$ is firms' distances to expressways, smaller $Dist_{it}$ implies higher traffic accessibility; traffic route density $Density_{it}$ is an alternative measurement of traffic accessibility to capture road network complexity; vector X_{it} is a control vector. Dependent variables $Channel_{it}$ include inventory and outsourcing level. This study uses the ratio of intermediate input to total output as the measurement of outsourcing, the same as Ding, Sun & Jiang (2016).

$$Channel_{it} = \alpha_0 + \alpha_1 Dist_{it} + \alpha_2 Density_{it} + \mu X_{it} + u_t + u_r + u_j + u_i + \epsilon_{it}. \tag{9b}$$

3.3.4 Proxies on Transportation Cost

Previous economic geography literature has proposed different proxies to capture transportation cost and infrastructure factors. Limao & Venables (2001) use country-level road, railway density, and telephones per person to indicate infrastructure development level; Poncet (2003) uses liberal distances between provincial capitals to capture domestic transportation costs; Baum-Snow et al. (2012) divide highways and railways into radial and ring lines, then count the quantity of radial and ring lines as the proxy of traffic accessibility. By contrast, there are also some studies that focus on firm-level trade activities, e.g., Holl (2016) uses firms' nearest distance to highways to indicate transportation cost; Liu et al. (2017) calculate firms' distance to highways and highway density around 20km or 30km radius of each firms as the proxies of traffic accessibility; Bougheas et al. (2000), Tang et al. (2019) use firms' location and regional data to capture transportation cost.

Table 3.3-2: Summary Statistics of Key Explanatory Variables.

	Obs	Full Sample Average (Std. Dev.)		1998		2003		2007
Dist	1,652,658	8.708 (1.514)	98,530	9.504 (1.556)	127,764	8.843 (1.526)	290,298	8.377 (1.420)
Road-Density	310	-1.47 (1.08)	31	-1.80 (1.03)	31	-1.49 (1.05)	31	-0.97 (1.11)
Rail-Density	302	-4.44 (0.87)	30	-4.51 (0.86)	30	-4.39 (0.85)	31	-4.41 (0.99)
River-Density	270	-4.31 (1.58)	26	-4.45 (1.69)	26	-4.21 (1.55)	27	-4.29 (1.58)
Road-Density (firm-merged)	1506093	-0.75 (0.65)	88402	-1.26 (0.55)	119029	-1.01 (0.52)	261415	-0.25 (0.56)
Rail-Density (firm-merged)	1506093	-4.20 (0.56)	88402	-4.33 (0.59)	119029	-4.22 (0.54)	261415	-4.10 (0.50)
River-Density (firm-merged)	1426985	-3.40 (1.49)	83639	-3.55 (1.74)	109121	-3.74 (1.23)	248097	-3.32 (1.46)
Establishment	1910160	1.20 (3.72)	138749	1.12 (1.40)	144117	1.13 (1.99)	312995	1.08 (1.66)

Note: Average and standard deviation of the full sample and the subsamples in different years (1998, 2003, 2007) are compared in this table. The unit of $Dist$ is log meters, the units of $Road-Density$, $Rail-Density$, $River-Density$ are the log length (km) of traffic route per unit of land area ($1km^2$).

In the third and fourth chapters, road density and firms' nearest distance to highways are applied as the key measurements of transportation cost. Table 3.3-2 illustrates the summary statistics of these variables, it can be found that firms' average minimum distances to highways gradually decreased during 1998 to 2007. However, expressways are also called limited-access highways because vehicles must use entrances to enter expressways, hence firms' line distance to expressways may fail to correctly reflect traffic accessibility. These biases tend to be negligible when firms' line distances are much larger than the distances between each pair of expressway entrances. In addition, some firms may have more than one plant, and firms' distance to highways may not correctly reflect their traffic accessibility. This information can be found at the bottom of Table 3.3-2, suggesting that there are some firms that have more than one establishment, even if the average number of establishments declines from 1.12 in 1998 to 1.08 in 2007. The dataset used in the study does not contain those firms with more than one establishment for eliminating multi-site bias. Road density is used to control the impacts of other ranked roads, which is a province-level indicator, measured by the length (km) of the ranked roads per unit of land area ($1km^2$). Ranked roads do not only include Expressway, First to Fourth Class Road. Compared with expressways and first class roads, other classes of ranked roads also have important impacts on traffic accessibility, because of their higher density. Table 3.3-2 shows that the province-level road density gradually increases over time. When this variable is merged with existing firms, the firms' road density is higher and increases faster than province-level road density, suggesting new entry firms and relocation firms tend to locate in those provinces with well-developed transportation infrastructure.

Besides the impacts of road infrastructure, Table 3.3-2 also provides the summary statistics of railway and waterway density. Similar as road density, both the province-level railway density and firm-level railway density tend to increase over time, while the latter increases faster than province-level railway density, suggesting new entry firms and relocation firms tend to locate in those provinces with well-developed transportation infrastructure. By contrast, the density of waterway is affected by rainfall and available rivers, it does not show a monotonic increase over the sample period. According to the data provided by Table 3.3-2, we can largely conclude that Chinese transport infrastructure increased very fast during 1998 to 2007. The increase of firm-level road density is faster than province-level railway density, compared with railway and waterway infrastructure, suggesting that road expansion has more significant impacts on firm entry and relocation.

However, the construction of high-quality roads may not always reduce transportation costs, because these roads normally charge toll fees, while other roads with lower quality are usually free to use. A possible explanation is that firms choose to use new constructed highways because the potential benefits are larger than payments to use highways. On the one hand, toll fee charging can increase firms transport costs; on the other hand, new-constructed roads tend to have higher quality, which can save firms' transportation time, reduce fueling costs, increase firms' turnover rate and induce them to sell products more smoothly. Extant empirical studies normally agree that domestic integration and trade are largely promoted by infrastructure development, implying that most firms tend to use highways even if they need to pay toll fees, as the benefits of using highways are larger than toll fee expenditures.

As an alternative proxy of traffic accessibility, entrance distance is used to replace $Dist$ to make robustness checks. The entrance distance is constructed to address the biases of limited-access highways. In the ACASIAN dataset, the information of expressways' entrances are only available for 2007. Entrance layers in other years (1998-2006) are constructed by deleting the entrances in 2007 that are not connected to the expressways in previous years.

3.3.5 Control Variables

The main control variables include population density, ownership, trade dummy variable, firm size, and age. First, the existing literature usually uses population density, employment density, and market potential to capture the population agglomeration effect (Faber 2014, Holl 2016, Gibbons et al. 2016). This effect is indicated by the population density (people per $1km^2$ of land) in our study, its data are collected from the China City Statistical Yearbook, which provides the registration population for each prefecture-level administrative region.

Second, the ownership is indicated by a continuous measurement, the same as Ding, Guariglia & Harris (2016), i.e., the fraction of four different capital types with respect to firms' total paid-in capital, including state-owned, foreign-invested, collective-owned, and private-invested capital. In empirical studies, the indicator of private invested firms is dropped, meaning this study selects private firms as the baseline ownership, compared with other types of ownership. However, there are some errors in the NBS dataset, sometimes the sum of different types of capital is not equivalent to the total paid-in capital, so all those firms with these errors are dropped.

Third, trade firms are expected to have higher efficiency (Melitz 2003), this binary variable is denoted by '1' for those firms that export their products to the international market. Fourth, firms with larger scale are expected to have higher productivity (Dunne et al. 1988). This study uses the logarithm of firms' total assets, deflated by the deflater constructed by Brandt et al. (2012), as the indicator of size effects. Forth, firms' age is the difference between current and firms' established years. However, firms' establishment years could be incorrect, according to this approach, as some firms are hundreds of years old, while firms' ownership may change significantly during these long periods, e.g., the nationalization during the 1950s, the reform of privatization in the 1980s, which may fundamentally change firms' management and decision-making model, so this study replaces any pre-1978 firm establishment date (the start point of national economic reform) with 1978, to exclude those outliers.

Table 3.3-3: Summary Statistics of Control Variables

	Obs	Full Sample Average (Std. Dev.)	1998		2003		2007	
Population Density	1,607,865	6.316	89,495	6.316	124,770	6.170	285,834	6.358
		(.624)		(.616)		(.616)		(.639)
Age	1,652,658	8.322	98,530	10.320	127,764	8.756	290,298	7.426
		(6.979)		(7.228)		(7.329)		(6.350)
Exporter	1,652,658	.268	98,530	.256	127,764	.274	290,298	.243
		(.443)		(.436)		(.446)		(.429)
Size	1,652,612	9.572	98,517	9.572	127,764	9.619	290,298	9.645
		(1.387)		(1.471)		(1.407)		(1.36)
State Share	1,644,442	.092	97,815	.265	126,952	.101	290,180	.021
		(.276)		(.421)		(.288)		(.134)
Collective Share	1,644,442	.114	97,815	.302	126,952	.106	290,180	.041
		(.298)		(.422)		(.288)		(.188)
Foreign Share	1,644,442	.154	97,815	.125	126,952	.150	290,180	.161
		(.334)		(.290)		(.329)		(.347)

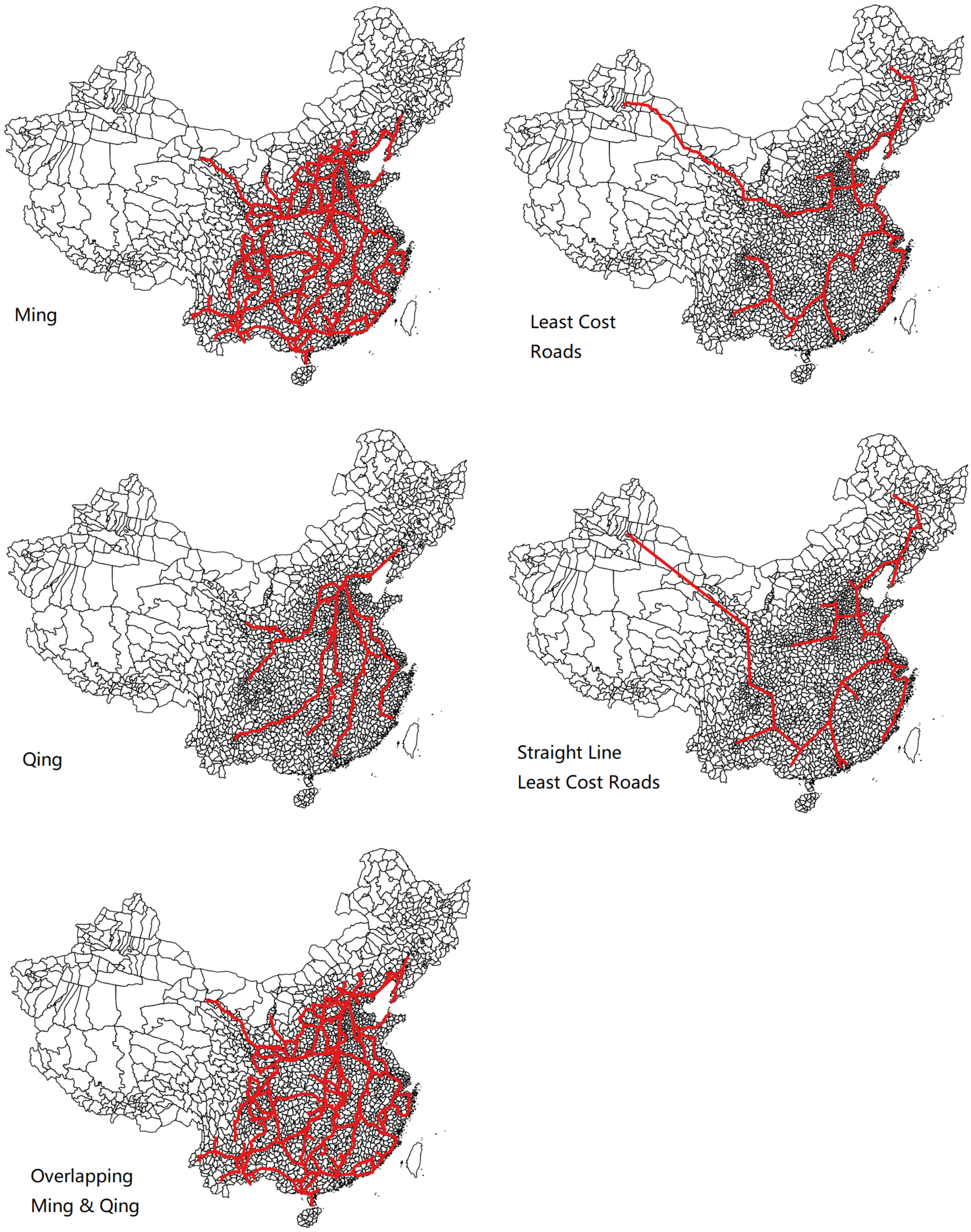
Note: Average and standard deviation of the full sample and the sub-samples in different years (1998, 2003, 2007) are compared in this table. The unit of $PopulationDensity$ is quantity of people per $1km^2$ of land (log), while the unit of Age is year, it equals the difference between current and firms' established years. $Size$ is firms' log asset value.

According to Table 3.3-3, three variables show monotonic decreases from 1999 to 2007, i.e., age, state-owned capital share, and collective-owned capital shares. The decrease of age suggests that there are lots of new entry firms in the manufacturing dataset, the significant decrease of state-owned and collective-owned capital share implies that private economies increase relatively much faster over the sample period.

3.3.6 Instrument Variable Construction

The third and fourth chapters use firms' nearest distance to highways as a key explanatory variable. In order to address the endogeneity issues, we firstly collect three historical road networks and generate two counter-factual road networks, then calculate the firms' nearest distance to these roads as instruments. The three historical roads include historical roads in the Ming and Qing dynasties, and overlapping networks of Ming and Qing dynasties. Following the approach of Faber (2014), the two counter-factual road networks are constructed by the least cost approach, one is based on a cost map, another is to use straight lines to connect node cities to generate the least cost path (Figure 3.3-1).

Figure 3.3-1: Five Road Network to Construct IV.

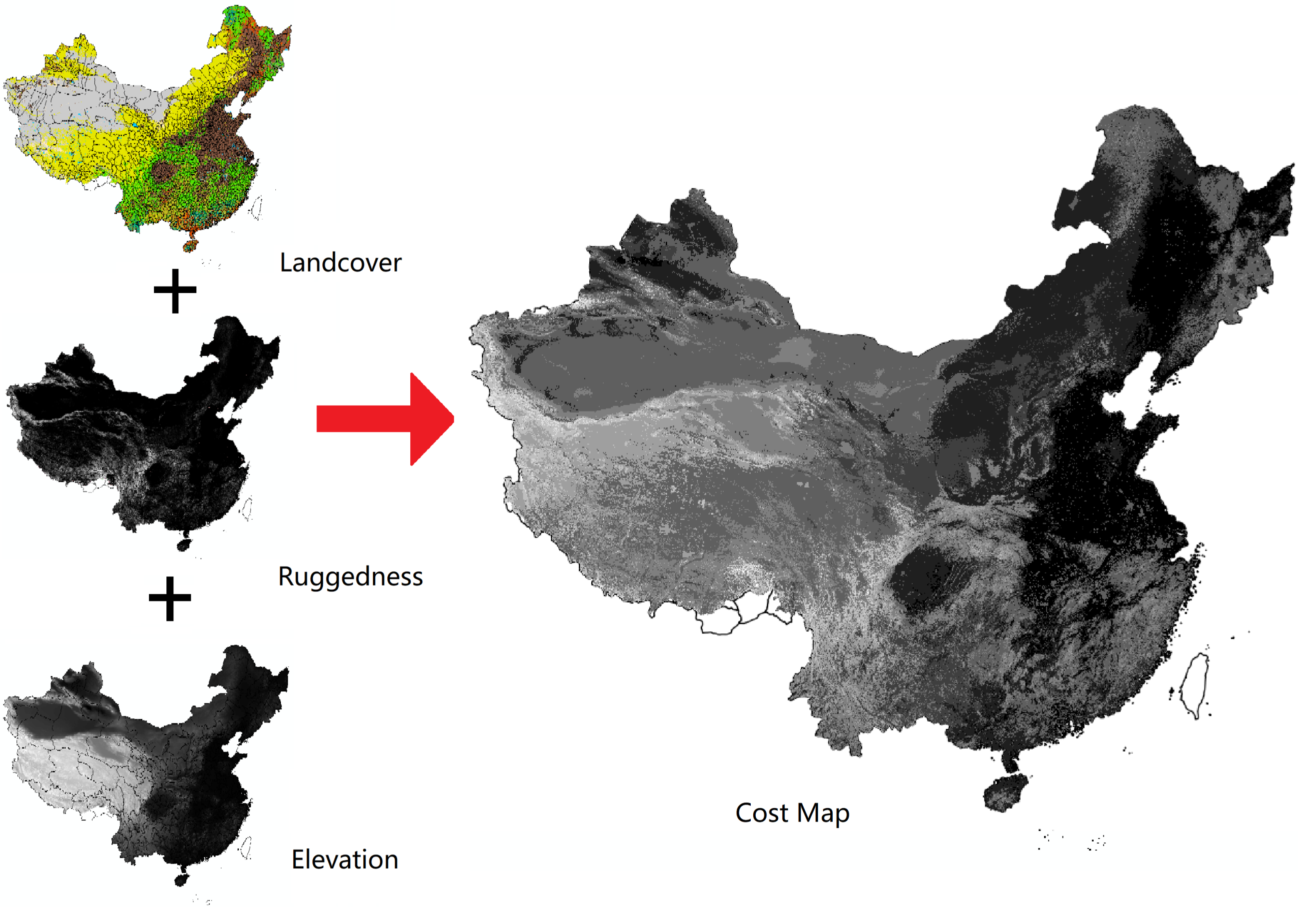


Least Cost Paths

To get the least cost roads, there are three steps: (1) define the node cities that the program needs to connect, (2) construct a cost map that defines the infrastructure construction cost in China, (3) construct the least cost roads on the basis of node cities and cost map.

In the first step, we define the cities according to China's highway construction plan. Due to our firm-level dataset (1998 to 2007) being able to exactly cover the construction period of '5-7' plan, the node cities in '5-7' plan are the initial cities that policy makers want to connect. The 1998 version of '5-7' plan selects 114 cities to connect, but there are only 60 cities' names provided in the published document.[4] These cities are selected according to their economic importance, which includes a series of indicators such as urbanization, population density, industrial output density, and traffic volumes.

Figure 3.3-2: The Construction of Cost Map.



On the basis of these node cities, the next step is to construct a cost map, it can reflect the highway construct cost along different types of surface, such as plain or mountain areas, urban or village areas. However, the construction cost across different types of surface cannot be reflected by just one variable, so previous studies tend to use different maps such as land-cover map or land ruggedness map to capture the construction cost. Following the study of

[4]The node cities include seven developed cities Beijing, Shanghai. Tianjin, Wuhan, Guangzhou, Shenyang, Dalian; and 57 median developed cities Tangshan, Taiyuan, Anshan, Changchun, Tsitsihar, Nanjing, Changzhou, Lianyungang, Yangzhou, Wenzhou, Jinan, Zaozhuang, Zhengzhou, Xinxiang, Fushun, Jinzhou, Liaoyang, Jilin, Daqing, Wuxi, Suzhou, Zhenjiang, Hangzhou, Fuzhou, Nanchang, Qingdao, Kaifeng, Shijiazhuang, Handan, Benxi, Haerbin, Xuzhou, Nantong, Yancheng, Ningbo, Huzhou, Hefei, Xiamen, Zibo, Luoyang, Changsha, Hengyang, Shenzhen, Liuzhou, Congqing, Kunming, Xian, Zhuhai, Guiyang, Urumchi, Nanning, Chengdu, Lanzhou.

Faber, the cost map is merged by three parts, i.e., land ruggedness, elevation, and land cover maps.

To construct the land ruggedness map, Faber uses China's elevation data from U.S. Digital Chart of the World (DCW) to compute the average slope for each 2x2 km^2 grid. Similarly, Nunn & Puga (2012) also use the average slope as the measurement of land ruggedness, but for each 30 arc-seconds grids (approximately 0.82x0.82 km^2). In addition, Nunn and Puga also construct an alternative ruggedness measurement (terrain ruggedness index), which depends on the elevation difference between each grid and its eight surrounding grids (north, northeast, east, southeast, south, southwest, west, and northwest), given by the square root of the sum of the squared differences in elevation, which is priority selection if this study.

The elevation map is sourced from the USGS National Elevation Dataset (NED), which provides global Elevation Raster maps with different resolutions, e.g., 1 to 1/3 arc-second. Due to these resolutions being much smaller than the land ruggedness map, this study selects 1 arc-second resolution.

Figure 3.3-3: Weight of different types of land covers.

Landcover	**Cost Index**
Water	9
Evergreen Needle leaf Forest	8
Evergreen Broadleaf Forest	8
Deciduous Needle leaf Forest	8
Deciduous Broadleaf Forest	8
Mixed Forests	8
Closed Shrublands	6
Open Shrublands	6
Woody Savannas	4
Savannas	3
Grasslands	3
Permanent Wetland	9
Croplands	2
Urban and Built-Up	1
Cropland/Natural Vegetation Mosaic	2
Snow and Ice	9
Barren or Sparsely Vegetated	2

The land cover map is collected from Broxton et al. (2014), the 0.5 km Global Land Cover Climatology. This dataset contains 17 different types of land covers, each type of land cover is given a 'cost index' (Figure 3.3-3) to indicate its influence on construction cost.[5] Similarly, raster values in land ruggedness map and elevation are also equally clustered into cost groups from 0 to 9, according to their raster values from small to large. At the end of this stage, the land ruggedness map (34% share), elevation map (33% share), and land cover map (33% share) are merged as the construction cost map (see Figure 3.3-2). The last step is to generate a continuous road network by the least cost approach, on the basis of node cities and cost map.

Corridor Approach to Generate Time-varying IVs

According to the previous discussion in the literature review, IVs based on historical roads and least cost approach are time-invariant, while the effects of time-invariant IVs could be absorbed by fixed effects, then induce the week IV issue. In order to convert time-invariant IVs to time-varying IVs, Hornung (2015) and Holl (2016) developed a corridor approach, which generates different types of buffer areas to filter ancient roads. These buffer areas are generated on the basis of current roads, the intuition is that when policy makers decide where to construct new traffic lines, they are basically connecting some nodes (important cities and sites) on the map, and some unimportant nodes (small cities or towns) are linked in the final project just because they are close to the projected lines rather than because of their importance. At the same time, if policy makers decide to link unimportant cities or towns into any potential project, they must modify the project, then the length of the projected routes and budgets may increase. Considering the budget constraints, a policy maker will not connect an unimportant node if the distance from an unimportant node is larger than a set upper limit. This upper limit is defined as the efficient distance for which the impact of traffic lines will decay to zero for any node further than this distance. In the study of Holl (2016), the efficient distance of roads in Spain is about 10km, so the radius buffer area is 10km, which means that those IVs outside this 10km distance have little impact on current roads. In this study, we try six different efficient distances, i.e., 10km, 20km, 30km, 40km, 50km, 60km, to generate time-variant IVs. The summary statistics of time-varying IVs can be found in Table 3.3-4.

[5]This dataset contains the global land use information from 2001 to 2010, available link: USGS Land Cover Institute.

Table 3.3-4: Summary Statistics of Time-varying IV.

Buffer Distances	Full Sample Average	Std. Dev.	1998 Average	Std. Dev.	2007 Average	Std. Dev.
			Distances to Least Cost Counterfactual Roads			
lc 10km	64504	93616	72382	94881	52480	70687
lc 20km	62560	92487	67829	90758	51689	70539
lc 30km	61394	91887	65972	89900	51413	70462
lc 40km	60934	91626	63773	89266	51315	70448
lc 50km	60605	91407	62798	89008	51277	70442
lc 60km	60373	91239	62199	88864	51263	70434
			Distances to Straight-line Least Cost Counterfactual Roads			
sl 10km	69580	98870	74808	100413	56808	73382
sl 20km	65976	94691	68899	90489	55310	72611
sl 30km	64380	92791	67234	90078	54510	71902
sl 40km	63722	92526	66066	89819	54239	71843
sl 50km	63345	92302	65353	89620	54131	71758
sl 60km	62998	92112	64803	89463	54097	71714
			Distances to Ming Historical Roads			
m10km	63470	153318	81316	190465	45634	91484
m20km	61290	151119	76986	185992	44718	90893
m30km	59970	150252	74432	184832	44397	90843
m40km	59245	146737	71238	183701	44246	90795
m50km	57692	148138	69257	182836	44152	90736
m60km	57010	147415	68037	182051	44028	90625
			Distances to Qing Historical Roads			
q10km	115189	193597	130127	232573	95602	138749
q20km	110486	180295	122886	206182	94782	138632
q30km	109062	180096	119501	205979	94227	138422
q40km	95236	173389	117898	205944	93826	138234
q50km	107560	179819	116750	205887	93742	138224
q60km	107090	179684	115541	205782	93584	138026
			Distances to Ming and Qing Roads			
mq10km	60811	153001	79275	190339	42316	90266
mq20km	58587	150791	74864	185858	41337	89648
mq30km	57197	149908	72084	184690	40970	89585
mq40km	56450	146187	68920	183543	40782	89523
mq50km	54880	147764	66891	182655	40654	89457
mq60km	54189	147031	65569	181859	40530	89340

Note: This table illustrates average and standard deviation of the five instruments, including the full sample and the sub-samples in 1998 and 2007.

3.4 Empirical Results

3.4.1 Baseline Regressions

Table 3.4-5: Relationship between Line Distance and Productivity (IV Regressions).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.025***	-0.034***	-0.021***	-0.033***	-0.027***	-0.027***	-0.022***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Age	-0.006***	-0.005***	-0.005***	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	-0.017**	-0.013***	-0.004	0.033***	0.035***	0.035***	0.035***
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Size	0.158***	0.155***	0.153***	0.165***	0.165***	0.165***	0.163***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.575***	-0.506***	-0.415***	-0.069***	-0.056***	-0.056***	-0.060***
	(0.012)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Collective-share	0.050***	0.032***	0.016***	-0.011**	-0.006	-0.006	-0.007
	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Foreign-share	-0.050***	-0.031***	-0.022***	0.014*	0.017*	0.016*	0.014*
	(0.010)	(0.006)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
Rail-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Road-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
River-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	1369177	1362969	1340814	1269369	1241601	1241354	1226767
Group	9512	45537	111944	314577	327471	327446	329817
Under-identification test	764.360	1014.196	1483.183	4829.538	4656.493	4638.951	4448.347
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5.5e+04	3.8e+04	2.3e+04	2.7e+04	2.6e+04	2.6e+04	2.6e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.133	1.263	1.837	1.487	1.031	1.098	0.408
P value	(0.715)	(0.261)	(0.175)	(0.222)	(0.310)	(0.295)	(0.523)
First-stage Results:							
Ming30km	0.216***						
	(0.008)						
Qing30km	0.057***						
	(0.008)						
SL30km		0.086***					
		(0.007)					
LC30km		0.220***	0.222***	0.278***	0.280***	0.279***	0.287***
		(0.008)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Qing50km			0.127***	0.155***	0.159***	0.159***	0.174***
			(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
F	681.069	580.091	596.047	2107.042	1966.565	1965.791	1871.750
R2	0.155	0.123	0.118	0.161	0.154	0.154	0.153
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. The full regression results are provided in Table B2. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 3.4-5 investigates the impacts of highway construction on firm-level productivity, under the control of different types of fixed effects. The fixed effects in columns (1) to (7) are controlled by two components, i.e., year fixed effects and time-invariant cross fixed effects. For the time-invariant cross fixed effects, column (1) controls time-invariant province-level and 4-digit industry fixed effects, column (2) controls time-invariant prefecture-level and 4-digit industry fixed effects, while column (3) controls time-invariant county-level and 4-digit industry fixed effects. Regional and industrial fixed effects can control persistent productivity gaps across regions and industries, and the impacts of geographical factors such as natural resource endowment, overall terrain ruggedness level, and local climate; while year fixed effects can capture nationwide productivity shocks in each year. Columns (4) to (7) further include firm-level fixed effects; column (4) only controls firm-level time-invariant fixed effects; column (5) controls time-invariant province-level, firm-level, and 4-digit industry fixed effects; column (6) controls time-invariant prefecture-level, firm-level, and 4-digit industry fixed effects; while column (7) controls time-invariant county-level, firm-level, and 4-digit industry fixed effects. Firm-level fixed effects can capture firm-level persistent productivity gaps; the cross fixed effects between firm and region can control the relocation effect, which means that if a firm changes its address over the sample period, the cross fixed effects between firm and region can capture firm-level fixed effects in different regions separately; the cross fixed effects between firm and industry can control the case that a firm changes its 4-digit industry over sample period. Due to columns (5) to (7) control firm-level, region-level, and industry-level fixed effects simultaneously, the firm-level persistent productivity gaps, relocation effect, and industry transition effect can be captured simultaneously.

However, our key explanatory variable, firms' distance to highways, could be affected by endogeneity issues because the spatial distribution of new roads could be affected by economic factors such as local income level. The intuition is that when policy makers decide the location of new constructed highways, they have the motivation to locate new infrastructure in those regions with higher potential economic growth, so their decisions could be affected by the spatial distribution of economic activities. At the same time, the spatial distribution of cities or economic activities is self-correlated over time, if a city was a regional capital fifty years ago, it is still very likely to be an important city at present. To solve the potential endogeneity issues, we construct IVs based on the counter-factual approach (applied by Faber (2014)) and historical roads in the Qing and Ming dynasties. Our results only illustrate those regressions can pass the Over-identification test and the Under-identification test.

For instance, Table 3.4-5 presents IV regressions with five different instruments (Ming30km, Qing30km, SL30km, LC30km and Qing50km).

The IV regression results in Table 3.4-5 across columns (1) to (7) keep consistent and significant, suggesting that highway construction has a casual impact on firms' productivity increase. The Under-identification test and Weak identification test are very significant, suggesting the instruments are correlated with the variable $Dist$; while the Over-identification test is insignificant under the 5% level, implying that this instrument combination are exogenous from simultaneous shocks. At the same time, due to our geographical spatial dataset only containing high-class highways, we also use $RoadDensity$, $RailDensity$, and $RiverDensity$ to test whether the density of low-class highways, railroads, and waterways can change the coefficient of $Dist$. The coefficient of $Dist$ is about -0.022 to 0.033, implying that a 10% decrease of $Dist$ is related to a 0.2% to 0.3% increase of firm-level productivity, this impact is a little stronger than in Spain Holl (2016), consistent with previous region-level studies such as Garcia-Mila et al. (1996) and Baum-Snow et al. (2012). If we compare the IV results in Table 3.4-5 with OLS results in Table B1, we can find the OLS results are biased and inconsistent.

Besides the transportation cost proxies, the coefficients of several firm-specific control variables also reveal noteworthy results. First, older and larger firms tend to have higher productivity. A 10% increase of firm size is related to 1.6% increase of productivity, while one year older is related to 0.02% increase of productivity. Second, state-owned firms are less productive than other types of firms, a 10% increase of state-owned capital shares is related to 0.6% decrease of productivity. At the same time, a 10% increase of foreign capital is related to about 0.15% increase of productivity. The productivity effects of foreign, private, and collective capital are relatively closer and more similar, but state-owned capital shares are systemically less productive than other capital shares. Third, the coefficient of $Exporter$ is about 0.035, suggesting that if a firm is an exporter, its productivity is 0.45% higher than other firms. However, this effect dose not consist across different columns, the relationship between highway development and trade activities are further specified in chapter five.

By contrast, the coefficients of region-level control variables seem to provide biased and inconsistent coefficients, when we control firm-level fixed effects. For example, the coefficients of population density tend to be positive when there is no firm fixed effect, then change to negative when we control both firm and region-level fixed effects. At the same time, the coefficient of $RoadDensity$, $RailDensity$, and $RiverDensity$ can be affected by the potential multi-collinearity issues. This issue could be induced by the inter-correlation across different types of transport routes, e.g., route density could be higher in coastal and higher income regions, route density could also be higher around some big cities or regional capitals. Due to the coefficients of these region-level variables are more likely to be biased, we do not provide further economic interpretation on these coefficients.

3.4.2 Robustness Checks

The baseline IV regressions can address the endogeneity issues of highway construction, but cannot solve potential firm-level selection and relocation biases. During our sample period, on the one hand, some firms may change their locations, on the other hand, some new firms may also enter the market; so there were only a small proportion of firms that never changed their addresses. According to Table 3.4-6, there are 98,530 firms recorded in 1998, while only 11,195 (11%) left in 2003 and 8,236 (8%) left in 2007 that have never changed their locations or exited the market. From 1998 to 2007, the average $Dist$ of address-unchanged firms decreases faster than the full sample and the subsample of new entry and relocation firms; suggesting highway development can reduce $Dist$ of address-unchanged firms on one side, one other side, highway development can also promote firms' entry and relocation, then reduce the overall $Dist$.

Table 3.4-6: Comparison with Address-unchanged Firms.

	Obs	Full Sample Average (Std. Dev.)	1998		2003		2007	
Full Sample:								
Dist	1,652,658	8.708	98,530	9.504	127,764	8.843	290,298	8.377
		(1.514)		(1.556)		(1.526)		(1.420)
Address-unchanged Firms since 1998:								
Dist	299,356	9.190	98,530	9.504	11,194	8.795	8,236	8.136
		(1.555)		(1.556)		(1.475)		(1.354)
Entry and Relocation Firms after 1998:								
Dist	1,353,302	8.602	-	-	116,570	8.848	282,062	8.384
		(1.484)		-		(1.531)		(1.422)

Note: Average and standard deviation of the full sample and the sub-samples in different years (1998, 2003, 2007) are compared in this table.

Table 3.4-6 shows that the decrease of $Dist$ can be induced by two mechanisms, i.e., one is highway expansion, another is firms' entry and relocation. Due to the use of highway could be more costly (due to toll fees), it may select more productive firms located closer to highways. If productive firms benefit more from locating closer to the highway, it will induce a positive relation between firm productivity and access to highways, which can induce potential endogeneity issues that cannot be captured by our counter-factual IVs and historical IVs. As the robustness checks, Table 3.4-7 illustrates sub-sample IV regressions that only consider those firms that have never changed their addresses since 1998, while the control variables and fixed effects are exactly the same as Table 3.4-5. The coefficients of $Dist$ are about -0.03, implying that a 10% decrease of $Dist$ is related to a 0.3% increase of firm-level productivity, consistent with previous IV regressions in Table 3.4-5, suggesting that even if the firm relocation and entry effects have been controlled, the positive influence of highway construction on productivity still holds.

Table 3.4-7: Robustness Checks on Relocation Effects.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	0.003	-0.019*	-0.018*	-0.031***	-0.029**	-0.029**	-0.031***
	(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Observation	234850	230950	222760	220632	216571	216567	212456
Group	6113	20348	36479	58388	60183	60182	60276
Under-identification test	572.515	802.285	740.108	1890.427	1904.805	1904.788	1877.849
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.9e+04	1.4e+04	1.2e+04	1.4e+04	1.4e+04	1.4e+04	1.4e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.024	0.591	5.779	0.208	0.071	0.071	0.203
P value	(0.878)	(0.442)	(0.016)	(0.649)	(0.790)	(0.791)	(0.653)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table illustrates sub-sample IV regressions that only consider those firms that have never changed their addresses since 1998, the control variables and fixed effects are the same as Table 3.4-5. The full regression results are provided in Table B3. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 3.4-8: Increase of Cargo Volume by Different Transport Modes.

Unit/10,000 tons	1998	2001	2004	2007
National Volumes	1,267,427	1,401,786	1,706,412	2,275,822
Highway Volumes	976,004	1,056,312	1,244,990	1,639,432
	(77.0%)	(75.4%)	(73.0%)	(72.0%)
Railway Volumes	164,309	193,189	249,017	314,237
	(13.0%)	(13.8%)	(14.6%)	(13.8%)
Waterway Volumes	109,555	132,675	187,394	281,199
	(8.6%)	(9.5%)	(11.0%)	(12.4%)
Total Share	(98.6%)	(98.7%)	(98.6%)	(98.2%)

The second part of robustness checks are about different types of transport modes. In our baseline IV regressions, road infrastructure is controlled by firm-level distance to highways $Dist$ and province-level road density $RoadDensity$, the first one can capture the impacts of high-class highways, the latter can control the impacts of both high-class and low-class roads. According to Table 4.4-15, there are more than 70% cargo volumes carried by road infrastructure from 1998 to 2007, if we include railway and waterway infrastructure, this share can increase to about 99%, implying that if we control the impacts of road, railway, and waterway infrastructure, they can largely represent Chinese carrying capacity of cargo.

However, if we control $RoadDensity$, $RailDensity$, and $RiverDensity$ simultaneously, it will induce two problems: (1) some northern provinces without waterways will be removed from regressions, it may induce selection biases; (2) variables $RoadDensity$, $RailDensity$, and $RiverDensity$ may suffer from multi-collinearity issue, and fail to provide meaningful regression coefficients. In order to solve these two problems, Table 3.4-9 tests the robustness of $Dist$ in two cases: one is the case that we only include $Dist$ and $RoadDensity$ to control the transport infrastructure, i.e., columns (1) to (4); another is that we include $Dist$, $RoadDensity$, and $RailDensity$ to control transport infrastructure, i.e., columns (5) to (8). We find the coefficient of $Dist$ is about -0.03 to -0.04, implying that a 10% decrease of $Dist$ is related to a 0.3% or 0.4% increase of firm-level productivity, a little larger than baseline regressions. In addition, the coefficients of $RoadDensity$ and $RailDensity$ are all positive, i.e., a 10% increase of $RoadDensity$ or $RailDensity$ is related to a 0.2% to 0.4% increase of firm-level productivity. These results confirm that the development of transport infrastructure can increase firms' productivity.

Table 3.4-9: Robustness on Different Types of Traffic Routes.

Dependent Variable: TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dist	-0.043***	-0.036***	-0.036***	-0.027***	-0.043***	-0.036***	-0.036***	-0.027***
	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
Road-Density	0.043***	0.040***	0.040***	0.038***	0.042***	0.038***	0.039***	0.037***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Rail-Density					0.021**	0.020**	0.020**	0.016*
					(0.007)	(0.007)	(0.007)	(0.007)
Observation	1339289	1309989	1309735	1294231	1339289	1309989	1309735	1294231
Group	329151	343180	343156	345625	329151	343180	343156	345625
Under-identification test	4075.504	3830.330	3812.747	3756.333	4079.313	3832.867	3815.228	3757.757
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.9e+04	1.8e+04	1.8e+04	1.9e+04	1.9e+04	1.8e+04	1.8e+04	1.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.697	0.822	0.807	0.383	0.749	0.856	0.841	0.407
P value	(0.404)	(0.365)	(0.369)	(0.536)	(0.387)	(0.355)	(0.359)	(0.523)
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry		4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The other control variables are the same as Table 3.4-5. The full regression results are provided in Table B4. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 3.4-10: Regressions with Lagged Dist.

Dependent Variable: TFP	Without Firm Fixed Effect (1)	(2)	(3)	With Firm Fixed Effect (4)	(5)	(6)	(7)
Main Results:							
Lag Dist	-0.027***	-0.026***	-0.026**	-0.098***	-0.083***	-0.084***	-0.077***
	(0.006)	(0.007)	(0.010)	(0.013)	(0.014)	(0.014)	(0.014)
Observation	910256	904696	886874	836897	819707	819531	812324
Group	8919	39289	89880	227986	235772	235737	237267
Under-identification test	665.455	780.839	737.319	1482.949	1207.257	1204.765	1203.870
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	2.7e+04	1.8e+04	6212.882	3274.749	2547.477	2536.444	2547.334
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.231	0.002	1.743	0.952	0.067	0.117	0.785
P value	(0.631)	(0.961)	(0.187)	(0.329)	(0.795)	(0.732)	(0.376)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B5. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 3.4-10 tests the possibility that new highways may take time to generate impacts on firms' productivity, so the explanatory variable is one-year-lagged firms' distance to highways. We find the coefficient of $LagDist$ is about -0.08 to -0.10, implying that a 10% decrease of lagged $Dist$ is related to a 0.8% or 1.0% increase of firm-level productivity, much larger than baseline regressions, implying that road expansion has long-term positive impacts on firms' productivity increase.

Table 3.4-11: Alternative Measurement: Entrance Distance.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist Inter	-0.027***	-0.032***	-0.023***	-0.036***	-0.030***	-0.030***	-0.024***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Observation	1369177	1362969	1340814	1269369	1241601	1241354	1226767
Group	9512	45537	111944	314577	327471	327446	329817
Under-identification test	737.527	1159.618	1969.677	7001.182	6565.355	6545.407	6181.729
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	8.4e+04	9.4e+04	4.9e+04	4.9e+04	4.6e+04	4.5e+04	4.5e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.592	0.221	2.171	2.198	1.685	1.765	0.835
P value	(0.442)	(0.892)	(0.141)	(0.138)	(0.194)	(0.184)	(0.361)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B6. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

In our baseline regressions, $Dist$ can only capture firms' distance to highways, On the other side, expressways are also named limited-access highways which require vehicles to enter expressways through entrances, the variable $Dist$ could reflect firms' distance to these entrances. For the robustness checks, Table 3.4-11 uses an alternative explanatory variable to substitute $Dist$, i.e., $EntranceDist$, while the other control variables are the same as baseline regressions. This variable is the firms' minimum distance to expressways' entrances. However, due to our dataset only containing the observations of expressways' entrances in 2007, the original $EntranceDist$ only has one-year observations. In order to convert $EntranceDist$ from cross-section to panel data, we compare the entrance map in 2007 with highway maps during 1998-2006, then delete the entrances in 2007 that are not connected to the expressways in each year (1998-2006). Results show that a 10% decrease of $EntranceDist$ is related to a increase in firm-level productivity by 0.25%-0.35%, also consistent with the previous baseline and IV regressions.

Table 3.4-12: Robustness Checks: Regressions without Big Cities.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.024***	-0.021***	-0.016**	-0.041***	-0.033***	-0.033***	-0.031***
	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Observation	1271649	1265474	1244441	1180652	1155614	1155368	1143357
Group	8362	44300	105266	291696	304010	303985	306414
Under-identification test	732.833	876.152	1366.370	6534.152	6381.905	6354.039	6199.557
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5.2e+04	3.6e+04	2.2e+04	3.9e+04	3.8e+04	3.8e+04	3.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.030	0.882	0.408	0.004	0.032	0.023	0.548
P value	(0.862)	(0.348)	(0.523)	(0.951)	(0.857)	(0.880)	(0.459)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B7. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 3.4-12 uses a sub-dataset without firms located in municipalities (Beijing, Tianjin, Shanghai, and Chongqing), while the other control variables are the same as baseline regressions. These regressions are designed to check city agglomeration effects, because much higher population density in these municipalities may cause productivity premiums. The results across all columns show that a 10% decrease of $Dist$ can increase firm-level productivity by 0.30%-0.40%, while the baseline results are 0.2%-0.3%, suggesting that highway expansion tends to promote a more significant increase of firm-level productivity in less-developed regions (non-municipality regions), consistent with Holl (2016)

In addition, our baseline and IV regressions use ACF TFP as the measurement of firm-level productivity, which is substituted by three other proxies (OLS, OP., and LP TFP) in Table 3.4-13. The coefficients of all in columns are significant and negative, a 10% decrease of $Dist$ is related to an increase in firm-level productivity by 0.35% (OLS TFP), 0.30% (OP TFP), 0.25% (LP TFP) respectively, suggesting firm-level productivity indicated by different proxies tends to provide consistent coefficients, the causality impacts of highway expansion on productivity can be confirmed. The Under-identification test and Weak identification test are very significant, suggesting the instruments are correlated with the variable $Dist$; while the Over-identification test is insignificant at the 5% level, implying their instrument combinations are exogenous from simultaneous shocks.

Table 3.4-13: Robustness Checks on Different Types of TFP.

Dependent Variable:	OLS TFP (1)	OP TFP (2)	LP TFP (3)	OLS TFP (4)	OP TFP (5)	LP TFP (6)
Main Results:						
Dist	-0.042***	-0.041***	-0.036***	-0.035***	-0.032***	-0.026***
	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
Observation	1244947	1269369	1269369	1203083	1226767	1226767
Group	310835	314577	314577	325357	329817	329817
Under-identification test	6632.761	6807.708	6807.708	6294.355	6464.496	6464.496
P value	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Weak identification test	3.8e+04	3.9e+04	3.9e+04	3.8e+04	3.9e+04	3.9e+04
P value	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Over-identification test	0.835	0.495	0.154	2.853	1.043	0.280
P value	(0.361)	(0.482)	(0.695)	(0.091)	(0.307)	(0.597)
Region				province	prefecture	county
Industry				4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table uses three alternative indicators of productivity (OLS, OP, and LP TFP), the full regression results are provided in Table B8. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Firm performance cannot only be reflected by productivity, but also is indicated by other variables such as value added, profit, and revenue. Table 3.4-14 regresses these three dependent variables (value added, profit, and revenue) on $Dist$, the dependent variables in columns (1) and (2) are value added, in columns (3) and (4) are profit, in columns (5) and (6) are revenue. Columns (1), (3), (5) control firm-level fixed effects and year fixed effects, while columns (2), (4), (6) control year fixed, and time-invariant county-level, firm-level, and 4-digit industry fixed effects. The results show that a 10% decrease of $Dist$ can increase firm-level value added by 0.35%-0.45%, increase firm-level profit by 0.45%-0.45%, increase firm-level revenue by 0.20%. These impacts are stronger than the impacts on productivity in baseline results are 0.2%-0.3%.

Table 3.4-14: The Impacts on Other Types of Firm Performances (Value Added, Profit, and Revenue).

Dependent Variable:	Value Added		Profit		Revenue	
	(1)	(2)	(3)	(4)	(5)	(6)
Main Results:						
Dist	-0.044***	-0.034***	-0.054***	-0.046***	-0.021***	-0.017***
	(0.006)	(0.006)	(0.010)	(0.010)	(0.005)	(0.005)
Observation	1084616	1045060	960296	923872	1269217	1226623
Group	301967	314424	275995	285378	314542	329780
Under-identification test	3740.332	3490.977	3179.254	2951.791	4360.739	4011.392
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.7e+04	1.7e+04	1.4e+04	1.4e+04	2.3e+04	2.2e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.331	0.760	0.869	1.798	0.014	0.708
P value	(0.565)	(0.384)	(0.351)	(0.180)	(0.905)	(0.400)
Fixed Effects:						
Region		county		county		county
Industry		4-digit		4-digit		4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B9, Table B10, and Table B11. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

3.4.3 Channel Studies and Heterogeneity Studies

The baseline regressions and robustness checks have verified the causal influence of traffic environment on productivity, this section investigates several possible channels that how highway expansion increases firm-level productivity. According to existing studies, productivity can be affected by the quality of capital input or human capital (Syverson 2011). If we consider international trade, productivity can also be increased by cheaper products or intermediate goods provided by foreign producers (Amiti & Konings 2007). The new new trade theories, represented by Melitz (2003), emphasize the roles of competition and market selection mechanism. The increase of trade exposure can be promoted by policy changes or improvement of transport infrastructure (Ding, Jiang & Sun 2016); during this process, productive companies tend to survive while less productive firms tend to leave the market (Melitz 2003).

For those firms with trade activities, their productivity can be increased by 'learning-by-doing effect' mechanism. It suggests that when firms try to solve problems in oversea markets with various types of business and institutional environment, they can learn how to operate more efficiently from their experience. Arrow (1971) points out that the learning process can induce knowledge accumulation, while knowledge acquisition plays an important role in productivity increase. The relationship across trade, infrastructure development, and productivity is studied in Chapter Five.

Besides the mechanisms input, market access, and trade; the baseline regressions also show that the coefficients of firm size are much larger than the coefficients of $Dist$ and other control variables, suggesting size-productivity relation is noteworthy. This relation could be involved with several mechanisms, such as inventory channel and outsourcing channel. Extant studies, such as Duranton & Puga (2004), Combes et al. (2012), Hashiguchi & Tanaka (2015), point out that infrastructure development can promote up- and down-stream industry cooperation, some firms will use cheaper intermediate inputs to substitute other inputs, so outsourcing activities will be promoted; meanwhile, the scales of these outsourcing firms will decrease because they can focus on a smaller range of production. By contrast, some firms maintaining high levels of inventory may have larger scales because they need to pay higher levels of inventory costs. The section investigates two mechanisms, i.e., inventory channel and outsourcing channel. Inventory level is a provided variable in our firm-level dataset, while outsourcing activities are indicated by the log ratio of intermediate input to total output, the same as Ding, Sun & Jiang (2016).

Table 4.4-12 regresses inventory level on the explanatory variable $Dist$, and control variables the same as baseline regressions. Regressions in columns (1) to (3) show that, when firm-level fixed effects are not controlled, a 10% decrease of $Dist$ can increase firms' inventory level by 0.40%-0.80%. Regressions in columns (4) to (7) show that, when firm-level fixed effects are controlled, a 10% decrease of $Dist$ can increase firms' inventory levels by 0.15%-0.25%. These results confirm that highway development tends to increase the firms' inventory level, a possible explanation is that highway construction tends to promote firm size increase while firm size is positively related to inventory level. This conjecture is supported by our previous regressions, i.e., highway construction can increase firms' profits and revenue, while the rapidly increased profits and revenues are expected to increase firm size.

Table 3.4-15: The Channel of Inventory.

Dependent Variable: Inventory	Without Firm Fixed Effect (1)	(2)	(3)	With Firm Fixed Effect (4)	(5)	(6)	(7)
Main Results:							
Dist	-0.076***	-0.083***	-0.041***	-0.026***	-0.019**	-0.016**	-0.014*
	(0.008)	(0.007)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
Observation	1316835	1310538	1288320	1216612	1189551	1189317	1174956
Group	9471	44921	109560	303835	316070	316049	318257
Under-identification test	581.598	1641.744	1492.790	4697.711	4537.538	4879.238	4334.301
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.5e+04	4.2e+04	2.2e+04	2.6e+04	2.5e+04	2.7e+04	2.5e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.217	2.559	1.158	1.917	2.372	3.958	2.624
P value	(0.641)	(0.110)	(0.282)	(0.166)	(0.124)	(0.047)	(0.105)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B12. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 3.4-16: The Channel of Inventory (Unchanged Address).

Dependent Variable: Inventory	Without Firm Fixed Effect (1)	(2)	(3)	With Firm Fixed Effect (4)	(5)	(6)	(7)
Main Results:							
Dist	-0.048***	-0.050***	-0.008	0.007	0.005	0.005	0.001
	(0.007)	(0.010)	(0.013)	(0.012)	(0.013)	(0.013)	(0.013)
Observation	229259	225349	217144	214723	210740	210736	206660
Group	6082	20158	35992	57289	59036	59035	59113
Under-identification test	578.990	1013.610	738.845	1058.622	1020.631	1020.624	1004.328
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	3.0e+04	1.4e+04	1.1e+04	7956.428	7949.028	7948.745	8158.926
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	2.637	2.850	0.129	0.768	0.723	0.723	0.692
P value	(0.104)	(0.091)	(0.720)	(0.381)	(0.395)	(0.395)	(0.406)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B13. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

As the robustness checks, Table 4.4-12 illustrates the regressions on the basis of a subsample, this sample only contains the firms that have never changed their addresses, the same as Table 3.4-7. Regressions in columns (1) to (3) show that, when firm-level fixed effects are not controlled, a 10% decrease of $Dist$ can increase firms' inventory level by 0.50%. Regressions in columns (4) to (7) show that, when firm-level fixed effects are controlled, the impacts of $Dist$ on firms' inventory level are insignificant. These results suggest that the inventory of address-unchanged firms is not sensitive to highway development. At the same time, the significant coefficients of $Dist$ in Table 4.4-12 reflect that new entry and relocation firms are more sensitive to highway construction. If they selection addresses that are close to highways, they tend to have higher levels of inventory.

Table 3.4-17: Inventory as Control Variable.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.025***	-0.033***	-0.021***	-0.032***	-0.026***	-0.026***	-0.021***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Inventory	-0.002	0.001	0.002	0.008***	0.009***	0.009***	0.010***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observation	1316835	1310538	1288320	1216612	1189551	1189317	1174956
Group	9471	44921	109560	303835	316070	316049	318257
Under-identification test	763.222	992.241	1492.758	4697.547	4537.349	4521.085	4334.216
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5.3e+04	3.7e+04	2.2e+04	2.6e+04	2.5e+04	2.5e+04	2.5e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.268	0.217	1.146	1.584	0.886	0.945	0.370
P value	(0.605)	(0.217)	(0.284)	(0.208)	(0.347)	(0.331)	(0.543)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B14. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 3.4-17 includes inventory as the control variable, while the dependent variable is productivity. Regressions in columns (1) to (3) show that, when firm-level fixed effects are not controlled, the coefficients of inventory are insignificant. Regressions in columns (4) to (7) show that, when firm-level fixed effects are controlled, a 10% increase of inventory is related to the increase of firm-level productivity by 0.10%. Although they are not IV results, the coefficients of inventory are consistent with our conjecture, i.e., inventory level is positively related to firm size and productivity. Further studies on firm size effects can be found in Chapter Four.

Table 4.4-13 regresses outsourcing (log ratio of intermediate input to total output) on the explanatory variable $Dist$, and the control variables are the same as baseline regressions. Regressions in columns (1) to (3) show that, when firm-level fixed effects are not controlled, a 10% decrease of $Dist$ can increase the outsourcing levels by 0.05%-0.10%. Regressions in columns (4) to (7) show that, when firm-level fixed effects are controlled, a 10% decrease of $Dist$ can increase the outsourcing levels by about 0.10%. These results confirm that highway development tends to promote firms' outsourcing activities, i.e., a higher proportion of intermediate inputs.

Table 3.4-18: The Impacts on Outsourcing.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Outsourcing	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.006***	-0.009***	-0.004	-0.007***	-0.009***	-0.009***	-0.009***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observation	1366418	1360228	1338105	1266582	1238883	1238636	1224100
Group	9506	45504	111767	314048	326898	326873	329221
Under-identification test	536.768	1550.466	1178.550	4348.172	4202.267	4187.265	3996.247
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.4e+04	2.8e+04	2.1e+04	2.3e+04	2.2e+04	2.2e+04	2.2e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	3.452	4.066	0.227	0.081	0.013	0.009	0.398
P value	(0.063)	(0.131)	(0.634)	(0.775)	(0.908)	(0.925)	(0.528)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B15. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

As the robustness checks, Table 3.4-19 illustrates the regressions on the basis of address-unchanged firms, the same as Table 3.4-7. Regressions in columns (1) to (3) show that, when firm-level fixed effects are not controlled, the impacts of $Dist$ on outsourcing are insignificant. Regressions in columns (4) to (7) show that, when firm-level fixed effects are controlled, a 10% decrease of $Dist$ can increase the firms' outsourcing levels by 0.20%.

These results suggest that the impacts of $Dist$ are more significant for address-unchanged firms.

Table 3.4-19: The Impacts on Outsourcing (Unchanged Address).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Outsourcing	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	0.001	-0.002	-0.003	-0.022***	-0.021***	-0.021***	-0.021***
	(0.002)	(0.002)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)
Observation	234199	230299	222106	219922	215867	215863	211758
Group	6108	20322	36397	58268	60051	60050	60136
Under-identification test	449.271	1027.736	626.377	1482.094	1546.317	1546.303	1522.351
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.0000)	(0.000)	(0.000)
Weak identification test	1.8e+04	9839.683	1.0e+04	9626.691	9474.999	9474.591	9499.398
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.0000)	(0.000)	(0.000)
Over-identification test	0.013	4.748	4.808	1.839	1.275	1.274	1.404
P value	(0.908)	(0.093)	(0.028)	(0.175)	(0.259)	(0.259)	(0.236)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B16. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 3.4-20 includes outsourcing as the control variable, while the dependent variable is productivity. Regressions in (1) to (3) columns show that, when firm-level fixed effects are not controlled, a 10% increase of outsourcing is related to the decrease of firm-level productivity by 0.10%. Regressions in columns (4) to (7) show that, when firm-level fixed effects are controlled, a 10% increase of outsourcing is related to the decrease of firm-level productivity by 0.50%. Although they are not IV results, the coefficients of inventory are consistent with our conjecture, i.e., outsourcing is negatively related to firm size and productivity, the smaller firms have lower productivity than larger firms.

Table 3.4-20: The Regression with Outsourcing Share as Control Variable.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.026***	-0.036***	-0.021***	-0.032***	-0.027***	-0.027***	-0.022***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Out	-0.107***	-0.105***	-0.090***	-0.052***	-0.051***	-0.051***	-0.051***
	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Observation	1366418	1360228	1338105	1266582	1238883	1238636	1224100
Group	9506	45504	111767	314048	326898	326873	329221
Under-identification test	763.228	1012.002	1480.664	4814.339	4641.439	4623.897	4430.648
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5.4e+04	3.8e+04	2.3e+04	2.7e+04	2.6e+04	2.6e+04	2.6e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.183	1.525	1.780	1.494	1.051	1.122	0.371
P value	(0.669)	(0.217)	(0.182)	(0.222)	(0.305)	(0.289)	(0.543)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B17. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 3.4-21: The Correlation of Inventory and Outsourcing with Firm Size.

Dependent	Employment Size							
Variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
Dist	-0.000	-0.000	-0.000	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Inventory	0.044***	0.041***	0.041***	0.039***				
	(0.001)	(0.001)	(0.001)	(0.001)				
Out					-0.012***	-0.009***	-0.009***	-0.009***
					(0.002)	(0.002)	(0.002)	(0.002)
R2	0.145	0.127	0.127	0.123	0.138	0.121	0.121	0.117
Observation	1479200	1461735	1461735	1461735	1529583	1511700	1511700	1511700
Group	458914	526920	527131	545396	469162	539247	539465	557962
F	2667.302	2232.192	2233.570	2148.188	2578.891	2202.011	2203.706	2125.104
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B18. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

As further robustness checks, Table 3.4-21 uses firm size (log employment) as the dependent variable. Regressions in (1) to (4) columns show that a 10% increase of inventory level is related to the increase of firm size by 0.40%. Regressions in columns (5) to (8) show that, when firm-level fixed effects are controlled, a 10% increase of outsourcing is related to the decrease of firm size by 0.10%. These results confirm that higher levels of inventory are related to larger size and higher productivity, while higher outsourcing level is related to smaller size and lower productivity. Due to highway development tends to increase both inventory and outsourcing level, the firm size dispersion is expected to increase. This size mechanism is further specified in the next chapter.

Table 3.4-22: The Heterogeneous Impacts across Coastal and Inland Regions.

Dependent Variable:	Coast				Inland			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
Dist	-0.061***	-0.050***	-0.051***	-0.049***	-0.011	-0.011	-0.010	-0.002
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Observation	951640	930076	929957	920367	317722	311525	311397	306400
Group	234105	245248	245239	246948	80470	82223	82207	82869
U test	3807.559	3616.240	3595.945	3567.282	3010.599	3078.295	3070.481	2955.723
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
W test	2.2e+04	2.1e+04	2.0e+04	2.1e+04	1.7e+04	1.7e+04	1.7e+04	1.7e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
O test	0.501	0.464	0.422	0.113	0.009	0.006	0.006	0.002
P value	(0.479)	(0.496)	(0.516)	(0.737)	(0.926)	(0.937)	(0.938)	(0.969)
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B19. The Under-identification test (U test) reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test (W test) reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test (O test) reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

This section studies the heterogeneous impacts of infrastructure development from two aspects. First, firms located in inland and coastal provinces face different factor endowments, e.g., coastal provinces have higher income levels, the business environments in these regions are more diversified and export-oriented. Second, firms normally focus on one or several goods as their main business, while different goods have various value-weight ratios, which may influence firms' unit transportation costs, then further affect their productivity. Syverson (2004) constructs value-weight ratio across 4-digit industries. Due to this industry-level proxy being likely to be absorbed by industrial fixed effects, we divide firms into two groups according to the median value-weight ratio, and run regressions for high and low value-weight ratio groups.[6]

Table 3.4-22 investigates the heterogeneous impacts of infrastructure development across inland and coastal provinces. Similar to the baseline regressions, observations are controlled by firm, regional, industrial, and year fixed effects. Columns (1) to (4) show that the impacts of highway construction on firms located in coastal provinces, a 10% decrease of $Dist$ can increase firm-level productivity by 0.50% to 0.60%. By contrast, columns (5) to (8) show that the impacts of highway construction on firms located in inland provinces, while the coefficients of $Dist$ are insignificant. These results suggest that firms located in coastal regions are more sensitive to highway expansion.

Table 3.4-23: The Heterogeneous Impacts across High or Low Value-Weight Ratio Products.

Dependent Variable:	High Value-weight Ratio				Low Value-weight Ratio			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
Dist	-0.033***	-0.025***	-0.025***	-0.022***	-0.032***	-0.031***	-0.032***	-0.028***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Observation	549372	538294	538219	530045	688082	680661	680492	674308
Group	149628	154139	154137	154426	185767	186815	186785	187510
U test	3171.266	3160.152	3142.789	3030.982	3124.326	3069.761	3058.966	3069.670
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
W test	1.6e+04	1.6e+04	1.6e+04	1.7e+04	2.1e+04	2.0e+04	2.0e+04	2.1e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
O test	0.000	0.037	0.019	0.508	0.146	0.053	0.030	0.010
P value	(0.984)	(0.847)	(0.890)	(0.476)	(0.703)	(0.818)	(0.863)	(0.919)
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B20. The Under-identification test (U test) reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test (W test) reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test (O test) reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 3.4-23 investigates the heterogeneous impacts across high and low value-weight ratio industries. Columns (1) to (4) show that, for industries with high value-weight ratio, a 10% decrease of $Dist$ can increase firm-level productivity by 0.20% to 0.30%. By contrast,

[6]Firms with value-weight ratio higher than the median are classified into high group, and vice verse.

columns (5) to (8) show that, for industries with low value-weight ratio, a 10% decrease of $Dist$ can increase firm-level productivity by about 0.30%. These results suggest that the impacts of highway construction industries with low value-weight ratio are slightly stronger than high value-weight ratio industries.

3.5 Conclusions

This study investigates the impacts of highway construction on firm-level productivity. In order to address potential endogeneity issues, we construct IVs based on historical roads and counter-factual roads approaches. The baseline regressions show that firms closer to newly constructed highways have productivity premiums, a 10% decrease of distance to highways can increase firm-level productivity by 0.2%-0.3%. The decrease of firms' distance to highways is induced by two mechanisms: on the one side, highway development can reduce $Dist$ of address-unchanged firms; on the other side, highway development can also promote firms' entry and relocation, then reduce the overall $Dist$. Highways have lagged impacts on productivity growth, if we replace $Dist$ as one-year-lagged $Dist$, a 10% decrease of lagged $Dist$ is related to a 0.8% or 1.0% increase of firm-level productivity. These results are robust across different types of transport modes (road, railway, waterway), different productivity measurements (OLS, OP, LP, and ACF productivity), and distance measurements (entrance distances).

Channel studies show that firms closer to highways have higher inventory level, while higher inventory level is related to larger firm size and higher productivity; by contrast, firms closer to highways have higher outsourcing level, while higher outsourcing level is related to smaller firm size and lower productivity, but their coefficients are smaller. The highways' impacts on inventory and outsourcing activities seem to conflict with each other, considering the very large and significant coefficients of firm size, a possible explanation is that the firm size increase is much faster even if the outsourcing promotion effect can reduce the average firm size. The increase of firm size is accompanied with productivity increases because large firms can normally reap more benefits, a potential consequence is that firm size dispersion will be increased during this process. This firm size mechanism is further studied in the next chapter.

Besides these key explanatory variables, the coefficients of several control variables also provide noteworthy results. First, a 10% increase of firm size is related to 1.6% increase of productivity, while one year older is related to 0.02% increase of productivity. Second, state-owned firms are less productive than other types of firms, a 10% increase of state-owned capital is related to 0.15% decrease of productivity. The productivity effects of foreign, private, and collective capital are relatively closer and more similar, but those of state-owned capital are systemically less productive than for other types of capital. Heterogeneity studies show that highway construction effects in coastal provinces are stronger than in inland provinces, firms' productivity in coastal regions shows a more significant increase after highway construction than inland regions.

These results provide more details to explain the spillover effects of Chinese highway expansion. Existing studies for China do not always support the view that there are positive effects of infrastructure improvement on economic growth, but they do tend to note more significant relocation effects (Faber 2014, Baum-Snow et al. 2017); while studies in western developed countries tend to conclude that road construction has positive effects on economic efficiency (Garcia-Mila et al. 1996, Holl 2016, Gibbons et al. 2016). Faber (2014) finds the improvement of transportation infrastructure opens up competition between rich and rural regions, it promotes population relocation and the development of scale economy in rich regions, but also accelerates the outflow of immigration in rural regions; similarly, Baum-Snow et al. (2017) also find road and railroad construction promote population decentralization, from city central areas toward suburban areas. These facts suggest that, due to road construction not always connecting rich cities or high productivity regions, it may compel economic activities to make more adjustments to accommodate the new traffic environment than in western developed countries. If the adjustments involve reinvestment, the adjustment period could be longer than the infrastructure construction period, e.g., when a new highway comes, a firm wishing to move to a site closer to its target market may take several years to build a new factory; therefore, the firm's original factory may lose its competitiveness during its adjustment period in the new traffic environment. More importantly, China's infrastructure expansion is a continuous process with high speed and has been largely independent of market logic since 1997, which means firms' adjustment can never reach an equilibrium state. A direct consequence is that the bonus of infrastructure expansion will take a longer time to realise. However, our results reveal that highway construction can generate observable and persistent impacts simultaneously after the completion of highway projects, suggesting that highway construction effects are stronger than our expectations. At the same time, the spillover effects of Chinese highway expansion are expected to increase over time, which can be further investigated by the following studies.

.1 Appendix A

Summary Statistics

Table A1: Firms Identified by GIS.

Year	Total Obs	Successfully Identified	Fail to be Identified	Identified Rate
1998	165118	164513	605	99.63%
1999	162033	161039	994	99.39%
2000	162885	160269	2616	98.39%
2001	171256	169144	2212	98.77%
2002	181557	176484	5073	97.21%
2003	196222	193781	2441	98.76%
2004	279092	276671	2421	99.13%
2005	271835	270116	1719	99.37%
2006	301961	300067	1894	99.37%
2007	336768	333911	2857	99.15%

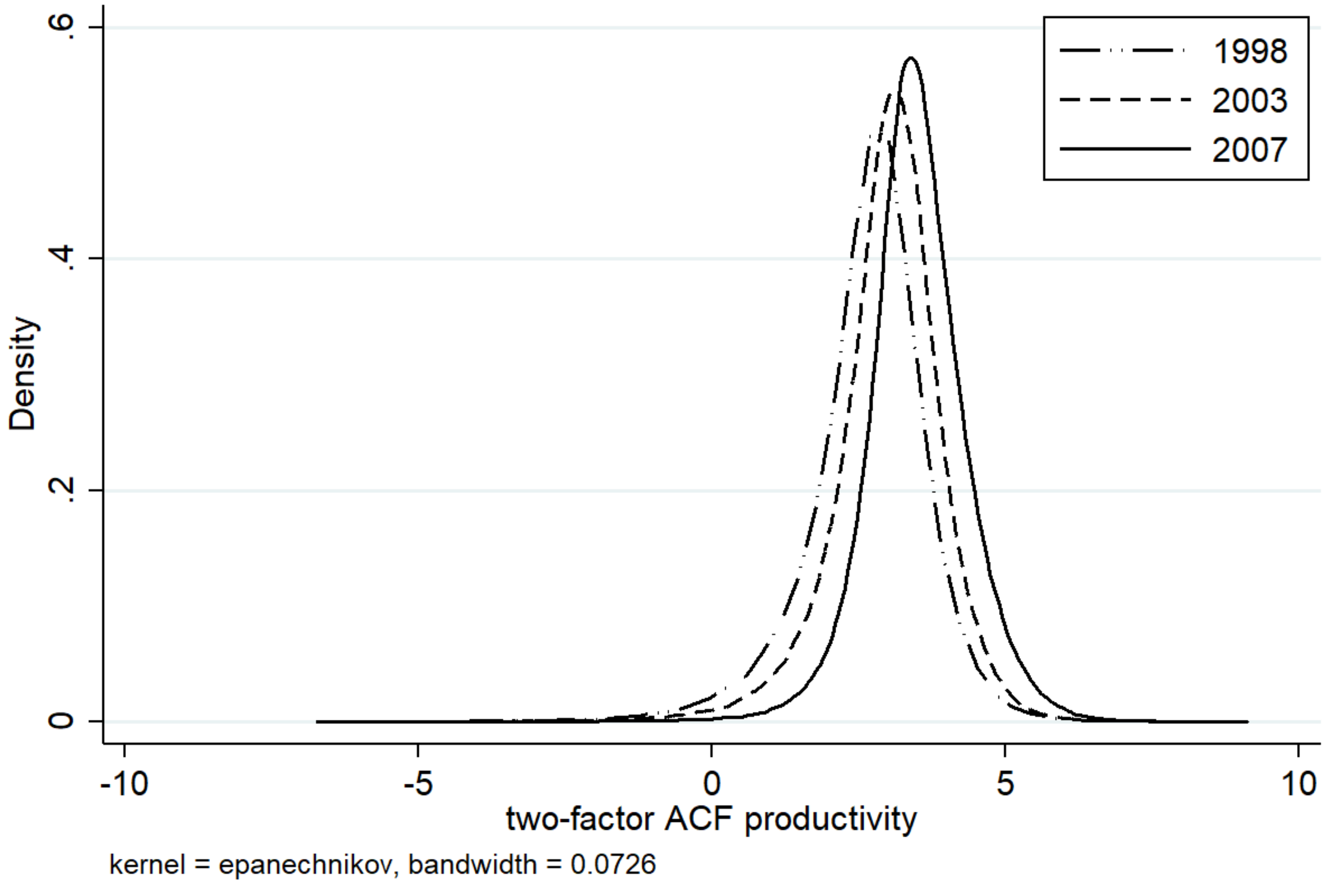
Note: This table illustrates how many firms' addresses are converted into longitude and latitude through Baidu Map Open Platform.

Table A2: Additional Summary Statistics

	Obs	Full Sample Average (Std. Dev.)	1998		2003		2007	
L	2,024,884	4.779	127,942	5.098	181,982	4.837	323,698	4.623
		(1.151)		(1.217)		(1.140)		(1.090)
K	2,025,949	8.287	128,421	8.348	181,982	8.327	323,728	8.287
		(1.720)		(1.767)		(1.733)		(1.660)
M	2,020,381	9.564	127,687	9.217	181,563	9.575	323,594	9.869
		(1.424)		(1.526)		(1.442)		(1.328)
Add	1,723,231	8.182	121,770	7.858	177,847	8.095	319,778	8.547
		(1.437)		(1.487)		(1.426)		(1.351)
OLS TFP	1,621,161	3.822	94,972	3.252	124,848	3.685	286,967	4.265
		(1.063)		(1.085)		(1.035)		(.965)
OP TFP	1,652,658	3.951	98,530	3.547	127,764	3.830	290,298	4.299
		(1.084)		(1.163)		(1.094)		(.998)
LP TFP	1,652,658	6.271	98,530	6.238	127,764	6.428	290,298	6.772
		(1.307)		(1.259)		(1.211)		(1.135)
ACF TFP	1,652,658	3.121	98,530	2.660	127,764	2.996	290,298	3.485
		(.945)		(1.020)		(.937)		(.848)

Note: Average and standard deviation of the full sample and the sub-samples in different years (1998, 2003, 2007) are compared in this table.

Figure A1: Kernel Density Estimate of acf Productivity



.2 Appendix B

Additional Regression Results

Table B1: Relationship between Line Distance and Productivity (OLS Results).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect				
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Dist	-0.007***	0.002*	0.001	0.001	0.002	0.002	0.001	
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	
Rail-Density	0.015	0.024*	0.031**	0.012	0.013	0.013	0.009	
	(0.018)	(0.012)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)	
Road-Density	0.012	0.018	0.006	0.021***	0.019***	0.019***	0.018**	
	(0.017)	(0.012)	(0.009)	(0.005)	(0.005)	(0.005)	(0.006)	
River-Density	-0.013	-0.031**	-0.023**	-0.025***	-0.020***	-0.020***	-0.020***	
	(0.013)	(0.010)	(0.009)	(0.005)	(0.006)	(0.006)	(0.006)	
Pop	0.089***	-0.039**	-0.026*	-0.035***	-0.038***	-0.039***	-0.043***	
	(0.005)	(0.013)	(0.013)	(0.006)	(0.006)	(0.006)	(0.006)	
Age	-0.006***	-0.005***	-0.005***	-0.002***	-0.002***	-0.002***	-0.002***	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Export	-0.016**	-0.012***	-0.004	0.033***	0.035***	0.035***	0.035***	
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	
Size	0.159***	0.156***	0.153***	0.165***	0.165***	0.165***	0.163***	
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	
State-share	-0.575***	-0.506***	-0.416***	-0.069***	-0.056***	-0.056***	-0.060***	
	(0.012)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	
Collective-share	0.050***	0.032***	0.016***	-0.010**	-0.006	-0.006	-0.007	
	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	
Foreign-share	-0.046***	-0.027***	-0.021***	0.014*	0.017*	0.017*	0.014*	
	(0.010)	(0.006)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)	
R2	0.153	0.140	0.127	0.107	0.107	0.107	0.107	
Observation	1369685	1369685	1369685	1385589	1385589	1385589	1385589	
Group	10020	52253	140815	430797	471459	471681	488639	
F	997.5	1521.8	1811.3	3466.9	3391.9	3392.2	3368.0	
Region	province	prefecture	county		province	prefecture	county	
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Firm				Yes				Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table illustrate OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B2: Relationship between Line Distance and Productivity.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.025***	-0.034***	-0.021***	-0.033***	-0.027***	-0.027***	-0.022***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Rail-Density	0.018	0.028*	0.033**	0.015*	0.015*	0.014*	0.010
	(0.018)	(0.012)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
Road-Density	0.012	0.017	0.006	0.023***	0.020***	0.020***	0.019***
	(0.017)	(0.012)	(0.009)	(0.005)	(0.005)	(0.005)	(0.006)
River-Density	-0.016	-0.035***	-0.025**	-0.028***	-0.022***	-0.022***	-0.022***
	(0.013)	(0.010)	(0.009)	(0.005)	(0.006)	(0.006)	(0.006)
Pop	0.079***	-0.049***	-0.032*	-0.042***	-0.043***	-0.043***	-0.046***
	(0.006)	(0.013)	(0.013)	(0.006)	(0.006)	(0.006)	(0.006)
Age	-0.006***	-0.005***	-0.005***	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	-0.017**	-0.013***	-0.004	0.033***	0.035***	0.035***	0.035***
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Size	0.158***	0.155***	0.153***	0.165***	0.165***	0.165***	0.163***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.575***	-0.506***	-0.415***	-0.069***	-0.056***	-0.056***	-0.060***
	(0.012)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Collective-share	0.050***	0.032***	0.016***	-0.011**	-0.006	-0.006	-0.007
	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Foreign-share	-0.050***	-0.031***	-0.022***	0.014*	0.017*	0.016*	0.014*
	(0.010)	(0.006)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
Observation	1369177	1362969	1340814	1269369	1241601	1241354	1226767
Group	9512	45537	111944	314577	327471	327446	329817
Under-identification test	764.360	1014.196	1483.183	4829.538	4656.493	4638.951	4448.347
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5.5e+04	3.8e+04	2.3e+04	2.7e+04	2.6e+04	2.6e+04	2.6e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.133	1.263	1.837	1.487	1.031	1.098	0.408
P value	(0.715)	(0.261)	(0.175)	(0.222)	(0.310)	(0.295)	(0.523)
First-stage Results:							
Ming30km	0.216***						
	(0.008)						
Qing30km	0.057***						
	(0.008)						
SL30km		0.086***					
		(0.007)					
LC30km		0.220***	0.222***	0.278***	0.280***	0.279***	0.287***
		(0.008)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Qing50km			0.127***	0.155***	0.159***	0.159***	0.174***
			(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
F	681.069	580.091	596.047	2107.042	1966.565	1965.791	1871.750
R2	0.155	0.123	0.118	0.161	0.154	0.154	0.153
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B3: Robustness Checks on Relocation Effects.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	0.003	-0.019*	-0.018*	-0.031***	-0.029**	-0.029**	-0.031***
	(0.007)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Rail-Density	0.037**	0.043***	0.031*	0.036***	0.040***	0.040***	0.036***
	(0.013)	(0.012)	(0.012)	(0.010)	(0.010)	(0.010)	(0.010)
Road-Density	0.074***	0.082***	0.051**	0.043**	0.043**	0.043**	0.045**
	(0.021)	(0.018)	(0.017)	(0.015)	(0.015)	(0.015)	(0.015)
River-Density	-0.052***	-0.050***	-0.030*	-0.019	-0.020	-0.020	-0.019
	(0.015)	(0.014)	(0.013)	(0.011)	(0.010)	(0.010)	(0.011)
Pop	0.131***	-0.005	0.008	-0.006	-0.002	-0.002	-0.002
	(0.011)	(0.021)	(0.019)	(0.017)	(0.017)	(0.017)	(0.017)
Age	-0.015***	-0.013***	-0.011***	-0.003***	-0.003***	-0.003***	-0.003***
	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Export	0.030***	0.020**	0.025***	0.044***	0.047***	0.047***	0.049***
	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)	(0.008)
Size	0.162***	0.161***	0.161***	0.199***	0.197***	0.197***	0.193***
	(0.004)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.006)
State-share	-0.613***	-0.523***	-0.385***	-0.075***	-0.070***	-0.070***	-0.073***
	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Collective-share	0.038***	0.009	-0.000	-0.019**	-0.018**	-0.018**	-0.018**
	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Foreign-share	-0.008	0.009	0.042**	0.026	0.033	0.033	0.027
	(0.021)	(0.014)	(0.014)	(0.019)	(0.019)	(0.019)	(0.019)
Observation	234850	230950	222760	220632	216571	216567	212456
Group	6113	20348	36479	58388	60183	60182	60276
Under-identification test	572.515	802.285	740.108	1890.427	1904.805	1904.788	1877.849
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.9e+04	1.4e+04	1.2e+04	1.4e+04	1.4e+04	1.4e+04	1.4e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.024	0.591	5.779	0.208	0.071	0.071	0.203
P value	(0.878)	(0.442)	(0.016)	(0.649)	(0.790)	(0.791)	(0.653)
First-stage Results:							
Ming30km	0.274***						
	(0.010)						
Qing30km	0.065***						
	(0.010)						
SL30km		0.106***					
		(0.010)					
LC30km		0.305***	0.357***	0.612***	0.607***	0.607***	0.598***
		(0.012)	(0.012)	(0.009)	(0.009)	(0.009)	(0.009)
Qing50km			0.170***	0.431***	0.430***	0.430***	0.433***
			(0.010)	(0.013)	(0.014)	(0.014)	(0.014)
F	471.648	315.341	294.587	1040.967	989.525	989.529	925.678
R2	0.253	0.194	0.197	0.332	0.330	0.330	0.325
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: TThe Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B4: Robustness on Different Types of Traffic Routes.

Dependent Variable: TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Main Results:								
Dist	-0.043***	-0.036***	-0.036***	-0.027***	-0.043***	-0.036***	-0.036***	-0.027***
	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)
Road-Density	0.043***	0.040***	0.040***	0.038***	0.042***	0.038***	0.039***	0.037***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Rail-Density					0.021**	0.020**	0.020**	0.016*
					(0.007)	(0.007)	(0.007)	(0.007)
Pop	-0.046***	-0.046***	-0.047***	-0.049***	-0.047***	-0.047***	-0.047***	-0.049***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Age	-0.002***	-0.001***	-0.001***	-0.002***	-0.002***	-0.001***	-0.001***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	0.032***	0.035***	0.035***	0.035***	0.032***	0.035***	0.035***	0.035***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Size	0.166***	0.166***	0.166***	0.165***	0.166***	0.166***	0.166***	0.165***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.059***	-0.048***	-0.048***	-0.052***	-0.060***	-0.048***	-0.048***	-0.052***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Collective-share	-0.011**	-0.007	-0.007	-0.007*	-0.011**	-0.007	-0.007	-0.007*
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Foreign-share	0.014*	0.016*	0.016*	0.014*	0.014*	0.016*	0.016*	0.013*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Observation	1339289	1309989	1309735	1294231	1339289	1309989	1309735	1294231
Group	329151	343180	343156	345625	329151	343180	343156	345625
Under-identification test	4075.504	3830.330	3812.747	3756.333	4079.313	3832.867	3815.228	3757.757
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.9e+04	1.8e+04	1.8e+04	1.9e+04	1.9e+04	1.8e+04	1.8e+04	1.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.697	0.822	0.807	0.383	0.749	0.856	0.841	0.407
P value	(0.404)	(0.365)	(0.369)	(0.536)	(0.387)	(0.355)	(0.359)	(0.523)
First-stage Results:								
Qing50km	0.151***	0.155***	0.155***	0.169***	0.151***	0.155***	0.155***	0.169***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
LC30km	0.280***	0.282***	0.281***	0.289***	0.280***	0.281***	0.281***	0.289***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
F	2367.984	2200.268	2198.870	2083.066	2270.001	2111.037	2110.233	1998.204
R2	0.158	0.151	0.151	0.149	0.158	0.151	0.151	0.149
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry		4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table B5: Regressions with Lagged Dist.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Lag Dist	-0.027***	-0.026***	-0.026**	-0.098***	-0.083***	-0.084***	-0.077***
	(0.006)	(0.007)	(0.010)	(0.013)	(0.014)	(0.014)	(0.014)
Rail-Density	0.001	0.020	0.032**	0.008	0.004	0.004	-0.004
	(0.021)	(0.014)	(0.012)	(0.008)	(0.008)	(0.008)	(0.008)
Road-Density	-0.001	0.005	-0.010	-0.006	-0.015*	-0.015*	-0.018**
	(0.018)	(0.012)	(0.010)	(0.006)	(0.006)	(0.006)	(0.006)
River-Density	-0.049***	-0.071***	-0.054***	-0.037***	-0.034***	-0.034***	-0.035***
	(0.015)	(0.012)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
Pop	0.065***	-0.055***	-0.043**	-0.074***	-0.069***	-0.070***	-0.076***
	(0.006)	(0.014)	(0.014)	(0.007)	(0.007)	(0.007)	(0.007)
Age	-0.009***	-0.009***	-0.008***	-0.003***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	-0.033***	-0.027***	-0.018***	0.007*	0.006	0.006	0.006
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Size	0.167***	0.165***	0.162***	0.141***	0.140***	0.140***	0.138***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
State-share	-0.536***	-0.470***	-0.380***	-0.061***	-0.049***	-0.049***	-0.054***
	(0.013)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.009)
Collective-share	0.034***	0.013*	-0.005	-0.036***	-0.029***	-0.029***	-0.028***
	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Foreign-share	-0.033**	-0.008	0.000	0.007	0.011	0.011	0.010
	(0.011)	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)
Observation	910256	904696	886874	836897	819707	819531	812324
Group	8919	39289	89880	227986	235772	235737	237267
Under-identification test	665.455	780.839	737.319	1482.949	1207.257	1204.765	1203.870
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	2.7e+04	1.8e+04	6212.882	3274.749	2547.477	2536.444	2547.334
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.231	0.002	1.743	0.952	0.067	0.117	0.785
P value	(0.631)	(0.961)	(0.187)	(0.329)	(0.795)	(0.732)	(0.376)
First-stage Results:							
Ming30km	0.183***			0.126***	0.116***	0.116***	0.117***
	(0.008)			(0.003)	(0.004)	(0.004)	(0.004)
Qing30km	0.063***						
	(0.008)						
SL30km		0.087***					
		(0.007)					
LC30km		0.182***	0.150***				
		(0.007)	(0.006)				
Qing50km			0.085***	-0.004	-0.004	-0.004	0.003
			(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
F	602.385	454.455	428.120	1377.029	1263.648	1263.235	1244.637
R2	0.142	0.109	0.100	0.119	0.110	0.109	0.109
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B6: Alternative Measurement: Entrance Distance.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist Inter	-0.027***	-0.032***	-0.023***	-0.036***	-0.030***	-0.030***	-0.024***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Rail-Density	0.019	0.028*	0.034***	0.015*	0.015*	0.014*	0.010
	(0.018)	(0.012)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
Road-Density	0.014	0.020	0.008	0.024***	0.021***	0.021***	0.020***
	(0.017)	(0.012)	(0.009)	(0.005)	(0.005)	(0.005)	(0.006)
River-Density	-0.012	-0.032**	-0.023*	-0.024***	-0.019***	-0.019***	-0.019***
	(0.013)	(0.010)	(0.009)	(0.005)	(0.006)	(0.006)	(0.006)
Pop	0.076***	-0.048***	-0.033*	-0.042***	-0.042***	-0.042***	-0.046***
	(0.006)	(0.013)	(0.013)	(0.006)	(0.006)	(0.006)	(0.006)
Age	-0.006***	-0.005***	-0.005***	-0.002***	-0.002***	-0.002***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	-0.017**	-0.013***	-0.004	0.033***	0.035***	0.035***	0.034***
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Size	0.158***	0.155***	0.153***	0.165***	0.165***	0.165***	0.164***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.576***	-0.508***	-0.415***	-0.068***	-0.055***	-0.055***	-0.059***
	(0.012)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Collective-share	0.049***	0.031***	0.016***	-0.011**	-0.007	-0.007	-0.007
	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Foreign-share	-0.050***	-0.030***	-0.022***	0.014*	0.017*	0.017*	0.014*
	(0.010)	(0.006)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
Observation	1369177	1362969	1340814	1269369	1241601	1241354	1226767
Group	9512	45537	111944	314577	327471	327446	329817
Under-identification test	737.527	1159.618	1969.677	7001.182	6565.355	6545.407	6181.729
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	8.4e+04	9.4e+04	4.9e+04	4.9e+04	4.6e+04	4.5e+04	4.5e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.592	0.221	2.171	2.198	1.685	1.765	0.835
P value	(0.442)	(0.892)	(0.141)	(0.138)	(0.194)	(0.184)	(0.361)
First-stage Results:							
Ming30km	0.179***						
	(0.007)						
Qing30km	0.079***						
	(0.007)						
SL30km		0.140***					
		(0.005)					
LC30km		0.192***	0.202***	0.258***	0.260***	0.259***	0.266***
		(0.006)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
Qing50km			0.126***	0.128***	0.126***	0.126***	0.134***
			(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
F	684.328	452.607	480.578	1962.580	1887.816	1886.926	1857.151
R2	0.193	0.168	0.140	0.182	0.178	0.177	0.179
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B7: Robustness Checks: Regressions without Big Cities.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.024***	-0.021***	-0.016**	-0.041***	-0.033***	-0.033***	-0.031***
	(0.005)	(0.005)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Rail-Density	0.025	0.031*	0.039***	0.020**	0.021**	0.021**	0.014
	(0.019)	(0.013)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)
Road-Density	-0.022	-0.020	-0.033***	-0.029***	-0.031***	-0.031***	-0.034***
	(0.018)	(0.012)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
River-Density	-0.069***	-0.099***	-0.086***	-0.086***	-0.069***	-0.069***	-0.070***
	(0.015)	(0.012)	(0.011)	(0.006)	(0.007)	(0.007)	(0.007)
Pop	0.079***	-0.048***	-0.034**	-0.047***	-0.045***	-0.046***	-0.049***
	(0.006)	(0.013)	(0.013)	(0.006)	(0.006)	(0.006)	(0.006)
Age	-0.005***	-0.004***	-0.004***	-0.002***	-0.001***	-0.001***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	-0.015**	-0.011**	-0.002	0.031***	0.033***	0.033***	0.032***
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Size	0.158***	0.155***	0.152***	0.161***	0.162***	0.162***	0.160***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.578***	-0.502***	-0.410***	-0.068***	-0.055***	-0.055***	-0.059***
	(0.013)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)	(0.008)
Collective-share	0.043***	0.022***	0.005	-0.019***	-0.014***	-0.014***	-0.014***
	(0.007)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Foreign-share	-0.070***	-0.052***	-0.042***	0.011	0.015*	0.015*	0.014*
	(0.010)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)
Observation	1271649	1265474	1244441	1180652	1155614	1155368	1143357
Group	8362	44300	105266	291696	304010	303985	306414
Under-identification test	732.833	876.152	1366.370	6534.152	6381.905	6354.039	6199.557
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5.2e+04	3.6e+04	2.2e+04	3.9e+04	3.8e+04	3.8e+04	3.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.030	0.882	0.408	0.004	0.032	0.023	0.548
P value	(0.862)	(0.348)	(0.523)	(0.951)	(0.857)	(0.880)	(0.459)
First-stage Results:							
Ming30km	0.213***			0.322***	0.328***	0.327***	0.331***
	(0.009)			(0.004)	(0.004)	(0.004)	(0.004)
Qing30km	0.0596***						
	(0.008)						
SL30km		0.0864***					
		(0.008)					
LC30km		0.228***	0.227***				
		(0.008)	(0.007)				
Qing50km			0.124***	-0.0102	-0.0108	-0.0111*	0.00124
			(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
F	633.6	504.8	534.7	1958.2	1842.0	1839.8	1810.8
R2	0.157	0.122	0.120	0.180	0.176	0.176	0.177
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B8: Robustness Checks on Different Types of TFP.

Dependent Variable:	OLS TFP (1)	OP TFP (2)	LP TFP (3)	OLS TFP (4)	OP TFP (5)	LP TFP (6)
Main Results:						
Dist	-0.042***	-0.041***	-0.036***	-0.035***	-0.032***	-0.026***
	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)
Rail-Density	-0.038***	-0.010	0.020**	-0.028***	-0.020**	0.006
	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)
Road-Density	0.137***	0.029***	0.007	0.133***	0.013*	0.003
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
River-Density	-0.017*	-0.016**	-0.018**	-0.004	-0.015**	-0.015*
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
Pop	-0.067***	-0.035***	-0.003	-0.068***	-0.030***	-0.002
	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
Age	-0.002***	-0.002***	-0.000	-0.002***	-0.002***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	0.054***	0.040***	0.083***	0.053***	0.040***	0.079***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Size	0.098***	0.222***	0.398***	0.093***	0.219***	0.382***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.118***	-0.056***	-0.015*	-0.102***	-0.042***	-0.014
	(0.008)	(0.008)	(0.007)	(0.008)	(0.008)	(0.007)
Collective-share	-0.026***	-0.013**	-0.012**	-0.025***	-0.005	-0.007
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Foreign-share	-0.002	0.011	0.028***	-0.001	0.012	0.025***
	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)
Observation	1244947	1269369	1269369	1203083	1226767	1226767
Group	310835	314577	314577	325357	329817	329817
Under-identification test	6632.761	6807.708	6807.708	6294.355	6464.496	6464.496
P value	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Weak identification test	3.8e+04	3.9e+04	3.9e+04	3.8e+04	3.9e+04	3.9e+04
P value	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Over-identification test	0.835	0.495	0.154	2.853	1.043	0.280
P value	(0.361)	(0.482)	(0.695)	(0.091)	(0.307)	(0.597)
First-stage Results:						
Ming30km	0.301***	0.301***	0.301***	0.310***	0.310***	0.310***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Qing30km	0.037***	0.037***	0.037***	0.051***	0.051***	0.051***
	(0.005)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
F	2236.122	2236.122	2236.122	2020.280	2020.280	2020.280
R2	0.181	0.181	0.181	0.176	0.176	0.176
Region				province	prefecture	county
Industry				4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: Observations are controlled by firm-level fixed effects and year fixed effects. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B9: The Impacts on Value Added.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Value Added	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	0.015	0.045***	-0.007	-0.044***	-0.037***	-0.038***	-0.034***
	(0.008)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Rail-Density	-0.057**	-0.048***	-0.021*	-0.001	-0.011	-0.011	-0.010
	(0.020)	(0.013)	(0.011)	(0.007)	(0.007)	(0.007)	(0.007)
Road-Density	-0.013	-0.007	-0.020*	-0.003	-0.003	-0.003	-0.001
	(0.016)	(0.012)	(0.010)	(0.006)	(0.006)	(0.006)	(0.006)
River-Density	0.006	-0.011	-0.012	-0.025***	-0.017**	-0.017**	-0.017**
	(0.016)	(0.012)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
Pop	0.058***	-0.000	-0.013	-0.035***	-0.059***	-0.060***	-0.067***
	(0.009)	(0.009)	(0.008)	(0.006)	(0.006)	(0.006)	(0.007)
Age	0.002***	0.002***	0.003***	-0.001	-0.000	-0.000	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	0.180***	0.182***	0.177***	0.093***	0.091***	0.091***	0.088***
	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Size	0.738***	0.736***	0.729***	0.570***	0.554***	0.554***	0.546***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
State-share	-0.605***	-0.533***	-0.409***	-0.027***	-0.015	-0.015*	-0.017*
	(0.013)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Collective-share	0.029***	0.007	-0.007	-0.026***	-0.023***	-0.023***	-0.024***
	(0.007)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Foreign-share	-0.102***	-0.077***	-0.054***	0.026***	0.027***	0.027***	0.024**
	(0.008)	(0.006)	(0.006)	(0.008)	(0.008)	(0.008)	(0.008)
Observation	1177896	1171826	1150381	1084616	1059396	1059164	1045060
Group	9432	44761	108752	301967	312759	312733	314424
Under-identification test	544.026	1680.482	1520.200	3740.332	3506.020	3486.682	3490.977
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5.3e+04	2.9e+04	1.7e+04	1.7e+04	1.6e+04	1.6e+04	1.7e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	3.267	2.661	0.266	0.331	0.555	0.503	0.760
P value	(0.071)	(0.103)	(0.606)	(0.565)	(0.456)	(0.478)	(0.384)
First-stage Results:							
Qing30km	0.091***	0.151***	0.155***				
	(0.005)	(0.006)	(0.005)				
SL30km	0.235***	0.199***	0.178***	0.219***	0.217***	0.217***	0.236***
	(0.005)	(0.006)	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)
Qing50km				0.166***	0.172***	0.171***	0.183***
				(0.005)	(0.005)	(0.005)	(0.005)
F	810.828	532.866	571.813	1961.916	1822.847	1821.441	1746.526
R2	0.159	0.116	0.112	0.146	0.140	0.139	0.139
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B10: The Impacts on Profit.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Profit	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	0.046***	0.073***	-0.014	-0.054***	-0.044***	-0.045***	-0.046***
	(0.010)	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)
Rail-Density	-0.044	-0.038*	0.006	0.004	0.004	0.005	0.013
	(0.026)	(0.019)	(0.016)	(0.011)	(0.012)	(0.012)	(0.012)
Road-Density	0.126***	0.140***	0.110***	0.136***	0.133***	0.133***	0.137***
	(0.023)	(0.018)	(0.015)	(0.009)	(0.009)	(0.009)	(0.010)
River-Density	0.041	0.013	0.011	-0.015	-0.003	-0.003	-0.006
	(0.023)	(0.018)	(0.015)	(0.009)	(0.009)	(0.009)	(0.010)
Pop	0.103***	0.099***	0.089***	-0.007	-0.049***	-0.050***	-0.060***
	(0.010)	(0.023)	(0.021)	(0.011)	(0.011)	(0.011)	(0.012)
Age	-0.005***	-0.004***	-0.003***	-0.002***	-0.002**	-0.002**	-0.002**
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Export	0.038***	0.058***	0.067***	0.062***	0.064***	0.064***	0.061***
	(0.010)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Size	0.761***	0.761***	0.755***	0.583***	0.568***	0.568***	0.559***
	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
State-share	-0.818***	-0.735***	-0.590***	-0.103***	-0.083***	-0.084***	-0.084***
	(0.023)	(0.014)	(0.013)	(0.012)	(0.012)	(0.012)	(0.012)
Collective-share	0.068***	0.034***	0.011	-0.040***	-0.037***	-0.037***	-0.036***
	(0.011)	(0.008)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
Foreign-share	-0.192***	-0.132***	-0.117***	0.031*	0.027*	0.026*	0.025
	(0.020)	(0.011)	(0.011)	(0.013)	(0.013)	(0.013)	(0.013)
Observation	1058243	1051911	1030479	960296	936427	936219	923872
Group	9211	42560	100760	275995	284578	284543	285378
Under-identification test	511.975	1507.636	1338.334	3179.254	2968.643	2951.036	2951.791
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	4.6e+04	2.5e+04	1.4e+04	1.4e+04	1.3e+04	1.3e+04	1.4e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	2.741	1.642	0.001	0.869	1.887	1.940	1.798
P value	(0.098)	(0.200)	(0.978)	(0.351)	(0.170)	(0.164)	(0.180)
First-stage Results:							
Qing30km	0.091***	0.151***	0.155***				
	(0.005)	(0.006)	(0.005)				
SL30km	0.235***	0.199***	0.178***	0.219***	0.217***	0.217***	0.236***
	(0.005)	(0.006)	(0.006)	(0.004)	(0.005)	(0.005)	(0.005)
Qing50km				0.166***	0.172***	0.171***	0.183***
				(0.005)	(0.005)	(0.005)	(0.005)
F	810.828	532.866	571.813	1961.916	1822.847	1821.441	1746.526
R2	0.159	0.116	0.112	0.146	0.140	0.139	0.139
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B11: The Impacts on Revenue.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Revenue	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.027***	0.014*	-0.005	-0.021***	-0.019***	-0.019***	-0.017***
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Rail-Density	-0.171***	-0.156***	-0.126***	-0.067***	-0.065***	-0.065***	-0.059***
	(0.016)	(0.011)	(0.010)	(0.006)	(0.006)	(0.006)	(0.006)
Road-Density	0.089***	0.101***	0.118***	0.211***	0.209***	0.209***	0.218***
	(0.016)	(0.010)	(0.009)	(0.005)	(0.005)	(0.005)	(0.005)
River-Density	-0.012	-0.027**	-0.032***	-0.026***	-0.015*	-0.015*	-0.016**
	(0.013)	(0.010)	(0.009)	(0.005)	(0.006)	(0.006)	(0.006)
Pop	0.010	-0.032***	-0.033***	-0.032***	-0.035***	-0.036***	-0.038***
	(0.007)	(0.007)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)
Age	-0.006***	-0.005***	-0.004***	-0.001***	-0.001*	-0.001*	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	0.212***	0.217***	0.213***	0.126***	0.124***	0.124***	0.120***
	(0.006)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Size	0.686***	0.688***	0.683***	0.513***	0.497***	0.497***	0.491***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.810***	-0.728***	-0.575***	-0.085***	-0.069***	-0.070***	-0.068***
	(0.012)	(0.010)	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)
Collective-share	0.025***	-0.005	-0.031***	-0.031***	-0.029***	-0.029***	-0.029***
	(0.007)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
Foreign-share	-0.115***	-0.081***	-0.066***	0.013*	0.011	0.011	0.011
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Observation	1369017	1362815	1340663	1269217	1241453	1241206	1226623
Group	9512	45535	111934	314542	327434	327409	329780
Under-identification test	536.768	980.721	1240.230	4360.739	4214.547	4199.572	4011.392
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.4e+04	3.8e+04	2.2e+04	2.3e+04	2.2e+04	2.2e+04	2.2e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	3.373	3.149	3.679	0.014	0.479	0.484	0.708
P value	(0.066)	(0.076)	(0.055)	(0.905)	(0.489)	(0.486)	(0.400)
First-stage Results:							
LC30km	0.226***	0.235***	0.189***	0.275***	0.281***	0.281***	0.278***
	(0.008)	(0.008)	(0.008)	(0.005)	(0.005)	(0.005)	(0.005)
SL50km	0.075***	0.063***		0.061***	0.057***	0.056***	0.076***
	(0.009)	(0.007)		(0.005)	(0.005)	(0.005)	(0.005)
SL30km			0.108***				
			(0.007)				
F	861.018	573.836	591.914	2061.829	1917.498	1917.106	1818.202
R2	0.166	0.122	0.117	0.155	0.147	0.147	0.145
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B12: The Channel of Inventory.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Inventory	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.076***	-0.083***	-0.041***	-0.026***	-0.019**	-0.016**	-0.014*
	(0.008)	(0.007)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
Rail-Density	-0.029	-0.045**	-0.041**	0.003	-0.001	-0.001	-0.002
	(0.023)	(0.015)	(0.013)	(0.009)	(0.009)	(0.009)	(0.009)
Road-Density	-0.128***	-0.118***	-0.116***	-0.052***	-0.051***	-0.051***	-0.050***
	(0.017)	(0.012)	(0.011)	(0.007)	(0.007)	(0.007)	(0.008)
River-Density	0.104***	0.088***	0.071***	0.041***	0.037***	0.037***	0.038***
	(0.015)	(0.012)	(0.010)	(0.006)	(0.007)	(0.007)	(0.007)
Pop	-0.075***	0.079***	0.090***	0.073***	0.053***	0.054***	0.060***
	(0.009)	(0.013)	(0.013)	(0.008)	(0.007)	(0.007)	(0.008)
Age	0.008***	0.007***	0.006***	0.001***	0.001**	0.001**	0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	0.171***	0.151***	0.146***	0.049***	0.049***	0.049***	0.047***
	(0.009)	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Size	0.935***	0.932***	0.930***	0.782***	0.774***	0.774***	0.773***
	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
State-share	-0.163***	-0.164***	-0.156***	-0.021**	-0.021**	-0.021**	-0.019*
	(0.011)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Collective-share	-0.036***	-0.041***	-0.044***	0.011*	0.008	0.008	0.006
	(0.007)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Foreign-share	0.153***	0.125***	0.127***	0.030**	0.023*	0.023*	0.023*
	(0.011)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
Observation	1316835	1310538	1288320	1216612	1189551	1189317	1174956
Group	9471	44921	109560	303835	316070	316049	318257
Under-identification test	581.598	1641.744	1492.790	4697.711	4537.538	4879.238	4334.301
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.5e+04	4.2e+04	2.2e+04	2.6e+04	2.5e+04	2.7e+04	2.5e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.217	2.559	1.158	1.917	2.372	3.958	2.624
P value	(0.641)	(0.110)	(0.282)	(0.166)	(0.124)	(0.047)	(0.105)
First-stage Results:							
Qing50km	0.080***		0.127***	0.155***	0.159***		0.174***
	(0.004)		(0.005)	(0.004)	(0.005)		(0.005)
LC30km	0.260***	0.244***	0.222***	0.278***	0.280***	0.270***	0.287***
	(0.004)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Qing30km		0.145***				0.199***	
		(0.005)				(0.004)	
F	854.321	608.272	596.047	2107.042	1966.565	2005.135	1871.750
R2	0.170	0.130	0.118	0.161	0.154	0.159	0.153
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B13: The Channel of Inventory (Unchanged Address).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Inventory	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.048***	-0.050***	-0.008	0.007	0.005	0.005	0.001
	(0.007)	(0.010)	(0.013)	(0.012)	(0.013)	(0.013)	(0.013)
Rail-Density	0.013	0.012	0.008	0.035**	0.031*	0.031*	0.029*
	(0.017)	(0.016)	(0.016)	(0.012)	(0.012)	(0.012)	(0.013)
Road-Density	-0.121***	-0.104***	-0.097***	-0.050**	-0.050**	-0.050**	-0.053**
	(0.026)	(0.023)	(0.022)	(0.018)	(0.019)	(0.019)	(0.019)
River-Density	0.035*	0.032*	0.027	0.024*	0.013	0.013	0.014
	(0.018)	(0.016)	(0.014)	(0.012)	(0.012)	(0.012)	(0.012)
Pop	-0.051***	0.015	0.011	0.001	-0.010	-0.010	-0.008
	(0.012)	(0.025)	(0.023)	(0.019)	(0.019)	(0.019)	(0.020)
Age	0.007***	0.006***	0.004***	-0.000	-0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Export	0.129***	0.110***	0.100***	0.065***	0.070***	0.070***	0.068***
	(0.012)	(0.011)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Size	0.926***	0.917***	0.911***	0.805***	0.798***	0.798***	0.796***
	(0.004)	(0.004)	(0.004)	(0.008)	(0.008)	(0.008)	(0.008)
State-share	-0.153***	-0.136***	-0.113***	-0.021	-0.024	-0.024	-0.028*
	(0.013)	(0.013)	(0.015)	(0.012)	(0.013)	(0.013)	(0.013)
Collective-share	-0.058***	-0.056***	-0.058***	-0.005	-0.008	-0.008	-0.010
	(0.010)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)
Foreign-share	0.086***	0.072***	0.094***	0.031	0.022	0.022	0.017
	(0.021)	(0.018)	(0.020)	(0.025)	(0.025)	(0.025)	(0.025)
Observation	229259	225349	217144	214723	210740	210736	206660
Group	6082	20158	35992	57289	59036	59035	59113
Under-identification test	578.990	1013.610	738.845	1058.622	1020.631	1020.624	1004.328
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	3.0e+04	1.4e+04	1.1e+04	7956.428	7949.028	7948.745	8158.926
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	2.637	2.850	0.129	0.768	0.723	0.723	0.692
P value	(0.104)	(0.091)	(0.720)	(0.381)	(0.395)	(0.395)	(0.406)
First-stage Results:							
Ming30km	0.209***						
	(0.006)						
LC30km	0.284***	0.342***	0.357***				
	(0.007)	(0.009)	(0.012)				
Qing50km		0.124***	0.170***	0.448***	0.448***	0.448***	0.452***
		(0.007)	(0.010)	(0.014)	(0.014)	(0.014)	(0.014)
SL30km				0.582***	0.568***	0.568***	0.557***
				(0.022)	(0.021)	(0.021)	(0.021)
F	578.768	324.180	294.587	805.631	752.206	752.208	680.281
R2	0.317	0.199	0.197	0.291	0.287	0.287	0.281
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B14: Inventory as Control Variable.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.025***	-0.033***	-0.021***	-0.032***	-0.026***	-0.026***	-0.021***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Inventory	-0.002	0.001	0.002	0.008***	0.009***	0.009***	0.010***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Rail-Density	0.021	0.031*	0.037***	0.017*	0.015*	0.015*	0.010
	(0.018)	(0.012)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
Road-Density	0.009	0.014	0.002	0.016**	0.014*	0.014*	0.012*
	(0.017)	(0.012)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
River-Density	-0.012	-0.032**	-0.021*	-0.026***	-0.021***	-0.021***	-0.021***
	(0.013)	(0.010)	(0.009)	(0.005)	(0.006)	(0.006)	(0.006)
Pop	0.080***	-0.046***	-0.028*	-0.039***	-0.041***	-0.041***	-0.044***
	(0.006)	(0.014)	(0.013)	(0.006)	(0.006)	(0.006)	(0.006)
Age	-0.006***	-0.005***	-0.005***	-0.002***	-0.001***	-0.001***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	-0.017**	-0.013***	-0.005	0.031***	0.033***	0.033***	0.033***
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Size	0.161***	0.156***	0.154***	0.162***	0.162***	0.162***	0.160***
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.571***	-0.502***	-0.410***	-0.066***	-0.053***	-0.054***	-0.058***
	(0.011)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Collective-share	0.048***	0.031***	0.015***	-0.012***	-0.008*	-0.008*	-0.008*
	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Foreign-share	-0.047***	-0.028***	-0.020***	0.017*	0.019**	0.019**	0.016*
	(0.010)	(0.006)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
Observation	1316835	1310538	1288320	1216612	1189551	1189317	1174956
Group	9471	44921	109560	303835	316070	316049	318257
Under-identification test	763.222	992.241	1492.758	4697.547	4537.349	4521.085	4334.216
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5.3e+04	3.7e+04	2.2e+04	2.6e+04	2.5e+04	2.5e+04	2.5e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.268	0.217	1.146	1.584	0.886	0.945	0.370
P value	(0.605)	(0.217)	(0.284)	(0.208)	(0.347)	(0.331)	(0.543)
First-stage Results:							
Ming30km	0.216***						
	(0.008)						
Qing30km	0.056***						
	(0.008)						
SL30km		0.086***					
		(0.007)					
LC30km		0.221***	0.224***	0.280***	0.282***	0.282***	0.290***
		(0.008)	(0.006)	(0.004)	(0.004)	(0.004)	(0.005)
Qing50km			0.126***	0.152***	0.157***	0.156***	0.172***
			(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
F	658.592	552.070	562.853	1940.714	1811.446	1810.459	1722.082
R2	0.156	0.124	0.119	0.162	0.155	0.155	0.153
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B15: The Impacts on Outsourcing.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Outsourcing	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.006***	-0.009***	-0.004	-0.007***	-0.009***	-0.009***	-0.009***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Rail-Density	0.014***	0.012***	0.011***	-0.002	-0.002	-0.002	-0.000
	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Road-Density	-0.010**	-0.012***	-0.012***	-0.010***	-0.012***	-0.012***	-0.011***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
River-Density	0.020***	0.023***	0.025***	0.029***	0.027***	0.027***	0.026***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
Pop	-0.003	-0.000	-0.001	-0.000	0.001	0.001	0.003
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Age	0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	-0.003	-0.002	-0.004**	-0.010***	-0.010***	-0.010***	-0.010***
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Size	0.002***	0.002***	0.003***	0.010***	0.010***	0.010***	0.010***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
State-share	-0.006*	-0.004	-0.004	0.000	0.001	0.001	0.002
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Collective-share	-0.000	-0.003*	-0.003*	0.002	0.003	0.003	0.003
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Foreign-share	-0.022***	-0.015***	-0.017***	0.006	0.006	0.006	0.006
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
Observation	1366418	1360228	1338105	1266582	1238883	1238636	1224100
Group	9506	45504	111767	314048	326898	326873	329221
Under-identification test	536.768	1550.466	1178.550	4348.172	4202.267	4187.265	3996.247
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.4e+04	2.8e+04	2.1e+04	2.3e+04	2.2e+04	2.2e+04	2.2e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	3.452	4.066	0.227	0.081	0.013	0.009	0.398
P value	(0.063)	(0.131)	(0.634)	(0.775)	(0.908)	(0.925)	(0.528)
First-stage Results:							
Qing50km	0.078***	0.120***					
	(0.004)	(0.005)					
SL30km	0.240***						
	(0.005)						
SL50km		0.042***	0.089***	0.061***	0.057***	0.056***	0.076***
		(0.007)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)
LC30km		0.224***	0.199***	0.275***	0.281***	0.281***	0.278***
		(0.008)	(0.008)	(0.005)	(0.005)	(0.005)	(0.005)
F	807.336	575.161	586.597	2061.829	1917.498	1917.106	1818.202
R2	0.158	0.129	0.116	0.155	0.147	0.147	0.145
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B16: The Impacts on Outsourcing (Unchanged Address).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Outsourcing	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	0.001	-0.002	-0.003	-0.022***	-0.021***	-0.021***	-0.021***
	(0.002)	(0.002)	(0.003)	(0.005)	(0.005)	(0.005)	(0.005)
Rail-Density	-0.017**	-0.019***	-0.012*	-0.016***	-0.015***	-0.014***	-0.012**
	(0.006)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)
Road-Density	-0.002	-0.002	0.001	0.003	0.004	0.004	0.005
	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
River-Density	0.007	0.006	0.004	0.004	0.001	0.001	0.002
	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.006)
Pop	0.000	-0.001	-0.003	0.006	0.004	0.004	0.004
	(0.003)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Age	-0.000*	-0.000**	-0.000*	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	0.002	0.002	0.001	-0.002	-0.005	-0.005	-0.006
	(0.002)	(0.002)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
Size	0.004***	0.003***	0.003**	0.006**	0.006**	0.006**	0.006*
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.004	-0.004	-0.004	-0.002	-0.000	-0.000	0.001
	(0.003)	(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.006)
Collective-share	0.000	-0.002	-0.000	0.001	0.001	0.001	0.002
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Foreign-share	-0.022***	-0.013**	-0.013**	-0.003	-0.002	-0.002	0.001
	(0.004)	(0.005)	(0.005)	(0.008)	(0.009)	(0.009)	(0.009)
Observation	234199	230299	222106	219922	215867	215863	211758
Group	6108	20322	36397	58268	60051	60050	60136
Under-identification test	449.271	1027.736	626.377	1482.094	1546.317	1546.303	1522.351
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.0000)	(0.000)	(0.000)
Weak identification test	1.8e+04	9839.683	1.0e+04	9626.691	9474.999	9474.591	9499.398
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.0000)	(0.000)	(0.000)
Over-identification test	0.013	4.748	4.808	1.839	1.275	1.274	1.404
P value	(0.908)	(0.093)	(0.028)	(0.175)	(0.259)	(0.259)	(0.236)
First-stage Results:							
Qing50km	0.084***	0.119***					
	(0.006)	(0.008)					
SL30km	0.305***						
	(0.006)						
SL50km		0.043***	0.049***	0.064***	0.059**	0.059**	0.061***
		(0.011)	(0.014)	(0.019)	(0.019)	(0.019)	(0.018)
LC30km		0.315***	0.370***	0.638***	0.636***	0.636***	0.627***
		(0.012)	(0.015)	(0.011)	(0.011)	(0.011)	(0.010)
F	471.253	312.262	283.564	969.167	918.328	918.332	853.680
R2	0.252	0.199	0.186	0.302	0.297	0.297	0.290
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B17: The Regression with Outsourcing Share as Control Variable.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.026***	-0.036***	-0.021***	-0.032***	-0.027***	-0.027***	-0.022***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
Out	-0.107***	-0.105***	-0.090***	-0.052***	-0.051***	-0.051***	-0.051***
	(0.005)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Rail-Density	0.020	0.029*	0.034***	0.014*	0.014*	0.014*	0.009
	(0.018)	(0.012)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
Road-Density	0.012	0.016	0.005	0.022***	0.019***	0.019***	0.018***
	(0.017)	(0.012)	(0.009)	(0.005)	(0.005)	(0.005)	(0.006)
River-Density	-0.013	-0.033**	-0.024**	-0.030***	-0.024***	-0.024***	-0.024***
	(0.013)	(0.011)	(0.009)	(0.005)	(0.006)	(0.006)	(0.006)
Pop	0.078***	-0.050***	-0.033*	-0.042***	-0.042***	-0.043***	-0.046***
	(0.006)	(0.013)	(0.013)	(0.006)	(0.006)	(0.006)	(0.006)
Age	-0.005***	-0.005***	-0.004***	-0.002***	-0.001***	-0.001***	-0.002***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	-0.019***	-0.015***	-0.006*	0.032***	0.034***	0.034***	0.033***
	(0.006)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Size	0.158***	0.155***	0.153***	0.165***	0.165***	0.165***	0.163***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.566***	-0.497***	-0.408***	-0.068***	-0.055***	-0.055***	-0.059***
	(0.012)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
Collective-share	0.049***	0.031***	0.015***	-0.010**	-0.006	-0.006	-0.006
	(0.007)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Foreign-share	-0.052***	-0.033***	-0.024***	0.015*	0.018**	0.017*	0.015*
	(0.010)	(0.006)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
Observation	1366418	1360228	1338105	1266582	1238883	1238636	1224100
Group	9506	45504	111767	314048	326898	326873	329221
Under-identification test	763.228	1012.002	1480.664	4814.339	4641.439	4623.897	4430.648
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5.4e+04	3.8e+04	2.3e+04	2.7e+04	2.6e+04	2.6e+04	2.6e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.183	1.525	1.780	1.494	1.051	1.122	0.371
P value	(0.669)	(0.217)	(0.182)	(0.222)	(0.305)	(0.289)	(0.543)
First-stage Results:							
Ming30km	0.216***						
	(0.008)						
Qing30km	0.057***						
	(0.008)						
SL30km		0.087***					
		(0.007)					
LC30km		0.220***	0.222***	0.278***	0.280***	0.279***	0.287***
		(0.008)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
Qing50km			0.127***	0.155***	0.159***	0.159***	0.174***
			(0.005)	(0.004)	(0.005)	(0.005)	(0.005)
F	666.069	554.708	569.416	2003.864	1870.108	1869.376	1779.770
R2	0.155	0.123	0.118	0.161	0.154	0.154	0.152
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B18: The Correlation of Inventory and Outsourcing with Firm Size.

Dependent Variable:	Employment Size							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
Dist	-0.000	-0.000	-0.000	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
lnventory	0.044***	0.041***	0.041***	0.039***				
	(0.001)	(0.001)	(0.001)	(0.001)				
Out					-0.012***	-0.009***	-0.009***	-0.009***
					(0.002)	(0.002)	(0.002)	(0.002)
Rail-Density	0.075***	0.048***	0.048***	0.041***	0.060***	0.034***	0.034***	0.028***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Road-Density	-0.047***	-0.038***	-0.038***	-0.037***	-0.053***	-0.043***	-0.043***	-0.040***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Pop	0.069***	0.049***	0.049***	0.048***	0.071***	0.052***	0.052***	0.051***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Age	0.000	-0.000	-0.000	-0.000	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	0.103***	0.093***	0.093***	0.090***	0.098***	0.088***	0.088***	0.085***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Size	0.308***	0.289***	0.288***	0.283***	0.330***	0.308***	0.308***	0.302***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	0.070***	0.066***	0.066***	0.064***	0.074***	0.071***	0.071***	0.067***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Collective-share	0.004	0.000	-0.000	-0.001	0.008**	0.004	0.004	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Foreign-share	0.025***	0.020***	0.020***	0.018***	0.029***	0.024***	0.024***	0.022***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
R2	0.145	0.127	0.127	0.123	0.138	0.121	0.121	0.117
Observation	1479200	1461735	1461735	1461735	1529583	1511700	1511700	1511700
Group	458914	526920	527131	545396	469162	539247	539465	557962
F	2667.302	2232.192	2233.570	2148.188	2578.891	2202.011	2203.706	2125.104
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B19: The Heterogeneous Impacts across Coastal and Inland Regions.

Cluster	Coast				Inland			
Dependent Variable: TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
Dist	-0.061***	-0.050***	-0.051***	-0.049***	-0.011	-0.011	-0.010	-0.002
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Rail-Density	0.047***	0.040***	0.040***	0.038***	-0.136***	-0.149***	-0.149***	-0.149***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.020)	(0.020)	(0.020)	(0.020)
Road-Density	-0.005	-0.013*	-0.012*	-0.016*	0.159***	0.161***	0.161***	0.167***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.014)	(0.014)	(0.014)	(0.014)
River-Density	-0.054***	-0.034***	-0.034***	-0.037***	0.072***	0.068***	0.068***	0.060***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.015)	(0.015)	(0.015)	(0.015)
Pop	-0.052***	-0.052***	-0.052***	-0.055***	-0.040*	-0.035	-0.038	-0.063*
	(0.006)	(0.006)	(0.006)	(0.006)	(0.019)	(0.019)	(0.019)	(0.025)
Age	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***	-0.001*	-0.001*	-0.002**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Export	0.022***	0.024***	0.024***	0.024***	0.088***	0.088***	0.088***	0.086***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.006)
Size	0.163***	0.163***	0.163***	0.162***	0.169***	0.169***	0.169***	0.168***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.004)	(0.005)	(0.005)	(0.005)
State-share	-0.042***	-0.029***	-0.029**	-0.032***	-0.100***	-0.088***	-0.089***	-0.093***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)	(0.012)
Collective-share	-0.002	0.001	0.001	0.002	-0.028***	-0.023**	-0.023**	-0.023**
	(0.004)	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.007)	(0.007)
Foreign-share	0.008	0.012	0.011	0.010	0.068**	0.062**	0.061**	0.053*
	(0.007)	(0.007)	(0.007)	(0.007)	(0.022)	(0.022)	(0.022)	(0.022)
Observation	951640	930076	929957	920367	317722	311525	311397	306400
Group	234105	245248	245239	246948	80470	82223	82207	82869
Under-identification test	3807.559	3616.240	3595.945	3567.282	3010.599	3078.295	3070.481	2955.723
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	2.2e+04	2.1e+04	2.0e+04	2.1e+04	1.7e+04	1.7e+04	1.7e+04	1.7e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.501	0.464	0.422	0.113	0.009	0.006	0.006	0.002
P value	(0.479)	(0.496)	(0.516)	(0.737)	(0.926)	(0.937)	(0.938)	(0.969)
First-stage Results:								
Ming30km	0.278***	0.281***	0.280***	0.282***	0.365***	0.352***	0.351***	0.362***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
Qing30km	0.022**	0.020**	0.020**	0.039***		0.046***	0.046***	0.052***
	(0.007)	(0.007)	(0.007)	(0.007)		(0.009)	(0.009)	(0.009)
Qing50km					0.000			
					(0.008)			
F	1562.005	1433.966	1431.917	1363.196	711.962	709.509	710.644	707.461
R2	0.166	0.158	0.158	0.156	0.228	0.231	0.230	0.235
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table B20: The Heterogeneous Impacts across High or Low Value-Weight Ratio Products.

Cluster	High Value-weight Ratio				Low Value-weight Ratio			
Dependent Variable:								
TFP	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
Dist	-0.033***	-0.025***	-0.025***	-0.022***	-0.032***	-0.031***	-0.032***	-0.028***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Rail-Density	0.022*	0.025**	0.025**	0.021*	-0.036**	-0.044***	-0.044***	-0.049***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)	(0.012)
Road-Density	0.017	0.011	0.011	0.007	0.035***	0.032***	0.033***	0.032***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)
River-Density	-0.018*	-0.029**	-0.029**	-0.025**	-0.025**	-0.014	-0.014	-0.013
	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)
Pop	-0.038***	-0.030***	-0.031***	-0.035***	-0.057***	-0.061***	-0.060***	-0.063***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)
Age	-0.002***	-0.001**	-0.001**	-0.002**	-0.002***	-0.001**	-0.001**	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	0.024***	0.025***	0.025***	0.023***	0.040***	0.041***	0.041***	0.041***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Size	0.173***	0.171***	0.170***	0.169***	0.162***	0.162***	0.162***	0.161***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
State-share	-0.064***	-0.056***	-0.056***	-0.063***	-0.059***	-0.054***	-0.055***	-0.058***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Collective-share	-0.010	-0.004	-0.004	-0.005	-0.003	-0.004	-0.004	-0.003
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Foreign-share	0.011	0.017	0.016	0.014	0.014	0.014	0.013	0.012
	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)	(0.009)	(0.010)
Observation	549372	538294	538219	530045	688082	680661	680492	674308
Group	149628	154139	154137	154426	185767	186815	186785	187510
Under-identification test	3171.266	3160.152	3142.789	3030.982	3124.326	3069.761	3058.966	3069.670
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.6e+04	1.6e+04	1.6e+04	1.7e+04	2.1e+04	2.0e+04	2.0e+04	2.1e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.000	0.037	0.019	0.508	0.146	0.053	0.030	0.010
P value	(0.984)	(0.847)	(0.890)	(0.476)	(0.703)	(0.818)	(0.863)	(0.919)
First-stage Results:								
Ming30km	0.309***	0.300***	0.300***	0.305***	0.337***	0.321***	0.320***	0.326***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Qing50km	-0.005				-0.013			
	(0.007)				(0.009)			
Qing30km		0.034***	0.033***	0.048***		0.035***	0.036***	0.048***
		(0.008)	(0.008)	(0.008)		(0.009)	(0.009)	(0.009)
F	964.272	910.076	909.986	884.247	1056.584	1036.531	1034.669	1025.033
R2	0.175	0.168	0.168	0.170	0.162	0.162	0.161	0.161
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

.3 Appendix C

.3.1 Determinants of Productivity

Productivity is the measurement of production efficiency, it is the part of economic output that cannot be explained by inputs differences. The determinants of productivity include the quality of capital inputs and labor inputs (human capital), management skills, and intangible capital such as patents and experiences of managers and workers. At the same time, competition may play its role outside firms, by promoting the exit of low-productivity firms and the entry of high-productivity firms, which is also called 'between effects'.

Input Heterogeneity

The quality of capital inputs and labor inputs are not captured by the standard productivity function, i.e., if some firms have more high-skilled workers or more efficient machines, the input heterogeneity may affect productivity differences. Here there is a debate called 'technological explanations of productivity', i.e., whether the technology process is embodied in physical capital, or human capital (Syverson 2011). If we assume that new technology is only embodied in new machines and new plants, we will find that productivity growth is mainly promoted by construction of new plants, and firms' entry and exit (Foster et al. 2001). It seems that this assumption is not easy to be satisfied in practice, but many productivity dispersion literatures tend to support this prediction (Dunne et al. 1988, Gibbons et al. 2016, Ding, Jiang & Sun 2016).

Van Biesebroeck (2003) assumes firms learn new technology through the transformation of earlier vintage plants. The results of that study support not only capital renovation but also that firms' entry and exit also promote labor productivity growth, consistent with much productivity dispersion literature. The difference in these branches of studies usually use their own approaches to indicate capital inputs, rather than single currency-valued indicators, because they want to estimate how much 'technologies' are embodied in new investments. For example, Sakellaris & Wilson (2004) focus on U.S. firms, use cumulated past investments, weighted by capital depreciation and technological process, as the indicator of capital stocks. Their results show that capital efficiency grows between $8\% - 17\%$ per year from 1972 to 1996. If these estimates are correct, technology-embodied indicators can account for two-thirds of changes in total factor productivity estimated by standard production function. Another study, Cummins & Violante (2002), uses a different approach, finding a smaller value (around 5% per year), but these results generally support that capital qualities play an important role in productivity growth, and these qualities failed to be captured by currency-valued indicators.

Besides capital quality, quality of labor inputs (human capital) is also widely discussed in productivity literature, e.g., the knowledge of workers and technical staff, and the experience of managers. Many studies try to evaluate labor quality from multiple dimensions, e.g., education, working experiences, or specialized training, and these labor quality indicators are normally shown to have significant influence on productivity growth (Ilmakunnas et al. 2004, Fox & Smeets 2011) or outputs growths (Moretti 2004). However, even these labor quality indicators are under controlled, there are still large amounts of productivity dispersion which fail to be explained (Ilmakunnas et al. 2004, Galindo-Rueda & Haskel 2005), and some studies also support the view that industries such as high-tech sectors benefit more from human capital measured by educational level because these industries reply more on intangible input (Moretti 2004), which highlights the importance of heterogeneity issues in productivity growth.

A more likely explanation is that technology is embodied in both capital and labor input, which highlights the influences of firms' entry and exit, and educational level of employees in productivity estimates. However, the fraction of these two mechanisms may vary across different industries, which emphasizes the importance of industry heterogeneity.

Management Structure and Ownership

The quality of mangers' practices to coordinate the inputs of capital, labor, and intermediate inputs also plays an important role. However, managerial inputs are too abstract, and not included in most datasets. To overcome these data constraints, Bloom & Van Reenen (2007), Bloom, Genakos, Sadun & Van Reenen (2009), Bloom & Van Reenen (2010) use management scores to indicate management quality, which include firms' day-to-day operations such as target making and monitoring. The results show a positive relationship between productivity and management score, while this positive relationship is weaker in fast-growing countries, suggesting some underlying factor such as market distortion in emerging countries. At the same time, scores are mainly affected by market competition and whether a company is family-owned company, highlighting the importance of ownership-decided management heterogeneity.

Management structure is another influential factor which affects firms' decision-making efficiency when facing unexpected conditions, e.g., vertically integrated firms have higher management efficiency, even if they are associated with higher management cost (Forbes & Lederman 2010); scale economy and scope economy are also proved to have positive effects on firms' productivity (Bertrand & Schoar 2003, Bernard et al. 2010).

Firms' quantity of plants is not only important in management studies, but also in economic geography literature. For management studies, Hortaçsu & Syverson (2007) find that the interrelationship between plants within a same company might be closer than those plants coming from different companies. However, the existing China' firm-level data set does not provide plant-level information, which means the cross-plant relationship cannot be captured. Further more, geography economy studies emphasize the importance of firms' location effects, which requires firm's unique address. To solve these problems, it is reasonable to control firms' quantity of plants, i.e., those firms with only one plant are more likely to be independent from the influence of heterogeneous management.

Competition and Trade

In competition studies, the influences of trade exposure and openness are normally considered separately. Trade competition can affect firms' productivity through within and between approaches (Pavcnik 2000), motivated by cheaper products or intermediate goods provided by foreign producers (Amiti & Konings 2007). When trade exposure increases, domestic firms have to face fiercer competition, which can promote intra-industry resource reallocation and increase of general welfare (Melitz 2003). In this process, similar to that of competition effects, high-efficiency companies tend to enter the export market while low-efficiency firms tend to leave the export market and stay in the domestic market, because most export sales and profits are gained by high-productivity producers (Melitz 2003). The increase of trade exposure can be affected by policy changes or improvement of transport infrastructure (Ding, Jiang & Sun 2016), but the cost reduction effects on China's international trade have not been fully understood.

However, sometimes high level of trade competition can also cause low-tech industries to shrink and exit (Bloom, Draca & Van Reenen 2009), because when an industry faces fierce foreign competition, producers may have multiple choices, e.g., they can increase their product quality or increase their efficiency. If most producers could not compete with foreign producers, this industry will tend to shrink, suggesting the potential bad influence of trade openness (Syverson 2011).

When a country relaxes its trade constrictions or improves its transportation system, it may also encourage more firms to decide to export. Syverson (2011) summarized that most studies tend to conclude that export firms have higher productivity, but export decisions and productivity seem to be a pair of endogenous variables. On the one hand, the higher productivity of exporters may reflect 'selection effects'; on the other hand, earlier studies also show that many exporters witnessed productivity growth after they entered the export market, i.e., 'learning-by-exporting' (Van Biesebroeck 2005, De Loecker 2007*a*). The latter suggests that when firms try to solve problems in oversea markets with various types of business and institutional environments, they can learn how to operate more efficiently from their experience. Arrow (1971) point out that the learning process can induce knowledge accumulation, while knowledge acquisition plays an important role in productivity increase. These two types of influences are both observed by Ding, Jiang & Sun (2016). They find that the development level of transport infrastructure, indicated by the density of highways, railways, and so on, has both a positive influence on companies' export decisions and their export volume, while the impact on export decisions is larger than on export volume. However, current studies rarely use geographical datasets. This project also investigates the simultaneous effects of infrastructure improvement.

Intangible Input

Another branch of studies focus on technology processes as embodied in new products, production processes, and innovation patents (Foster et al. 2001), e.g., technology, management structure, distribution channel, or information for consumers' tastes. These studies use R&D expenditures (Research & Development) as a key indicator of technology inputs, because technology processes could be the consequence of R&D inputs.

Doraszelski & Jaumandreu (2013) focus on Spanish firms, and find positive effects of R&D expenditures on productivity, but the outcomes of R&D vary a lot for different firms. Aw et al. (2008) find that those firms in Taiwan with high R&D expenditures also tend to be export firms, while export firms usually have higher productivity than domestic firms. Jefferson et al. (2006) find large and capital-intensive firms in China tend to make more R&D investments, especially for those state-owned enterprises (SOEs), because of marginal profits differences from DFs to FIEs (foreign invested enterprises). These findings suggest although existing studies generally support R&D effects on productivity, heterogeneity problems across firms with different industries/ownership still make large differences.

Besides R&D expenditures, workers' experience seems to increase over time, which can also increase work efficiency, i.e., learning-by-doing process, determined by workers' 'learning rate' (how fast they can become familiar with new machines or processes) and 'forgetting rate' (depreciation of this knowledge and experience) (Benkard 1999).

For the aircraft manufacturing industry, Benkard uses a 'learning curve' to illustrate the decrease of the minimum required working hours for each new produced plane. Similar to the shapes of cost functions, this learning curve decreases very fast at the beginning, e.g., for a new type of aircraft, the required working hours are cut in half when the 30th plane is produced, and halved again when the 100th one is produced; after the 130th plane, the required working hours begin to increase. Benkard argues that when an aircraft is substituted by a new one, the accumulated changes embodied in this model are large enough to make earlier work experiences outdated. Similarly, for the shipbuilding industry, Thornton & Thompson (2001) find that during World War II, if a shipyard changes its ship design, its production efficiency would decrease, if the design comes from another shipyard, the decrease would be larger. Similarly, Levitt et al. (2011) also find consistent results in the automobile manufacturing industry. When a plant is involved in the production of a new model, it begins a new learning curve, the production efficiency will increase gradually along the learning curve. However, these branches of studies usually focus on monopolistic competitive industries, using labor productivity rather than total factor productivity. It suggests those industries that rely on automated production lines might benefit more from learning-by-doing process. The effects of working experiences could be different in those industries with high updating speed of their products, or in traditional industries which hardly change their products.

Last but not least, there are also some studies that pay attention to IT applications of less than 20 years. These branches of studies are characterized by closely watching any new process in high-tech industries and the application of information technology.

Existing literature (Jorgenson et al. 2008, Oliner et al. 2008) tends to find that labor productivity for U.S. companies from 1995 to 2000 increased faster than in the next decade or early 1990s, which seems to be motivated by investments in intangible assets, especially for IT investments, in the late 1990s. With the tech bubble bursting in 2001, private productivity growth in the U.S. as well as IT investment both decreased. After 2000, productivity growth in the U.S. mainly relied on cost cutting and industry restructuring (Oliner et al. 2008). Similarly, cross-country comparisons between America and the European Union during the same period also supports the view that smaller size of IT investment in Europe than the U.S. lead to lower productivity growth (Van Ark et al. 2008). However, the benefits from IT shifts vary from countries with different legal restrictions, e.g., employment protection (Bartelsman et al. 2011), implying the importance of underlying factors, e.g., institution effects.

Knowledge Spillover

In general, spillover effects indicate the spillover of intangible inputs such as patents, specialized knowledge, producers' strategies, or even skilled labors; they can happen between different plants within same firms, or among different firms (Syverson 2011). The consequence of productivity spillover would lead to productivities of all producers converging to a same level in the long term, when productivity growth of those market leaders is slower than the speed of productivity spillovers.

According to the finding of Bartelsman et al. (2008), domestic productivity convergences are faster than cross-country convergences. Similarly, Hu et al. (2005) demonstrate that foreign invested enterprises in China are usually isolated from other creative activities held by domestic companies. However, most productivity spillover literature has to solve the 'reflection problem', i.e., how to correctly measure 'knowledge spillovers'. According to the definition of 'spillovers', technology spillovers will be more significant for those firms within relatively shorter distances (geographical distance, technological distance). Bloom et al. (2013) use patents to indicate intangible input, and find a positive effect. Another branch of studies, such as Crespi et al. (2008), use a more detailed dataset, which allow them to directly trace 'information flow'. They use U.K. CIS dataset, which contains detailed information of knowledge and patent cations in firm-level innovation. Their results show that most information flow are embodied in patents, which can explain 50% productivity growth.

These findings suggest that domestic spillover might be the main channel in knowledge spillover, while knowledge spillover seems to have various speeds between firms with different ownerships, scales, or in different industries. In our research, geographical distance is considered as an essential channel for knowledge spillover, under the control of these firm-level effects.

.3.2 Biases in Estimation

Endogeneity Bias and Simultaneity Bias: Endogeneity problems are very common in productivity estimates, caused by correlation between productivity shocks and inputs usage, i.e., $E(x_{it}, \epsilon_{it}) \neq 0$, where $x_{it} = (k_{it}, l_{it})$; $\epsilon_{it} = (\omega_{it}, I_{it}, e_{it})$ (De Loecker 2007*b*). Productivity estimated by OLS approach is very likely to suffer from endogeneity problems, e.g., endogeneity can be caused by positive or negative correlation between inputs usage x_{it} and prices of raw materials, market competition in different regions, and firm-specific issues such as management quality and multi-product firms.

Selection Bias and Survivorship Bias: Standard production functions do not consider the roles of firms' entry and exit in productivity growth, while many empirical studies agree that market selection plays an important role, while large enterprises have higher productivity and survival rates (Van Beveren 2012, Dunne et al. 1988). These facts highlight the importance of resource allocation and selection effects. So it is necessary to consider firms' entry and exit in productivity research, otherwise the econometric results may just reflect the characteristics of "selected" firms. At the same time, Survivorship bias may also cause endogeneity bias, if there is any market selection factor correlated with input usage, e.g., if markets select high-productivity firms, while higher-productivity firms tend to have larger scales, productivity will be endogenous of input usage due to the positive correlation between input scale x_{it} and productive $(\alpha_0 + \epsilon_{it})$.

The Influences of Input Prices and Output Prices: The estimates of firm-level productivity can be based on revenue or physical-quantity, but many studies, due to data limitation, just use revenue-based productivity or use industry-level price to deflate prices fluctuation (Foster et al. 2008, Van Beveren 2012). However, applying these two types of estimate approaches may make great differences, e.g., physical productivity show more significant and persistent dispersion than revenue productivity; for a market allowing firms' entry and exit, entrants tend to have higher physical productivity than revenue productivity (Foster et al. 2008). Price factors are also very likely to induce the endogeneity problem, if firms tend to adjust their input usage when input and output prices rise or fall (Van Beveren 2012). This study only estimates revenue-based productivity, so price effect details should be taken into account.

Firms' Heterogeneity and Multi-product Firms: Homogeneity assumption means all producers face identical demand, and produce products with similar technology. In practice, many firms might be multi-product producers, they produce goods in the same industry but these goods are different enough and fail to be substitutable products, this is also called the aggregate problem (Bernard et al. 2009). This problem would lead two consequences; first, causing difficulty in defining industries of multi producers, unless we have a plant-level dataset; second, if producers' heterogeneity is endogenous of input usage, e.g., some high-tech products may require more intangible inputs, we need efficient instruments to eliminate such kind of endogeneity.

This study uses a semi-parametric approach, which can largely solve endogeneity bias; price changes are eliminated by price deflators; firms' heterogeneity is partly solved by two-digit cross-industry estimates in the first stage and four-digit industry fixed effects in the second stage; selection effects on surviving firms are currently neglected because this study mainly focuses on existing firms.

.3.3 Productivity Estimation Results

Two-factor Production Function

Productivity across 29 two-digit manufacturing industries (non-resource industries) are estimated by four different approaches, i.e., OP, LP, ACF, and Wooldridge GMM approaches. [7] [8] For the same industry, different approaches have different sample scales due to various indicators used in regressions. OP approach tends to report the least number of observations, because OP approach uses investment as the proxy for productivity shocks. According to perpetual inventory approach, investment is defined as the sum of the changes of capital stocks and capital depreciation, while capital stocks may not always increase over time. If there is a negative investment in any year, these observations will be removed in log-transformed production function and cause selection biases. By contrast, LP approach is less likely to suffer from selection biases because it uses intermediate input as the proxy of productivity shocks. However, LP approach is more likely to suffer from multi-collinearity problem than OP approach (Van Beveren 2012), because of the limitation of its proxy (intermediate input). In order to check how serious multi-collinearity problem is, we also add acf-correct LP results for comparison.

Table C2 compares the results of LP approach (left) and acf-correct LP approach (right), with fourth-order polynomial regressions and 200 bootstrap repetitions for each regression. Column 'L' includes the coefficients of labor input, column 'K' contains the coefficients of capital input; column 't(L)' includes the corresponding t-statistics of labor input coefficients, column 't(K)' indicates the corresponding t-statistics for the coefficients of capital input. The last three columns calculate the difference between the results estimated by LP and ACF approach respectively, i.e., difference of coefficient 'L', coefficient 'K' and industry-level productivity. Some values are removed because the production functions for some industries failed to be estimated (insignificant coefficients), this could be caused by small sample scales, or suggest that these industries have non-linear production functions, but the discussion on the shapes of those abnormal production functions is beyond the scope of this study. The differences between the coefficients of capital input estimated by LP and ACF approaches are not significant, but the differences on labor input are very large, suggesting that LP approach suffers from collinearity problem, and the coefficient of labor input is underestimated.

Table C3 has the same format as Table C2, it compares the results of OP and ACF approach. Productivity estimated by OP approach is slightly larger than ACF approach, because larger and more comprehensive companies tend to report continuous and positive investment series, while large enterprises also tend to have higher productivity due to size effects. OP approach only cover half the amount of observations than ACF approach due to the selection issue; compared with ACF results, more industries in OP results fail to report significant coefficients on capital inputs. However, the estimation gaps between OP and ACF results are much smaller than the gaps between LP and ACF results, the gaps between labor coefficients in Table C3 are also much smaller than in Table C2, suggesting ACF approach can solve multi-collinearity problem more successfully than LP approach.

Table C4 compares the results of LP and Wooldridge GMM approach (Wooldridge 2009), it combines the two-stage estimate into only one step, which is expected to offset the multi-collinearity issue. However, the GMM results are very close to LP results, suggesting the multi-collinearity issue has not been completely solved, so the time-lagged intermediate input is not a suitable instrument. A possible explanation is that, due to Wooldridge GMM

[7] Stata module 'prodest' is applied to estimate productivity (Mollisi & Rovigatti 2017).
[8] The industry definition comes from Chinese Standard Industrial Classification (CSIC).

converting all assumptions from LP approach into moment conditions, the endogeneity of productivity shocks cannot be completely solved. In the two-stage LP approach, the first stage uses polynomial regressions to exhaust the information of productivity shocks; by contrast, Wooldridge GMM is based on LP's assumptions so it has to add a polynomial term in its one-stage estimation, while the polynomial term may over-exhaust productivity shock and lead the input coefficients to be down-ward estimated. In other words, the assumptions behind the moment conditions are designed for a two-stage approach, they cannot hold simultaneously. According to Table C4, the coefficients of 'L' and the return to scale provided by LP and Wooldridge GMM are both very small (around -0.5), suggesting the downward biases of input coefficients.

According to Table C2, C3 and C4, ACF results are largely consistent with OP results but different from LP and Wooldridge GMM results. The consistent results of Wooldridge GMM and LP approach and their small return to scales suggest downward biases of input coefficients. Compared with LP approach, ACF approach can more successfully solve the collinearity issue than LP approach; compared with OP approach, ACF approach does not suffer from selection issue; so ACF approach is the best choice for the given dataset.

Three-factor Production Function

According to productivity estimated by two-factor production function, ACF approach is an ideal choice to estimate productivity for the given dataset. This section tries to make similar comparison across different approaches for three-factor production function, to check whether they can report more significant coefficients or consistent productivity estimates, even if a three-factor estimate may report more insignificant coefficients. Similar to Table C2, C3, and C4, Table C5 and C6 compare LP and ACF results, OP and ACF results. Gaps between OP and ACF results are much smaller than gaps between LP and ACF results, suggesting the pattern revealed in two-factor production function still holds in three-factor production function, i.e., ACF approach can more successfully solve the collinearity issue than LP approach, and does not suffer from the selection issue.

However, there are still obvious differences, e.g., compared with two-factor productivity, productivity estimated by the three-factor function is much smaller, and more industries fail to report statistical significant coefficients (Table C7). One possible explanation is intermediate input is closely related to productivity to productivity shocks because it is more adjustable than capital input and labor input, thus the three-factor approach may overestimated the coefficients of intermediate and induce low productivity estimates. Consequently, two-factor productivity is the better choice than three-factor productivity for the given dataset.

Table C1: ID of 2-digit Industries.

13	Manufacturing of agricultural and non-staple foodstuff industry
14	Foodstuff manufacturing industry
15	Beverage manufacturing industry
16	Tobacco industry
17	Textile industry
18	Manufacturing industry of textile costume, shoes, and caps
19	Manufacturing industry of leather, fur, feather (cloth with soft nap) and their products
20	Wood processing and manufacturing industry of wood, bamboo, rattan, palm, and straw-made article
21	Cabinetmaking industry
22	Papermaking and paper products industry
23	Printing and record medium reproduction industry
24	Manufacturing industry for culture, education and sports goods
25	Petroleum processing and cooking industry
26	Chemical materials and chemical products manufacturing industry
27	Pharmaceutical manufacturing industry
28	Manufacture of chemical fibers industry
29	Rubber product industry
30	Plastic products industry
31	Non-metallic mineral products industry
32	Ferrous metal smelting and rolling processing industry
33	Non-ferrous metal smelting and rolling processing industry
34	Metalwork industry
35	General machinery manufacturing industry
36	Specialized facility manufacturing industry
37	Transport and communication facilities manufacturing industry
39	Electric machinery of communication equipment manufacturing industry
40	Manufacturing industry of communication equipment, computer and other electronic equipment
41	Manufacturing industry of instrument and meters, and machinery for culture and office
42	Artwork and other manufacturing industry

Table C2: Comparison between LP and ACF approaches

L	K	t(L)	t(K)	Num	P-value Wald	Return to Scale	Industry TFP	Industry	L(acf)	K(acf)	t(L,acf)	t(K,acf)	Num	P-value Wald	Return to Scale	Industry TFP	L-L(acf)	K-K(acf)	TFP-acfTFP
0.311	0.252	66.663	39.986	112164	0.000	-0.438	4.172	13	0.650	0.190	77.036	9.784	112164	0.000	-0.161	3.204	-0.339	0.062	0.968
0.269	0.233	37.930	20.977	45426	0.000	-0.499	4.600	14	0.719	0.179	184.854	7.262	45426	0.000	-0.103	2.953	-0.450	0.054	1.647
0.315	0.216	31.011	15.578	31382	0.000	-0.469	4.783	15	0.705	0.260	74.384	6.167	31382	0.341	-0.035	2.506	-0.390	-0.044	2.277
0.242	0.441	5.335	3.182	2281	0.034	-0.316	4.518	16	0.873	-0.112	3.987	-0.348	2281	0.481	-0.239	6.700	-0.631		
0.295	0.207	91.035	44.867	161661	0.000	-0.497	4.940	17	0.538	0.280	95.426	41.860	161661	0.000	-0.182	3.124	-0.243	-0.073	1.816
0.461	0.159	92.802	32.235	92908	0.000	-0.380	4.545	18	0.717	0.137	842.420	24.843	92908	0.000	-0.146	3.382	-0.255	0.021	1.163
0.439	0.165	75.416	21.975	45467	0.000	-0.396	4.686	19	0.687	0.161	31.835	6.011	45467	0.000	-0.153	3.441	-0.247	0.004	1.246
0.318	0.208	42.082	23.152	38183	0.000	-0.473	4.648	20	0.607	0.252	334.988	21.800	38183	0.000	-0.141	2.989	-0.288	-0.043	1.660
0.363	0.176	38.613	16.049	21833	0.000	-0.462	4.876	21	0.743	0.142	29.781	7.880	21833	0.000	-0.115	3.336	-0.380	0.034	1.541
0.258	0.201	40.324	22.013	56508	0.000	-0.541	5.103	22	0.652	0.212	32.355	5.851	56508	0.000	-0.137	3.172	-0.394	-0.011	1.931
0.270	0.181	37.278	16.945	38754	0.000	-0.549	4.969	23	0.595	0.278	125.315	9.242	38754	0.000	-0.127	2.746	-0.325	-0.097	2.223
0.424	0.156	56.451	13.459	25034	0.000	-0.420	4.742	24	0.685	0.152	26.031	8.694	25034	0.000	-0.164	3.450	-0.261	0.004	1.292
0.317	0.286	21.762	10.843	13413	0.000	-0.398	4.502	25	0.562	0.395	12.811	7.583	13413	0.000	-0.044	2.356	-0.245	-0.109	2.147
0.244	0.250	52.790	31.487	140246	0.000	-0.506	4.901	26	0.544	0.287	32.758	14.557	140246	0.000	-0.170	3.244	-0.300	-0.036	1.657
0.240	0.256	25.257	17.967	37699	0.000	-0.504	4.963	27	0.710	0.317	21.267	8.015	37699	0.000	0.028	2.079	-0.470	-0.062	2.884
0.302	0.229	21.213	11.497	10117	0.000	-0.470	4.891	28	0.671	0.254	38.160	8.968	10117	0.000	-0.075	2.866	-0.369	-0.025	2.025
0.280	0.200	26.593	14.384	22528	0.000	-0.520	5.159	29	0.559	0.315	36624.890	18195.020	22528	0.000	-0.125	2.823	-0.279	-0.115	2.335
0.334	0.227	65.917	27.633	88025	0.000	-0.440	4.580	30	0.644	0.239	51.236	12.674	88025	0.000	-0.118	3.083	-0.310	-0.012	1.497
0.283	0.182	83.256	39.565	162089	0.000	-0.535	5.196	31	0.556	0.279	75.866	32.114	162089	0.000	-0.165	3.021	-0.273	-0.097	2.175
0.305	0.243	41.774	20.260	42931	0.000	-0.452	4.742	32	0.674	0.252	25.694	7.577	42931	0.000	-0.074	2.879	-0.369	-0.010	1.863
0.351	0.257	43.361	21.949	35463	0.000	-0.392	4.509	33	0.608	0.227	175.704	10.279	35463	0.000	-0.164	3.584	-0.257	0.029	0.925
0.358	0.204	85.846	29.596	106062	0.000	-0.438	4.683	34	0.613	0.282	667.670	48.140	106062	0.000	-0.105	2.900	-0.255	-0.078	1.784
0.278	0.211	73.507	30.051	141712	0.000	-0.511	5.071	35	0.539	0.306	217651.200	19116.040	141712	0.000	-0.155	3.083	-0.262	-0.095	1.988
0.215	0.227	37.064	28.316	78225	0.000	-0.558	5.209	36	0.507	0.314	588.986	57.053	78225	0.000	-0.178	3.103	-0.292	-0.087	2.106
0.292	0.215	47.773	20.291	84499	0.000	-0.492	5.039	37	0.655	0.234	99.980	5.618	84499	0.002	-0.110	3.096	-0.363	-0.019	1.943
0.369	0.206	71.369	25.992	75674	0.000	-0.425	4.907	39	0.676	0.216	79.109	15.403	75674	0.000	-0.108	3.391	-0.307	-0.009	1.516
0.326	0.221	54.492	22.946	77407	0.000	-0.454	4.998	40	0.646	0.239	278.885	16.375	77407	0.000	-0.115	3.218	-0.320	-0.018	1.780
0.258	0.197	32.057	19.060	37154	0.000	-0.544	5.549	41	0.561	0.235	82.550	5.423	37154	0.000	-0.204	3.717	-0.303	-0.038	1.832
0.386	0.152	53.675	18.842	33996	0.000	-0.463	5.049	42	0.615	0.219		17762.130	33996	0.000	-0.166	3.408	-0.229	-0.068	1.641
Average	**Average**			**Sum**		**Average**	**Average**		**Average**	**Average**			**Sum**		**Average**	**Average**	**Average**	**Average**	**Average**
0.314	0.219			1858841		-0.467	4.846		0.638	0.232			1858841		-0.129	3.202	-0.324	-0.033	1.781

Table C3: Comparison between OP and ACF approache

L	K	t(L)	t(K)	Num	P-value Wald	Return to Scale	Industry TFP	Industry	L(acf)	K(acf)	t(L,acf)	t(K,acf)	Num	P-value Wald	Return to Scale	Industry TFP	L-L(acf)	K-K(acf)	TFP-acfTFP
0.568	0.279	81.065	5.733	49852	0.002	-0.154	2.831	13	0.650	0.190	77.036	9.784	112164	0.000	-0.161	3.244	-0.082	0.089	-0.413
0.600	0.238	49.934	3.426	21232	0.020	-0.162	3.008	14	0.719	0.179	184.854	7.262	45426	0.000	-0.103	3.584	-0.118	0.059	-0.576
0.600	0.235	35.022	3.609	14083	0.013	-0.165	3.240	15	0.705	0.260	74.384	6.167	31382	0.341	-0.035	3.204	-0.105	-0.025	0.036
0.454	0.310	7.081	2.977	1262	0.032	-0.235	4.651	16	0.873	-0.112	3.987	-0.348	2281	0.481	-0.239	2.879	-0.419		
0.479	0.165	117.798	11.023	76921	0.000	-0.357	4.386	17	0.538	0.280	95.426	41.860	161661	0.000	-0.182	2.506	-0.060	-0.115	1.880
0.664	0.137	97.976	9.238	44847	0.000	-0.199	3.661	18	0.717	0.137	842.420	24.843	92908	0.000	-0.146	3.124	-0.053	-0.001	0.537
0.634	0.202	83.623	5.999	21855	0.000	-0.165	3.396	19	0.687	0.161	31.835	6.011	45467	0.000	-0.153	2.953	-0.053	0.041	0.442
0.558	0.201	47.603	3.380	15967	0.000	-0.242	3.613	20	0.607	0.252	334.988	21.800	38183	0.000	-0.141	2.900	-0.049	-0.051	0.714
0.686	0.179	52.666	2.895	10375	0.032	-0.135	3.309	21	0.743	0.142	29.781	7.880	21833	0.000	-0.115	3.218	-0.057	0.038	0.091
0.470	0.246	40.838	6.222	27886	0.000	-0.284	3.739	22	0.652	0.212	32.355	5.851	56508	0.000	-0.137	3.391	-0.182	0.034	0.348
0.492	0.257	38.750	5.219	18464	0.000	-0.251	3.371	23	0.595	0.278	125.315	9.242	38754	0.000	-0.127	3.096	-0.103	-0.021	0.275
0.641	0.118	66.862	1.572	12658	0.002	-0.241	3.935	24	0.685	0.152	26.031	8.694	25034	0.000	-0.164	3.083	-0.044		
0.364	0.262	16.306	4.771	6335	0.000	-0.374	4.492	25	0.562	0.395	12.811	7.583	13413	0.000	-0.044	3.441	-0.198	-0.132	1.052
0.372	0.247	54.068	11.130	69535	0.000	-0.382	4.353	26	0.544	0.287	32.758	14.557	140246	0.000	-0.170	3.083	-0.172	-0.040	1.269
0.517	0.245	34.843	9.921	19994	0.000	-0.239	3.699	27	0.710	0.317	21.267	8.015	37699	0.000	0.028	3.450	-0.193	-0.073	0.249
0.446	0.339	23.480	5.515	4739	0.001	-0.215	3.178	28	0.671	0.254	38.160	8.968	10117	0.000	-0.075	2.746	-0.225	0.085	0.432
0.486	0.227	33.537	4.401	11098	0.000	-0.288	3.926	29	0.559	0.315	36624.890	18195.020	22528	0.000	-0.125	3.627	-0.074	-0.089	0.299
0.517	0.227	81.910	7.940	42776	0.000	-0.256	3.751	30	0.644	0.239	51.236	12.674	88025	0.000	-0.118	3.021	-0.127	-0.012	0.730
0.447	0.182	82.441	1.907	75277	0.000	-0.371	4.389	31	0.556	0.279	75.866	32.114	162089	0.000	-0.165	3.103	-0.109		
0.480	0.204	42.958	2.022	20240	0.002	-0.316	4.234	32	0.674	0.252	25.694	7.577	42931	0.000	-0.074	3.172	-0.194	-0.048	1.063
0.499	0.325	41.431	7.300	16873	0.000	-0.175	3.260	33	0.608	0.227	175.704	10.279	35463	0.000	-0.164	3.408	-0.109	0.098	-0.148
0.553	0.171	84.355	3.745	51740	0.000	-0.277	4.060	34	0.613	0.282	667.670	48.140	106062	0.000	-0.105	3.717	-0.060	-0.111	0.343
0.475	0.229	69.202	5.705	71300	0.000	-0.296	4.008	35	0.539	0.306	217651.200	19116.040	141712	0.000	-0.155	3.336	-0.064	-0.077	0.672
0.446	0.169	47.235	1.671	37841	0.000	-0.385	4.593	36	0.507	0.314	588.986	57.053	78225	0.000	-0.178	2.989	-0.061		
0.532	0.158	60.407	1.588	43568	0.002	-0.309	4.342	37	0.655	0.234	99.980	5.618	84499	0.002	-0.110	3.382	-0.123		
0.589	0.207	88.513	13.148	40937	0.000	-0.205	3.875	39	0.676	0.216	79.109	15.403	75674	0.000	-0.108	2.823	-0.087	-0.009	1.052
0.570	0.178	83.316	2.016	39954	0.005	-0.252	4.120	40	0.646	0.239	278.885	16.375	77407	0.000	-0.115	2.079	-0.075	-0.061	2.041
0.476	0.184	42.573	5.886	18938	0.000	-0.341	4.578	41	0.561	0.235	82.550	5.423	37154	0.000	-0.204	2.866	-0.086	-0.051	1.712
0.564	0.226	62.710	5.330	16546	0.000	-0.210	3.602	42	0.615	0.219		17762.130	33996	0.000	-0.166	2.356		0.007	
Average 0.523	**Average** 0.219			**Sum** 903093		**Average** -0.258	**Average** 3.848		**Average** 0.638	**Average** 0.232			**Sum** 1858841		**Average** -0.129	**Average** 3.096	**Average** -0.117	**Average** -0.019	**Average** 0.613

Table C4: Comparison between LP and Wooldridge GMM approaches

L	K	t(L)	t(K)	Num	P-value Wald	Return to Scale	Industry TFP	Industry	L(gmm)	K(gmm)	t(L,gmm)	t(K,gmm)	Num	P-value Wald	Return to Scale	Industry TFP	L-L(gmm)	K-K(gmm)	TFP-TFP(gmm)
0.311	0.252	66.663	39.986	112164	0.000	-0.438	4.172	13	0.309	0.260	83.774	41.829	112164	0.000	-0.431	4.112	0.002	-0.008	0.059
0.269	0.233	37.930	20.977	45426	0.000	-0.499	4.600	14	0.265	0.242	48.078	25.146	45426	0.000	-0.493	4.539	0.004	-0.010	0.062
0.315	0.216	31.011	15.578	31382	0.000	-0.469	4.783	15	0.311	0.221	48.294	18.748	31382	0.000	-0.468	4.754	0.004	-0.006	0.029
0.242	0.441	5.335	3.182	2281	0.034	-0.316	4.518	16	0.233	0.338	7.697	3.663	2281	0.000	-0.429	5.707	0.010	0.103	-1.189
0.295	0.207	91.035	44.867	161661	0.000	-0.497	4.940	17	0.290	0.209	129.126	50.032	161661	0.000	-0.500	4.949	0.005	-0.002	-0.009
0.461	0.159	92.802	32.235	92908	0.000	-0.380	4.545	18	0.464	0.158	132.483	34.539	92908	0.000	-0.379	4.539	-0.002	0.001	0.006
0.439	0.165	75.416	21.975	45467	0.000	-0.396	4.686	19	0.441	0.165	109.766	24.886	45467	0.000	-0.393	4.672	-0.002	-0.001	0.015
0.318	0.208	42.082	23.152	38183	0.000	-0.473	4.648	20	0.315	0.215	58.432	23.900	38183	0.000	-0.470	4.612	0.003	-0.006	0.036
0.363	0.176	38.613	16.049	21833	0.000	-0.462	4.876	21	0.364	0.183	49.675	17.312	21833	0.000	-0.453	4.814	-0.001	-0.007	0.062
0.258	0.201	40.324	22.013	56508	0.000	-0.541	5.103	22	0.249	0.203	54.636	24.797	56508	0.000	-0.548	5.132	0.009	-0.002	-0.030
0.270	0.181	37.278	16.945	38754	0.000	-0.549	4.969	23	0.260	0.198	46.618	20.794	38754	0.000	-0.542	4.875	0.010	-0.017	0.093
0.424	0.156	56.451	13.459	25034	0.000	-0.420	4.742	24	0.427	0.148	77.203	15.999	25034	0.000	-0.425	4.789	-0.003	0.008	-0.047
0.317	0.286	21.762	10.843	13413	0.000	-0.398	4.502	25	0.307	0.268	32.643	14.948	13413	0.000	-0.425	4.706	0.010	0.017	-0.203
0.244	0.250	52.790	31.487	140246	0.000	-0.506	4.901	26	0.237	0.244	81.192	48.426	140246	0.000	-0.519	4.983	0.007	0.006	-0.082
0.240	0.256	25.257	17.967	37699	0.000	-0.504	4.963	27	0.227	0.265	34.732	25.369	37699	0.000	-0.508	4.945	0.013	-0.009	0.018
0.302	0.229	21.213	11.497	10117	0.000	-0.470	4.891	28	0.295	0.236	29.930	12.009	10117	0.000	-0.469	4.861	0.007	-0.007	0.030
0.280	0.200	26.593	14.384	22528	0.000	-0.520	5.159	29	0.275	0.226	41.844	17.894	22528	0.000	-0.499	4.970	0.005	-0.026	0.189
0.334	0.227	65.917	27.633	88025	0.000	-0.440	4.580	30	0.330	0.222	96.345	35.387	88025	0.000	-0.448	4.633	0.004	0.004	-0.053
0.283	0.182	83.256	39.565	162089	0.000	-0.535	5.196	31	0.275	0.195	118.417	45.728	162089	0.000	-0.531	5.130	0.009	-0.013	0.066
0.305	0.243	41.774	20.260	42931	0.000	-0.452	4.742	32	0.296	0.228	56.268	25.448	42931	0.000	-0.476	4.917	0.009	0.015	-0.175
0.351	0.257	43.361	21.949	35463	0.000	-0.392	4.509	33	0.344	0.249	57.602	22.248	35463	0.000	-0.407	4.608	0.007	0.008	-0.099
0.358	0.204	85.846	29.596	106062	0.000	-0.438	4.683	34	0.354	0.211	111.953	38.260	106062	0.000	-0.435	4.651	0.005	-0.007	0.032
0.278	0.211	73.507	30.051	141712	0.000	-0.511	5.071	35	0.270	0.218	97.105	46.801	141712	0.000	-0.513	5.054	0.008	-0.007	0.017
0.215	0.227	37.064	28.316	78225	0.000	-0.558	5.209	36	0.205	0.243	50.559	32.989	78225	0.000	-0.552	5.127	0.009	-0.015	0.082
0.292	0.215	47.773	20.291	84499	0.000	-0.492	5.039	37	0.283	0.220	70.568	34.314	84499	0.000	-0.497	5.045	0.009	-0.005	-0.006
0.369	0.206	71.369	25.992	75674	0.000	-0.425	4.907	39	0.367	0.198	99.590	33.796	75674	0.000	-0.435	4.983	0.002	0.008	-0.076
0.326	0.221	54.492	22.946	77407	0.000	-0.454	4.998	40	0.324	0.222	80.249	30.188	77407	0.000	-0.454	4.996	0.001	-0.001	0.002
0.258	0.197	32.057	19.060	37154	0.000	-0.544	5.549	41	0.252	0.221	43.977	23.153	37154	0.000	-0.527	5.384	0.006	-0.023	0.165
0.386	0.152	53.675	18.842	33996	0.000	-0.463	5.049	42	0.384	0.161	76.395	21.038	33996	0.000	-0.455	4.982	0.002	-0.010	0.067
Average	**Average**			**Sum**		**Average**	**Average**		**Average**	**Average**			**Sum**		**Average**	**Average**	**Average**	**Average**	**Average**
0.314	0.219			1858841		-0.467	4.846		0.309	0.220			1858841		-0.472	4.878	0.005	0.000	-0.032

Table C5: Comparison between LP and ACF approaches (Three-factor production function)

L	K	M	t(L)	t(K)	t(M)	Num	P-value Wald	Return to Scale	Industry TFP	Industry	L(acf)	K(acf)	M(acf)	t(L,acf)	t(K,acf)	t(M,acf)	Num	P-value Wald	Return to Scale	Industry TFP	L-L(acf)	K-K(acf)	M-M(acf)	TFP-TFP(acf)
0.032	0.061	0.781	10.806	4.808	60.169	118393	0.000	-0.126	1.666	13	0.064	0.041	0.905	62.760	8.505	413.465	118393	0.000	0.010	0.505	-0.032	0.019	-0.123	1.161
0.033	0.064	0.785	6.846	15.049	63.137	47479	0.000	-0.117	1.537	14	0.058	0.042	0.929	44.462	10.546	532.108	47479	0.000	0.028	0.282	-0.025	0.023	-0.143	1.255
0.050	0.073	0.776	8.804	10.486	46.287	32722	0.000	-0.101	1.463	15	0.076	0.053	0.907	35.319	5.337	174.430	32722	0.000	0.036	0.290	-0.027	0.021	-0.131	1.173
0.103	0.313	0.672	3.097	3.666	13.006	2348	0.463	0.088	0.141	16	0.142	0.179	0.831	9.033	3.410	24.332	2348	0.000	0.152	-0.387	-0.038	0.134	-0.159	0.527
0.026	0.051	0.787	13.341	19.649	73.789	164729	0.000	-0.136	1.733	17	0.061	0.008	0.901	23.712	0.798	317.982	164729	0.000	-0.029	0.811	-0.036		-0.113	
0.098	0.058	0.737	29.262	9.667	60.466	93857	0.000	-0.107	1.819	18	0.132	0.021	0.827	78.633	9.594	549.446	93857	0.000	-0.020	1.072	-0.034	0.037	-0.089	0.747
0.079	0.044	0.800	22.387	7.975	51.515	46176	0.000	-0.077	1.437	19	0.104	0.012	0.882	46.558	1.421	370.412	46176	0.710	-0.002	0.757	-0.025		-0.082	
0.043	0.051	0.789	9.841	8.289	41.245	39051	0.000	-0.116	1.606	20	0.067	0.035	0.895	45.794	5.638	585.401	39051	0.353	-0.003	0.646	-0.024	0.017	-0.106	0.960
0.068	0.058	0.749	10.847	5.527	25.877	22238	0.000	-0.124	1.838	21	0.101	0.038	0.878	54.645	4.677	324.792	22238	0.000	0.017	0.621	-0.033	0.020	-0.129	1.217
0.021	0.062	0.787	6.124	4.277	46.491	57698	0.000	-0.130	1.642	22	0.045	0.043	0.905	13.513	4.131	401.891	57698	0.210	-0.007	0.560	-0.024	0.019	-0.118	1.082
0.021	0.104	0.724	3.897	5.988	51.995	39760	0.000	-0.151	1.751	23	0.045	0.056	0.912	22.951	10.281	445.461	39760	0.000	0.014	0.396	-0.024	0.048	-0.188	1.355
0.073	0.077	0.713	17.084	3.880	32.849	25431	0.000	-0.137	2.014	24	0.103	0.056	0.847	35.352	4.660	273.704	25431	0.379	0.006	0.749	-0.030	0.021	-0.134	1.265
0.015	0.044	0.792	2.177	1.984	37.698	13831	0.000	-0.148	1.944	25	0.042	0.018	0.911	20.553	3.177	754.714	13831	0.000	-0.028	0.826	-0.027	0.026	-0.120	1.118
0.017	0.073	0.771	8.316	9.192	74.861	143997	0.000	-0.139	1.780	26	0.048	0.028	0.895	19.119	3.265	648.946	143997	0.000	-0.029	0.814	-0.031	0.045	-0.124	0.966
0.011	0.098	0.696	1.852	10.631	36.273	39298	0.000	-0.195	2.230	27	0.067	0.039	0.881	19.550	4.393	325.107	39298	0.006	-0.013	0.700		0.059	-0.185	
0.016	0.036	0.881	2.581	2.759	44.791	10457	0.001	-0.068	1.020	28	0.032	0.025	0.950	15.133	3.081	573.272	10457	0.094	0.007	0.322	-0.017	0.011	-0.069	0.698
0.031	0.072	0.769	6.028	7.465	29.792	22949	0.000	-0.129	1.717	29	0.074	0.058	0.853	24.794	4.974	302.342	22949	0.019	-0.015	0.809	-0.042	0.013	-0.084	0.907
0.053	0.066	0.756	21.217	29.781	56.424	89890	0.000	-0.125	1.781	30	0.083	0.048	0.869	47.052	6.808	541.039	89890	0.912	0.000	0.717	-0.030	0.018	-0.113	1.063
0.034	0.061	0.790	15.806	44.678	89.393	165231	0.000	-0.115	1.533	31	0.048	0.042	0.915	33.834	5.510	224.206	165231	0.069	0.004	0.462	-0.014	0.020	-0.125	1.071
0.015	0.052	0.854	4.411	6.028	44.859	44839	0.000	-0.079	1.199	32	0.051	0.009	0.934	54.481	2.826	1730.128	44839	0.000	-0.007	0.582	-0.035	0.043	-0.080	0.617
0.028	0.048	0.831	7.643	12.117	38.165	36612	0.000	-0.094	1.424	33	0.049	0.034	0.920	14.421	2.769	289.057	36612	0.720	0.002	0.539	-0.021	0.014	-0.089	0.885
0.040	0.071	0.743	16.725	18.389	55.044	108182	0.000	-0.147	1.937	34	0.073	0.051	0.869	42.885	7.549	592.937	108182	0.047	-0.007	0.744	-0.033	0.020	-0.127	1.192
0.009	0.041	0.755	4.137	4.692	59.443	144286	0.000	-0.195	2.188	35	0.037	0.051	0.892	46.552	14.121	963.932	144286	0.000	-0.019	0.677	-0.028	-0.011	-0.137	1.512
-0.006	0.069	0.785	-1.713	5.822	56.685	80415	0.000	-0.151	1.733	36	0.023	0.028	0.919	27.477	5.495	331.892	80415	0.000	-0.030	0.676		0.042	-0.134	
0.040	0.072	0.745	10.948	13.238	56.968	87082	0.000	-0.143	1.926	37	0.074	0.051	0.880		9154.443	159263.100	87082	0.000	0.005	0.637	-0.035	0.021	-0.135	1.289
0.055	0.043	0.719	22.175	2.967	45.451	77006	0.000	-0.183	2.443	39	0.112	0.025	0.838	70.103	4.484	927.365	77006	0.000	-0.025	1.150	-0.057	0.018	-0.118	1.293
0.059	0.076	0.733	17.600	9.608	50.722	80062	0.000	-0.132	1.969	40	0.096	0.033	0.870	29.403	2.458	260.457	80062	0.840	-0.001	0.790	-0.037	0.043	-0.136	1.178
0.037	0.085	0.674	8.501	5.767	32.325	38526	0.000	-0.204	2.577	41	0.070	0.058	0.848	29.960	6.472	449.133	38526	0.000	-0.024	0.947	-0.032	0.027	-0.174	1.630
0.065	0.050	0.782	16.337	5.261	52.519	34540	0.000	-0.104	1.644	42	0.086	0.035	0.872	181.193	22.837	3929.123	34540	0.000	-0.006	0.799	-0.021	0.014	-0.091	0.845
Average	**Average**	**Average**				**Sum**		**Average**	**Average**		**Average**	**Average**	**Average**				**Sum**		**Average**	**Average**	**Average**	**Average**	**Average**	**Average**
0.040	0.071	0.765				1907085		-0.124	1.713		0.071	0.042	0.887				1907085		0.000	0.638	-0.030	0.029	-0.123	1.080

Table C6: Comparison between OP and ACF approaches (Three-factor production function)

L	K	M	t(L)	t(K)	t(M)	Num	P-value Wald	Return to Scale	Industry TFP	Industry	L(acf)	K(acf)	M(acf)	t(L,acf)	t(K,acf)	t(M,acf)	Num	P-value Wald	Return to Scale	Industry TFP	L-L(acf)	K-K(acf)	M-M(acf)	TFP-TFP(acf)
0.058	0.051	0.897	13.582	6.155	140.367	51619	0.557	0.005	0.532	13	0.064	0.041	0.905	62.760	8.505	413.465	118393	0.000	0.010	0.505	-0.006	0.009	-0.008	0.027
0.042	0.041	0.920	6.617	2.158	119.347	21821	0.872	0.004	0.440	14	0.058	0.042	0.929	44.462	10.546	532.108	47479	0.000	0.028	0.282	-0.015	0.000	-0.009	0.158
0.069	0.008	0.883	8.589	0.192	71.194	14454	0.322	-0.040	0.946	15	0.076	0.053	0.907	35.319	5.337	174.430	32722	0.000	0.036	0.290	-0.007		-0.024	
0.084	0.094	0.772	1.861	1.303	17.121	1275	0.522	-0.050	1.544	16	0.142	0.179	0.831	9.033	3.410	24.332	2348	0.000	0.152	-0.387			-0.059	
0.047	0.027	0.882	17.406	2.621	170.566	77722	0.000	-0.044	0.905	17	0.061	0.008	0.901	23.712	0.798	317.982	164729	0.000	-0.029	0.811	-0.014	0.019	-0.019	0.095
0.121	0.030	0.800	24.526	2.157	101.154	45102	0.000	-0.050	1.320	18	0.132	0.021	0.827	78.633	9.594	549.446	93857	0.000	-0.020	1.072	-0.011	0.008	-0.027	0.248
0.085	0.029	0.872	19.335	5.357	100.194	22052	0.028	-0.015	0.827	19	0.104	0.012	0.882	46.558	1.421	370.412	46176	0.710	-0.002	0.757	-0.020	0.018	-0.011	0.070
0.065	0.046	0.875	7.959	3.865	65.901	16191	0.248	-0.014	0.752	20	0.067	0.035	0.895	45.794	5.638	585.401	39051	0.353	-0.003	0.646	-0.002	0.011	-0.019	0.106
0.094	0.041	0.867	8.201	3.458	52.051	10472	0.876	0.002	0.740	21	0.101	0.038	0.878	54.645	4.677	324.792	22238	0.000	0.017	0.621	-0.007	0.003	-0.011	0.119
0.041	0.038	0.883	7.543	3.259	83.086	28199	0.007	-0.038	0.826	22	0.045	0.043	0.905	13.513	4.131	401.891	57698	0.210	-0.007	0.560	-0.005	-0.005	-0.022	0.266
0.041	0.038	0.874	6.332	2.238	131.531	18708	0.005	-0.046	0.890	23	0.045	0.056	0.912	22.951	10.281	445.461	39760	0.000	0.014	0.396	-0.004	-0.018	-0.038	0.494
0.098	0.005	0.828	11.502	0.406	56.014	12772	0.000	-0.069	1.357	24	0.103	0.056	0.847	35.352	4.660	273.704	25431	0.379	0.006	0.749	-0.005		-0.019	
0.003	0.020	0.874	0.312	1.414	45.580	6466	0.000	-0.103	1.376	25	0.042	0.018	0.911	20.553	3.177	754.714	13831	0.000	-0.028	0.826			-0.037	
0.020	0.028	0.876	6.451	1.808	128.550	70551	0.000	-0.076	1.129	26	0.048	0.028	0.895	19.119	3.265	648.946	143997	0.000	-0.029	0.814	-0.028		-0.019	
0.032	0.019	0.850	3.673	0.674	63.426	20503	0.001	-0.100	1.362	27	0.067	0.039	0.881	19.550	4.393	325.107	39298	0.006	-0.013	0.700	-0.035		-0.032	
0.025	0.025	0.933	3.163	2.534	79.642	4833	0.085	-0.018	0.538	28	0.032	0.025	0.950	15.133	3.081	573.272	10457	0.094	0.007	0.322	-0.008	0.000	-0.017	0.216
0.050	0.044	0.842	4.716	1.865	35.823	11196	0.010	-0.064	1.152	29	0.074	0.058	0.853	24.794	4.974	302.342	22949	0.019	-0.015	0.809	-0.023	-0.014	-0.012	0.343
0.078	0.041	0.838	17.357	2.717	89.967	43257	0.003	-0.043	1.098	30	0.083	0.048	0.869	47.052	6.808	541.039	89890	0.912	0.000	0.717	-0.005	-0.007	-0.032	0.380
0.042	0.038	0.895	14.838	4.371	156.590	76093	0.005	-0.025	0.702	31	0.048	0.042	0.915	33.834	5.510	224.206	165231	0.069	0.004	0.462	-0.005	-0.003	-0.020	0.240
0.027	0.031	0.928	6.175	3.612	130.992	20792	0.081	-0.014	0.566	32	0.051	0.009	0.934	54.481	2.826	1730.128	44839	0.000	-0.007	0.582	-0.024	0.022	-0.006	-0.015
0.036	0.038	0.900	6.174	3.452	85.061	17165	0.022	-0.026	0.763	33	0.049	0.034	0.920	14.421	2.769	289.057	36612	0.720	0.002	0.539	-0.013	0.004	-0.020	0.225
0.067	0.021	0.839	14.162	1.023	80.986	52315	0.001	-0.073	1.292	34	0.073	0.051	0.869	42.885	7.549	592.937	108182	0.047	-0.007	0.744	-0.006		-0.030	
0.026	0.035	0.871	6.681	1.833	103.848	71932	0.001	-0.069	1.070	35	0.037	0.051	0.892	46.552	14.121	963.932	144286	0.000	-0.019	0.677	-0.011		-0.022	
0.006	0.025	0.894	1.204	0.821	109.292	38371	0.016	-0.075	1.021	36	0.023	0.028	0.919	27.477	5.495	331.892	80415	0.000	-0.030	0.676			-0.025	
0.053	0.024	0.859	10.565	1.015	126.076	44247	0.007	-0.064	1.168	37	0.074	0.051	0.880		9154.443	159263.100	87082	0.000	0.005	0.637	-0.021		-0.021	
0.084	0.056	0.809	15.343	4.010	65.686	41272	0.003	-0.051	1.309	39	0.112	0.025	0.838	70.103	4.484	927.365	77006	0.000	-0.025	1.150	-0.028	0.032	-0.028	0.158
0.076	0.012	0.834	15.221	0.453	73.799	40725	0.006	-0.078	1.426	40	0.096	0.033	0.870	29.403	2.458	260.457	80062	0.840	-0.001	0.790	-0.020		-0.036	
0.056	0.061	0.833	8.993	2.816	68.887	19342	0.057	-0.050	1.132	41	0.070	0.058	0.848	29.960	6.472	449.133	38526	0.000	-0.024	0.947	-0.014	0.004	-0.015	0.185
0.077	0.020	0.855	16.149	1.316	106.318	16676	0.003	-0.047	1.118	42	0.086	0.035	0.872	181.193	22.837	3929.123	34540	0.000	-0.006	0.799	-0.009		-0.017	
Average	**Average**	**Average**				**Sum**		**Average**	**Average**		**Average**	**Average**	**Average**				**Sum**		**Average**	**Average**	**Average**	**Average**	**Average**	**Average**
0.055	0.033	0.865				924466.000		-0.046	1.009		0.071	0.042	0.887				1926448.000		0.000	0.637	-0.013	0.005	-0.022	0.195

Table C7: Comparison between two-factor and three-factor production function (ACF results)

L	K	t(L)	t(K)	Num	P-value Wald	Return to Scale	Industry TFP	Industry	L(3)	K(3)	M(3)	t(L)	t(K)	t(M)	Num	P-value Wald	Return to Scale	Industry TFP	L(2)-L(3)	K(2)-K(3)	TFP(2)-TFP(3)
0.650	0.190	77.036	9.784	112164	0.000	-0.161	3.204	13	0.064	0.041	0.905	62.760	8.505	413.465	118393	0.000	0.010	0.505	0.586	0.148	2.699
0.719	0.179	184.854	7.262	45426	0.000	-0.103	2.953	14	0.058	0.042	0.929	44.462	10.546	532.108	47479	0.000	0.028	0.282	0.661	0.137	2.671
0.705	0.260	74.384	6.167	31382	0.341	-0.035	2.506	15	0.076	0.053	0.907	35.319	5.337	174.430	32722	0.000	0.036	0.290	0.629	0.208	2.216
0.873	-0.112	3.987	-0.348	2281	0.481	-0.239	6.700	16	0.142	0.179	0.831	9.033	3.410	24.332	2348	0.000	0.152	-0.387	0.732		
0.538	0.280	95.426	41.860	161661	0.000	-0.182	3.124	17	0.061	0.008	0.901	23.712	0.798	317.982	164729	0.000	-0.029	0.811	0.477		
0.717	0.137	842.420	24.843	92908	0.000	-0.146	3.382	18	0.132	0.021	0.827	78.633	9.594	549.446	93857	0.000	-0.020	1.072	0.585	0.116	2.310
0.687	0.161	31.835	6.011	45467	0.000	-0.153	3.441	19	0.104	0.012	0.882	46.558	1.421	370.412	46176	0.710	-0.002	0.757	0.582		
0.607	0.252	334.988	21.800	38183	0.000	-0.141	2.989	20	0.067	0.035	0.895	45.794	5.638	585.401	39051	0.353	-0.003	0.646	0.539	0.217	2.343
0.743	0.142	29.781	7.880	21833	0.000	-0.115	3.336	21	0.101	0.038	0.878	54.645	4.677	324.792	22238	0.000	0.017	0.621	0.642	0.104	2.714
0.652	0.212	32.355	5.851	56508	0.000	-0.137	3.172	22	0.045	0.043	0.905	13.513	4.131	401.891	57698	0.210	-0.007	0.560	0.606	0.169	2.612
0.595	0.278	125.315	9.242	38754	0.000	-0.127	2.746	23	0.045	0.056	0.912	22.951	10.281	445.461	39760	0.000	0.014	0.396	0.550	0.222	2.350
0.685	0.152	26.031	8.694	25034	0.000	-0.164	3.450	24	0.103	0.056	0.847	35.352	4.660	273.704	25431	0.379	0.006	0.749	0.582	0.096	2.701
0.562	0.395	12.811	7.583	13413	0.000	-0.044	2.356	25	0.042	0.018	0.911	20.553	3.177	754.714	13831	0.000	-0.028	0.826	0.519	0.377	1.529
0.544	0.287	32.758	14.557	140246	0.000	-0.170	3.244	26	0.048	0.028	0.895	19.119	3.265	648.946	143997	0.000	-0.029	0.814	0.496	0.258	2.430
0.710	0.317	21.267	8.015	37699	0.000	0.028	2.079	27	0.067	0.039	0.881	19.550	4.393	325.107	39298	0.006	-0.013	0.700	0.643	0.279	1.379
0.671	0.254	38.160	8.968	10117	0.000	-0.075	2.866	28	0.032	0.025	0.950	15.133	3.081	573.272	10457	0.094	0.007	0.322	0.639	0.229	2.544
0.559	0.315	36624.890	18195.020	22528	0.000	-0.125	2.823	29	0.074	0.058	0.853	24.794	4.974	302.342	22949	0.019	-0.015	0.809	0.486	0.257	2.014
0.644	0.239	51.236	12.674	88025	0.000	-0.118	3.083	30	0.083	0.048	0.869	47.052	6.808	541.039	89890	0.912	0.000	0.717	0.561	0.191	2.366
0.556	0.279	75.866	32.114	162089	0.000	-0.165	3.021	31	0.048	0.042	0.915	33.834	5.510	224.206	165231	0.069	0.004	0.462	0.509	0.238	2.560
0.674	0.252	25.694	7.577	42931	0.000	-0.074	2.879	32	0.051	0.009	0.934	54.481	2.826	1730.128	44839	0.000	-0.007	0.582	0.623	0.244	2.297
0.608	0.227	175.704	10.279	35463	0.000	-0.164	3.584	33	0.049	0.034	0.920	14.421	2.769	289.057	36612	0.720	0.002	0.539	0.559	0.194	3.046
0.613	0.282	667.670	48.140	106062	0.000	-0.105	2.900	34	0.073	0.051	0.869	42.885	7.549	592.937	108182	0.047	-0.007	0.744	0.540	0.231	2.155
0.539	0.306	217651.200	19116.040	141712	0.000	-0.155	3.083	35	0.037	0.051	0.892	46.552	14.121	963.932	144286	0.000	-0.019	0.677	0.502	0.254	2.407
0.507	0.314	588.986	57.053	78225	0.000	-0.178	3.103	36	0.023	0.028	0.919	27.477	5.495	331.892	80415	0.000	-0.030	0.676	0.484	0.287	2.427
0.655	0.234	99.980	5.618	84499	0.002	-0.110	3.096	37	0.074	0.051	0.880		9154.443	159263.100	87082	0.000	0.005	0.637		0.184	
0.676	0.216	79.109	15.403	75674	0.000	-0.108	3.391	39	0.112	0.025	0.838	70.103	4.484	927.365	77006	0.000	-0.025	1.150	0.564	0.191	2.240
0.646	0.239	278.885	16.375	77407	0.000	-0.115	3.218	40	0.096	0.033	0.870	29.403	2.458	260.457	80062	0.840	-0.001	0.790	0.550	0.206	2.427
0.561	0.235	82.550	5.423	37154	0.000	-0.204	3.717	41	0.070	0.058	0.848	29.960	6.472	449.133	38526	0.000	-0.024	0.947	0.492	0.178	2.770
0.615	0.219		17762.130	33996	0.000	-0.166	3.408	42	0.086	0.035	0.872	181.193	22.837	3929.123	34540	0.000	-0.006	0.799		0.184	
Average	**Average**			**Sum**		**Average**	**Average**		**Average**	**Average**	**Average**				**Sum**		**Average**	**Average**	**Average**	**Average**	**Average**
0.638	0.232			1858841		-0.129	3.202		0.071	0.042	0.887				1907085		0.000	0.638	0.568	0.207	2.393

Chapter 4

The Impacts of Highways on Firm Size Distribution

4.1 Introduction

The relative strengths of large and small enterprises largely decide the market structure, a market occupied by a few large firms tends to have higher market concentration, and vice versa. The relative growth rates of large and small firms will change the firm size distribution (FSD) and market structure in the long run; however, the competition between large and small firms is asymmetric. On the positive side, the market selection mechanism will select productive firms to survive, these survival firms tend to be larger, more productive, and produce diversified products (Bernard et al. 2012, Freund & Pierola 2015, Foster et al. 2016); on the other side, the high productivity of some large firms also reflect their market influences and pricing-setting power rather than their production efficiency, sometimes the idiosyncratic shocks of a few giant enterprises can even cause aggregate fluctuations (Freund & Pierola 2015, Gabaix 2011, Magerman et al. 2016). A more concentrated market maintained by monopolistic power is expected to have lower efficiency of resource allocation than a concentrated market derived from productivity competition, so the market concentration could be productive or less productive. In the long run, the continuous screenshots of market concentration or dispersion will provide a description of FSD evolution, which reveals the changes of market structure and competition level. During the process of FSD evolution, those firms can also choose cooperation rather than competition according to their comparative advantages. Firms are not likely to be autarkists that produce all necessary goods, they tend to specialize in a small category of products and outsource all other secondary business to upstream suppliers, while the final decision is decided by the tradeoff between the costs of specialization and autarky. When transaction efficiency increases and average transaction costs decreases, firms tend to be specialized producers outsourcing their secondary business to upstream suppliers, which can change the hierarchical market structure (Yang & Ng 2015). There are many studies focusing on the effects of transaction cost reduction, e.g., the impacts of tariff decline and market deregulation; a recent new trend is to use spatial analysis to investigate the role of infrastructure development in transaction cost decline.

Extant studies point out that transportation infrastructure developments do not always generate positive effects on firm-level productivity growth and regional output growth, but more significant relocation effects on population and economic activities. The development of transportation infrastructure can reduce firms' inventory costs, increase their logistics efficiency, then encourage more firms to choose to be specialized producers and to outsource their secondary business to upstream suppliers (Yang & Ng 2015, Datta 2012). At the same time, infrastructure development can also attract the entry of new firms and extant firms to move into the regions with well-developed infrastructure because of lower setup costs and transaction costs; this process can strengthen the connections between upstream and downstream firms, and generate agglomeration economies in those regions with well-developed infrastructure, but the development gaps between rich and rural regions, large and small enterprises will also widen (Holtz-Eakin & Schwartz 1995, Combes et al. 2012, Faber 2014). However, there is limited evidence about road construction effects in China, few studies use detailed spatial data to investigate the impacts of firm-level accessibility changes on firm size and market structure evolution. The proposal of this study is to investigate the impacts of transportation infrastructure development on firm size evolution and market concentration, which are essential to understand the mechanism of road construction effects, to explain why different studies tend to get opposite results.

The expansion of the road network will change the balance of power across market selection, relocation, and industrial agglomeration effects, as well as reshaping the competition structure. Firstly, road expansion will connect more cities into national and global markets, the increase of competition level will strengthen market selection effects, which can induce high-productive firms to grow faster and replace other less-productive firms (Pagano & Schivardi 2003, Lentz & Mortensen 2008), while the increase of market selection effects will reshape firm productivity and size distribution to be more left-truncated (Combes et al.

2012). Secondly, the increase of traffic accessibility can also attract the entry of new firms and incumbent firms to move into the regions with well-developed infrastructure because of lower setup and transaction costs (Holtz-Eakin & Schwartz 1995, Faber 2014). This process will promote industry agglomeration, increase firm size and productivity dispersion, and induce the size distribution to be right-skewed; because the decrease of unit intermediate input costs will encourage up- and down-stream industry cooperation, and outsourcing activities, while outsourcing activities allow some small scale outsourcing firms to co-exist with the small amount of large enterprises (Duranton & Puga 2004, Combes et al. 2012). Thirdly, the development of transportation infrastructure will affect the growth of large and small firms in different ways; due to large enterprises owning more comparative advantages than small firms (Pagano & Schivardi 2003, Hottman et al. 2016), they can also exploit more profits from traffic accessibility shocks, then change the market structure.

This study finds that expressway construction has significant agglomeration effects on FSD evolution; outsourcing, specialization, and firm entry play important roles in this process. Due to China's rapid road expansion, firms' traffic accessibility gradually decreased from 1998 to 2007, which further reduced transaction costs. The rise of traffic accessibility promotes the growth of firms' average scales, but this impact varies from large to small firms, i.e., large enterprises tend to growth faster than other firms, while small firms gradually shrink, then the overall size dispersion tends to increase; a 10% decrease of firms' distance to high-class highways can increase firm size dispersion by 0.4% to 1.3%, increase market concentration by 1.4% to 2.0%. The increase of size dispersion and market concentration are motivated by the rapid expansion of large firms, outsourcing activities, and the establishment of new and small firms. Better traffic conditions tend to increase outsource and inventory levels, but new entry firms tend to choose addresses around a certain brand of distance from highways, i.e., not very close and not very far from highways. Higher inventory levels are correlated with larger firm size, while new entry and outsourcing firms tend to be smaller. To solve the endogeneity problems in road projection and construction, two types of instrument variables are constructed, i.e., historical roads and counterfactual roads, the same as in the third chapter. The two-stage instrumental regressions provide consistent results with baseline regressions, suggesting there is a robust causal relationship between road construction and firm size evolution.

This chapter is organized as follows. Section 2 reviews the literature about infrastructure construction effects, relevant theories, and empirical evidence about firm size evolution. Section 3 introduces the data sources and research methodology. Section 4 analyzes the empirical results. Section 5 concludes.

4.2 Literature Review

4.2.1 Theoretical Background

Transportation Infrastructure and Agglomeration Economies

The development of transportation infrastructure will reshape international and domestic trade network, then strengthen the connections between upstream and downstream firms. According to the framework of new classical microeconomic theories, specialization and economic organization are given by the solution of market equilibrium. The solution contains a series of outcomes, i.e., at the one extreme, an agent can be an autarkist that produces all necessary goods; at the opposite extreme, an agent can choose to specialize in only one good and outsource all other secondary business to upstream suppliers. Agents' specialization is decided by the tradeoff between autarky and transaction costs (Yang & Ng 2015). When transaction efficiency increases and average transaction costs decrease, firms tend to be specialized producers and outsource their secondary business to upstream suppliers, which can change hierarchical market structure. In order to specify this story, it is necessary to come down to firms' transaction cost structure. Firms always need to hold appropriate amounts of raw materials to ensure they can produce goods smoothly; meanwhile, their finished goods may not be sold immediately, when these inventories are maintained in storehouses, firms must pay for inventory costs, e.g., costs of storage and depreciation. When firms make their decisions about inventory management, they need to consider the logistic efficiency and transport costs. Sometimes, firms use new technologies such as IT to optimize their inventory management so that they are able to maintain smaller amounts of inventory for their normal operation.[1] At the same time, the development of transportation infrastructure can provide better traffic conditions which allow vehicles to move faster and reach farther destinations, then help firms to take less time to transport their products to target markets, and get raw materials faster. Then firms find they no longer needed to maintain a high level of inventory for their daily operation, they can enjoy the decline of inventory costs. During this

[1]The impacts of telecommunications infrastructure see (Roller & Waverman 2001)

process, firms also face the tradeoff between specialization and transaction costs. Due to the speed of vehicles increasing and inventory costs declining significantly, firms find they can more easily find suitable suppliers; which encourages more firms choosing to be specialized producers, and the connections of upstream and downstream agents will be promoted (Datta 2012).

However, the positive externality of the road network tends to decay over distances, which encourages new firms to establish near to new highways, and promotes incumbent firms to relocate their sites to get higher traffic accessibility. Due to budget constraints, governments cannot build highways to connect all cities, the returns of road construction can only be obtained by a portion of firms and regions. In China, the expressway network was initially designed to connect important cities such as provincial capitals and port cities, and vehicles on these newly constructed roads can move faster than on normal roads. However, these efficient expressways cannot always provide straight-line routes between firms and their destinations, compared with normal roads which are free or charged low toll fees during 2000 to 2010, expressways charged higher toll fees because of their higher construction costs; so firms need to weigh the pros and cons of using expressways. A firm near to expressways has more motivation to use expressways than a firm far from expressways, because the latter has to firstly arrive at expressways through normal roads before they use expressways, which will further increase logistic costs. This issue induces an important phenomenon in spatial analysis, i.e., the road network has spillover effects on surrounding regions, while the spillover distances are affected by firms' tradeoff between using or not using these newly constructed roads. Existing evidence supports the view that the spillover distances are not so far as our expectation; in most cases, only regions connected by new roads will show significant productivity or output increases, which suggests that few firms are willing to travel more than tens of kilometers to use new highways (Holtz-Eakin & Schwartz 1995, Faber 2014). The unbalanced spatial distribution of new roads will encourage firms to establish or move into the regions with well-developed infrastructure because of lower setup costs and transaction costs (Holtz-Eakin & Schwartz 1995, Faber 2014).

Agglomeration economies will be strengthened in those regions with well-developed infrastructure, with the continuous firm and population relocation; which can largely change firm-level productivity and size distribution; but the relationship between agglomeration and its benefits shows a U-shaped curve. As previously discussed, transportation infrastructure development will encourage firm and population immigration, while industrial agglomeration can promote knowledge sharing and spillover, and reduce firms' transaction costs because firms can more easily find suitable suppliers and potential consumers. Further more, industrial agglomeration is normally accompanied with population concentration, which can provide a local labour pool, increase matching efficiency and employment. The agglomeration of population and economic activities will make them more productive, but large and more productive enterprises tend to reap more benefits from agglomeration economies and grow faster than small or less productive firms; so the increase of agglomeration economies will also dilate productivity and size distribution, and induce the distribution to be right-skewed (Pagano & Schivardi 2003, Combes et al. 2012). Meanwhile, the increase of firm size dispersion is not only motivated by the fast expansion of large and productivity firms, but also the rise of small firms, because the decrease of transaction cost will induce specialization and outsourcing activities. Some firms will use cheaper intermediate inputs to substitute other inputs, these outsourcing firms tend to have smaller scales than vertical integration enterprises, while those industries that are highly dependent on intermediate inputs will benefit more from agglomeration effects (Hashiguchi & Tanaka 2015, Combes et al. 2012). On the opposite side, congestion of population and economic activities will increase land and inputs prices, then reduce the benefit of agglomeration effects. When more firms get access to the national market, competition level tend to increase and magnify market selection effects (Duranton & Puga 2004, Combes et al. 2012), which can induce high-productive firms to grow faster and replace other less-productive firms (Pagano & Schivardi 2003, Lentz & Mortensen 2008). This process will generate a opposite reaction to agglomeration effects, because selection effects tend to shrink firm size dispersion and induce firm size distribution to be more left-truncated.

Scale Economies and Market Structure

The development of infrastructure will deeply affect transaction costs and FSD; while large and productive enterprises tend to reap more benefits than small and less-productive firms, their comparative advantages will induce unbalanced reactions to external shocks. These shocks can be policy shocks or the development of infrastructure, while the unbalanced reactions will then motivate the evolution of market structure. When these comparative advantages are very significant in an industry, this sector tends to be dominated by small amounts of large enterprises and show significant market concentration level, while high concentration is normally accompanied with low competition, slower investment and output growth

(Pagano & Schivardi 2003).

Early studies of FSD tend to find that firm size follows Gibrat's Law (Hart & Prais 1956, Simon & Bonini 1958, Mansfield 1962, Ijiri & Simon 1964), which frequently emerges from many social studies, e.g., city size distribution, firm size and productivity distribution, firm entry, exit, and random growth. Gibrat's Law assumes the growth rate of firm size is independent from firm size, and predicts FSD follow log-normal distribution. However, the following studies tend to find that large firms have slower growth rates than young and small firms, which leads to the conclusion that Pareto distribution is the better description of FSD (Luttmer 2007, Guo et al. 2013, Amador & Opromolla 2013, Gao et al. 2015, Yuan et al. 2016). Gabaix (2011) interprets this facts as the residuals of idiosyncratic shocks of large firms, i.e., large firms are so influential on macro fluctuations, e.g., GDP growth may be affected by export of Boeing, innovation of Wal-Mart, and the difficulties of Nokia; hence large firms are incompressible 'grains' in aggregated economic fluctuations, this is also named the 'granular' hypothesis.

Current studies about size evolution mostly acknowledge that firms tend gradually to grow larger over time but growth rates tend to decline with the increase of firm scales, so new business are usually more in number but systematically smaller than established businesses. During firms' lifetime, firms always face the possibility of exiting the market; sometimes firms' exit is triggered by wrong decision-making or macro-level shocks, while survival firms tend to be larger, more productive and capital-intensive, hire skilled workers and produce diversified products, sometimes they are also more likely to involve in international trade (Bernard et al. 2012, Freund & Pierola 2015, Foster et al. 2016). However, these phenomena did not always exist in history; at least the size-skill relationship has been negative for a very long time, i.e., small firms hire a higher fraction of skilled workers (Holmes & Mitchell 2008). Due to firms facing the tradeoff between hiring a large quantity of workers or hiring a small number of skilled workers, their decisions depend on the relative marginal returns of quantity or quality modes; e.g., firms in retailing sectors tend to hire great amounts of employees, and use information technologies to improve their management efficiency; while firms in law and consulting sectors value the importance of human capital (Coles & Mortensen 2016), the positive size-productivity relationship is not always the truth in the long run. At the same time, the slowdown of long-term productivity growth in the U.S. reflects the massive physical investment and that skilled workers in these large enterprises failed to promote productivity growth as before (Gordon 2000). During the 1950's and 1960's, the correlation between public investment such as transportation infrastructure investment and aggregate productivity growth is very strong, but the long-term productivity slowdown during 1970's and 1980's cannot be explained by public investment (Fernald 1999). These facts induce a question, i.e., if large firms have higher Solow residual, survival rate, and accounting profit, while the long-term aggregate productivity tend to slow down, how do large firms maintain their Solow residual at a high level?

One interpretation is about scale economies, i.e., the unit costs of production decrease with the increased size of operating units (Chandler et al. 2009). However, the dramatic increases of industrial output are not always equal to economies of scale, because firms' scale and quantity are limited by market scale. During the period of the first and second industrial revolution, the expansion of large enterprises can be distinguished into two groups, i.e., capital-intensive and labour-intensive industries. The expansion of labor-intensive industries was motivated by more machines and workers, i.e., textile, lumber, printing and publishing, larger firms gained cost advantages but these advantages were not significant enough to lead to industry concentration. By contrast, significant scale economies rose in those so-called 'capital-intensive' industries, i.e., petroleum, chemicals, primary metals, and transportation equipment industries, because their capital-labor ratio was relatively easier to increase with the continuous investment in new machines and equipment, standardized parts, carefully organized production processes, and new energy such as fossil fuels, while capital deepening can significantly decrease unit costs in that period. However, when several enterprises reached their minimum efficient scale (the point of profit maximization under monopolistic competition), their industrial capacity was so huge that the national demand could be easily satisfied by a small number of enterprises. For example, from 1880 to 1885, the Standard Oil Trust reduced their average production cost from 2.5 to 1.5 cent per gallon, meanwhile, they had to raise their daily production from 1,500-2,000 barrels to 5,000-6,500 barrels to maintain their cost advantages (Chandler et al. 2009).

Another explanation is about market power, based on the assumption that most firms operate in detailed diversified monopolistic competition markets, or the firms themselves are large enough to gain price-setting power. These cases yield the pricing rule $p = \frac{MC}{1+\mu}$ under the principle of profit maximization, where MC is marginal cost, μ is the price elasticity of demand (Dixit & Stiglitz 1977). For each monopolistic competition market, term $\frac{1}{1+\mu}$ is also called markup on marginal costs, it is well defined by price-demand elasticity, indicating how influential a firm is on price in the corresponding monopolistic competition market. At the same time, the market power that firms can exploit from segment markets varies a lot from large to small firms (Hottman et al. 2016). Large enterprises are obviously able to exploit higher markups, sometimes their influence is so significant that firm-specific idiosyncratic shocks fail to average out on the aggregate level (Gabaix 2011). Such 'large firm residuals' were demonstrated by recent empirical studies, e.g., the top 100 firms cause 90% aggregate fluctuations, when taking account of supplier-buyer linkages (Magerman et al. 2016); a small number of large firms can affect trade structure and comparative advantage, one third variation of export-GDP ratio stem from top firms (Freund & Pierola 2015); the idiosyncratic shocks of the largest 100 U.S. firms can explain one third variations of aggregate output and the Solow residual (Gabaix 2011).

According to existing literature, it can be concluded that large and productive firms tend to grow faster and reap more benefits than small and less-productive firms, which lead to a positive relationship between firm size and productivity. The development of infrastructure will generate a pair of opposite forces; stronger agglomeration effects will encourage specialization and outsourcing activities, increase size dispersion and induce firm productivity and size distribution to be right-skewed; on the other side, congestion will reduce the benefit of agglomeration effect and strengthen market selection effects, while stronger market selection tends to reshape firm productivity and size distribution to be more left-truncated. At the same time, agglomeration effects will be strengthened by competitive advantages of regions with better natural conditions, e.g., coastal regions have better access to international markets, and national municipalities and provincial capital may have advantages in attracting foreign investment; hence the development gaps between rich and rural regions tend to increase during this process. However, it is too ambitious for this study to investigate all theoretical productions.

4.2.2 Empirical Evidence

Introduction of FSD Evolution and Infrastructure Bonus

Firm size distributions vary across industries and countries throughout history, while large enterprises only appear in certain industries, implying that industrial heterogeneity plays an important role in FSD evolution. When few firms become very large and influential in some industries, their firm-level shocks may also affect aggregate fluctuations. If we come back to the world of 1870s, a majority of industrial products were produced by four industrialized countries, i.e., Great Britain 32%, United States 23%, Germany 13%, and France 10%. On the eve of the World War I (1913), the fraction of industrial output of Great Britain decreased from 32% to 14%, while this decline was taken by United States, whose fraction increased from 23% to 36%. Before World War II (1938), the top-four industrialized countries also changed significantly, i.e., United States 32%, USSR 19%, Germany 11%, and Great Britain 9%. From 1917 to 1973, the 200 largest industrial enterprises in the United States, Germany, and Great Britain tended to aggregate in some typical industries. In the United States, 70% of its 200 largest industrial enterprises were in food manufacturing, chemicals, petroleum, primary metals, fabricated metals, machinery, and transportation equipment industries; in Germany, 75% of its 200 largest enterprises were in food manufacturing, chemicals, textiles, primary metals, machinery, transportation equipment industries; similarly, in Great Britain, 70% of the largest enterprises aggregated in food manufacturing, chemicals, textiles, primary metals, transportation equipment industries (Chandler et al. 2009). As a more detailed comparision, Lotti & Santarelli (2004) investigate firm size evolution of four representative industries in Italy (Food, Footwear and Clothing, Instruments, Electrical and Electronic Engineering) from 1986 to 1994. Food, Footwear and Clothing industries are considered to be traditional sectors because technological and productivity progress play less important roles than Instruments, Electrical and Electronic Engineering industries. These authors find that firms' age plays an important role in FSD evolution of the latter two industries, firms older than two or three years tend to follow lognormal distribution, because small entrants can rapidly increase their capacity through the investment on new technologies, and converge to the capacity of incumbent firms. By contrast, some traditional industries such as Footwear and Clothing industries are characterized by slower learning processes and weaker selection effects, their convergence processes on lognormal distribution is hard to find, suggesting the importance of cross-industry heterogeneity in FSD evolution.

Extant studies find firm size variation is largely affected by demand-side factors, suggesting that if highway construction can promote firm expansion and productivity growth, costs reduction effects are expected to play a secondary important role, the more direct and important channel is specification and agglomeration effects. Di Giovanni et al. (2014) use French firm-level database from 1999 to 2007, which contains firms' domestic sales and destination-specific exports, this data sample allows us to trace the linkages of heterogeneous firms with their multiple export markets. They decompose the growth rate of firm sale to each destination market (domestic and foreign destinations) into three parts: macroeconomic shocks, industrial shocks, and firm-level shocks. The results show that firm-level shocks can explain 80% of aggregated sale variation, while these firm idiosyncratic shocks are highly affected by input-output linkages. However, that study did not further distinguish the differences between supplier-producer and producer-consumer linkages. Bernard et al. (2019) decompose firms' revenue scale into upstream supply, downstream demand, and firms' production capacity, then find firm size can be largely explained by downstream demand (around 80%), upstream component and production capacity only explain 20% firm size variation. Downstream demand can be further decomposed into final demand and demand of intermediate input from downstream industries, and their results also show that almost all downstream component is explained by network sales rather than final demand. Similarly, Hottman et al. (2016) decompose firms' revenue size into marginal costs, markups, appeal (product heterogeneity such as quality and taste), and product scope; they find 50-75% firm size variation can be explained by appeal, 20-25% can be explained by product scope; by contrast, only less 20% size variation can be attributed to marginal costs. These results suggest that the demand of down-stream firms and final consumers will largely decide output decisions of their suppliers, the benefits of infrastructure improvements mainly come from the closer connection between firms and their downstream cooperators and final consumers.

Studies of transportation infrastructure tend to conclude that infrastructure development can promote productivity and output growth, but the relocation effects widen the gaps of aggregate output between rich and rural regions. Holl (2016) finds firms' distances to highways have negative influence on firms' productivity (especially for manufacturing firms), under the control of agglomeration and location effects, historical trend and firm specific effects. Similarly, Garcia-Mila et al. (1996) find highways have significant positive effects on private output, slightly smaller than the effects of private capital; but these effects become insignificant when they take one-year or two-year lags on these public capital variables, implying public investment have lagged influences on economic output. For the study of the U.K., Gibbons et al. (2016) investigate both micro- and aggregated level evidence; for the aggregated level study, they find road expansion in the U.K. promotes the entry of new firms and employment growth; for the firm-level evidence, they find the effects on employment and output of extant firms close to zero, suggesting road improvements influence employment mainly through firm entry and exit rather than firm-level employment increases. By contrast, Faber (2014) finds rural regions connected by new highways witnessed slower industry output growth than non-connected regions, while those counties with more population or close to provincial capitals/ metropolises can grow faster than others. Consistent with Faber's study, Baum-Snow et al. (2012) also use historical roads as instruments, find infrastructure improvement promotes the decentralization of population and economic activities, while the effects of ring lines are much larger than radial lines, and infrastructure investments in western regions are not as efficient as in eastern regions. The strong population relocation effects are consistent with Faber's research, however, population decentralization may not be a unique consequence of infrastructure investment, slower population growth could also be affected by income level, which is supported by the differences between western and eastern regions.

Extant evidence do not always support positive effects of transportation infrastructure improvement on firm-level productivity growth or regional output growth, but more significant relocation effects (relocation of population and economic activities). The theories of agglomeration and transaction costs predict that infrastructure development will promote industrial and population agglomeration, consistent with extant evidence. However, the implication of industrial and population agglomeration has not been well investigated; extant theories predict agglomeration effects will increase size dispersion and induce firm productivity and size distribution to be right-skewed, while congestion and market selection effects will promote the exit of less-productive firms. This study aims to provide stylized description and interpretation of FSD evolution during rapid road expansion period in China, to explain how FSD respond to infrastructure development and transaction cost decline. China is one of the most important emerging economies, and its development experiences should provide important evidence to explain how rapid road expansion affect FSD evolution.

Some Stylized Facts of FSD Evolution

Extant evidence supports the view that large firms show slower growth rates, so FSD of all firms are not likely to follow Gibran's law. When firms are divided into different size and age groups, FSDs of these narrowly defined groups tend to follow Gibran's law. Samples that only include large and old tend to have more symmetric and decentralized FSD than young and small firm samples; while samples including both large and small firms tend to show more right-skewed FSD. These FSD variation across different size and age groups also have important policy implication. The policy shocks that can promote new firm entry and growth will change FSD significantly, because firm entry will increase competition level and increase exit rates of incumbent firms due to market selection effects, which induces a trade-off between survival rate of incumbent firms, and the entry and growth of young and small firms.

Bee et al. (2017) study size and scope evolution of French export firms, but this dataset only include large scale firms.[2] Results show that lognormal distribution is a better description of French FSD than Pareto distribution.[3] At the same time, firms in the lower and upper tail fail to follow lognormal distribution, some upper tail firms with diversified products seem to follow Pareto distribution, these firms also tend to be exporters with multiple destination markets. These results confirm that samples containing large firms tend to have symmetric distribution, and Pareto distribution sometimes can describe top enterprises; with the expansion of sample size, FSD tends to follow lognormal distribution. Young and small firms tend to growth faster than old ones. When a firm is newly established, its survival rate is positively related to the size when it was established; firms with less than 5 employees when they start up only tend to have a 10% survival to 15 years, this proportion rise to 20% if the established firm has more than 20 employees. When firms' age is more than 15 years, their annual risk of dying decreases sharply and is independent from size at start up (less than 10 percent), but it is still closely correlated with current sizes. These results suggest market selection effects are closely connected with firm aging and their initial scales.

There are also studies including more small and medium size firms to investigate FSD evolution, and they find different but more colourful results. Cabral & Mata (2003) compare two samples of Portuguese firms, a small sample including top 1,000 large enterprises, and a larger sample covering 33,678 firms, provided by the Portuguese Ministry of Employment. They find FSD of large firms is more symmetric and bell-shaped, while for a large sample, FSD is right-skewed. In addition, firm age also plays an important role in firm size evolution. When firms are divided into different age groups, their average sizes tend to grow larger while their size dispersions tend to become larger and more symmetric over time. However,

[2] their sample does not include firms exporting value less than 100,000 euros within EU and exporting value less than 1,000 euros outside EU.

[3] However, lognormal distribution only contains a second-order moment of firm size, while the increase of the order of moment can increase the matching degree on the FSD.

even for the firms in the largest age group, i.e., over 30 years, their FSD is still far from symmetric distribution. In addition, these authors also use longitudinal data to check the robustness of previous results, the results support that FSD is right-skewed but this skewness tends to decrease, the divergence tends to increase, and the overall distribution evolves towards lognormal distribution over time. These results suggest that even though the overall FSD does not follow lognormal distribution, this asymmetry is partly caused by distribution variation across different size and age groups.

Segarra & Teruel (2012) focus on Spanish manufacturing firms, using a sample of firms with more than two employees from 2001 to 2006. Whatever firm size is indicated by employees or sales, FSDs are normally right-skewed; when firms become older, FSD tends to move rightwards and become less concentrated. They also use Gini index to indicate firm size inequality, and find firm size inequality keeps stable before firms are 40 years old, then firm size Gini index become more volatile and tend to decrease, implying firm size inequality varies across different age groups. When firms are divided into small (firms with less than 50 employees) and large firm(more than 250 employees) groups, size inequality of the small group is higher than in large group, while Gini index of large firms is more volatile over different ages. The relationship between sample size (ordered by firm size) and firm size inequality is non-linear and cannot be well approximated by Pareto distribution; Pareto coefficients of employees size are larger than one (superlinear relationship) for small samples, then gradually converge to one with sample growth; Pareto coefficients of employees size show a superlinear relationship for small samples, and show sublinear relationship for larger samples. Similarly, sample size order by firm age is positively correlated with size inequality. When firm size is measured by sales, Pareto coefficients are larger than one for small samples with a small amount of old firms, then converge to one and reduce to less than one, implying the relationship transforms from superlinear to sublinear; while the firm size measured by sale always shows a sublinear relationship across different age samples. These results suggest that firm age and size variation can affect firms size evolution and induce selection biases if the database cannot cover all firms with different scales.

By contrast, Huber et al. (2013) develop a specialized model to study size evolution of small and medium size enterprises (SMEs) in Austria. They use the Austrian Social Security Database (ASSD), which contains EU-defined manufacturing firms with less than 250 employees from 1972 to 2005. Average firm sizes measured by employees decrease from

11.57 to 10.35 from 1999 to 2004, while the growth rates of market for different industries decrease from 4% (1988-1998) to 1% (1999-2004), implying firm size inequality decreases under the background of gradually market size expansion. Compared with some extant studies which tend to conclude large and old firms follow Gibran's law, their results show that firm size evolution of SMEs does not satisfy Gibran's law; smaller firms tend to grow faster than large firms. If a new established firm has larger scale, it usually has longer life expectation. The authors also evenly divide SMEs into four size groups, they find firms' probability to change their size groups increased after 2000; the probability to transact from large- to small-size group increases with firm aging, while younger and smaller firms are more likely to be Gazelle companies (the companies witness sharp explosion of their size). The shrinking of incumbent firms (specifically for large firms) have destructive impact on employment, e.g., if a firm move from a large-scale group to a smaller group, it has significant influence on number of jobs (the downward transaction of size group accounts for two-third job loss.). However, the previous results mainly focus on small and medium firms, firms with more than 259 employees and starting up earlier than 1972 are excluded. After including these outliers as robustness checks, they find the distribution of all firms still follows the similar pattern, the job creation effect is even amplified by those large and old firms. Their results suggest that the policy to increase firms' life expectation can reduce firms' exit rate, but will also reduce the average growth rates of new established firms; by contrast, the policy to encourage firm entry will increase competition level, and increase exit rates of incumbent firms. So there is a trade-off between small and medium size firms growth and lift expectation of incumbent firms, and a trade-off between firm entry and exit; due to the exit of large and old enterprises, sharp expansion or shrink of incumbent firms have significant influence on employment, there is also a trade-off between short-term and long-term job creation.

The Role of Financial Constraints, Openness, and Deregulation in FSD Evolution

Financial Constraint is a widely acknowledge factor in FSD evolution, because large firms normally have financing advantages than smaller and new established firms. Cabral & Mata (2003) test the influences of wealth constraints and human capital (work experience) on firm size in Portugal, while wealth constraints and human capital are indicated by entrepreneurs' age and educational level. Results show that if a firm is established by a young entrepreneur, its firm size tends to be 30% lower other firms; the entrepreneur's age plays a very important role during firm's first seven startup years; when firms become older, the effect of entrepreneur's age become insignificant. By contrast, entrepreneur's educational level has positive influence on firm size, and this effect gradually becomes stronger over time. Angelini & Generale (2008) study the influence of financial constraints using the database of small-scale Italian manufacturing firms. According to their baseline definition of financial constraints, if a firm reports that its loan application was turn down by a bank, it is considered a financially constrained firm; if a firm is always a financially constrained firms throughout the sample period, it is labeled as a persistently constrained firm. They find that financially constrained firms tend to have smaller scales than other firms; when firms grow older, their probability to be financially constrained firms will decrease. The FSD of financial constrained firms is slightly more concentrated and leftward than other firms, the FSD of persistently constrained firms is significantly higher and more concentrated than other financial constrained firms. Due to the difference of FSD between financial constrained firms and other firms, while the proportion of financial constrained firms is higher in young firm groups; Angelini & Generale (2008) further show that, for those firms younger than six years, the FSD of financial constrained firms is much more concentrated than for the other firms, suggesting financial constrained firms tend to locate at the left tail of overall FSD.

Trade liberalization and deregulation will affect transaction costs, while extant evidence supports the view that firms in industries with higher trade openness level tend to have lower average costs and marginal cost dispersion (Del Gatto et al. 2008). Such kind of cost reduction effect will further affect competition level and enforce market selection effects, then change market structure and FSD. Alfaro & Chari (2014) investigate the influences of deregulation on firm size evolution and resource relocation in India. They show that firm sizes measured by assets and sales both decrease from 1989 to 2005; the overall market concentration of all industries tend to decline, while concentration in deregulated industries are generally lower than restricted industries.[4] At the same time, firm sizes measured by assets and sales are both negatively related to deregulation effects, while the entry of new small firms tend to increase with deregulation. However, the further quantile regressions show that deregulation effect is non-linear across different size of firms, i.e., large enterprises are more likely to expand their scale after deregulation, new entrants with bigger size also benefit more after deregulation than small-size entrants, suggesting small and medium firms still face some constraints when they try to expand their scale, the most benefits of deregulation effect are gained by large incumbent firms and large new entrants. On the other hand, market open-

[4]market concentration is indicated by three measurements, i.e., Herfindahl index based on sales, the variation of assets and sales.

ness and deregulation may have destructive effects if the economy cannot properly cope with new challenges. Braguinsky et al. (2011) investigate the relationship between labour market distortion and firm size distribution. They developed a structural model to explain that employment protection policies tax wages, cause labour market distortion, and finally affect FSD evolution. Their results show that the proportion of small firms with only 1-10 workers keeps increasing in Portugal from 1986 to 2009, while other industrialized countries such as U.S. and Denmark show the opposite process, i.e., the increases of average firm size. At the same time, the shrinkage of firm scales also associates with productivity reduction, suggesting that, the more resources are mis-located to smaller and less-productive firms, the more Portuguese firms could be crowded out by productive overseas firms.

The changes of market concentration can also affect market power, then increase overall allocation efficiency, and decrease firm-level profitability due to higher competition level and the decrease of market power. Hutchinson & Persyn (2012) examine the relationship between trade openness, economic integration and concentration. They show that trade openness indicated by relative international and domestic trade flows keeps increasing from 1980 to 2003 in most Western European countries. By contrast, the average market concentration of all sectors tend to increase from 1991 to 2005, several largest firms gradually grow larger, while the market share of medium size firms shrink and taken by smaller firms, the overall FSD become more symmetric and less skewed.[5] Their results show that trade openness is negatively correlated with market concentration, while higher market concentration is related to high level of markups. The evolution of firm size is accompanied with the decrease of the share of labour costs in industrial value added with the integration of E.U., suggesting firms do not need to hire so many employees as before; there is a substitution process from unskilled employees to skilled employees and capital formation with trade integration. Similarly, Konings et al. (2005) investigate the effects of privatization and trade openness on competition and market power in Bulgaria and Romania (firm-level data from 1994 to 1998) and find that state-owned firms have lower markups than private firms, while foreign invested firms have the highest markups than others, suggesting state-owned enterprises set prices closer to their marginal costs and more likely to maximize social welfare (‘allocation efficiency’, page.3). When trade openness increase, the overall profitability will decrease, especially for those highly concentrated industries; while the privatization of large-scale state-owned enterprises normally induces higher monopoly and lower allocation efficiency.

[5]Market concentration is measured by Herfindahl–Hirsch index, Lerner index, and the market share of top firms.

Empirical Evidence of Chinese FSD Evolution

China is a rapid growth emerging economy, and its large enterprises show similar or even faster growth rates than small firms, which leads to Pareto firm size distribution, while FSD in developed countries tend to follow lognormal distribution. Due to both the scales of small and large firms increasing very fast, firm size inequality tends to increase with the economic growth of China. Financial constraints and market openness are also found to have important influences on FSD evolution; better financing conditions are much beneficial for small and medium firms, while the process of Chinese trade liberalization will decrease market distortion and affect firms' profitability.

Gao et al. (2015) focus on Chinese list companies from 2001 to 2013, and find a significant drop of Pareto coefficients from 2001 to 2008 and slight decline after 2008, suggesting the increase of firm size dispersion, but this effect weakens a lot after the 2008 World Financial crisis. Whether firm size is measured by assets or equities, firm scales kept increasing during 2001 to 2013, while firm size dispersion measured by Gini index has also increased significantly, implying the rise of firm-level inequality (For robustness checks, they use revenues to indicate firm size and get consistent results). From 2001 to 2008, Pareto coefficients decreased significantly; however, Pareto coefficients kept stable after 2008, implying firm size inequality increases significant before the 2008 crisis. This conclusion is consistent with another inequality measurement, i.e., Gini Index; the increase of Gini coefficients (measured by asset and equities respectively) was very fast before 2008, then turned to slowed down. In order to investigate the mechanism of firm size evolution, they estimate the impacts of firm size on firm size growth, and find larger firms tend to growth faster. The same as the structural break of firm size inequality after 2008, firm size effect on firm growth also decreases a lot after the financial crisis. Results show that the growth rate gap of assets between large and small firms disappeared after 2008, while the growth rate gap of equities still existed.[6] These results suggest that, even for listed companies, large firms still own advantages to expand their scales, which is further reflected at the unbalanced growth between large and small firms.

[6] The growth rate gap of equities decreased by about 60%, these results stay robust after excluding newly listed firms. Newly list firm have limited number of observations, they may cause selection bias on final distribution.

Zhang et al. (2009) also focus on large firms, and find revenue scales of top 500 enterprises follow Pareto distribution from 2002 to 2007, with an exponent of 1.[7] By contrast, Guo et al. (2013) investigate FSD of top enterprises in U.S., China, and the world, and find size (measured by revenues) of top 500 companies gradually increases from 2001 to 2009, but their growth rates slow down after 2008; FSD follows Pareto distribution and keeps stable over time, with an exponent of 1.[8] At the same time, top firms of U.S. show much slower growth, their FSD is closer to lognormal distribution rather than Pareto distribution, suggesting higher firm size inequality. Then they use Gini index to measure firm size dispersion, and find that size inequality of Chinese top firms is higher than U.S. but keeps declining during investigation period, while size dispersion in the U.S. keeps increasing and reaches the level of China by 2008. However, whether listed companies or top 500 enterprises, they only represent a small part of economic activities in China, while many extant studies using different sample sizes in developed countries get opposite conclusions. To investigate this issue, Segarra & Teruel (2012) focus on the impacts of sample size on FSD, and find FSD changes significantly across different size of samples, some medium-size sample seem to be be well approximated by Pareto distribution, implying the studies only using large firm samples may suffer from selection bias.

By contrast, Yuan et al. (2016) investigate China's FSD, using the China Annual Survey of Industrial Firms Dataset (1998-2007). They use several factors, including financial constraints, market concentration, ownership effect, and trade openness, to explain Chinese FSD evolution. They find FSD of huge amounts of Chinese manufacturing firms is right-skewed but this skewness tends to decrease over time. Eastern region has larger Pareto coefficient than central and western regions, but the increase rate of Pareto coefficient in central and western China are higher than Eastern region, implying firm size inequality of eastern China is lower than other regions but the decrease rate of size inequality in central and western China are faster than eastern China; theses three Pareto coefficients curves intersect around 2006, suggesting China's firm size inequality of Central and Western China converged with Eastern region around 2006. The regression results show that lower financing constraints are beneficial for the survival of small and medium enterprises, larger market is associated with

[7]They find top 500 enterprises follow Zipf's Law, while Zipf distribution is the discrete distribution of Pareto distribution.

[8]Data of Chinese top 500 enterprises is collected from the annual survey of the Development of China's Enterprises (2002-2010), issued by the China Enterprise Confederation, while data of top 500 U.S. and world companies (1995-2010) is collected from Fortune 500, they provide the information firms' employees, assets, and rank.

lower firm size inequality, the share of SOEs and foreign firms may have crowd-out effect on small and medium firms, trade openness can increase export exposure then also induce crowd-out effect. As the robustness checks, the authors use law enforcement, based on the fact that judicial efficiency can protect property right and promote firm development, and time-lagged financing constraints as the instruments of financing constraints, and substitute the measurement of market concentration as Herfindahl index, then get consistent results; which implies that there is a casual relationship between financing constraints and firm size inequality, better financing conditions can promote the development of small and medium size enterprises.

Lu & Yu (2015) investigate the impacts of Chinese trade liberalization on resource misallocation after China joined the WTO in 2001, while Chinese extensive tariff reduction after 2001 was normally believed to be an important contributing factor on market competition, which can reduce market distortion and resource misallocation. Market distortion in their research is indicated by firm-level markups, which is estimated based on ACF approach and DLW approach (Olley & Pakes 1992, Ackerberg et al. 2015, De Loecker & Warzynski 2012). They use Poisson regression developed by Silva & Tenreyro (2006) to estimate the relationship between trade openness (tariff reduction) and firm performance, the results show that import volume and markup are positively correlated with trade openness, while markup dispersion is negatively correlated with trade openness. For robustness checks, they use Gini index, coefficient of variation (CV), and mean deviation (RMD) to indicate size dispersion, while the coefficient of tariff decline keeps negative significant; these results confirm that market distortion tends to decrease with tariff reduction, the decrease if transaction costs will deeply change market structure and concentration.

Summary of Existing Empirical Studies

Firms' average scales tend to gradually grow larger over time; when firms become larger, firm size dispersion also tend to increase. Due to market selection mechanism, less-productive firms are not likely to survive, while these firms are normally have small scales, which induces a fact that the size distribution of young and small firms are right-skewed, such kind of skewness tends to decrease when firms become larger and older (Cabral & Mata 2003, Segarra & Teruel 2012). Extant evidence shows that FSD in developed countries, such as the U.S. and France, tends to follow lognormal distribution, while studies of emerging economies such as China tend to find large enterprises follow Pareto distribution (Zhang et al. 2009, Segarra & Teruel 2012, Gao et al. 2015, Bee et al. 2017); implying that large firms in developed countries usually have lower growth rate than smaller firms, while large firms in developing countries have similar growth rates as small firms, the growth rates of those large firms in developing are more likely to be independent from firm size.

Extant studies also have identified several key factors to explain firm size evolution. Financial constraint is a frequently discussed factor of financial cooperate studies, the results tend to support that financial constraint is an important but not the unique factor to promote firm growth (Cabral & Mata 2003, Angelini & Generale 2008, Alfaro & Chari 2014, Yuan et al. 2016). Due to the impossibility of financial constraint being directly measured, there have been some recent moves to external factors such as trade openness, protection and deregulation effect (through transaction costs channels), and market concentration effect, to study the external mechanism of firms size evolution (Lotti & Santarelli 2004, Braguinsky et al. 2011, Hutchinson & Persyn 2012, Alfaro & Chari 2014, De Loecker et al. 2016). The empirical results support the view that deregulation and market openness can promote FSD to move from left to the right, but the changing shape of FSD suggests the benefits of deregulation tend to be gained by large enterprises rather than small firms; on the contrary, the impact of market protection may promote the decline of market inequality at the expense of the shrink of large firms, suggesting that the impacts of rising of decreasing market protection and openness are asymmetric. Furthermore, these non-linear relationships vary from different age and size groups, and are affected by ownership, market concentration effects, and capital intensity across different industries, implying the colorful structure inside FSD evolution.

However, there is some research that has used spatial analysis to investigate how firm size evolution responds to infrastructure improvement, which can reduce transaction costs and strengthen spatial connections across regions. Extant evidence, industrial organization, and trade theories all point out that there are positive relationships between firm size, productivity and growth rate (Pagano & Schivardi 2003, Lentz & Mortensen 2008), while firms' productivity and size distribution are also shaped by market selection and spatial industry agglomeration (Combes et al. 2012). The expansion of road network will change the balance between market selection and industrial agglomeration effects. Transportation infrastructure development can reduce firms' logistics costs (Yang & Ng 2015) and time consumption (Datta 2012), it can promote up- and down-stream industry cooperation, and reduce unit costs of intermediate inputs such as inventory costs (Fernald 1999, Duranton & Puga 2004, Combes et al. 2012). These effects will encourage firms to use the cheaper intermediate inputs to substitute other inputs; then those industries that highly dependent on intermediate inputs will benefit more from agglomeration effects. At the same time, congestion will reduce the benefit of agglomeration effects because population and industries concentration will increase land and inputs prices (Hashiguchi & Tanaka 2015) which would cause a potential non-linear relationship between traffic accessibility and agglomeration benefits. Compared with small and medium size firms, larger enterprises own more competitive or monopolistic advantages than smaller firms, and they can normally grow faster than smaller firms (Combes et al. 2012); these effects tend to increase competition and firms size dispersion. In order to fill these research gaps, the first aim of this study is to investigate the size effects and distributional effects generated by expressway construction; then the second purpose is to identify the channels that traffic accessibility affects firms size evolution, e.g., outsourcing channel and market selection mechanism.

4.3 Methodology

4.3.1 Data and Descriptive Statistics

This chapter is based on exactly the same data and instrument variables as the previous chapter, the main explanatory and dependent variables are specified in Table 4.3-1. We use employee quantity and real total assets (deflated by the tangible asset deflators constructed by Brandt et al. (2012)) as the indicators of firm size; these two measurements are log-transformed in regressions. Some literature using the NBS dataset usually selects industrial output and revenue to indicate firm size, but this dataset cannot correctly reflect the revenue and output distribution because it does not include non-SOEs with less than five million RMB revenue, which leads to left-truncated distribution. Therefore, revenue and industrial output are not selected as firm size measurements.

According to Table 4.3-1, firms' quantity gradually increases from 1998 to 2007, but the average size measured by employees tends to decline, while the average size measured by total assets keeps stable over that period; suggesting the share of younger firms with less employees increased from 1998 to 2007, this process is related to capital density increase, which promotes substitution of labour for capital inputs. At the same time, the firm size dispersion, indicated by standard deviation (calculated based on employment and total assets), gradually decreases from 1998 to 2007. These patterns are also confirmed from the kernel density estimate plots (Figure D1). In Figures D4 and D5, firms are classified into four age groups, it can be found that the firm size and size dispersion tend to be larger for old firms, and this pattern is consistent from 1998 to 2007 (Figure D6 and D7). For financial constraints, extant studies have developed multiple indices, e.g., Ding et al. (2013) using the ratio of cash flow to tangible fixed assets to indicate financial constraints; similarly, Yuan et al. (2016) using cash flows and the share of interest expense to fixed assets to indicate financial constraints. This study uses three measurements to capture the impacts of financial constraints on firm size evolution, i.e., Cash Flow Ratio CFK, Short-term Debt SDK, and Long-term Debt LDK.

Besides the firm-level variables, this study also constructs several region-level variables to investigate the impacts of highway expansion on firm size dispersion and market concentration. These firm-level variables are aggregated by each 2-digit industry in each prefecture, weighted by firms' employment size. The inter-quantile index tends to decrease over the sample period, whatever the IQ index is calculated from employment size $L : Size$ or deflated asset size $RealA : Size$. By contrast, the Herfindahl–Hirschman Index tends to increase from 1998 to 2007, suggesting the increase of market concentration. $Dist$ is firms' weighted distance to highways, it tends to decrease over time. Road density, rail density and population density are used to control the impacts of other transport routes and population agglomeration effects, respectively. However, road density and railway density tend to decrease over time, suggesting that new entry and relocation firms tend to select the provinces with lower transport route density. The waterway density is not controlled because most northern provinces do not have waterways, it may induce selection biases in region-level regressions. At the same time, the previous chapter also shows that once we control road density, other types of transport modes have very slight impacts on the explanatory variable $Dist$, suggesting the impacts of waterway density can be largely absorbed by road density and railway density. The share of state-owned capital and collective capital gradually decreases over the sample period, while the share of foreign and private capital tends to increase. At the same time, firms' average age tends to decrease over time, suggesting the entry of younger firms are very fast, they are more likely to be foreign and private firms.

Table 4.3-1: Summary Statistics

	Obs	Full Sample Average (Std. Dev.)	Obs	1998 Average (Std. Dev.)	Obs	2003 Average (Std. Dev.)	Obs	2007 Average (Std. Dev.)
Firm-level Variables:								
Log Labour	1,826,633	4.705	131,358	4.929	138,575	4.788	305,258	4.588
		(1.133)		(1.213)		(1.146)		(1.067)
Log Assets	1,758,442	4.927	109,053	4.966	136,152	4.955	301,777	4.992
		(1.395)		(1.481)		(1.425)		(1.362)
Cash Flow Ratio	1552878	72.699	-	-	135932	39.827	303784	93.813
(CFK)	-	(7252.072)		-		(9012.875)		(2209.415)
Short-term Debt	1749940	825.4241	108762	2578.792	135280	676.896	301253	636.418
Ratio (SDK)		(41772.94)		(131195.1)		(12380.68)		(10898.27)
Long-term Debt	1749940	109.398	108762	911.808	135280	48.034	301253	19.528
Ratio (LDK)		(28197.66)		(99859.89)		(4915.569)		(471.376)
Firm-level Variables:								
L: IQ	69,795	.428	6,909	.507	6,732	.452	7,353	.357
(log)		(.343)		(.360)		(.351)		(.299)
Real A: IQ	69,666	.737	6,894	.776	6,712	.771	7,344	.682
(log)		(.397)		(.412)		(.407)		(.366)
L: HHI	70,064	-3.508	6,931	-3.368	6,757	-3.368	7,378	-3.787
(log)		(1.297)		(1.186)		(1.250)		(1.37)
Real A: HHI	69,924	-2.820	6,914	-2.725	6,735	-2.700	7,368	-3.034
(log)		(1.169)		(1.138)		(1.118)		(1.216)
Road-Density	70,064	-1.268	6,931	-1.594	6,757	-1.366	7,378	-.722
(log)		(.786)		(.733)		(.701)		(.810)
Rail-Density	69,976	-4.424	6,926	-4.504	6,736	-4.384	7,378	-4.319
(log)		(.625)		(.668)		(.608)		(.592)
Pop-Density	62,203	5.854	5,353	5.962	6,191	5.751	6,906	5.846
(log)		(.792)		(.721)		(.817)		(.819)
Dist	70,064	9.808	6,931	10.481	6,757	9.804	7,378	9.155
(log)		(1.381)		(1.345)		(1.382)		(1.136)
Sate-share	70,064	.254	6,931	.469	6,757	.230	7,378	.074
		(.328)		(.356)		(.306)		(.180)
Collective-share	70,064	.123	6,931	.212	6,757	.102	7,378	.054
		(.209)		(.259)		(.186)		(.139)
Foreign-share	70,064	.102	6,931	.076	6,757	.102	7,378	.123
		(.199)		(.170)		(.207)		(.210)
Age	70,064	19.611	6,931	48.979	6,757	14.696	7,378	10.191
		(77.815)		(182.574)		(11.604)		(8.650)

Note: Mean and standard deviation of the full sample and the sub-samples in different years (1998, 2003, 2007) are compared in this table. Labour is quantity of employment, Assets is total asset value. The unit of $Dist$ is log meters, the units of $Road - Density$ and $Rail - Density$ are log length (km) of traffic route per unit of land areas ($1km^2$). The unit of Pop is quantity of people per $1km^2$ of land (log), while the unit of Age is year, it equals the difference between current and firms' established years.

4.3.2 Model Specification and Hypothesis

Extant evidence, industrial organization, and trade theories all point out that there are positive relationships between firm size, productivity and growth rate (Pagano & Schivardi 2003, Lentz & Mortensen 2008). Firms' productivity and size distribution are shaped by market selection and agglomeration effects. The stronger market selection effects induce the firm size distribution to be left-truncated, while industrial agglomeration can increase size and productivity dispersion and make FSD right-skewed (Combes et al. 2012). The expansion of road network will change the balance of power between market selection and industrial agglomeration effects. Transportation infrastructure development can reduce firms' inventory costs (Fernald 1999), logistics costs (Yang & Ng 2015), time consumption (Datta 2012), then increase firms' profits, productivity, scales, and promote the entry of small firms; at the same time, larger enterprises have more competitive or monopolistic advantages than smaller firms; they can normally grow faster than smaller firms (Combes et al. 2012). These effects tend to increase competition and firm size dispersion.

Firm size Y_{it} is indicated by employment and real total assets; while firm size dispersion Y_{jrt} is calculated based on firm-level employment and assets in each 2-digit industry by prefecture in each year, indicated by the interquartile range (IQ Range) and Herfindahl–Hirschman Index (HHI). IQ range is defined as the difference between the 75th and 25th quantiles, adjusted by median ($\frac{Q75-Q25}{Q50}$), this measurement is always positive and tends to increase when the firm size dispersion increases (Martin 2008, Ding, Jiang & Sun 2016). HHI is another widely used indicator of dispersion and market competition (Hutchinson & Persyn 2012, Alfaro & Chari 2014), it is defined as the sum of the squares of the market shares $s_{j,i}$ of the firms in same industries J ($HHI_j = \sum_{i\in J} s_{j,i}^2$). If an industry is dominated by a single firm, HHI is equal to 1; while an industry composed by diversified firms has a small HHI. The coefficients of $Dist_{it}$ are expected to be negative because higher traffic accessibility will promote the entry of small firms and the expansion of large enterprises, which will increase the firm size dispersion (Combes et al. 2012).

$$Y_{it} = \alpha_0 + \beta_1 Dist_{it} + \beta_2 Density_{it} + \mu X_{it} + \epsilon_r + \epsilon_j + \epsilon_t + u_{it}. \tag{4.1}$$

$$Y_{jrt} = \alpha_0 + \beta_1 Dist_{jrt} + \beta_2 Density_{jrt} + \mu X_{jrt} + \epsilon_r + \epsilon_j + \epsilon_t + u_{rjt}. \tag{4.2}$$

Equation 4.1 regresses the dependent variables Y_{it} on the explanatory variable (firms' distance to expressways) $Dist_{it}$, $Density_{it}$ and the control vector X_{it}, observations are grouped by year fixed effects ϵ_t and the cross fixed effects of 4-digit industries ϵ_j and county-level regions ϵ_r. Firms closer to expressways have larger traffic accessibility. The impacts of $Dist_{it}$ on firm size are expected to be positive, if road expansion can significantly promote outsourcing activities. However, for those firms with very few outsourcing activities, $Dist_{it}$ should have negative impacts on firm size growth. However, $Dist_{it}$ cannot capture road network complexity, i.e., a firm located in a region with low road density will have smaller traffic accessibility than a firm located in a region with intensive traffic lines. Therefore, it is necessary to use both province-level traffic route density $Density_{it}$ and firm-level road distance $Dist_{it}$ to capture traffic accessibility; different from $Dist_{it}$ the coefficients of $Density_{it}$ are expected to be positive. The control vector X_{it} includes population density to control population agglomeration effects, export dummy to capture international market shocks, age effect because firm size tends to gradually increase over time, ownership effects and financial constraints, u_{it} represents residuals. When the dependent variable Y_{it} is the firm size, both ϵ_r and ϵ_r are used to capture time-invariant fixed effects in each county r for each 4-digit industry j, while year fixed effects ϵ_t is controlled separately to capture macroeconomic shocks in all industry. For the robustness checks, the cross fixed effects of year, 4-digit industries, and country IDs ϵ_{rjt} are applied as a stricter specification of industry level shocks; they can control the annual 4-digit industry shocks in each county, so the impacts of industrial agglomeration and market structure shocks can also be captured by these fixed effects. Equation 4.1 shows that when we run region-industry regressions, the dependent variable Y_{jrt} is the firm size dispersion, only the 2-digit industrial fixed effects ϵ_j, prefecture-level regions ϵ_r, and year dummy ϵ_t are controlled.

Extant theories claim that infrastructure development can increase traffic accessibility and magnify industrial agglomeration effects, which can promote up- and down-stream industry cooperation and reduce unit input costs (Duranton & Puga 2004, Combes et al. 2012). This effect will encourage firms to use cheaper intermediate inputs to substitute other inputs; then those industries that highly dependent on intermediate inputs will benefit more from agglomeration effects. Even though congestion may reduce the benefit of agglomeration effects because population and industry concentration will increase land and input prices (Hashiguchi & Tanaka 2015), the overall benefit of infrastructure development should be promoted by outsourcing and the increase of intermediate input share. This study uses the ratio of intermediate input to total output as the measurement of outsourcing, the same as Ding, Sun & Jiang (2016). At the same time, firms will also be established in or attracted to migrate to the regions with well-developed infrastructure because of lower setup costs and transaction costs; this process can promote the relocation of population and economic activities, then encourage knowledge sharing, generate agglomeration economies and scale economy (Holtz-Eakin & Schwartz 1995, Roller & Waverman 2001, Faber 2014). If a firm has an observation in year t but did not have an observation in year $t-1$, this firm in year t would be defined as a new entry firm. According to extant theories, firms' entry probability is expected to be positively correlated with traffic accessibility, i.e., those younger and smaller firms are expected to be established around expressways.

$$Channel_{it} = \alpha_0 + \beta_1 Dist_{it} + \beta_2 Density_{it} + \mu X_{it} + \epsilon_r + \epsilon_j + \epsilon_t + u_{it}. \tag{4.3}$$

Equation 4.3 is designed to investigate the channels of how infrastructure development affects firm size distribution and competition structure. The same as baseline regressions, the variable $Dist_{it}$ is firms' distance to expressways, the smaller $Dist_{it}$ implies higher traffic accessibility; road density $Density_{it}$ is an alternative measurement of traffic accessibility to capture road network complexity; vector X_{it} is a control vector. Dependent variables $Channel_{it}$ include outsourcing and entry dummy variable. At the same time, the fixed effects terms ϵ_r, ϵ_j, and ϵ_t are used to capture time-invariant fixed effects in each county r of each 4-digit industry j, and macroeconomic shocks.

4.4 Empirical Results

4.4.1 Baseline Regressions

Extant theories predict that firm size distribution will be affected by market selection, agglomeration effect, and scale economy; stronger market selection will induce FSD to be left-truncated, while industry agglomeration and scale economies tend to increase the firm size dispersion and reshape size distribution to be right-skewed, because of the entry or new and small firms and the comparative advantages of large enterprises (Duranton & Puga 2004, Combes et al. 2012). The firm-level baseline regressions find that firms closer to the expressway tend to have larger scales, while quantile regressions further show that expressway construction has heterogeneous impacts on large and small firms. Large firms tend to increase faster than medium-size firms, while small firms tend to have smaller scales if they are close to expressways. The regional baseline regressions further show that highway development can increase firm size dispersion. This process could be motivated by the rapid expansion of large forms. entry of new and small firms, specialization and outsourcing activities, which were discussed in the section of channel studies.

Table 4.4-2: IV Regressions of Size Effects (Employment Size).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
L	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dist	-0.022**	-0.018*	-0.042***	-0.007	-0.007	-0.007	-0.009
	(0.008)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)	(0.006)
Rail-Density	0.112***	0.092***	0.063***	0.091***	0.070***	0.070***	0.065***
	(0.020)	(0.014)	(0.011)	(0.006)	(0.006)	(0.006)	(0.006)
Road-Density	-0.169***	-0.162***	-0.131***	-0.004	0.003	0.003	0.004
	(0.016)	(0.010)	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)
River-Density	0.069***	0.056***	0.052***	0.025***	0.027***	0.027***	0.029***
	(0.016)	(0.012)	(0.009)	(0.005)	(0.005)	(0.005)	(0.005)
Pop	0.011	0.052***	0.035***	0.053***	0.043***	0.043***	0.040***
	(0.011)	(0.010)	(0.010)	(0.004)	(0.004)	(0.004)	(0.004)
Export	0.623***	0.608***	0.567***	0.118***	0.104***	0.104***	0.102***
	(0.011)	(0.007)	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	0.311***	0.341***	0.390***	0.106***	0.099***	0.099***	0.097***
	(0.016)	(0.014)	(0.012)	(0.006)	(0.006)	(0.006)	(0.006)
Collective-share	0.130***	0.125***	0.115***	0.014***	0.012***	0.012***	0.010**
	(0.008)	(0.007)	(0.006)	(0.003)	(0.003)	(0.003)	(0.003)
Foreign-share	0.269***	0.243***	0.250***	0.045***	0.037***	0.037***	0.035***
	(0.012)	(0.009)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)
Age	0.002***	0.002***	0.002***	0.000**	0.000*	0.000*	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	1254245	1247772	1225588	1146617	1088381	1088175	1077752
Group	9481	44458	107003	294126	305855	305820	307056
Under-identification test	719.287	1092.693	1471.839	2647.006	2282.029	2271.889	2300.264
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.7e+04	4.4e+04	2.4e+04	1.2e+04	1.0e+04	1.0e+04	1.1e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.001	0.000	1.107	2.381	1.922	2.326	2.234
P value	(0.971)	(0.995)	(0.293)	(0.123)	(0.166)	(0.127)	(0.135)
First-stage Results:							
Ming30km	0.174***	0.213***	0.198***				
	(0.006)	(0.006)	(0.006)				
SL30km	0.198***	0.136***	0.122***	0.109***	0.107***	0.107***	0.130***
	(0.007)	(0.006)	(0.007)	(0.005)	(0.006)	(0.006)	(0.006)
LC50km				0.180***	0.178***	0.178***	0.178***
				(0.005)	(0.006)	(0.006)	(0.006)
F	557.798	404.101	472.446	1448.035	1271.905	1272.487	1255.407
R2	0.151	0.107	0.090	0.110	0.102	0.102	0.104
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table E2. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. However, not all regressions can pass the Over-identification and Under-identification tests, suggesting some instruments combinations are not efficient in corresponding regressions, so the provided results are those regressions with efficient IVs. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 4.4-2 investigates the impacts of highway construction on firm-level productivity under the control of different types of fixed effects. The fixed effects in columns (1) to (7) are controlled by two components, i.e., year fixed effects and time-invariant cross fixed effects. For the time-invariant cross fixed effects, column (1) controls time-invariant province-level and 4-digit industry fixed effects, column (2) controls time-invariant prefecture-level and 4-digit industry fixed effects, while column (3) controls time-invariant county-level and 4-digit industry fixed effects. Regional and industrial fixed effects can control persistent scale gaps across regions and industries, and the impacts of geographical factors such as natural resource endowment, overall terrain ruggedness level, and local climate; while year fixed effects can capture nationwide shocks on firm size in each year. Columns (4) to (7) further include firm-level fixed effects; column (4) only controls firm-level time-invariant fixed effects; column (5) controls time-invariant province-level, firm-level, and 4-digit industry fixed effects; column (6) controls time-invariant prefecture-level, firm-level, and 4-digit industry fixed effects; while column (7) controls time-invariant county-level, firm-level, and 4-digit industry fixed effects. Firm-level fixed effects can capture firm-level persistent scale gaps; the cross fixed effects between firm and region can control the relocation effect, which means that if a firm changes its address over the sample period, the cross fixed effects between firm and region can capture firm-level fixed effects in different regions separately; the cross fixed effects between firm and industry can control the case that a firm changes its 4-digit industry over sample period. Due to columns (5) to (7) control firm-level, region-level, and industry-level fixed effects simultaneously, the firm-level persistent size gaps, relocation effect, and industry transition effect can be captured simultaneously.

However, our key explanatory variable, firms' distance to highways, could be affected by endogeneity issues because the spatial distribution of new roads could be affected by economic factors such as local income level. The intuition is that when policy makers decide the location of newly constructed highways, they have the motivation to locate new infrastructure in those regions with higher potential economic growth, so their decisions could be affected by the spatial distribution of economic activities. At the same time, the spatial distribution of cities or economic activities is self-correlated over time, if a city was a regional capital fifty years ago, it is still very likely to be an important city at present. To solve potential endogeneity issues, we construct IVs based on the counter-factual approach (applied by Faber (2014)) and historical roads in the Qing and Ming dynasties. However, not all instruments can pass the Over-identification test, suggesting that some instruments are inefficient. Our results only illustrate those regressions with efficient IVs. For instance, Table 3.4-5 presents IV regressions with five different instruments (Ming50km, SL50km and Qing50km), this does not mean that we only have these five IVs because the reported regressions are just the most represented results.

Table 4.4-3: IV Regressions of Size Effects (Unchanged Addresses).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
L	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dist	-0.116***	-0.081***	-0.022	0.024	0.025	0.025	0.030
	(0.014)	(0.017)	(0.025)	(0.034)	(0.040)	(0.040)	(0.041)
Observation	132839	128088	119886	115136	107031	107028	105789
Group	5600	15104	23254	31572	31701	31700	31820
Under-identification test	404.054	610.865	344.303	69.727	58.614	58.614	57.374
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	7206.946	6102.002	3587.261	886.474	762.234	762.213	738.347
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	2.806	0.058	0.250	0.001	0.002	0.002	0.031
P value	(0.094)	(0.810)	(0.617)	(0.977)	(0.967)	(0.967)	(0.859)
Dependent Variable:							
A	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dist	-0.197***	-0.192***	-0.043	0.066*	0.075*	0.075*	0.076*
	(0.015)	(0.022)	(0.031)	(0.027)	(0.033)	(0.033)	(0.034)
Observation	132839	128088	119886	115136	107031	107028	105789
Group	5600	15104	23254	31572	31701	31700	31820
Under-identification test	471.358	610.865	344.303	69.727	58.614	58.614	57.374
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.0e+04	6102.002	3587.261	886.474	762.234	762.213	738.347
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	3.327	1.281	0.029	0.412	0.465	0.465	0.379
P value	(0.068)	(0.258)	(0.866)	(0.521)	(0.496)	(0.495)	(0.538)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table E4 and Table E5. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 4.4-2 shows that highway construction has positive impacts on firm size growth when there is no firm-level fixed effects, i.e., a 10% decrease of $Dist$ is related to 0.20% to 0.40% increase of firm size. When the firm-level fixed effects are controlled, the coefficients of $Dist$ become insignificant. To address the endogeneity issues induced by firms' relocation and entry, Table 4.4-3 uses a subsample that only includes firms who have never changed their address since 1998. When there is no firm-level fixed effects, a 10% decrease of $Dist$ is related to 0.20% to 1.16% increase of firm size. However, when the firm-level fixed effects are controlled, the coefficients of $Dist$ become insignificant. These results suggest that highways' impacts are absorbed by the cross fixed effects of firm, 4-digit, and region fixed effect. Some firms closer to highways have larger scales because they are selected. The similar pattern can also be found in Table E3, this table uses asset size to substitute employment. The coefficients of $Dist$ keep significant when there is no firm-level fixed effects.

Besides the key explanatory variable $Dist$, the coefficients of control variables in Table 4.4-2 also reveal noteworthy results. First, a 10% increase of $RoadDensity$ is related to 0.60% to 1.10% increase of firm size, while a 10% increase of $RiverDensity$ is related to 0.20% to 0.70% increase of firm size. Second, a 10% increase of state-owned capital shares are related to about 1.0% increase of firm scale, while a 10% increase of collective capital shares and foreign capital shares are just related to 0.1% and 0.4 increase of firm scale respectively. Older firms tend to be slightly larger than young firms, consistent with extant studies. Third, a 10% increase of population density is related to 0.30% to 0.50% increase of firm size, while export firms tend to have 10% larger scales than non-exporters. At the same time, once a firms' address was decided, highway expansion has insignificant impacts on firm size variation.

Table 4.4-4: Quantile Regressions (Size Effects).

Dependent Variable: Log Employment	(1) 10th	(2) 20th	(3) 30th	(4) 40th	(5) 50th
Dist	0.004***	0.0008	-0.002*	-0.004***	-0.006***
	(0.001)	(0.0009)	(0.0007)	(0.0007)	(0.0006)
	(6) 60th	(7) 70th	(8) 80th	(9) 90th	
Dist	-0.008***	-0.010***	-0.012***	-0.016***	
	(0.0007)	(0.0008)	(0.0010)	(0.0013)	

Note: Observations are grouped by 4-digit industry and year fixed effects. The key explanatory variable $Dist$ is firms' minimum distance to highways, while the common control variables are road density, prefectural population density, export binary variabl, firms' age, cash flow ratio (CFK), and firms' ownership variables. The full regression results are provided in Table E1. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Existing studies show that large enterprises tend to have more competitive advantages than young and small firms (Hottman et al. 2016). Table 4.4-4 further confirms that the increase of traffic accessibility has heterogeneous impacts on large and small firms. This table uses the 10th to 90th quantile regressions to examine the potential heterogeneous impacts of traffic accessibility on large and small firms. Firm size is indicated by firm-level employment, grouped by the 4-digit industry and year fixed effects, under the control of population density, age, financial constraints, export, and ownership effects. The coefficients of $Dist$ keep negative in columns (3) to (9), while the absolute values of these coefficients tend to increase, suggesting larger enterprises tend to grow faster than medium and small firms under the same traffic accessibility conditions. By contrast, the coefficients of $Dist$ are positive in column (1), suggesting a region close to highways tends to have a small fraction of small firms.

4.4.2 Distributional Effects of Expressway Construction.

The previous firm-level regressions show that highway expansion has heterogeneous impacts on large and small firms, while this section focuses on the impacts of highway construction on market structure. The dependent variable in Table 4.4-5 is 75th-25th inter-quantile range index of each 2-digit industry in each prefecture-level region, a higher value of IQ range index reflects a higher level of firm size dispersion. The dependent variable in Table 4.4-6 is Herfindahl-Hirschman Index (HHI) of each 2-digit industry in each prefecture-level region, where a higher value of HHI reflects a higher level of market concentration of large firms. Due to the dependent variables are region-level indicators, all firm-level explanatory variables, control variables, and instrument variables are also aggregated to region-level variables for each 2-digit industry in each prefecture-level region, weighted by firms' employment scales.

For Table 4.4-5 investigates the impacts of highway construction on firm size dispersion, instrumented by aggregated IVs (Qing50km, Ming50km, and SL50km). The fixed effects in columns (1) and (2) are controlled by year fixed effects and regional fixed effects. The regional fixed effects can control persistent productivity gaps across different provinces or prefectures, and the impacts of geographical factors such as natural resource endowment, overall terrain ruggedness level, and local climate; while year fixed effects can capture nationwide shocks in each year. Columns (3) to (5) further include 2-digit industrial fixed effects, they can control time-invariant region-level fixed effects for each province or prefecture. Columns (1) to (5) show that, when there are no 2-digit industry-level fixed effects, a 10% decrease of $Dist$ is related to 0.80% to 1.17% increase of firm dispersion; when the industry-level fixed effects are controlled, a 10% decrease of $Dist$ is related to 0.40% to 1.30% increase of firm dispersion. These results suggest that those regions with better traffic conditions tend to have higher firm size dispersion in each 2-digit industry.

Table 4.4-5: Highways' Impacts on Firm Size Dispersion (75-25th IQ Range).

Dependent Variable: L IQ Range	(1)	(2)	(3)	(4)	(5)
Main Results:					
Dist	-0.117***	-0.081***	-0.125***	-0.130***	-0.038*
	(0.020)	(0.020)	(0.017)	(0.015)	(0.017)
Observation	54026	54026	53103	54011	53657
Group	31	290	30	892	6776
Under-identification test	12.993	94.771	27.725	303.546	1001.752
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5387.854	4193.597	4446.237	4679.707	5673.647
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.385	3.266	0.018	1.227	1.714
P value	(0.535)	(0.071)	(0.892)	(0.268)	(0.190)
First-stage Results:					
Qing50km	0.112*	0.300***	0.049***	0.103***	0.073***
	(0.041)	(0.029)	(0.006)	(0.010)	(0.017)
SL50km	0.307***	0.337***	0.306***	0.293***	
	(0.033)	(0.038)	(0.006)	(0.009)	
Ming50km					0.427***
					(0.013)
F	479.185	48.989	1891.626	560.997	542.937
R2	0.332	0.305	0.300	0.324	0.411
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table E6. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 4.4-6 investigates the impacts of highway construction on market concentration, instrumented by aggregated IVs (Qing50km, Ming30km, Ming50km, and SL50km). When there are no 2-digit industry-level fixed effects, a 10% decrease of $Dist$ is related to 1.5% to 6.2% increase of firm dispersion; when the industry-level fixed effects are controlled, a 10% decrease of $Dist$ is related to 1.4% to 2.2% increase of firm dispersion. These results suggest that those regions with better traffic conditions tend to have higher levels of market concentration. These results are consistent with Alfaro & Chari (2014), they find that large enterprises are more likely to expand their scales after deregulation and trade openness policies, while the entry of new small firms tends to be promoted, suggesting the most benefits of the deregulation effect are gained by large incumbent firms and large new entrants. The highway expansion is expected to reduce the transaction cost, the increase of market share of large firms is consistent with previous studies.

Table 4.4-6: Highways' Impacts on Firm Size Dispersion (HHI).

Dependent Variable: L HHI	(1)	(2)	(3)	(4)	(5)
Main Results:					
Dist	-0.154**	-0.622***	-0.141***	0.005	-0.217***
	(0.050)	(0.050)	(0.028)	(0.017)	(0.017)
Observation	62203	62203	61165	62192	61901
Group	31	290	30	919	7612
Under-identification test	15.650	96.230	28.165	384.421	690.069
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	7722.211	3865.700	5316.548	8712.423	2674.526
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.370	0.388	3.805	10.408	0.882
P value	(0.543)	(0.533)	(0.051)	(0.001)	(0.348)
First-stage Results:					
Qing50km	0.052	0.300***	-0.020**	0.025*	0.323***
	(0.042)	(0.029)	(0.006)	(0.010)	(0.014)
Ming50km	0.383***				
	(0.052)				
SL50km		0.337***			0.330***
		(0.038)			(0.018)
Ming30km			0.308***	0.419***	
			(0.009)	(0.012)	
F	361.633	48.989	1814.244	632.420	460.421
R2	0.381	0.305	0.325	0.399	0.349
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table E7. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

4.4.3 Robustness Checks

As robustness checks, Table 4.4-7 and Table 4.4-8 use another measurement to investigate the distributional effects of road construction, i.e., firm size dispersion is calculated based on real assets (total assets deflated by the tangible asset deflators constructed by Baum-Snow et al. (2012)). The same as Table 4.4-5 and Table 4.4-6, the dependent variables and control variables are region-level indicators, they are aggregated by each 2-digit industry in each prefecture-level region, weighted by firms' employment scales.

Table 4.4-7: Highways' Impacts on Firm Size Dispersion (75-25th IQ Range, Asset Size).

Dependent Variable: A IQ Range	(1)	(2)	(3)	(4)	(5)
Main Results:					
Dist	-0.157***	-0.043*	-0.175***	-0.178***	-0.058*
	(0.033)	(0.020)	(0.022)	(0.017)	(0.025)
Observation	53320	53320	52424	53312	52963
Group	30	289	30	889	6728
Under-identification test	12.861	94.662	27.641	298.500	788.089
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5301.666	4170.294	4412.806	4609.381	2796.589
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.105	0.030	0.316	0.902	0.969
P value	(0.746)	(0.862)	(0.574)	(0.342)	(0.325)
First-stage Results:					
Qing50km	0.112*	0.300***	0.049***	0.103***	0.323***
	(0.041)	(0.029)	(0.006)	(0.010)	(0.014)
SL50km	0.307***	0.337***	0.306***	0.293***	0.330***
	(0.033)	(0.038)	(0.006)	(0.009)	(0.018)
F	479.185	48.989	1891.626	560.997	460.421
R2	0.332	0.305	0.300	0.324	0.349
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table F1. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

For Table 4.4-7 investigates the impacts of highway construction on firm size dispersion (deflated asset size). The fixed effects in columns (1) and (2) are controlled by year fixed effects and regional fixed effects. They show that a 10% decrease of $Dist$ is related to 0.40% to 1.57% increase of firm dispersion. Columns (3) to (5) further include 2-digit industrial fixed effects, they can control time-invariant region-level fixed effects for each province or prefecture. Columns (3) to (5) show that a 10% decrease of $Dist$ is related to 0.60% to 1.80% increase of firm dispersion. These results suggest that those prefectures with better traffic conditions tend to have higher firm size dispersion in each 2-digit industry, consistent with our baseline results.

Table 4.4-8: Highways' Impacts on Firm Size Dispersion (HHI, Asset Size).

Dependent Variable: A HHI	(1)	(2)	(3)	(4)	(5)
Main Results:					
Dist	-0.135***	-0.507***	-0.117***	-0.005	-0.172***
	(0.041)	(0.041)	(0.024)	(0.014)	(0.016)
Observation	61680	61680	60656	61677	61398
Group	30	289	30	919	7586
Under-identification test	15.605	96.421	28.105	381.552	681.837
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	7598.824	3848.028	5229.548	8567.711	2666.126
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.520	0.185	3.905	11.792	1.143
P value	(0.471)	(0.667)	(0.048)	(0.001)	(0.285)
First-stage Results:					
Qing50km	0.052	0.300***	-0.020**	0.025*	0.323***
	(0.042)	(0.029)	(0.006)	(0.010)	(0.014)
Ming50km	0.383***				
	(0.052)				
SL50km		0.337***			0.330***
		(0.038)			(0.018)
Ming30km			0.308***	0.419***	
			(0.009)	(0.012)	
F	361.633	48.989	1814.244	632.420	460.421
R2	0.381	0.305	0.325	0.399	0.349
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table F2. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

For Table 4.4-8 investigates the impacts of highway construction on market concentration (deflated asset size). The fixed effects in columns (1) and (2) are also controlled by year fixed effects and regional fixed effects. They show that a 10% decrease of $Dist$ is related to 1.30% to 5.10% increase of firm dispersion. Columns (3) to (5) further include 2-digit industrial fixed effects, they can control time-invariant region-level fixed effects for each province or prefecture. Columns (3) to (5) show that a 10% decrease of $Dist$ is related to 1.20% to 1.70% increase of firm dispersion. These results suggest that those prefectures with better traffic conditions tend to have higher market concentration in each 2-digit industry, consistent with our baseline results.

Table 4.4-9: Highways' One-year-lagged Impacts (75-25th IQ Range).

Dependent Variable: L IQ Range	(1)	(2)	(3)	(4)	(5)
Main Results:					
Lag-Dist	-1.149***	-0.980***	-1.173***	-1.285***	-0.906*
	(0.189)	(0.262)	(0.184)	(0.167)	(0.396)
Observation	47755	47755	47029	47739	47426
Group	31	290	30	884	6627
Under-identification test	11.772	82.737	27.472	256.430	427.036
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	3698.739	1713.505	3214.985	3084.636	719.881
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.831	1.927	0.051	2.237	0.143
P value	(0.362)	(0.165)	(0.821)	(0.135)	(0.706)
First-stage Results:					
Qing50km	0.009*	0.023***	0.003***	0.008***	0.006**
	(0.004)	(0.003)	(0.001)	(0.001)	(0.002)
SL50km	0.032***	0.029***	0.032***	0.030***	
	(0.004)	(0.003)	(0.001)	(0.001)	
Ming50km					0.020***
					(0.001)
F	1361.652	63.821	1014.141	560.795	438.168
R2	0.298	0.229	0.267	0.287	0.270
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table F3. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

For Table 4.4-9 uses the one-year-lagged $Dist$ as explanatory variables to test whether highway construction has lagged impacts on firm size dispersion. The fixed effects in columns (1) and (2) are controlled by year fixed effects and regional fixed effects. They show that a 10% decrease of lagged $Dist$ is related to the 10% to 11.5% increase of firm dispersion. Columns (3) to (5) further include 2-digit industrial fixed effects, they can control time-invariant region-level fixed effects for each province or prefecture. Columns (3) to (5) show that a 10% decrease of lagged $Dist$ is related to the 9% to 13% increase of firm dispersion. These results suggest that highway expansion has very significant lagged impacts on size dispersion.

Table 4.4-10: Highways' One-year-lagged Impacts (HHI).

Dependent Variable: L HHI	(1)	(2)	(3)	(4)	(5)
Main Results:					
Lag-Dist	-1.795**	-7.990***	-1.566***	0.050	-4.686***
	(0.574)	(0.644)	(0.336)	(0.208)	(0.433)
Observation	53603	53603	52794	53593	53327
Group	31	290	30	915	7323
Under-identification test	14.851	82.869	27.585	326.269	212.448
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	3948.120	1660.313	2658.391	4029.395	431.794
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.478	1.724	3.209	9.911	1.191
P value	(0.489)	(0.189)	(0.073)	(0.002)	(0.275)
First-stage Results:					
Qing50km	0.006	0.023***	-0.001	0.004***	0.017***
	(0.005)	(0.003)	(0.001)	(0.001)	(0.002)
Ming50km	0.033***				
	(0.006)				
SL50km		0.029***			0.017***
		(0.003)			(0.002)
Ming30km			0.026***	0.035***	
			(0.001)	(0.001)	
F	846.836	63.821	790.703	545.572	380.718
R2	0.310	0.229	0.259	0.313	0.260
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table F4. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

For Table 4.4-10 also uses the one-year-lagged $Dist$ as explanatory variables, to test whether highway construction has lagged impacts on market concentration. The fixed effects in columns (1) and (2) are controlled by year fixed effects and regional fixed effects. They show that a 10% decrease of lagged $Dist$ is related to the 18% to 80% increase of firm dispersion. Columns (3) to (5) further include 2-digit industrial fixed effects, they can control time-invariant region-level fixed effects for each province or prefecture. Columns (3) to (5) show that a 10% decrease of lagged $Dist$ is related to the 17% to 47% increase of firm dispersion. These results suggest that highway expansion has very stronger lagged impacts on size dispersion.

Table 4.4-11: Highways' One-year-lagged Impacts (Different Inter-percentiles).

Dependent Variable: L IQ Range	60-40th (1)	70-30th (2)	80-20th (3)	90-10th (4)
Main Results:				
Lag-Dist	-1.545**	-1.510*	-1.426**	-1.420*
	(0.562)	(0.594)	(0.553)	(0.621)
Observation	43217	47378	47428	47428
Group	6409	6627	6627	6627
Under-identification test	261.803	292.207	292.305	292.305
P value	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	405.731	413.939	413.623	413.623
P value	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.059	0.398	2.721	2.788
P value	(0.808)	(0.528)	(0.099)	(0.095)
First-stage Results:				
Qing50km	0.017***	0.017***	0.017***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
SL50km	0.017***	0.017***	0.017***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
F	380.718	380.718	380.718	380.718
R2	0.260	0.260	0.260	0.260
Fixed Effects:				
Region	prefecture	prefecture	prefecture	prefecture
2digit Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table F5. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 4.4-11 uses four types of percentile range index to test the potential heterogeneous impacts of expressway construction on size dispersion, the explanatory variable is still the one-year-lagged $Dist$. Column (1) uses the 60th to 40th IQ index, column (2) uses the 70th to 30th IQ index, column (3) uses the 80th to 20th IQ index, while column (4) uses the 90th to 10th IQ index as the measurement of size dispersion. Results show that, when the inter-percentile range increases from 60-40th to 90-10th, the coefficients of $Dist$ keep stable and negative, but the absolute values of these coefficients gradually decrease from 1.5 to 1.4, suggesting the distributional effects of traffic accessibility are more significant for medium size firms. These results are consistent with Hutchinson & Persyn (2012) and Alfaro & Chari (2014), they show that if we reduce the transaction costs, the market shares of medium size firms tend to be taken by large firms and small new entry firms.

4.4.4 Channel Studies

The baseline regressions and robustness checks have verified the causal influence of highway construction on size dispersion and market concentration, this section investigates how highway expansion increases can affect firm size dispersion and market concentration. According to existing studies, infrastructure development can increase traffic accessibility, promote up- and down-stream industry cooperation, and reduce unit input costs (Duranton & Puga 2004, Combes et al. 2012). This effect will encourage firms to focus on a smaller range of production, promoting outsourcing activities. Those firms with more significant outsourcing activities are expected to have smaller scales (Hashiguchi & Tanaka 2015). At the same time, Cabral & Mata (2003), Hutchinson & Persyn (2012), Alfaro & Chari (2014) show that large and productive firms tend to increase faster than small and medium size firms. The strong size-productivity relationship is also found in our previous chapter, suggesting that the increase of firm size dispersion and market concentration could be motivated by the the faster expansion of large firms and the rapid entry of new and small firms. At the same time, this process could also be accompanied with the shrink of the market share of medium size firms (supported by Table 4.4-11), and by a previous empirical study such as Alfaro & Chari (2014). Table 4.4-12 regresses inventory levels on the explanatory variable $Dist$, while Table 4.4-13 regresses the outsourcing levels on the explanatory variable $Dist$. Inventory level is a provided variable in our firm-level dataset, while outsourcing activities are indicated by the log ratio of intermediate input to total output, the same as Ding, Sun & Jiang (2016). Results in Table 4.4-12 show that, when firm-level fixed effects are not controlled, a 10% decrease of $Dist$ can increase firms' inventory level by 0.40%-0.80%; when firm-level fixed effects are controlled, a 10% decrease of $Dist$ can increase firms' inventory level by 0.15%-0.25%. Results Table 4.4-13 in show that, when firm-level fixed effects are not controlled, a 10% decrease of $Dist$ can increase firms' outsourcing level by 0.06%-0.10%; when firm-level fixed effects are controlled, a 10% decrease of $Dist$ can increase firms' outsourcing level by 0.07%-0.10%. These results confirm that firms close to highways tend to have more significant outsourcing activities, which is expected to reduce the firm average scales. At the same time, firms close to highways also have higher inventory levels, while higher inventory levels could increase firm scales, because the previous chapter shows that highway construction can increase firms' profit and revenue. These results reveal that some firms may choose the outsourcing strategy, some others may increase their scales, so their overall size dispersion tends to increase (see Table 4.4-16).

Table 4.4-12: The Channel of Inventory.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Inventory	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.076***	-0.083***	-0.041***	-0.026***	-0.019**	-0.016**	-0.014*
	(0.008)	(0.007)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
Observation	1316835	1310538	1288320	1216612	1189551	1189317	1174956
Group	9471	44921	109560	303835	316070	316049	318257
Under-identification test	581.598	1641.744	1492.790	4697.711	4537.538	4879.238	4334.301
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.5e+04	4.2e+04	2.2e+04	2.6e+04	2.5e+04	2.7e+04	2.5e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.217	2.559	1.158	1.917	2.372	3.958	2.624
P value	(0.641)	(0.110)	(0.282)	(0.166)	(0.124)	(0.047)	(0.105)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B12. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 4.4-13: The Impacts on Outsourcing.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Outsourcing	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
Dist	-0.006***	-0.009***	-0.004	-0.007***	-0.009***	-0.009***	-0.009***
	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Observation	1366418	1360228	1338105	1266582	1238883	1238636	1224100
Group	9506	45504	111767	314048	326898	326873	329221
Under-identification test	536.768	1550.466	1178.550	4348.172	4202.267	4187.265	3996.247
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.4e+04	2.8e+04	2.1e+04	2.3e+04	2.2e+04	2.2e+04	2.2e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	3.452	4.066	0.227	0.081	0.013	0.009	0.398
P value	(0.063)	(0.131)	(0.634)	(0.775)	(0.908)	(0.925)	(0.528)
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table B15. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 4.4-14: Channel Studies: Firm Entry as Dependent Variable.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Firm Entry	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dist	0.017***	0.012***	0.013***	0.023***	0.020***	0.021***	0.016***
	(0.001)	(0.001)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
Observation	1254245	1247772	1225588	1146617	1088381	1088175	1077752
Group	9481	44458	107003	294126	305855	305820	307056
Under-identification test	719.287	1092.693	1471.839	2647.006	2282.029	2271.889	2300.264
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.7e+04	4.4e+04	2.4e+04	1.2e+04	1.0e+04	1.0e+04	1.1e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.065	1.652	2.636	0.413	0.412	0.408	0.432
P value	(0.799)	(0.199)	(0.105)	(0.520)	(0.521)	(0.523)	(0.511)
First-stage Results:							
Ming30km	0.174***	0.213***	0.198***				
	(0.006)	(0.006)	(0.006)				
SL30km	0.198***	0.136***	0.122***	0.109***	0.107***	0.107***	0.130***
	(0.007)	(0.006)	(0.007)	(0.005)	(0.006)	(0.006)	(0.006)
LC50km				0.180***	0.178***	0.178***	0.178***
				(0.005)	(0.006)	(0.006)	(0.006)
F	557.798	404.101	472.446	1448.035	1271.905	1272.487	1255.407
R2	0.151	0.107	0.090	0.110	0.102	0.102	0.104
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table G3. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Firm entry could also be an important mechanism to increase firm size dispersion. Table 4.4-14 investigates the impacts of $Dist$ on firm entry, because road expansion can induce new entry firms to choose the regions with well-developed infrastructure (Holtz-Eakin & Schwartz 1995, Roller & Waverman 2001, Faber 2014). If a firm firstly appears in our dataset, it is considered as an new entry firm. Table 4.4-16 shows that, when firm-level fixed effects are not controlled, a 10% decrease of $Dist$ can decrease firms' entry probability by 0.12%-0.17%; when firm-level fixed effects are controlled, a 10% decrease of $Dist$ can decrease firms' entry probability by about 0.20%. These results suggest that new entry firms tend to choose addresses not very close to highways, a possible explanatory is that the competition level around highways is too high to allow these new entry firms to survive. However, the measurement of firm entry also has a problem, i.e., the NBS dataset includes all SOEs and non-SOEs with more than five million RMB revenue, those non-SOEs firms firstly appearing in our dataset are actually the firms that can earn exceeding five million RMB. Therefore, the measurement of firm entry for non-SOEs firms can also be explained as the probability of non-SOEs firms to enter the group of firms with the designated size (more than five million RMB).

Table 4.4-15: Comparison of Entry Firms with Address-unchanged Firms.

	Obs	Full Sample Average (Std. Dev.)	1999		2003		2006	
Full Sample:								
Dist	1,815,165	8.736 (1.529)	119,855	9.461 (1.575)	137,900	8.850 (1.538)	268,814	8.408 (1.430)
Address-unchanged Firms since 1998:								
Dist	381,038	9.178 (1.574)	76,139	9.455 (1.540)	14,196	8.754 (1.472)	11,519	8.136 (1.379)
Entry Firms								
Dist	351,539	8.781 (1.514)	20,943	9.411 (1.622)	34,570	8.970 (1.533)	51,272	8.570 (1.437)

Note: Average and standard deviation of the full sample and the sub-samples in different years (1998, 2003, 2007) are compared in this table.

Table 4.4-15 provides the summary statistics of $Dist$ in the three groups. The first one is our baseline sample that contains all firms, the second group only contains firms that have never changed their addresses since 1998, while the third group only contains those firms labeled by 'entry firms', i.e., the scales of those firms firstly appear in our dataset. The average $Dist$ of the full-sample group gradually decreases from 1999 to 2006, while this trend is faster in address-unchanged group, suggesting that those entry and relocation firms tend to choose addresses not very close to highways, which can explain the positive coefficients of $Dist$ in Table 4.4-14. For the new entry group, the average $Dist$ is more close to the full-sample group, suggesting new entry firms also choose addresses not very far from highways. These results confirm that highway expansions tend to promote firm entry, but new entry firms tend to choose addresses around a certain brand of distance from highways, i.e., not very close and not very far from highways.

The previous results confirm that highway expansion can promote outsourcing activities and increase firms' inventory level, while outsourcing activities are expected to be negatively related to firm size, inventory levels are expected to be positively related to firm size. These two explanations are also indirectly supported by our previous region-level regressions, i.e., the increase of firm size dispersion and market concentration, and supported by Cabral & Mata

(2003), Huber et al. (2013), Alfaro & Chari (2014). These researches also find that deregulation and trade openness policies tend to reduce transaction costs, then induce the market share of medium-size firms to be replaced by large firms and new entry firms, so the size dispersion tends to increase. Table 4.4-16 provides further evidence of these mechanisms, it includes $Inventory$, $Outsourcing$, and $Entry$ as control variables, the other variables are the same as baseline variables. Columns (1), (3), (5) control year and firm fixed effects, while columns (2), (4), (6) control year, and the cross fixed effects across firm, county, and 4-digit industry. Due to we do not have IVs of $Inventory$, $Outsourcing$, and $Entry$, these results can only reflect a correlation rather causality relationship. The coefficients of $Inventory$ are positive, suggesting that higher inventory levels are correlated with larger firm size. The coefficients of $Outsourcing$ and $Entry$ are negative, suggesting new entry and outsourcing firms tend to be smaller, then the firm size dispersion and market concentration tend to increase.

Table 4.4-16: The Correlation of Inventory and Outsourcing with Firm Size.

Dependent Variable: L	(1)	(2)	(3)	(4)	(5)	(6)
Inventory	0.044***	0.039***				
	(0.001)	(0.001)				
Out			-0.012***	-0.009***		
			(0.002)	(0.002)		
Entry					-0.051***	-0.054***
					(0.001)	(0.001)
R2	0.145	0.123	0.138	0.117	0.138	0.119
Observation	1479200	1461735	1529583	1511700	1541996	1523935
Group	458914	545396	469162	557962	471677	561388
F	2667.3	2148.2	2578.9	2125.1	2595.4	2084.0
Fixed Effects:						
Region		county		county		county
Industry		4-digit		4-digit		4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table G4. This table provides OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

4.5 Conclusions

The evolution of firm size distribution is affected by the agglomeration economy and market selection effects, agglomeration effects tend to increase size dispersion and induce FSD to be right-skewed, while stronger market selection effects will reshape FSD to be left-truncated (Duranton & Puga 2004, Combes et al. 2012). This study finds that highway expansions tend to increase firm size, but this impact varies from large to small firms, i.e., most large and medium enterprises tend to expand their scales when traffic accessibility increases, especially for large firms; a few small firms located at the far left of the size distribution tend to decrease their scales when traffic accessibility increase, implying large enterprises tend to get more benefits form highway expansion. Due to the heterogeneous growth of firm scales, the overall size dispersion tends to increase with traffic accessibility increase, suggesting that road expansions generate very significant industrial agglomeration effects. A 10% decrease of firms' distance to high-class highways can increase the firm size dispersion by 0.4% to 1.3%, increase the market concentration by 1.4% to 2.0%. The increase of size dispersion and market concentration are motivated by the rapid expansion of large firms, outsourcing activities, and the establishment of new and small firms. Channel studies show that firms closer to expressways are more likely to outsource their intermediate inputs and have higher inventory levels, but new entry firms tend to choose addresses around a certain brand of distance from highways, i.e., not very close and not very far from highways. Higher inventory levels are correlated with larger firm size, while new entry and outsourcing firms tend to be smaller. These results are consistent with Cabral & Mata (2003), Huber et al. (2013), Alfaro & Chari (2014), they find that deregulation and trade openness policies tend to reduce transaction costs, then induce the market share of medium-size firms to be replace by large firms and new entry firms, so the size dispersion tends to increase.

Several policy implications are noteworthy: firstly, firms will be attracted to move or establish around new constructed highways, consistent with the findings of Faber (2014). At the same time, new entry firms tend to choose the addresses around a certain brand of distance from highways, those addresses very close to highways are not attractive for new entry firms. Secondly, even for those regions that are linked by new expressways, it is not certain that they will benefit from infrastructure development. Due to the increase of traffic accessibility encouraging outsourcing activities, and upstream- and downstream-linkages, the comparative advantages of industrial clusters will become much more significant than in the past. However, the nature of agglomeration economies induces industrial clusters to emerge in those regions with the best endowments of resources and factors, then the concentration of population and economic activities will widen the gaps between rich and rural regions. Thirdly, large firms tend to benefit more from infrastructure development and transaction costs decline. This phenomenon is confirmed again in this study and it presents an issue for policy makers, i.e., to regulate the market appropriately and make sure the decline of transaction costs will encourage productivity-motivated expansion rather than the expansion fueled by market influence.

.1 Appendix D

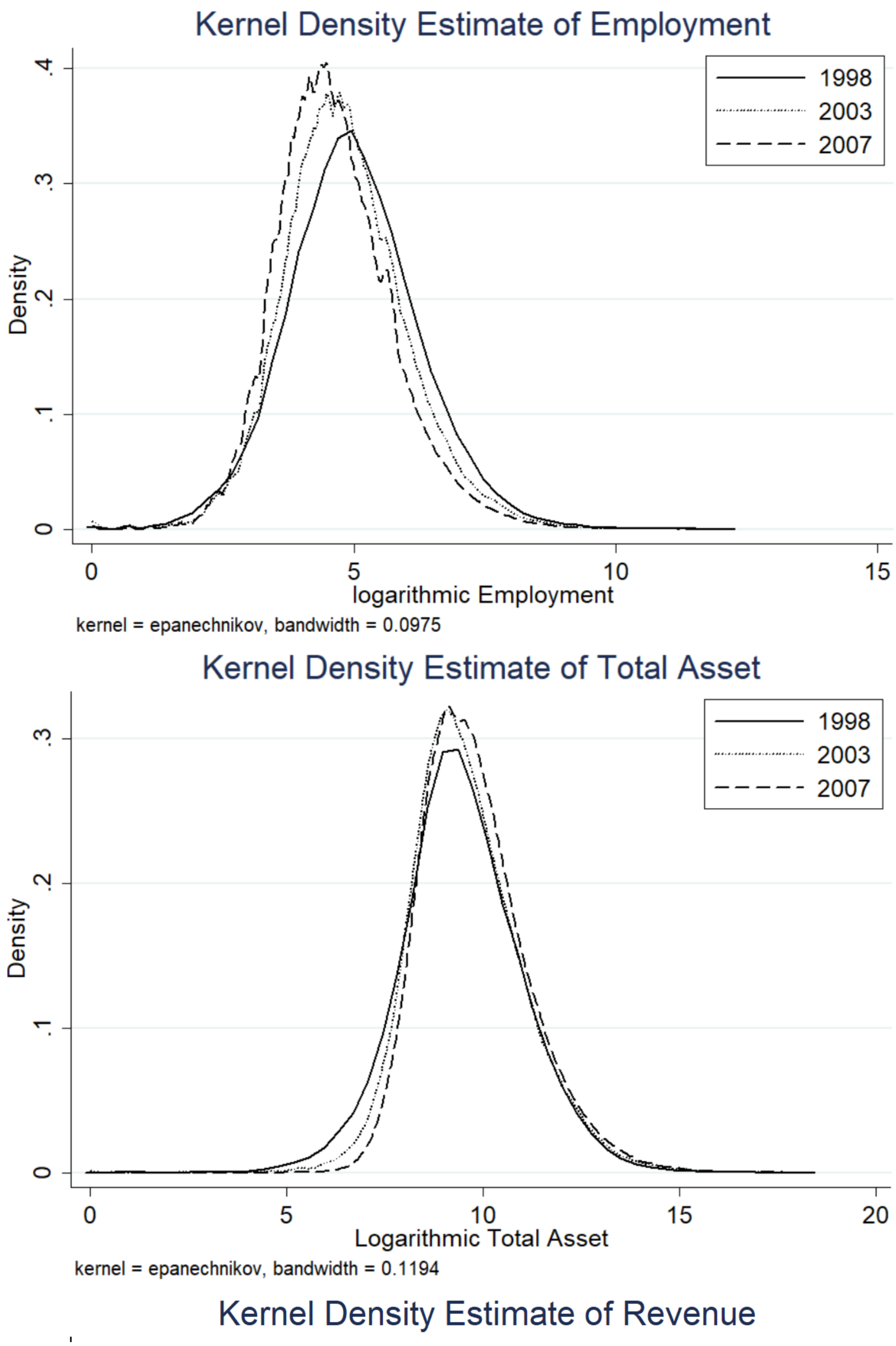


Figure D1: Firm size is measured by employment and total assets. The plot above illustrates employment distribution; the plot below illustrates total asset distribution.

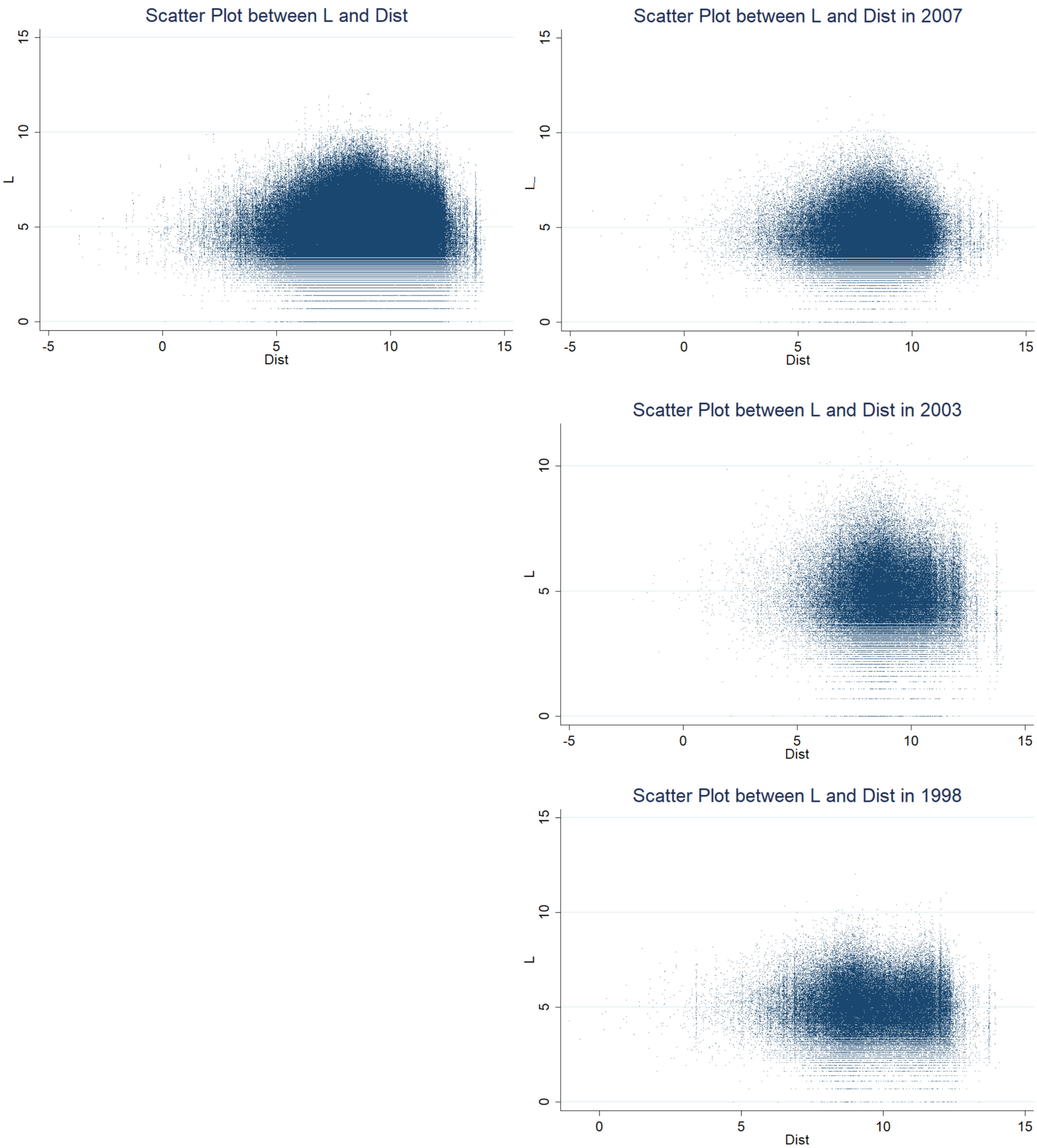


Figure D2: The left plot illustrates the relationship between $Dist$ and L; the right three plots show time-variant scatter plots in 1998, 2003, and 2007.

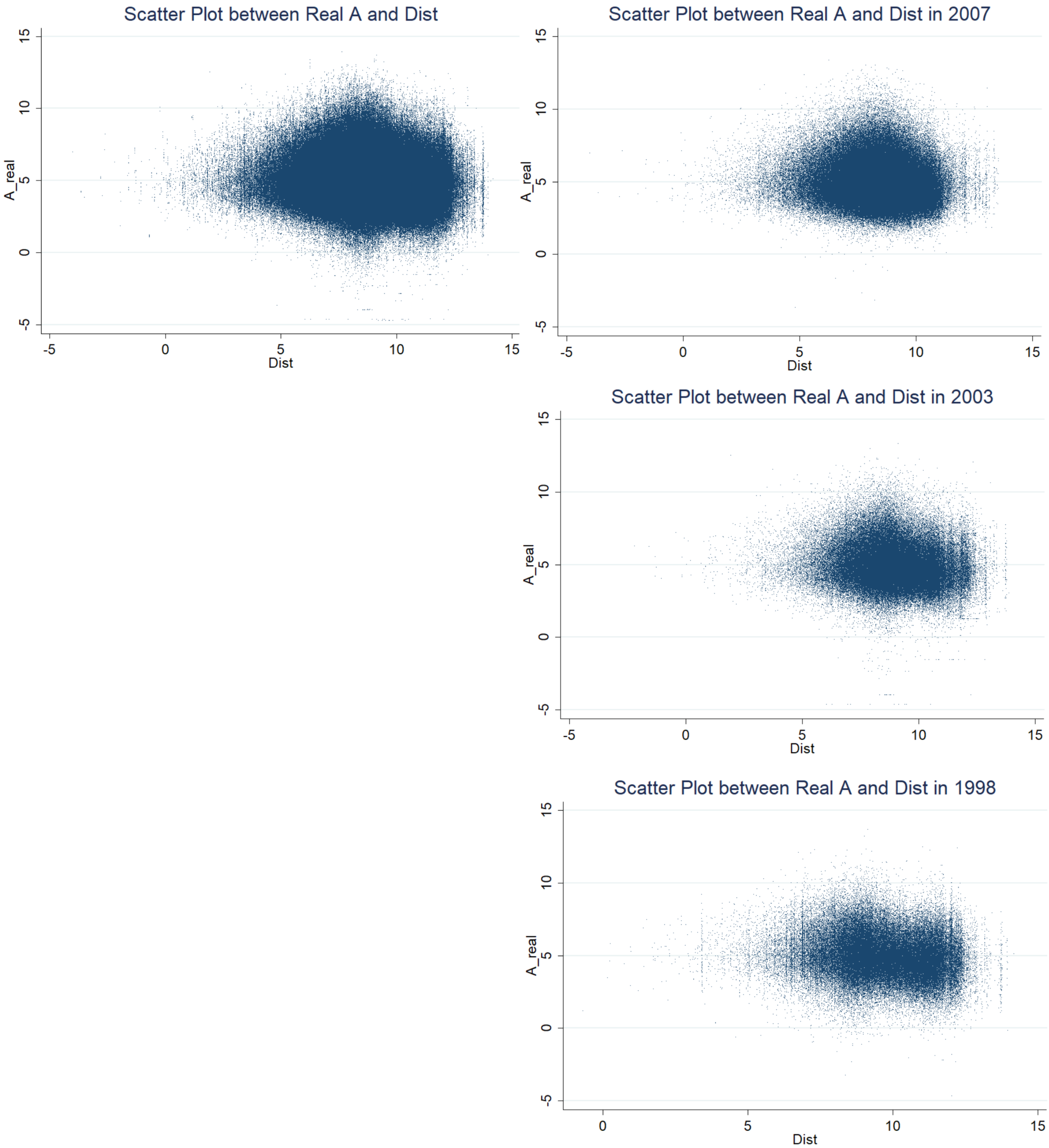


Figure D3: The left plot illustrates the relationship between $Dist$ and $Real A$; the right three plots show time-variant scatter plots in 1998, 2003, and 2007.

Figure D4: Older firms tend to employ more employees, and have higher dispersion.

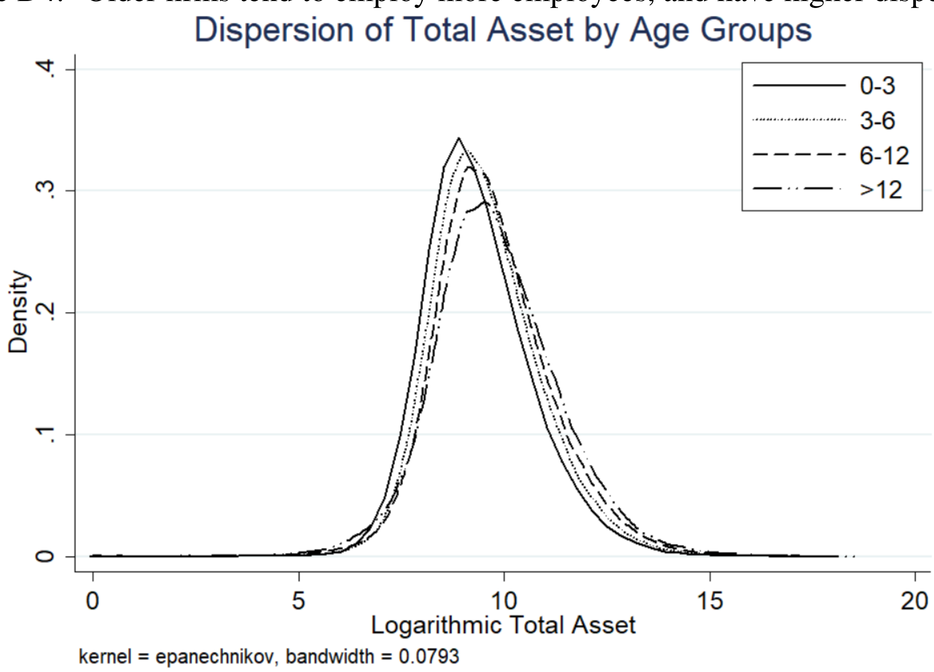


Figure D5: Older firms tend to have more assets, and have higher dispersion.

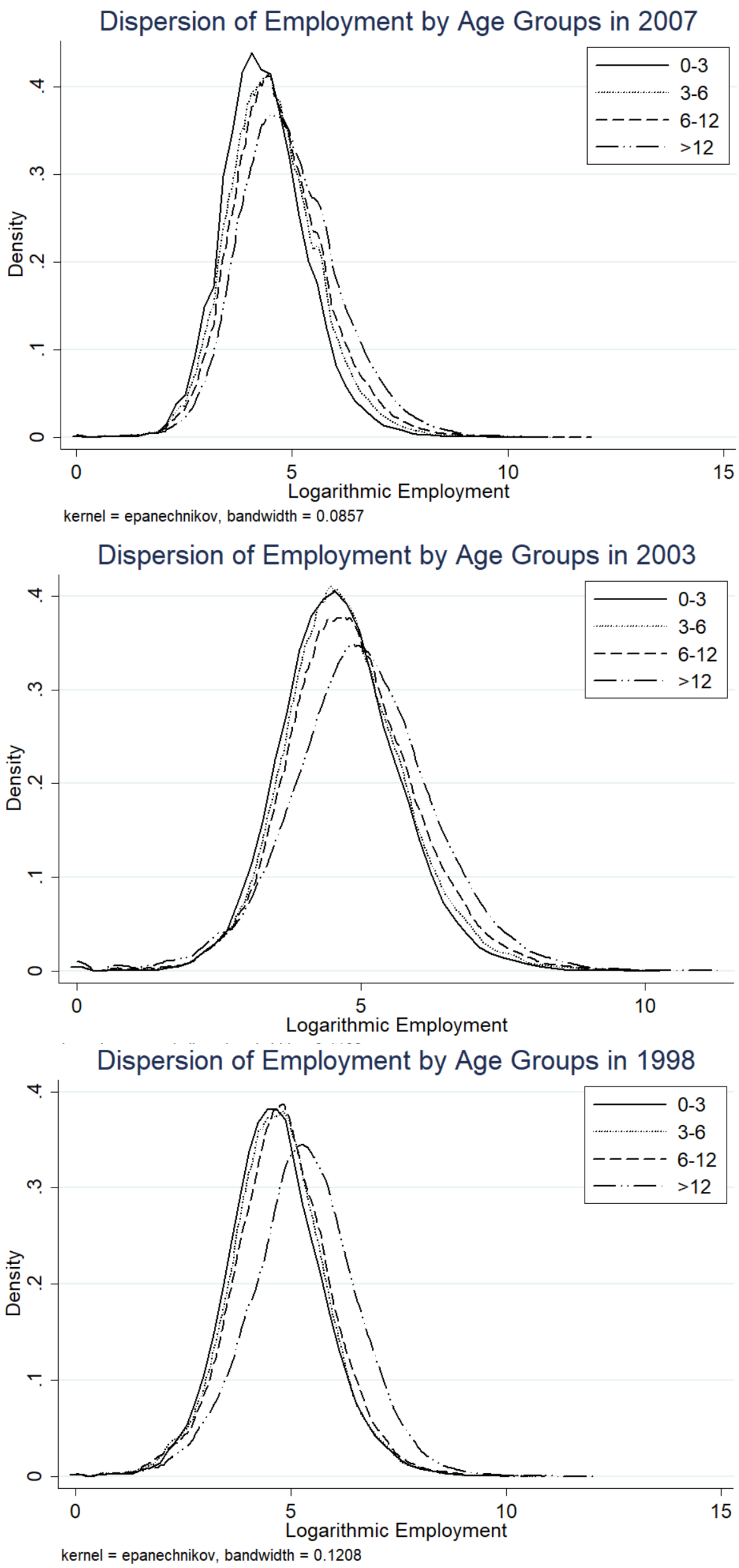


Figure D6: Firm size is indicated by total asset in 2007, 2003, and 1998; firms are divided into four age groups.

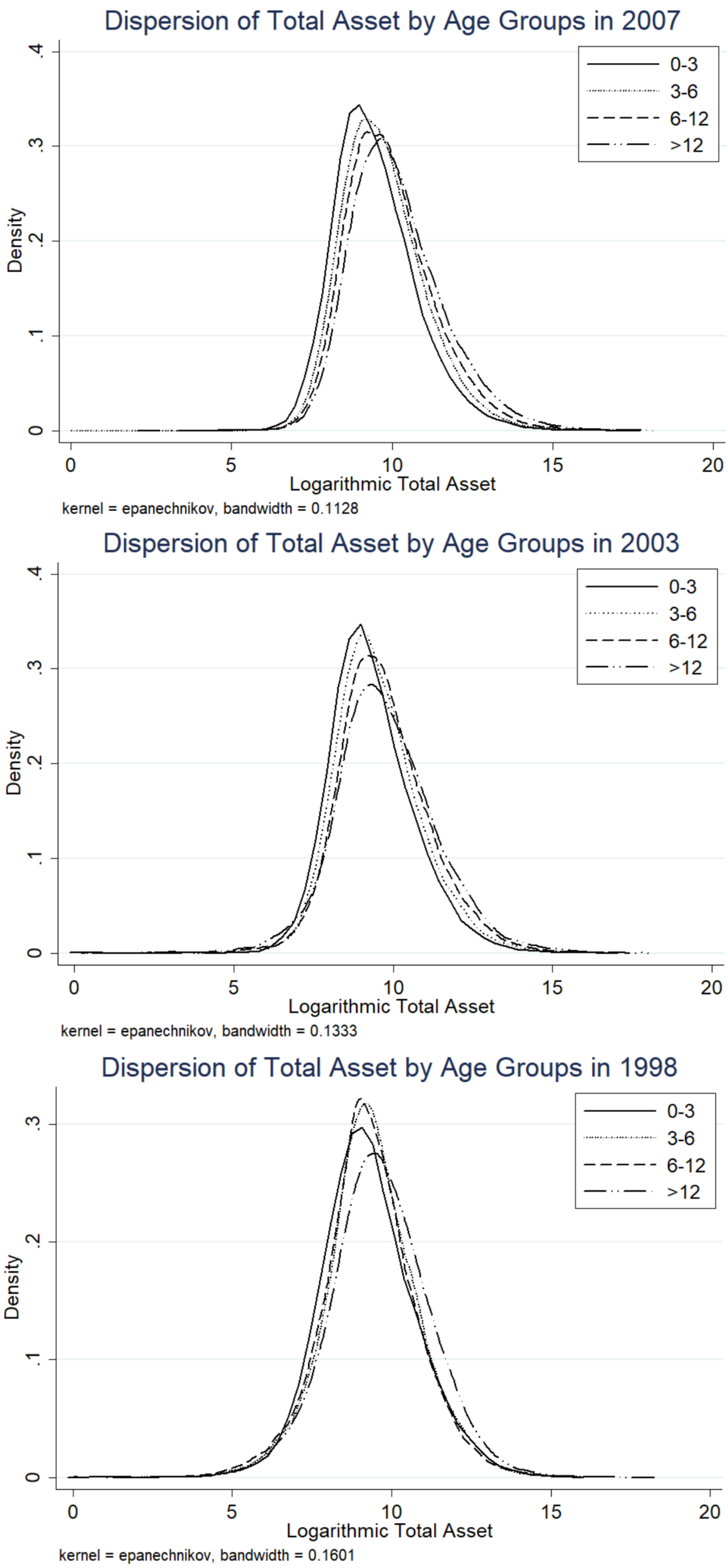


Figure D7: Firm size in indicated by total asset in 2007, 2003, and 1998; firms are divided into four age groups.

.2 Appendix E

Additional Results

Table E1: Quantile Regressions of Size Effects.

Dependent Variables:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
L	10th	20th	30th	40th	50th	60th	70th	80th	90th
Dist	0.00400***	0.000770	-0.00154*	-0.00357***	-0.00551***	-0.00752***	-0.00973***	-0.0124***	-0.0163***
	(0.00114)	(0.000889)	(0.000746)	(0.000665)	(0.000641)	(0.000678)	(0.000780)	(0.000962)	(0.00128)
Rail-Density	-0.187***	-0.166***	-0.151***	-0.137***	-0.124***	-0.111***	-0.0968***	-0.0791***	-0.0536***
	(0.00472)	(0.00368)	(0.00309)	(0.00275)	(0.00266)	(0.00281)	(0.00323)	(0.00398)	(0.00531)
Road-Density	0.00836	-0.00407	-0.0130***	-0.0208***	-0.0282***	-0.0359***	-0.0444***	-0.0548***	-0.0698***
	(0.00490)	(0.00382)	(0.00321)	(0.00286)	(0.00276)	(0.00291)	(0.00335)	(0.00414)	(0.00551)
River-Density	-0.0524***	-0.0582***	-0.0623***	-0.0659***	-0.0694***	-0.0730***	-0.0769***	-0.0817***	-0.0887***
	(0.00146)	(0.00114)	(0.000957)	(0.000852)	(0.000822)	(0.000869)	(0.001000)	(0.00123)	(0.00164)
Pop	0.0537***	0.0552***	0.0563***	0.0573***	0.0582***	0.0592***	0.0602***	0.0615***	0.0633***
	(0.00323)	(0.00252)	(0.00212)	(0.00189)	(0.00182)	(0.00192)	(0.00221)	(0.00273)	(0.00363)
Export	0.432***	0.485***	0.524***	0.558***	0.590***	0.624***	0.661***	0.706***	0.771***
	(0.00420)	(0.00328)	(0.00275)	(0.00245)	(0.00237)	(0.00250)	(0.00288)	(0.00355)	(0.00473)
State-share	-0.0650***	0.0591***	0.148***	0.226***	0.300***	0.377***	0.462***	0.566***	0.715***
	(0.0106)	(0.00825)	(0.00693)	(0.00618)	(0.00596)	(0.00630)	(0.00724)	(0.00893)	(0.0119)
Collective-share	0.137***	0.144***	0.149***	0.153***	0.157***	0.162***	0.166***	0.172***	0.181***
	(0.00641)	(0.00500)	(0.00420)	(0.00374)	(0.00361)	(0.00381)	(0.00439)	(0.00541)	(0.00721)
Foreign-share	0.188***	0.226***	0.254***	0.279***	0.302***	0.326***	0.353***	0.385***	0.432***
	(0.00529)	(0.00412)	(0.00346)	(0.00309)	(0.00298)	(0.00315)	(0.00362)	(0.00446)	(0.00595)
Age	-0.000439	0.000494	0.00116***	0.00175***	0.00231***	0.00288***	0.00352***	0.00430***	0.00542***
	(0.000386)	(0.000301)	(0.000253)	(0.000225)	(0.000217)	(0.000230)	(0.000264)	(0.000326)	(0.000435)
CFK	-0.000000799	-0.000000649	-0.000000542	-0.000000447	-0.000000357	-0.000000265	-0.000000162	-3.67e-08	0.000000144
	(0.000000650)	(0.000000507)	(0.000000425)	(0.000000379)	(0.000000365)	(0.000000386)	(0.000000445)	(0.000000548)	(0.000000731)
Observation	1254753	1254753	1254753	1254753	1254753	1254753	1254753	1254753	1254753
Fixed Effect	ind, year	ind, year	ind, year	ind, year	ind, year	ind, year	ind, year	ind, year	ind, year

Note: Observations are grouped by 4-digit industry and year fixed effects. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table E2: IV Regressions of Size Effects (Employment Size).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
L	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dist	-0.022**	-0.018*	-0.042***	-0.007	-0.007	-0.007	-0.009
	(0.008)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)	(0.006)
Rail-Density	0.112***	0.092***	0.063***	0.091***	0.070***	0.070***	0.065***
	(0.020)	(0.014)	(0.011)	(0.006)	(0.006)	(0.006)	(0.006)
Road-Density	-0.169***	-0.162***	-0.131***	-0.004	0.003	0.003	0.004
	(0.016)	(0.010)	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)
River-Density	0.069***	0.056***	0.052***	0.025***	0.027***	0.027***	0.029***
	(0.016)	(0.012)	(0.009)	(0.005)	(0.005)	(0.005)	(0.005)
Pop	0.011	0.052***	0.035***	0.053***	0.043***	0.043***	0.040***
	(0.011)	(0.010)	(0.010)	(0.004)	(0.004)	(0.004)	(0.004)
Export	0.623***	0.608***	0.567***	0.118***	0.104***	0.104***	0.102***
	(0.011)	(0.007)	(0.006)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	0.311***	0.341***	0.390***	0.106***	0.099***	0.099***	0.097***
	(0.016)	(0.014)	(0.012)	(0.006)	(0.006)	(0.006)	(0.006)
Collective-share	0.130***	0.125***	0.115***	0.014***	0.012***	0.012***	0.010**
	(0.008)	(0.007)	(0.006)	(0.003)	(0.003)	(0.003)	(0.003)
Foreign-share	0.269***	0.243***	0.250***	0.045***	0.037***	0.037***	0.035***
	(0.012)	(0.009)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)
Age	0.002***	0.002***	0.002***	0.000**	0.000*	0.000*	0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	1254245	1247772	1225588	1146617	1088381	1088175	1077752
Group	9481	44458	107003	294126	305855	305820	307056
Under-identification test	719.287	1092.693	1471.839	2647.006	2282.029	2271.889	2300.264
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.7e+04	4.4e+04	2.4e+04	1.2e+04	1.0e+04	1.0e+04	1.1e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.001	0.000	1.107	2.381	1.922	2.326	2.234
P value	(0.971)	(0.995)	(0.293)	(0.123)	(0.166)	(0.127)	(0.135)
First-stage Results:							
Ming30km	0.174***	0.213***	0.198***				
	(0.006)	(0.006)	(0.006)				
SL30km	0.198***	0.136***	0.122***	0.109***	0.107***	0.107***	0.130***
	(0.007)	(0.006)	(0.007)	(0.005)	(0.006)	(0.006)	(0.006)
LC50km				0.180***	0.178***	0.178***	0.178***
				(0.005)	(0.006)	(0.006)	(0.006)
F	557.798	404.101	472.446	1448.035	1271.905	1272.487	1255.407
R2	0.151	0.107	0.090	0.110	0.102	0.102	0.104
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. However, not all regressions can pass the Over-identification and Under-identification tests, suggesting some instruments combinations are not efficient in corresponding regressions, so the provided results are those regressions with efficient IVs. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table E3: IV Regressions of Size Effects (Asset Size).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect				
A	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Dist	-0.165***	-0.165***	-0.110***	0.005	0.008	0.008	0.003	
	(0.010)	(0.014)	(0.011)	(0.006)	(0.006)	(0.006)	(0.006)	
Rail-Density	-0.125***	-0.114***	-0.133***	-0.047***	-0.044***	-0.044***	-0.047***	
	(0.022)	(0.017)	(0.014)	(0.006)	(0.007)	(0.007)	(0.007)	
Road-Density	-0.023	-0.027*	-0.001	0.104***	0.104***	0.104***	0.107***	
	(0.018)	(0.012)	(0.010)	(0.005)	(0.005)	(0.005)	(0.005)	
River-Density	0.076***	0.053***	0.054***	0.020***	0.018***	0.019***	0.020***	
	(0.020)	(0.015)	(0.011)	(0.005)	(0.005)	(0.005)	(0.005)	
Pop	0.026*	-0.062***	-0.060***	-0.012*	-0.008	-0.008	-0.014**	
	(0.012)	(0.012)	(0.011)	(0.005)	(0.005)	(0.005)	(0.005)	
Export	0.602***	0.600***	0.574***	0.092***	0.079***	0.079***	0.076***	
	(0.017)	(0.011)	(0.008)	(0.002)	(0.002)	(0.002)	(0.002)	
State-share	0.527***	0.569***	0.605***	0.130***	0.124***	0.124***	0.127***	
	(0.022)	(0.017)	(0.015)	(0.006)	(0.006)	(0.006)	(0.006)	
Collective-share	0.081***	0.086***	0.097***	0.018***	0.019***	0.019***	0.018***	
	(0.011)	(0.009)	(0.008)	(0.004)	(0.004)	(0.004)	(0.004)	
Foreign-share	0.715***	0.683***	0.654***	0.062***	0.053***	0.052***	0.050***	
	(0.018)	(0.012)	(0.010)	(0.006)	(0.006)	(0.006)	(0.006)	
Age	0.002***	0.002***	0.002***	0.000	0.000	0.000	0.000	
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
CFK	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
LDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Observation	1254245	1247772	1225588	1146617	1088381	1088175	1077752	
Group	9481	44458	107003	294126	305855	305820	307056	
Under-identification test	719.287	1160.631	1471.839	2647.006	2282.029	2271.889	2300.264	
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Weak identification test	6.7e+04	2.2e+04	2.4e+04	1.2e+04	1.0e+04	1.0e+04	1.1e+04	
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Over-identification test	0.619	0.552	0.083	1.074	1.057	0.790	0.000	
P value	(0.432)	(0.458)	(0.773)	(0.300)	(0.304)	(0.374)	(0.984)	
First-stage Results:								
Ming30km	0.174***		0.198***					
	(0.006)		(0.006)					
SL30km	0.198***	0.185***	0.122***	0.109***	0.107***	0.107***	0.130***	
	(0.007)	(0.006)	(0.007)	(0.005)	(0.006)	(0.006)	(0.006)	
Qing50km		0.120***						
		(0.006)						
LC50km				0.180***	0.178***	0.178***	0.178***	
				(0.005)	(0.006)	(0.006)	(0.006)	
F	557.798	345.166	472.446	1448.035	1271.905	1272.487	1255.407	
R2	0.151	0.075	0.090	0.110	0.102	0.102	0.104	
Fixed Effects:								
Region	province	prefecture	county		province	prefecture	county	
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit	
Firm				Yes	Yes	Yes	Yes	
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table E4: Robustness Checks of Size Effects (Unchanged Addresses).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
L	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dist	-0.116***	-0.081***	-0.022	0.024	0.025	0.025	0.030
	(0.014)	(0.017)	(0.025)	(0.034)	(0.040)	(0.040)	(0.041)
Rail-Density	0.156***	0.142***	0.117***	0.121***	0.099***	0.099***	0.091***
	(0.022)	(0.020)	(0.017)	(0.010)	(0.011)	(0.011)	(0.011)
Road-Density	-0.010	0.012	0.043*	0.068***	0.071***	0.071***	0.067***
	(0.027)	(0.024)	(0.022)	(0.014)	(0.014)	(0.014)	(0.014)
River-Density	0.035	0.053*	0.086***	0.094***	0.106***	0.106***	0.114***
	(0.024)	(0.022)	(0.019)	(0.012)	(0.014)	(0.014)	(0.014)
Pop	-0.056**	0.110**	0.090**	0.075***	0.069***	0.068***	0.064***
	(0.018)	(0.035)	(0.031)	(0.016)	(0.016)	(0.016)	(0.016)
Export	0.688***	0.650***	0.540***	0.143***	0.131***	0.131***	0.130***
	(0.017)	(0.015)	(0.015)	(0.009)	(0.010)	(0.010)	(0.010)
State-share	0.085***	0.116***	0.166***	0.062***	0.057***	0.057***	0.053***
	(0.023)	(0.021)	(0.023)	(0.012)	(0.012)	(0.012)	(0.012)
Collective-share	-0.050***	-0.044***	-0.042**	-0.014	-0.012	-0.012	-0.014
	(0.014)	(0.013)	(0.014)	(0.008)	(0.008)	(0.008)	(0.008)
Foreign-share	0.085***	0.120***	0.135***	0.067***	0.060**	0.060**	0.055**
	(0.023)	(0.022)	(0.024)	(0.018)	(0.019)	(0.019)	(0.018)
Age	0.001*	0.001*	0.001	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	132839	128088	119886	115136	107031	107028	105789
Group	5600	15104	23254	31572	31701	31700	31820
Under-identification test	404.054	610.865	344.303	69.727	58.614	58.614	57.374
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	7206.946	6102.002	3587.261	886.474	762.234	762.213	738.347
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	2.806	0.058	0.250	0.001	0.002	0.002	0.031
P value	(0.094)	(0.810)	(0.617)	(0.977)	(0.967)	(0.967)	(0.859)
First-stage Results:							
SL30km	0.142***			0.705***	0.612***	0.612***	0.596***
	(0.012)			(0.084)	(0.074)	(0.074)	(0.074)
LC50km	0.161***	0.175***	0.187***	0.032	0.147	0.147	0.157
	(0.012)	(0.009)	(0.017)	(0.113)	(0.108)	(0.108)	(0.107)
Ming30km		0.216***	0.224***				
		(0.009)	(0.013)				
F	240.314	166.027	131.160	222.816	198.302	198.290	198.242
R2	0.195	0.140	0.129	0.170	0.164	0.164	0.166
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table E5: Robustness Checks of Size Effects (Unchanged Addresses, Asset Size).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
A	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dist	-0.197***	-0.192***	-0.043	0.066*	0.075*	0.075*	0.076*
	(0.015)	(0.022)	(0.031)	(0.027)	(0.033)	(0.033)	(0.034)
Rail-Density	-0.030	-0.014	-0.020	0.000	0.003	0.003	-0.003
	(0.027)	(0.023)	(0.020)	(0.010)	(0.010)	(0.010)	(0.010)
Road-Density	0.131***	0.120***	0.141***	0.141***	0.129***	0.129***	0.130***
	(0.033)	(0.030)	(0.026)	(0.013)	(0.014)	(0.014)	(0.014)
River-Density	0.004	0.019	0.022	0.028**	0.030**	0.030**	0.035***
	(0.030)	(0.028)	(0.022)	(0.009)	(0.010)	(0.010)	(0.010)
Pop	0.010	0.060	0.016	0.029*	0.031*	0.031*	0.028*
	(0.021)	(0.034)	(0.029)	(0.014)	(0.014)	(0.014)	(0.014)
Export	0.671***	0.637***	0.523***	0.094***	0.089***	0.089***	0.086***
	(0.024)	(0.021)	(0.020)	(0.008)	(0.008)	(0.008)	(0.008)
State-share	0.234***	0.286***	0.310***	0.071***	0.070***	0.070***	0.072***
	(0.031)	(0.028)	(0.029)	(0.013)	(0.013)	(0.013)	(0.013)
Collective-share	-0.152***	-0.118***	-0.100***	-0.025**	-0.024**	-0.024**	-0.024**
	(0.017)	(0.017)	(0.017)	(0.008)	(0.009)	(0.009)	(0.009)
Foreign-share	0.533***	0.545***	0.477***	0.043*	0.038	0.038	0.032
	(0.033)	(0.029)	(0.032)	(0.019)	(0.019)	(0.019)	(0.019)
Age	0.001*	0.001*	0.001	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	132839	128088	119886	115136	107031	107028	105789
Group	5600	15104	23254	31572	31701	31700	31820
Under-identification test	471.358	610.865	344.303	69.727	58.614	58.614	57.374
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.0e+04	6102.002	3587.261	886.474	762.234	762.213	738.347
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	3.327	1.281	0.029	0.412	0.465	0.465	0.379
P value	(0.068)	(0.258)	(0.866)	(0.521)	(0.496)	(0.495)	(0.538)
First-stage Results:							
Ming30km	0.180***	0.216***	0.224***				
	(0.008)	(0.009)	(0.013)				
LC50km	0.216***	0.175***	0.187***	0.032	0.147	0.147	0.157
	(0.007)	(0.009)	(0.017)	(0.113)	(0.108)	(0.108)	(0.107)
SL60n30km				0.705***	0.612***	0.612***	0.596***
				(0.084)	(0.074)	(0.074)	(0.074)
F	261.085	166.027	131.160	222.816	198.302	198.290	198.242
R2	0.228	0.140	0.129	0.170	0.164	0.164	0.166
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table E6: Highways' Impacts on Firm Size Dispersion (75-25th IQ Range).

Dependent Variable: L IQ Range	(1)	(2)	(3)	(4)	(5)
Main Results:					
Dist	-0.117***	-0.081***	-0.125***	-0.130***	-0.038*
	(0.020)	(0.020)	(0.017)	(0.015)	(0.017)
Road-Density	-0.028	-0.026	-0.040**	-0.030	-0.042
	(0.043)	(0.029)	(0.015)	(0.023)	(0.022)
Rail-Density	-0.063	-0.058	0.102***	-0.060*	-0.045
	(0.039)	(0.034)	(0.013)	(0.026)	(0.025)
Pop	-0.020	0.002	-0.003	-0.022	0.012
	(0.019)	(0.021)	(0.014)	(0.015)	(0.023)
State-share	0.433***	0.417***	0.372***	0.406***	0.355***
	(0.044)	(0.032)	(0.042)	(0.029)	(0.027)
Collective-share	-0.068*	-0.069	-0.031	-0.032	-0.034
	(0.029)	(0.035)	(0.040)	(0.033)	(0.033)
Foreign-share	0.184***	0.176***	0.229***	0.229***	0.197***
	(0.043)	(0.047)	(0.047)	(0.041)	(0.049)
Age	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFK	-0.000***	-0.000**	-0.000**	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SDK	0.000**	0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LDK	-0.000	-0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observation	54026	54026	53103	54011	53657
Group	31	290	30	892	6776
Under-identification test	12.993	94.771	27.725	303.546	1001.752
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5387.854	4193.597	4446.237	4679.707	5673.647
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.385	3.266	0.018	1.227	1.714
P value	(0.535)	(0.071)	(0.892)	(0.268)	(0.190)
First-stage Results:					
Qing50km	0.112*	0.300***	0.049***	0.103***	0.073***
	(0.041)	(0.029)	(0.006)	(0.010)	(0.017)
SL50km	0.307***	0.337***	0.306***	0.293***	
	(0.033)	(0.038)	(0.006)	(0.009)	
Ming50km					0.427***
					(0.013)
F	479.185	48.989	1891.626	560.997	542.937
R2	0.332	0.305	0.300	0.324	0.411
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: This table illustrates IV results. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table E7: Highways' Impacts on Firm Size Dispersion (HHI).

Dependent Variable: L HHI	(1)	(2)	(3)	(4)	(5)
Main Results:					
Dist	-0.154**	-0.622***	-0.141***	0.005	-0.217***
	(0.050)	(0.050)	(0.028)	(0.017)	(0.017)
Road-Density	-0.017	0.049	-0.240***	-0.042	0.008
	(0.085)	(0.074)	(0.038)	(0.034)	(0.020)
Rail-Density	-0.027	-0.069	-0.004	-0.005	-0.034
	(0.075)	(0.078)	(0.037)	(0.035)	(0.022)
Pop	-0.425***	-0.123	-0.396***	-0.365***	-0.069***
	(0.058)	(0.086)	(0.029)	(0.021)	(0.017)
State-share	0.316***	0.259***	0.398***	0.244***	0.034*
	(0.046)	(0.034)	(0.045)	(0.025)	(0.017)
Collective-share	0.049	0.058	0.098*	0.043	0.047*
	(0.050)	(0.039)	(0.050)	(0.032)	(0.019)
Foreign-share	-0.364*	0.080	-0.838***	-0.433***	0.093**
	(0.177)	(0.072)	(0.080)	(0.058)	(0.032)
Age	-0.000	-0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFK	0.000	0.000	-0.000	-0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SDK	0.000*	0.000*	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LDK	-0.000***	-0.000***	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observation	62203	62203	61165	62192	61901
Group	31	290	30	919	7612
Under-identification test	15.650	96.230	28.165	384.421	690.069
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	7722.211	3865.700	5316.548	8712.423	2674.526
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.370	0.388	3.805	10.408	0.882
P value	(0.543)	(0.533)	(0.051)	(0.001)	(0.348)
First-stage Results:					
Qing50km	0.052	0.300***	-0.020**	0.025*	0.323***
	(0.042)	(0.029)	(0.006)	(0.010)	(0.014)
Ming50km	0.383***				
	(0.052)				
SL50km		0.337***			0.330***
		(0.038)			(0.018)
Ming30km			0.308***	0.419***	
			(0.009)	(0.012)	
F	361.633	48.989	1814.244	632.420	460.421
R2	0.381	0.305	0.325	0.399	0.349
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: This table illustrates IV results. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

.3 Appendix F

Additional Robustness Checks

Table F1: Highways' Impacts on Firm Size Dispersion (75-25th IQ Range, Asset Size).

Dependent Variable: A IQ Range	(1)	(2)	(3)	(4)	(5)
Main Results:					
Dist	-0.157***	-0.043*	-0.175***	-0.178***	-0.058*
	(0.033)	(0.020)	(0.022)	(0.017)	(0.025)
Road-Density	-0.011	-0.020	-0.002	-0.016	-0.025
	(0.040)	(0.029)	(0.020)	(0.027)	(0.026)
Rail-Density	-0.059	-0.045	0.081***	-0.057	-0.042
	(0.044)	(0.033)	(0.016)	(0.031)	(0.031)
Pop	-0.033	0.026	-0.020	-0.038*	0.021
	(0.029)	(0.027)	(0.015)	(0.017)	(0.027)
State-share	0.392***	0.398***	0.299***	0.344***	0.281***
	(0.049)	(0.036)	(0.046)	(0.035)	(0.031)
Collective-share	-0.181***	-0.173***	-0.157***	-0.151***	-0.091*
	(0.039)	(0.037)	(0.047)	(0.038)	(0.037)
Foreign-share	0.105*	0.079	0.216***	0.234***	0.217***
	(0.053)	(0.045)	(0.040)	(0.046)	(0.057)
Age	0.000	0.000	0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFK	-0.000	-0.000	-0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SDK	-0.000	-0.000	-0.000	-0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LDK	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observation	53320	53320	52424	53312	52963
Group	30	289	30	889	6728
Under-identification test	12.861	94.662	27.641	298.500	788.089
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	5301.666	4170.294	4412.806	4609.381	2796.589
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.105	0.030	0.316	0.902	0.969
P value	(0.746)	(0.862)	(0.574)	(0.342)	(0.325)
First-stage Results:					
Qing50km	0.112*	0.300***	0.049***	0.103***	0.323***
	(0.041)	(0.029)	(0.006)	(0.010)	(0.014)
SL50km	0.307***	0.337***	0.306***	0.293***	0.330***
	(0.033)	(0.038)	(0.006)	(0.009)	(0.018)
F	479.185	48.989	1891.626	560.997	460.421
R2	0.332	0.305	0.300	0.324	0.349
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: This table illustrates IV results. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table F2: Highways' Impacts on Firm Size Dispersion (HHI, Asset Size).

Dependent Variable: A HHI	(1)	(2)	(3)	(4)	(5)
Main Results:					
Dist	-0.135***	-0.507***	-0.117***	-0.005	-0.172***
	(0.041)	(0.041)	(0.024)	(0.014)	(0.016)
Road-Density	0.007	0.063	-0.196***	-0.016	0.029
	(0.070)	(0.061)	(0.031)	(0.029)	(0.020)
Rail-Density	-0.012	-0.043	-0.002	0.005	-0.006
	(0.052)	(0.066)	(0.030)	(0.029)	(0.022)
Pop	-0.348***	-0.080	-0.319***	-0.300***	-0.036*
	(0.049)	(0.067)	(0.025)	(0.019)	(0.016)
State-share	0.306***	0.246***	0.367***	0.235***	0.015
	(0.041)	(0.030)	(0.041)	(0.022)	(0.016)
Collective-share	0.016	0.010	0.052	-0.002	-0.012
	(0.042)	(0.034)	(0.045)	(0.028)	(0.018)
Foreign-share	-0.339*	0.007	-0.691***	-0.351***	0.055
	(0.173)	(0.066)	(0.077)	(0.056)	(0.030)
Age	-0.000	-0.000	0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFK	0.000	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SDK	0.000***	0.000**	0.000*	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LDK	-0.000***	-0.000***	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observation	61680	61680	60656	61677	61398
Group	30	289	30	919	7586
Under-identification test	15.605	96.421	28.105	381.552	681.837
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	7598.824	3848.028	5229.548	8567.711	2666.126
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.520	0.185	3.905	11.792	1.143
P value	(0.471)	(0.667)	(0.048)	(0.001)	(0.285)
First-stage Results:					
Qing50km	0.052	0.300***	-0.020**	0.025*	0.323***
	(0.042)	(0.029)	(0.006)	(0.010)	(0.014)
Ming50km	0.383***				
	(0.052)				
SL50km		0.337***			0.330***
		(0.038)			(0.018)
Ming30km			0.308***	0.419***	
			(0.009)	(0.012)	
F	361.633	48.989	1814.244	632.420	460.421
R2	0.381	0.305	0.325	0.399	0.349
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: This table illustrates IV results. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table F3: Highways' One-year-lagged Impacts (75-25th IQ Range).

Dependent Variable: L IQ Range	(1)	(2)	(3)	(4)	(5)
Main Results:					
Lag-Dist	-1.149***	-0.980***	-1.173***	-1.285***	-0.906*
	(0.189)	(0.262)	(0.184)	(0.167)	(0.396)
Road-Density	-0.035	-0.031	-0.035*	-0.039	-0.042
	(0.043)	(0.031)	(0.016)	(0.025)	(0.023)
Rail-Density	-0.070	-0.072*	0.100***	-0.073*	-0.060*
	(0.039)	(0.036)	(0.013)	(0.029)	(0.028)
Pop	-0.025	0.010	-0.002	-0.028	0.016
	(0.019)	(0.022)	(0.015)	(0.016)	(0.026)
State-share	0.424***	0.403***	0.364***	0.407***	0.351***
	(0.045)	(0.035)	(0.041)	(0.031)	(0.030)
Collective-share	-0.068*	-0.069	-0.037	-0.032	-0.046
	(0.030)	(0.038)	(0.043)	(0.036)	(0.037)
Foreign-share	0.190***	0.184***	0.224***	0.234***	0.191***
	(0.045)	(0.050)	(0.050)	(0.044)	(0.053)
Age	0.000	0.000	0.000*	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFK	-0.000**	-0.000	-0.000*	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SDK	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LDK	0.000***	0.000***	0.000**	0.000**	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observation	47755	47755	47029	47739	47426
Group	31	290	30	884	6627
Under-identification test	11.772	82.737	27.472	256.430	427.036
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	3698.739	1713.505	3214.985	3084.636	719.881
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.831	1.927	0.051	2.237	0.143
P value	(0.362)	(0.165)	(0.821)	(0.135)	(0.706)
First-stage Results:					
Qing50km	0.009*	0.023***	0.003***	0.008***	0.006**
	(0.004)	(0.003)	(0.001)	(0.001)	(0.002)
SL50km	0.032***	0.029***	0.032***	0.030***	
	(0.004)	(0.003)	(0.001)	(0.001)	
Ming50km					0.020***
					(0.001)
F	1361.652	63.821	1014.141	560.795	438.168
R2	0.298	0.229	0.267	0.287	0.270
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: This table illustrates IV results. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table F4: Highways' One-year-lagged Impacts (HHI).

Dependent Variable: L HHI	(1)	(2)	(3)	(4)	(5)
Main Results:					
Lag-Dist	-1.795**	-7.990***	-1.566***	0.050	-4.686***
	(0.574)	(0.644)	(0.336)	(0.208)	(0.433)
Road-Density	-0.051	0.045	-0.228***	-0.072*	0.021
	(0.080)	(0.100)	(0.038)	(0.034)	(0.026)
Rail-Density	-0.054	-0.111	-0.008	-0.015	-0.063*
	(0.080)	(0.124)	(0.037)	(0.037)	(0.029)
Pop	-0.454***	-0.128	-0.402***	-0.373***	-0.102***
	(0.065)	(0.126)	(0.031)	(0.024)	(0.027)
State-share	0.396***	0.309***	0.457***	0.288***	0.059**
	(0.043)	(0.040)	(0.049)	(0.027)	(0.022)
Collective-share	0.066	0.049	0.152**	0.070	0.011
	(0.057)	(0.046)	(0.052)	(0.036)	(0.026)
Foreign-share	-0.363	0.099	-0.846***	-0.424***	0.097*
	(0.193)	(0.079)	(0.083)	(0.061)	(0.041)
Age	-0.000	-0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFK	0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
SDK	0.000***	0.000***	0.000**	0.000*	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
LDK	-0.000***	-0.000***	-0.000***	-0.000***	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observation	53603	53603	52794	53593	53327
Group	31	290	30	915	7323
Under-identification test	14.851	82.869	27.585	326.269	212.448
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	3948.120	1660.313	2658.391	4029.395	431.794
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.478	1.724	3.209	9.911	1.191
P value	(0.489)	(0.189)	(0.073)	(0.002)	(0.275)
First-stage Results:					
Qing50km	0.006	0.023***	-0.001	0.004***	0.017***
	(0.005)	(0.003)	(0.001)	(0.001)	(0.002)
Ming50km	0.033***				
	(0.006)				
SL50km		0.029***			0.017***
		(0.003)			(0.002)
Ming30km			0.026***	0.035***	
			(0.001)	(0.001)	
F	846.836	63.821	790.703	545.572	380.718
R2	0.310	0.229	0.259	0.313	0.260
Fixed Effects:					
Region	province	prefecture		province	prefecture
2digit Industry			Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes

Note: This table illustrates IV results. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table F5: Highways' One-year-lagged Impacts (Different Inter-percentiles).

Dependent Variable: L IQ Range	60-40th (1)	70-30th (2)	80-20th (3)	90-10th (4)
Main Results:				
Lag-Dist	-1.545**	-1.510*	-1.426**	-1.420*
	(0.562)	(0.594)	(0.553)	(0.621)
Road-Density	0.012	-0.001	-0.055*	-0.061*
	(0.028)	(0.026)	(0.024)	(0.029)
Rail-Density	-0.009	-0.056	-0.077*	-0.087**
	(0.037)	(0.034)	(0.030)	(0.034)
Pop	0.043	0.042	-0.010	-0.020
	(0.031)	(0.029)	(0.027)	(0.031)
State-share	0.220***	0.264***	0.355***	0.434***
	(0.033)	(0.032)	(0.031)	(0.036)
Collective-share	-0.031	-0.092*	-0.050	-0.084
	(0.043)	(0.041)	(0.039)	(0.045)
Foreign-share	0.136*	0.144*	0.208***	0.308***
	(0.058)	(0.057)	(0.057)	(0.067)
Age	-0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)
CFK	-0.000*	-0.000	-0.000	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
SDK	-0.000*	-0.000	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
LDK	0.000**	0.000	0.000*	0.000*
	(0.000)	(0.000)	(0.000)	(0.000)
Observation	43217	47378	47428	47428
Group	6409	6627	6627	6627
Under-identification test	261.803	292.207	292.305	292.305
P value	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	405.731	413.939	413.623	413.623
P value	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.059	0.398	2.721	2.788
P value	(0.808)	(0.528)	(0.099)	(0.095)
First-stage Results:				
Qing50km	0.017***	0.017***	0.017***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
SL50km	0.017***	0.017***	0.017***	0.017***
	(0.002)	(0.002)	(0.002)	(0.002)
F	380.718	380.718	380.718	380.718
R2	0.260	0.260	0.260	0.260
Fixed Effects:				
Region	prefecture	prefecture	prefecture	prefecture
2digit Industry	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes

Note: This table illustrates IV results. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. This table illustrates OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

.4 Appendix G

Channel Studies

Table G1: Summary Statistics of Firm Age.

	age(min) Percentiles	Smallest			age(max) Percentiles	Smallest		
1%	0	-530			-2	-7963		
5%	1	-6			0	-7962		
10%	2	-5	Obs	1845097	1	-7961	Obs	1845097
25%	4	-3	Sum of Wgt.	1845097	3	-7960	Sum of Wgt.	1845097
50%	7		Mean	19.95461	5		Mean	8.950067
		Largest	Std. Dev.	130.2284		Largest	Std. Dev.	38.44301
75%	14	2007			10	2002		
90%	29	2007	Variance	16959.43	20	2002	Variance	1477.865
95%	42	2007	Skewness	14.42475	33	2005	Skewness	-49.51463
99%	55	2007	Kurtosis	213.9021	49	2006	Kurtosis	15892.35

	age(med) Percentiles	Smallest			age Percentiles	Smallest		
1%	0	-5305			0	-7960		
5%	1	-99			1	-7959	Obs	
10%	1	-98	Obs	1845097	1	-7883	Obs	1843909
25%	3	-97	Sum of Wgt.	1845097	3	-7882	Sum of Wgt.	1843909
50%	6		Mean	10.89742	6	Mean	12.11244	
		Largest	Std. Dev.	38.59637		Largest	Std. Dev.	67.05038
75%	12	2002			12	2006		
90%	25	2002	Variance	1489.68	25	2006	Variance	4495.754
95%	38	2005	Skewness	37.40112	38	2006	Skewness	21.8372
99%	50	2006	Kurtosis	1972.56	50	2006	Kurtosis	1274.462

Note: Firm age is calculated basing one four approaches. age equals current year minus start up year. However, some firms report different start up year across sample period, e.g a firm may report its start up year as 1978 in 1998, but report 0 in 2000, suggesting the start up year is inconsistent over the panel dataset. $age(min)$ equals to current year minus a firm's minimum reported start up year over sample period; similarly, $age(max)$ equals to current year minus the largest reported start up year, $age(med)$ equals to current year minus the median reported start up year over sample period.

Table G2: Summary Statistics of Channel Indicators.

	Observations	Full Sample Mean (Std. Dev.)	Observations	1999 Mean (Std. Dev.)	Observations	2003 Mean (Std. Dev.)	Observations	2006 Mean (Std. Dev.)
Channel Indicators:								
Outsourcing	1810756	.8379348	118261	.829476	136955	.9137875	269970	.7591956
		(35.48325)		(5.518223)		(18.53354)		(1.717594)
Entry	1845428	.1952875	123318	.1799494	139442	.251151	271983	.1916149
		(.396422)		(.3841469)		(.4336768)		(.3935724)

Note: Mean and standard deviation of the full sample and the sub-samples in different years (1999, 2003, 2006) are compared in this table.

Table G3: Channel Studies: Firm Entry as Dependent Variable.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Firm Entry	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dist	0.017***	0.012***	0.013***	0.023***	0.020***	0.021***	0.016***
	(0.001)	(0.001)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
Rail-Density	-0.008	-0.002	0.019*	0.068***	0.087***	0.087***	0.083***
	(0.009)	(0.008)	(0.007)	(0.005)	(0.005)	(0.005)	(0.005)
Road-Density	-0.039***	-0.038***	-0.037***	-0.100***	-0.118***	-0.118***	-0.113***
	(0.005)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
River-Density	-0.006	-0.009**	-0.009**	-0.006*	-0.004	-0.004	-0.004
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Pop	0.007**	0.038***	0.044***	0.015***	0.005	0.004	0.005
	(0.002)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Export	-0.080***	-0.084***	-0.087***	-0.057***	-0.056***	-0.056***	-0.056***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.168***	-0.168***	-0.166***	-0.085***	-0.079***	-0.079***	-0.076***
	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Collective-share	-0.110***	-0.108***	-0.104***	-0.037***	-0.032***	-0.032***	-0.033***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
Foreign-share	-0.035***	-0.037***	-0.037***	-0.008*	-0.008	-0.008	-0.008*
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)	(0.004)
Age	-0.001***	-0.001***	-0.001***	-0.000**	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
CFK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LDK	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	1254245	1247772	1225588	1146617	1088381	1088175	1077752
Group	9481	44458	107003	294126	305855	305820	307056
Under-identification test	719.287	1092.693	1471.839	2647.006	2282.029	2271.889	2300.264
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.7e+04	4.4e+04	2.4e+04	1.2e+04	1.0e+04	1.0e+04	1.1e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.065	1.652	2.636	0.413	0.412	0.408	0.432
P value	(0.799)	(0.199)	(0.105)	(0.520)	(0.521)	(0.523)	(0.511)
First-stage Results:							
Ming30km	0.174***	0.213***	0.198***				
	(0.006)	(0.006)	(0.006)				
SL30km	0.198***	0.136***	0.122***	0.109***	0.107***	0.107***	0.130***
	(0.007)	(0.006)	(0.007)	(0.005)	(0.006)	(0.006)	(0.006)
LC50km				0.180***	0.178***	0.178***	0.178***
				(0.005)	(0.006)	(0.006)	(0.006)
F	557.798	404.101	472.446	1448.035	1271.905	1272.487	1255.407
R2	0.151	0.107	0.090	0.110	0.102	0.102	0.104
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table G4: The Correlation of Inventory and Outsourcing with Firm Size.

Dependent Variable: L	(1)	(2)	(3)	(4)	(5)	(6)
Inventory	0.044***	0.039***				
	(0.001)	(0.001)				
Out			-0.012***	-0.009***		
			(0.002)	(0.002)		
Entry					-0.051***	-0.054***
					(0.001)	(0.001)
Dist	-0.000	-0.001	-0.001	-0.001	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Rail-Density	0.075***	0.041***	0.060***	0.028***	0.091***	0.056***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Road-Density	-0.047***	-0.037***	-0.053***	-0.040***	-0.049***	-0.039***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Pop	0.069***	0.048***	0.071***	0.051***	0.076***	0.054***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Age	0.000	-0.000	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Export	0.103***	0.090***	0.098***	0.085***	0.107***	0.093***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Size	0.308***	0.283***	0.330***	0.302***	0.338***	0.310***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	0.070***	0.064***	0.074***	0.067***	0.060***	0.053***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
Collective-share	0.004	-0.001	0.008**	0.004	0.000	-0.003
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Foreign-share	0.025***	0.018***	0.029***	0.022***	0.029***	0.022***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
R2	0.145	0.123	0.138	0.117	0.138	0.119
Observation	1479200	1461735	1529583	1511700	1541996	1523935
Group	458914	545396	469162	557962	471677	561388
F	2667.3	2148.2	2578.9	2125.1	2595.4	2084.0
Fixed Effects:						
Region		county		county		county
Industry		4-digit		4-digit		4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: This table provides OLS results. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Chapter 5

Impacts of Highways on Firms' Export Activities

5.1 Introduction

Since its accession into the World Trade Organization in 2001, China has witnessed significant trade explosions. Its total export value (RMB) increased by 276% from 2000 to 2006, while its recorded export enterprises (above the designated scale) rose from 62,746 to 171,205. At the same time, China's trade structure also changed dramatically; the proportion of imported-material export flow gradually decreased from 50% to 30%, while provided-material flow declined from 9% to 5%, implying that China's dramatic export exploration was accompanied with an endogenous transition or import substitution, from processing trade to ordinary and other trade modes. Meanwhile, China's total highway length increased by 300% from 1990 to 2007, while its expressway length also increased from zero to about 150,000 kilometers until 2019. However, the research of infrastructure construction, especially firm-level and spatial analysis, was limited. A wide range of trade literature concluded that tariff reductions and trade facilitation policy can encourage firms' trade activities, promote innovation and productivity increase.[1] However, firms' trade costs are not only decided by tariff and policy changes, but also the quantity and quality of physical infrastructure stocks. This chapter investigates the impacts of China's infrastructure explosion on firms' export activities and export structure transition.

The contribution of this chapter includes the following four aspects: First, most trade litera-

[1]Three channels can promote trade activities, i.e., quality adaptation, variety improvement, and knowledge spillover (Fan et al. 2015).

ture uses aggregate infrastructure data, the firm-level interpretation on the impacts of infrastructure development is insufficient. Similar to many existing trade literature, this chapter investigates export activities at the firm and transaction level, but adds firm-level geographical factors to study the role of infrastructure development. Three comprehensive nationwide datasets are merged, including a spatial dataset of China's expressway network, a customs transaction-level dataset, and a manufacturing firm dataset.

Second, we construct a novel measure of transportation costs based on China' road network data and terrain data. Early studies tended to emphasize the roles of distance, remoteness, and scales of destination markets (Baldwin & Harrigan 2011), while the infrastructure factor was usually indicated by aggregate proxies such as road density. Recent gravity models have developed more efficient spatial measurements, such as accumulated distance along traffic lines, the least cost routes, or road connection dummy variables. These studies tended to distinguish traffic accessibility from trade costs rigorously, some of them have developed new approaches to infer trade costs from price gaps, e.g., Donaldson (2018). However, most of these studies using aggregate data may neglect the impacts of firm-level characteristics, e.g., firms' ownership and firm-level location effects. Our new proxy is constructed by the accumulated travel cost approach, it can reflect firms' transportation costs along the road network to the nearest trade ports.

Third, to address potential endogeneity issues during China's highway expansion, we use China's historical roads and counter-factual roads to construct instrument variables, following the approaches of Holl (2016) and Faber (2014). The historical road network includes China's ancient roads in the Qing and Ming Dynasties, this exogenous factor is expected to have causal impacts on the spatial distribution of current roads. Counter-factual roads are constructed according to the cost-minimized rule (Faber 2014), this factor is also exogenous to the current road distribution. Fourth, based on the comprehensive firm-level data, we test the heterogeneous impacts of highway construction on firms with varying size, age, and ownership, and the impacts on different types of trade modes, e.g., processing trade or ordinary trade. As the unit weight of goods may affect the unit transportation cost, we also investigate the role of value-weight ratio across different manufacture industries. These heterogeneous impacts have not been fully specified by previous infrastructure literature.

Our empirical conclusions can be summarized from the following aspects. First, our baseline regressions show that, from 2000 to 2006, firms closer to seaports (or coastal firms) are more likely to involve in international trade, have higher firm-level export values, and higher exports-sales ratios. When we consider the export intensive and extensive margins, coastal firms normally export more types of different products, but focus on a smaller number of destinations, these results stay robust even if the proxy of transportation cost is substituted by the proxies used by Huang & Xu (2012). Consequently, firms closer to seaports tend to have lower export intensive margins, i.e., export volume of products, export volume of products in each destination. This fact is also partly consistent with the prediction of Krugman-style love-of-variety models rather than with Armington models, i.e., if an economy doubles its export value with its economic growth, its export extensive margin is expected to increase faster than its intensive margin (Hummels & Klenow 2005). Second, our data also show that new exporters tend to emerge in inland regions. When we control the firm-region-industry fixed effects, the results suggest that highway expansion can increase the export value and export intensive margin of those firms not very close to seaports. A possible explanation is that those potential exporters close to seaports have already started exporting. Third, new entry processing exporters tend to choose addresses not very close to seaports. A possible explanation is that those potential processing exporters close to seaports have already started processing exporting because of its lower entry barriers than ordinary exports. When we strictly control firm-industry-region time-invariant heterogeneity, the decrease of travel cost to seaports can reduce firms' processing share in export value, suggesting that highway expansion can induce the processing exporters to transform into ordinary or other exporters. This explanation is also consistent with the study of Egger & Falkinger (2003), they show that more efficient road networks can lower the trade barriers and promote trade structure transition from processing-oriented modes to more diversified structures. Forth, high-tech industries expanded much more significantly than other sectors and dominated China's export products (more than 50% export value) at the end of our sample period. At the same time, the regressions results show that high-tech industries are not sensitive to highway expansion, highways' impacts on textile industries are stronger than other industries.

This chapter is structured as follows. Section 2 reviews the theoretical background and empirical evidence. Section 3 introduces the data sources, discusses summary statistics, and the methods to construct geographical transportation costs. Section 4 analyzes the empirical results. Section 5 concludes.

5.2 Literature Review

5.2.1 Theoretical Background

Infrastructure and Gravity Model

Many trade theories predict that firms' export probability is negatively correlated with distance but positively related to the size and remoteness of destinations (Baldwin & Harrigan 2011).[2] According to the EK trade model (Eaton & Kortum 2002), the probability of a firm to export goods to a destination market decreases with the bilateral distance to, and market scale of, its destination but is positively related to the remoteness of its destination. Similarly, some monopolistic competition models (Ottaviano et al. 2002) also predict that the probability of export decision decreases with the distance to, and the scale of, destination markets. When these distance and market scale factors are controlled, remoteness is positively related to export probability. By contrast, trade models with heterogeneous firms provide different explanations for firms' trade decisions (Melitz 2003, Melitz & Ottaviano 2008). In these models, no export to destination X means that, for a given product classification, no variety is exported to X. A firms' product-level export decision toward a destination depends on a cutoff of marginal costs in destination X, i.e., a firm will start to export goods to a foreign market X only if this firm's marginal cost is lower than a certain threshold. A firm's export probability would be larger when its bilateral distance is small because of lower trade cost. In the model of Melitz (2003), firms' export probability is positively related to the scale of its destination market; however, in a following model (Melitz & Ottaviano 2008), export probability is negatively correlated with market size because larger economies have lower marginal cost thresholds and higher competition. To summarize, extant trade models have provided many stylized descriptions on the determinants of trade cost and firms' trade decisions.These models tend to emphasize the role of distance, or use iceberg trade cost assumption, because real trade costs are difficult to measure directly.

$$Trade_{o,d} = F[Direct = f[Underlying]]$$

Gravity models are the most extensively used theoretical framework in trade literature, they try to link trade flows with distance and economic scales, while some limitations also arise when researchers try to build bridges form the theoretical to the empirical world. In a more generalized form, gravity models are normally expressed as two-stage functions, i.e., trade

[2] Not only the bilateral distance, but the distances to third trade partners can also affect bilateral trade flows, some researchers construct remoteness index to capture third country effects (Anderson & Van Wincoop 2003).

flows are decided by directly-related factors (or trade barrier function), which are decided by underlying factors. The underlying factors could include distance, infrastructure, fuel costs, and transport technology; while the directly-related factors could include economic size, income, monetary transport costs, and tariffs (Behar & Venables 2011). However, reduced gravity models are normally rewritten in liner form, but the mechanisms of how underlying factors affect trade flows are not specified in this form.[3] On the left-hand side of the gravity equation, the value of trade flows are decided by the variety, quantity, and prices of goods, while most empirical studies do not strictly distinguish these three factors due to data constraints (Baldwin & Harrigan 2011).

The Role of Trade Cost

During the last decade, there were increasing number of studies using gravity models to investigate the impacts of infrastructure on trade activities, while trade cost was the most intensively investigated channel. These studies tended to focus on country- or region-level trade activities, most of them concluded that infrastructure development and trade cost reduction could promote export entry, increase trade flows and welfare significantly. For example, Behar & Venables (2011) show that GDP and bilateral distance can explain 70% fluctuations of country-level trade flows; Limao & Venables (2001) study the transportation and communication infrastructure, and find infrastructure quantity can increase trade flow through the channel of trade costs. Some following studies turned their attentions to region-level impacts, Michaels (2008) finds highway connection in rural areas can increase retail sales and makes trucks become the main vehicles used on these regions. These facts suggest that highways can reduce trade barriers and encourage inter-regional trade, because goods are mainly transported by trucks on highways. Donaldson (2018) uses trade cost channel to interpret domestic trade and welfare gains. That study focuses on railroad construction during India's colonial era, and finds India's railroad construction could reduce trade costs, shrink price gaps across regions, and increase inter-regional trade volumes. At the same time, agricultural incomes in railway-connected regions were promoted by the decline of trade costs, these regions' incomes became less dependent on natural factors such as rainfall; their income volatility was also decreased by railway construction. These facts imply that firms' access to trade ports will be increased by infrastructure construction though trade cost

[3] For example, Behar & Venables (2011) provide a typical reduced gravity model: $Trade_{o,d} = \beta_1 GDP_o + \beta_2 GDP_d + \tau_1 Distance_{o,d} + \tau_2 Landlocked_{o,d} + \tau_3 Infrastructure_o + \tau_4 Tradefacilitation_o + u_{o,d}$, where τ_3 is the elasticity of trade flows with respect to infrastructure.

channel, while trade costs can be calculated according to price gaps of homogeneous commodities produced in unique regions.

Unbalanced infrastructure investment across countries can partially explain why newly generated trade flows are mostly short-distance flows, which are higher than theoretical predictions. The elasticity of trade flows with respect to bilateral distance tends to change over time, some countries' trade decisions have become more sensitive to distance and trade costs while some others not, because of country-specific factors such as infrastructure development can significantly reduce unit trade costs. Cross-country evidence shows that the trade-distance elasticity (absolute values) tended to increase from 1962 to 1996, because of the increase in the elasticity of trade costs with respect to distance. When dividing countries into different income groups, the trade-distance elasticity tended to increase for low-income countries, but decrease for high-income countries, implying that it was country-specific characteristics such as infrastructure investment rather than global shock such as fuel price increases that made the difference. These facts reveal that most newly generated trade flows are short-distance trade, only high-income countries become less sensitive to trade distance, because the intensive infrastructure investment on information and communication equipment mainly happened in these economies (Brun et al. 2005).

Many trade models allow for deriving arbitrary trade costs under the framework of general equilibrium.[4] Basing on Armington's assumption, the gravity model developed by Anderson & Van Wincoop (2003) assumes countries produce and trade differentiated products, while consumers' preferences have constant elasticity of substitution. Their model provides a gravity equation of international trade (Equation 5.1) to specify the relationship between trade flows and trade costs, where x_{id} represents the bilateral trade flows form i to d, y_i and y_d is the economic output of origin and destination country i and d, y^w is the economic output of the whole world. $\sigma > 1$ is the elasticity of substitution, Π_i denotes trade barriers of i with all countries (outward multilateral resistance), P_d represents trade barriers of d with other countries (outward multilateral resistance), while trade costs $t_{id} \geq 1$ captures bilateral trade costs from i to d. If the origin country exports goods with domestic price p_i to d, the destination

[4]The gravity model with multilateral resistance developed by Anderson & Van Wincoop (2003), heterogeneous firm models (Melitz 2003, Melitz & Ottaviano 2008), and the Ricardian trade model (Eaton & Kortum 2002)

price p_{id} satisfies $p_{id} = p_i t_{id}$.

$$x_{id} = \frac{y_i y_d}{y^w} \left(\frac{t_{id}}{\Pi_i P_d} \right)^{1-\sigma}, \tag{5.1}$$

Although arbitrary trade costs can be derived from the log-linear version of equation 5.1, its expressions lack strict definitions on multilateral resistance. Some following studies have derived more detailed methods to specify arbitrary trade cost function from aggregate trade flows, e.g., Novy (2013).[5] However, an emerging problem is that aggregate-level measurements only capture trade costs for each country's single industry, while firm-level factors such as location effects are neglected in these measurements.

$$t_{id,t} = Dist_{id}^{\beta_1} \, e^{\beta_0 + \beta_2 X_{id,t}}.$$

Compared with referred or derived trade costs, implied trade costs are more flexible in imposing underlying factors in empirical studies. β_1 is the elasticity of trade costs $t_{id,t}$ with respect to bilateral distance $Dist_{id}$, while the term $X_{id,t}$ includes a series of covariates which vary across different studies. For example, Jacks et al. (2011) investigate the roles of trade costs in global trade flow variation over the last two centuries. Their implied trade costs are decided by bilateral distance, dummy variables of fixed exchange rate regimes, common language, EU membership, and shared border. Jacks et al. (2010), the covariate term includes tariff payment (log product of tariff rate to trade flows for each trade partner), volatility of bilateral exchange rate (standard deviation), rail road density, British Empire membership, and whether it uses the gold standard. Chen & Novy (2018) control whether a country is a member of IMF, OCED, and WTO in their covariate term. In order to investigate micro-level impacts of highway construction, the present study uses the implied trade cost function by assuming firm-level trade costs following log-linear form, the further specifications are discussed in the Methodology section.

Relocation and Agglomeration Effects

Besides the trade cost channel, two other important channels are relocation and agglomeration effects, which can promote the relocation of population and economic activities, and

[5] Novy (2013) develops an analytical solution of trade costs with multilateral resistance factors controlled by relative trade flows, i.e., $x_{id}x_{di} = x_{ii}x_{dd}(\frac{t_{id}t_{di}}{t_{ii}t_{dd}})^{1-\sigma}$. Arbitrary trade costs are given by $\tau_{id} = (\frac{t_{id}t_{di}}{t_{ii}t_{dd}})^{\frac{1}{2}} - 1 = [(\frac{x_{id}x_{di}}{x_{ii}x_{d}})^{\frac{1}{2(\sigma-1)}}] - 1$, which captures the geometric average relative bilateral trade costs with respect to domestic trade costs of i and d.

strengthen the market potential of few core areas, while those firms closer to core areas would have higher competitiveness due to agglomeration effects. Infrastructure development can increase factor mobility and elasticity of labour supply and promote population migration from poor toward rich regions (Donaldson 2018), while this process is normally accompanied with the relocation of economic activities. However, the relocation of population and firms are not usually goods news for rural areas. Heckscher-Ohlin models predict that if infrastructure development reduces trade barriers, it can cause trade to shift from labour-intensive to capital-intensive industries, promoting production specialization, and generating higher demand for high-skilled workers, but these effects are unbalanced across different regions. For example, Michaels (2008) finds infrastructure construction causes the relative incomes of high-skilled workers to increase faster than low-skilled workers in rural areas, Faber (2014) finds that the rural regions connected by highways witnessed slower economic growth than rich regions, suggesting low-skilled jobs were crowded out while new jobs were mostly created in rich rather than in rural areas. Consequently, population and real income in rich regions tend to increase faster than rural areas in the long run, if there is no government intervention. The agglomeration of population and economic activities can generate significant scale economies, and significantly increase market scales, which make rural regions become more reliant on the market of rich regions.

These impacts of infrastructure predicted by trade theories are also consistent with some predictions from new economic geography theories. The latter emphasize that firms' location selection is largely affected by the scale of local markets. When trade barriers are not low enough, the rational choice for firms is to be closer to the nearest big markets, while the decline of trade barriers will increase factor mobility, and lead to a core-periphery patterns, whether for final goods markets or intermediate goods markets (Chuan 2007). The emerging of core-periphery patterns suggests that economic activities need to concentrate in some core areas, where there are well-developed infrastructure, more complex industry structure, and higher competitiveness. Firms that involved in international trade are more likely to locate in those core regions with larger local markets, while new entry firms will follow similar patterns because of local market attractions.

New economic geography literature tends to discuss agglomeration economies with three scopes, i.e., industrial, geographic, and temporal dimensions (Okubo 2009). (1) Industrial is the most intensively discussed dimension, which refers to the economic externalities arising from agglomeration across industries. In some narrowly and strictly defined agglomeration effects, such kind of cross-industry externalities are distinguished from within-industry externalities. The latter are also called localization economies, referring to the phenomenon that firms from the same industries concentrate in same regions and benefit from productivity premiums induced by agglomeration; while the externalities arising from all industries are defined as urbanization economies. The mechanisms of industrial agglomeration are specified by the core-periphery model (Krugman 1991) and vertical linkage models (Venables 1994, 1996, Krugman & Venables 1995). Although agglomeration effects in the core-periphery model are motivated by labour migration, by input-output linkage in vertical linkage models, they still have the isomorphic equilibrium properties (Okubo 2009). (2) Geographic dimension indicates that physical distance or traffic accessibility across agencies can affect spatial interaction and integration. (3) Temporal dimension is designed to capture time-lagged impacts and time series auto-correlation. A general specification of agglomeration effects is given by the following function (equation 5.2):

$$A_j = \Sigma_{k \in K} \big(q(x_j, x_k) a(d^I_{jk} d^G_{jk} d^T_{jk}) \big), \tag{5.2}$$

$$y_j = g(A_j) f(x_j). \tag{5.3}$$

All establishments $k \in K$ potentially interacting with establishment j can generate agglomeration effects on j through five factors. Term $q(x_j, x_k)$ captures the scales of economic activities of establishments j and k, while term $a(d^I_{jk} d^G_{jk} d^T_{jk})$ captures distance effects, including geographic distance d^G_{jk}, industrial distance d^I_{jk}, and temporal distance d^T_{jk}.[6] The overall agglomeration economies will generate productivity premiums on the production function of establishment j (equation 5.3).

According to these specifications, firm-level agglomeration economies are directly decided by market scales and firms' distance to destination markets. Both the concentration of same-industry and different-industry firms can generate agglomeration effects, due to localization and urbanization economies. Firms' size can also affect agglomeration economies because larger firms may have more connections with various firms from different industries. In empirical studies, the direct way to evaluate agglomeration effects is based on the estimation of production function, but that approach has much stricter requirements on the data quality. There are also a greater number of studies using indirect measurements: Shefer (1973),

[6] Industrial distance d^I_{jk} will equal zero for any pair of agencies from the same industry, while d^I_{jk} tends to increase when the products of j and k become further differentiate from each other.

Tabuchi (1986), Ciccone & Hall (1996), Ciccone (2002), Jordaan & Rodriguez-Oreggia (2012) use population, population density, human capital density measured by education level, and density of economic output as the indicators of agglomeration effects. Nakamura (1985), Henderson (1986) uses the share of employment to capture urbanization economies, using industry-level employment to indicate localization effects, and finds the impacts of localization are stronger; consistent with some recent studies such as Rosenthal & Strange (2003), Henderson (2003). Henderson et al. (1995), Rosenthal & Strange (2003) calculate the employment Herfindahl-Hirschman index (HHI) to indicate agglomeration economies induced by industry diversity. Lall et al. (1999), Graham (2007), Holl (2011) estimate market potential index based on cities' size and travel time along real road network. These empirical studies provide essential examples of measurement construction.

Trade Intensive and Extensive Margin

Intensive margin is associated with the expansion of existing trade flows, which is relatively a more straightforward concept and also intensively specified by traditional trade theories. The new trade models since Krugman (1980) and heterogeneous firms trade models represented by Melitz (2003) pay more attention to extensive margin, and assume that international trade flows of differentiated goods are largely motivated by consumers' preference on goods variety. These models were proposed on the background of a series of new emerging trend in recent decades, i.e., the acceleration of globalization after the Cold War, the spread of supply chains around emerging countries, and the booming of intermediate products trade. In the model of Krugman, identical firms trade heterogeneous goods with a given level of substitution elasticity. In case of lower substitution elasticity, trade activities are less sensitive to trade barriers but more sensitive to goods' variety; and vice versa for the case of high substitution elasticity. The 'New' new trade theories represented by Melitz, and the following studies, such as Helpman et al. (2008) and Chaney (2008), take firm-level heterogeneity in to consideration, and pay more attention to the role of extensive margin. These theories establish an explicit paradigm in which trade flow expansion can be decomposed into two dimensions, i.e., the intensive margin, and the extensive margin attributable to firms' entry/exit and scope expansion.

On the basis of Melitz, Chaney (2008) introduces country-level heterogeneity and asymmetric trade barriers and, using this model, shows that trade barriers can increase aggregate trade flow and encourage more firms to start exporting; these impacts are more significant in heterogeneous firms models than new trade models. The trade barriers can be decomposed to a fixed entry cost component and a variable component associated with transportation cost. When goods are more substitutable, the intensive margin is more sensitive to trade barriers. Due to higher substitution elasticity being related to a higher level of cross-sector competition, the new entry firms are normally smaller and less-productive firms than incumbent firms, they have difficulties to compete with exciting firms, inducing a more significant intensive margin expansion than extensive margin expansion. Equation 5.4 explicitly defines the the relationship between trade flows/variety $Export_{ab}$ with trade barriers $TradeBarriers_{ab}$ across a and b, where σ represents their elasticity.

$$Export_{ab} = Constant\frac{GDP_a GDP_b}{(TradeBarriers_{ab})^{\sigma}}, \tag{5.4}$$

Manova & Zhang (2009) uses the relative forces of scale economy and scope economy to explain the change of export intensity and variety by assuming that a firm in a given industry needs some resources and products to maintain its operation, e.g., financial supports or special intermediate inputs that must be imported from overseas. When resource and product availability are identical across firms with heterogeneous profitability and cost structure, firms' decisions on intensive or extensive expansion are motivated by the relatively force of firm-level scale economies and product-level scale economies. When firm-level scale economies can more significantly increase firms‘ profitability than product-level scale economies, large firms tend to increase both production/export intensive and extensive margins. On the other side, when product scale economies are more important than firm scale economies, firms will face diminishing return to scope, leading to a negative correlation between intensive or extensive margin.

We can also loosen this restriction, and assume that resource allocation across firms are relevant to firms' heterogeneous profitability and productivity; this description tries to reproduce the case in the real world, i.e., large enterprises have market influences, or productive firms are more affiliated with upstream producers.[7] In the model of Bernard et al. (2011), firms recieve firm-country-product level productivity shocks; productive and profitable firms have more resources to allocate, they can also increase export scope without considering the profitability of individual products, which induces a positive correlation between intensive and

[7] Debaere & Mostashari (2010) summarizes plenty of empirical studies, and concludes that trade liberalization is proved to have positive impact on extensive margin, though the channel of vertical specialization.

extensive margin.[8] In an extended model of Manova (2012), firms would firstly choose to export profitable goods. However, when firms decide to expand their export scope, they will add new products in order from high-profitable to low-profitable goods, until they exhaust their available financial resources. In the cases of heterogeneous firm-level resource availability, the increase of external financial supports or the decline of trade cost both tend to induce a positive correlation between intensive and extensive margin.

The Roles of Processing Trade and Outsourcing Activities

Infrastructure development can affect firms' processing or outsourcing decisions, through both trade costs and market selection channels. Processing trade accounts for about 50% of China' imports during the early years of China's WTO accession. Similar to many emerging economies, China provides tariff exemption for processing trade, so processing trade is not directly affect by tariff reduction. These processing firms tend to be less productive than non-processing firms (Yu 2015, Dai et al. 2016). On the one hand, processing trade is more likely to be affected by infrastructure development, because both input and output goods need to be transported through traffic lines from or toward foreign markets; on the other hand, infrastructure development can induce the entry of intermediate producers in the long run; then these processing firms will become gradually more dependent on domestic rather than foreign suppliers.[9] For example, Egger & Falkinger (2003) finds that infrastructure development can lower trade barriers, increase the quantity of intermediate producers, and make home countries become less dependent on aboard intermediate goods. At the same time, Chuan (2007) also points out that outward processing trade in EU is more dependent on infrastructure, compared with inward processing trade. These impacts suggest that economic structure or industry-specific characteristics also make differences, i.e., the EU mainly produces capital-intensive goods with higher quality and productivity gains, EU outward processing trade is more sensitive to trade barriers and changes of comparative ad-

[8] Bernard et al. (2011) states that productivity can be decomposed into firm-specific productivity (firms' ability) and product-level productivity (expertise in productivity). Although these two components are assumed to be independent and unknown before firms paying entry costs. The higher firm-level productivity will also be associated with higher product-specific productivity across all products, leading to a positive correlation between intensive and extensive margins.

[9] Extant theoretical and empirical studies of international trade show that ordinary exporters tend to be more productive and capital-intensive, hire more skilled employees, and earn higher revenues than non-exporters, because only these can cover the high transaction costs from international trade (Bernard et al. 2012). Many extant studies have found consistent evidence of this; e.g., evidence from developed economies such as the U.S. (Bernard & Jensen 1999, Bernard et al. 2005), Germany (Bernard & Wagner 1997), and France (Eaton et al. 2004, 2011), evidence from developing and emerging economies such as Colombia, Mexico and Morocco (Clerides et al. 1998), Korea and Taiwan (Aw et al. 2000).

vantages.

At the same time, infrastructure is not the unique factors in firms' outsourcing decisions. According to Heckscher-Ohlin models, firms tend to outsource intermediate inputs to economies with comparative advantages decided by factor endowments. By contrast, new trade theories emphasize that comparative advantages are not so important for the trade flows across industrialized economies, because industrialized countries are more involved in the production of differentiated intermediate goods. In these cases, country- and industry-specific variables such as market size, factor endowments, exchange rates and taxes are more likely to affect outsourcing and processing decisions (Chuan 2007).

5.2.2 Empirical Studies of Infrastructure Construction Effects

Cross-country Evidence

Extant studies usually agree that trade costs are largely affected by policy barriers, transport costs, wholesale and retail costs, while transportation costs are further affected by geographical locations, infrastructure, market structure of logistic industries, and administrative barriers (Limao & Venables 2001, Anderson & Van Wincoop 2004). Many previous studies find that focusing on infrastructure and trade activities tend to use aggregate trade data, a recent new trend is to calculate firms' accessibility to the nearest ports along existing road network Matthee & Naudé (2008), Cavallo et al. (2013), or the distance to the border of their trade parteners Costa-Campi & Viladecans-Marsal (1999). These cross-country studies tend to conclude that infrastructure development can encourage trade activities and firms' specification, while infrastructure development seems to follow diminishing return in cross-country evidence.

Many aggregate-level studies about infrastructure conclude that infrastructure investment and construction can reduce trade cost, encourage trade activities and regional integration. Limao & Venables (2001) investigate the impacts of geography and infrastructure on trade flows.They use the data from shipping companies and IMF; the former can provide land and sea transport costs from U.S. to other destination countries, the latter provide the shares of bilateral transport cost in free on board prices. They find land costs are about ten times as much as sea transport costs, while landlocked economies face 50% higher transport cost than coastal economies. Their infrastructure measures include country-level road, railway density, and telephone per person; based on these proxies, they find the development of infrastructure can significantly reduce transport costs, the elasticity of trade volumes with respect to transportation costs are about -3. Similarly, Vijil & Wagner (2012) find the aid for transport, communicate, and energy infrastructure can increase developing countries' export shares in GDP, the 10% increase of aid associates with 2.34% increase of export share, this effect is equivalent to the 2.71% decline of tariff or non-tariff barriers. Coşar & Demir (2016) estimate region-level transportation costs based on geographical locations and infrastructure; in order to solve the endogeneity issues, they use expressway stocks in the initial year as instrument. Empirical results show that the intensive Turkish investment on transportation infrastructure can reduce trade costs and promote regions' connectivity to international trade.

Duranton et al. (2014) investigate the impacts of travel costs on bilateral trade across U.S. cities. Based on city-level export data, they use bilateral highway distances across these cities to capture trade costs. In order to solve the endogeneity issues of highway length, they use exploration routes during 1528 to 1850, railway in 1989, and 1947 planned highways as instruments. Results show that trade volumes decrease with highway distance, i.e., the 10% additional highway distance associates with 5% decrease of exports (measured by weight), while the impacts on exports measured by values are insignificant. This fact implies that the increase of highway density makes those cities gain competitive advantages of heavy product exporting, and induce firms to specialize in heavy and low unit price goods.

Chuan (2007) focuses on EU processing trade, and finds economic scales of trade partners only have insignificant impacts on processing trade, while trade costs and factor endowments play much important roles. Outward processing trade is more likely to be observed in labour-abundant economies, while infrastructure stocks including electricity, telephone, and road network tend to encourage both outward and inward processing trade. Compared with outward trade, inward processing trade are more sensitive to cost factors such as relative prices, exchange rates, and taxes.

Some recent literature has combined more spatial analysis in trade studies and revealed more details. Bonfatti & Poelhekke (2017) investigate the roles of mine-related infrastructure on trade and regional integration. Most developing countries inherited infrastructure from colonial periods, this infrastructure was mostly designed to carry resources to trade ports, they are expected to generate fundamental influences on current regional trade integration because those mine-related roads can also be used to transport other goods. They combine several datasets, i.e., country-level UN trade, infrastructure data, global mines and ports data. However, due to the infrastructure only providing road and railway length rather than any quality and direction information of each traffic line, the authors use the quantity of operating mines as the indicator of infrastructure quality that serve mining activities. Their results show that exports of operating mines are affected by mine-related transport infrastructure through trade cost effects, while this effect becomes weaker if a mine is located in a landlocked country because of administrative barriers.

Donaldson (2018) investigates district-level effects of transportation infrastructure construction and, in this reduced gravity model, trade costs are affected by infrastructure term, trend term and time-fixed effects of origin and destination regions. The infrastructure term reflects two types of measures, i.e., (1) a dummy variable to indicate whether a origin-destination pair is connected by railroad network, or (2) lowest cost routes calculated based on transport routes, which include road, railroad, river, and costs. The transport cost of unit distance on railroads is set as 1, while other modes of transport are represented as the relative unit cost to railroad. In order to solve the endogenous issue in railway construction with respect to local income, the author uses information of Indian road construction plan and climate such as rainfall to construct instruments. Empirical results include two dimensional implications; firstly, railroad connection can reduce trade costs and price gaps across regions, then increase inter-regional trade flows; secondly, railroad connection can increase real income, and reduce income volatility, which have important welfare implications. Due to the fact that India was an agricultural economy during the colonial age, overall economic incomes can be largely represented by agricultural incomes, which are sensitive to natural factors such as weather and rainfall. The increase of inter-regional trade induced by the development of transportation infrastructure can reduce uncertainty of agricultural incomes, make more regions gain benefits form their comparative advantages, then increase overall welfare in those districts connected by railroads.

Asturias et al. (2019) focus on the Golden Quadrilateral (GQ), the most important highway network in India which connects four major metropolises (Delhi, Kolkata, Mumbai and Chennai); and find that GQ construction induces 2.72% real income increase in the manufacturing sector, while the initial investment on CQ was covered within just two years. This highway network data is combined with extremely detailed micro perspective data, i.e., the Annual Survey of Industries and the National Sample Survey. These datasets provide lots of firm-level proxies such as Indian manufacturing plant-level sales value, physical quantity, employment, wage payment, intermediate inputs, and capital stocks, which makes it possible to identify transportation cost from input-output relationship.

The theoretical arguments of these authors build on a static general equilibrium internal trade model, which allows plant-level heterogeneous markups. Three channels are identified from their model (Ricardian gains, allocation efficiency, and relative markup), while their empirical analysis mainly focuses on the allocation efficiency channel. Building on the achievement of Holmes et al. (2014) and Donaldson (2018), they apply a two-step approach to identify transportation cost and estimate export demand elasticity. At the first step, transportation cost is derived from price gaps across regions, similar to the approach used by Donaldson (2018). This approach assumes that prices charged in different regions by a nationwide monopolist only depend on the relative transportation cost toward these regions, i.e., $\frac{Price_a}{Price_b} = \frac{TransCost_a}{TransCost_b}$. A monopolist is defined as a plant selling more than 95% value in India (5-digit ASICC product code). According to this definition, there were 165 products sold by monopolists during 2001 to 2006. Then they regress prices of these monopolists on effective distances (measured by least cost route approach) from origin to destinations, where the coefficients of effective distances represent the unit transportation cost. In order to address potential endogeneity issues, they use straight lines to connect node cities to generate instrument variable, similar as the approach applied by Faber (2014). Building on the proxies estimated at the first stage, at the second step, they find the decrease of transportation cost can induce the increase of inter-state trade flow. In addition, due to firms with higher market power charging higher markups, this mechanism can push their prices even higher than their marginal production cost, while their labour and capital input share tend to be lower in their overall revenue. These predictions are also strongly supported by empirical results.

Duranton (2015) focuses on three impacts of spatial road network on trade flows and competition in Colombia, using city-level data and road network data (collected from Google Maps), and historical roads as instruments to solve the endogeneity issues. Results show that bilateral trade flows can be promoted if cities are close to traffic lines, a 10% decline of bilateral distance can increase bilateral trade values by 7%, or trade weight by 6%. In the city areas, a 10% increase of road density can also promote export flow by 3-5%, implying that region-level road connectivity can also encourage trade activities. Similarly, Martincus & Blyde (2013) investigate the impacts of transport infrastructure on exports in Chile, by using earthquake data as an exogeneous factor. According to Chile's Ministry of Infrastructure, the continuing earthquakes in Chile keep causing damage on road networks, this is expected to increase trade costs and generate negative impacts on trade activities. They combine earthquake data with the geography-referenced transaction-level export data of Chilean manufacturing firms from 2008 to 2011, which allows tracing the origin and destination of each export. Results show that the damage on road network by earthquakes can reduce shipments and exports, this effect is stronger for large firms that export homogeneous goods.

However, extant evidence also supports the view that the impacts of infrastructure development seems to follow diminishing return. Bougheas et al. (2000) find the development of core infrastructure (Transportation and Communication) promotes firms' specification, because of the decline of transaction costs and intermediate costs, while there is also a U-shape correlation between infrastructure accumulation and economic growth. This research uses U.S. manufacturing firm data and cross-country output data, while their infrastructure data comes from the World Bank, providing the physical measure of infrastructure stocks. Similar to other extant studies, our study uses lagged terms as IVs to solve the endogeneity issues (lagged shares of government consumption, investment to GDP etc.). Their conclusion is consistent with the theoretical statement of Imbs & Wacziarg (2003).

China-specific Studies

Many China-specific studies use geographical distance as the proxy of trade costs, they tend to find trade cost decline can promote aggregate exports and import growth. Poncet (2003) shows that transportation costs and administrative barriers play important roles in China's provincial trade flows. The border barriers across different administrative regions are affected by geographical locations, which are assumed to be zero inside provinces but take positive values between provinces. This author uses input-output table and provincial trade data to identify domestic and international trade volumes. Domestic transportation costs are indicated by liberal distances between provincial capitals, while international transportation costs are estimated as their provincial capitals to their foreign trade partners' capitals. The export prices in destination markets are decomposed into original prices, transportation costs, and trade barriers. The results show that, from 1987 to 1997, China's internal trade barriers declined to similar magnitudes to those within E.U. countries, and the magnitude between the U.S. and Canada. The decline of border barriers promotes China's domestic integration and increase their consumption.

Tang et al. (2019) investigate the impacts of high-speed railway construction on firms' export activities. Traffic accessibility is measured as whether a city is connected by railways; in order to solve endogeneity issues, they use the approach developed by Faber (2014) to construct least cost paths. Results show that firms in those cities connected by high-speed railways witness export growth by 12.7%, because of the decline of face-to face negotiation costs. This effect remains significant only if new constructed railway stations are located within 30km of cities. In addition, railway connection can also increase firms' export scopes, this impact is stronger for eastern provinces, capital- and technology-intensive sectors. Differently from road construction, high-speed railways are constructed for passenger transport, charging higher prices but providing faster transport services. The positive impacts of high-speed railway construction suggest that the benefits of knowledge spillover and faster speeds should generate larger benefits than higher service prices.

On the import side, Sun et al. (2019) study the interrelationship between high-speed railway construction and tariff reduction, and their impacts on import activities. They use price data of 75 types of products in 139 Chinese cities, while infrastructure development for each city is captured by road length and the dummy variable to indicate whether the city is connected by high-speed railways. Results show that those cities with high-speed railway connection benefit more from tariff reduction, their import volume and scope tend to increase faster than those cities without connection, while their retail prices of final goods also tend to decrease more significantly. Compared with landlocked cities, those cities closer to trade ports are more affected by infrastructure development. These facts confirm that the mechanism of trade costs reduction can be affected by tariff decline, similar to much extant trade literature; however, these results also highlight the roles of infrastructure in tariff decline, which can promote information dissemination and knowledge spillover, then amplify costs reduction effects.

Luo (2004) focuses on the impacts of infrastructure investment on regional inequality. The measurement of trade costs is calculated according to geographical distances between provinces, but these geographical distances are adjusted by road and railway density in these provinces.

Results show that geographical distance has negative impacts on the growth of rural areas, but the continuous infrastructure investment can reduce trade costs, promote interregional trade and shrink the development gaps between rich and rural areas. This effect is confirmed by Li & Li (2013), they find that both provincial and firm-level inventory decreased by transportation infrastructure investment from 1998 to 2007.

Li et al. (2018) investigate the impacts of nine international railways, that connect China and Europe or China and Central Asia, on trade activities from 2011 to 2015. They use spatial analysis to capture countries' geographical conditions and traffic accessibility; the line distances between China and its trade partners' capitals are calculated by Google maps as the geographical transportation costs, while railway connection and landlocked countries are indicated by dummy variables. Results show that railway connection can promote exports from China to Europe and Central Asia, but the rise of exports from Europe and Central Asia to China are limited. The export increases for China mainly concentrate in equipment and manufactured products, while China's imports are mainly agricultural goods.

A recent study has used more comprehensive approaches to measure trade costs. Guo & Yang (2019) focus on the impacts of transportation infrastructure and traffic accessibility on aggregate trade activities, by dividing China into eight regions and the world into eight overseas regions. Transportation costs in this study include three parts; i.e., land, ocean, and port operation costs. They use spatial analysis to estimate region-level land costs to China's top-ten trade ports, by using charges data and spatial data of China's road and railway network.[10] Ocean costs are estimated according to shipping costs and speed from China's main sea ports to eight global regions. Port operation costs are calculated based on port charges and scales. Empirical results show that traffic accessibility is positively correlated with the demand of maritime transport, those regions with lower transportation costs have higher export flows.

Firm-level measurements on trade and geographical locations are applied by Liu et al. (2017). They investigate firm-level impacts of highway construction on exports from 2000 to 2006, using China's manufacturing firm, Customs, and ACASIAN spatial datasets. They use firms' distance to highways and highway density around 20km or 30km radius of each firm as the proxies of traffic accessibility. However, these two measurements can only capture local traffic accessibility rather than accessibility to trade port, and can only capture domestic

[10] According to the study of Tiwari et al. (2003), distance to trade port can largely affect freight volumes and port choice.

transportation costs rather than the costs affected by different destination markets. In order to solve the endogeneity issues, they calculate firm-level land ruggedness index, constructed based on average slope of the land that surrounding firms within a certain radius. However, due to they use of only one instrument, the Over-identification test can not be carried out in this paper. Their results show that the increase of traffic accessibility can promote firms' exports, and increase firms' import and export scopes, while high-productivity firms benefit more than low-productivity firms.

To summarize, most China-specific studies use aggregate data to investigate the relationship between infrastructure and trade. The recent new trend is to use spatial analysis to investigate the firm-level impacts of infrastructure development on specialization, domestic or international trade, but there are limited studies providing China-specific firm-level explanation on the mechanism of infrastructure construction. The study of Liu et al. (2017) has combined firm-level and transact-level data, but their measurement of traffic accessibility is incomplete, because they do not include the information of destination countries and original ports. By contrast, Guo & Yang (2019) measure traffic accessibility in three dimensions, including domestic trade cost, ports' quality and overseas trade costs, but their study only focuses on aggregate-level impacts.

5.3 Background and Stylized Facts

This section introduces the background of China's trade reform and the stylized facts emerging from its export development. China's foreign trade was managed by a highly-centralized planned foreign trade regime before 1978; this system was established based on the background of China's import substitution and socialist industrialization strategy. China's trade reform is 'a process of crossing the river by feeling the stones', it cannot be simply described as a unilateral movement toward trade liberalization and protection reduction (Ianchovichina & Martin 2001, p. 3). During the 1950s, China temporarily adopted instruments such as tariff and license to manage foreign trade, but then transformed into a more centralized planned trade regime. In the 1970s, China's foreign trade was centralized and managed by 10 to 16 state-owned trade enterprises, each one of them could effectively monopolize the import and export of particular range of products (Lardy 1993).[11] These enterprises were authorized by

[11]For example, the imports and exports of agricultural products were managed by China National Grain, Oil and Food Import and Export Corporation, China National Native and Animal Products Import and Export

the Ministry of Foreign Trade to determine the imports of specified products according to the projected domestic demand-supply gaps, and the exports to cover the currency payments of planned imports, while the trade policy instruments such as tariffs, quotas and licenses were rarely used during the 1960s and 1970s.

The early trade reform from 1979 to 1987 gradually decentralized this foreign trade operation system and established more than 2,200 new trade corporations. Some of them were controlled by central government, while more of them were established by local government. Most production firms still could not trade directly with foreign companies, except for firms located in special economic zones and sino-foreign joint ventures. At the same time, the mandatory planning trade system was progressively reformed as a hybrid system with mandatory planning, guidance plan, and market adjustment (Huang & Chen 1999). Trade instruments such as a quota and reformed license system, and export rebate policies, were introduced and gradually replaced government's direct intervention after 1980. An export rebate system was first adopted as an experimental strategy in 1983, then the range of commodities was extended in 1985 from 17 final commodities to all commodities except zero-tex commodities, refined oil and crude oil. In 1986, intermediate commodities were also added into the basket of tax rebate. The quota and license system managed the import and export of most commodities, while two groups of commodities which are essential for national security were still highly restricted: agricultural products, such as grain, vegetable oils, sugar, tobacco, and cotton etc.; and industrial raw materials, such as crude oil, refined oil, coal, rubber, and chemical fertilizer etc. Import of equipment and technology was also highly restricted. A full export rebate system was established in 1988, the product-specific and value-added taxes were fully rebated.

During the 1990s, more reforms on the foreign exchange market were adopted, while tariffs were also dramatically reduced before China's WTO accession. In order to stimulate export expansion, firms and local governments were allowed to retain a proportion of foreign exchange earned from their exports, and to buy and sell their foreign exchange more flexibly after 1988. This system was replaced by the exchange bank settlement system in 1994 (Huang & Chen 1999). Meanwhile, Chinese government established several foreign exchange swap centers in province capitals and special economic zones to serve both domestic and foreign-invested enterprises, allowing them to transact foreign exchange with floating

Corporation, and China Textiles Import and Export Corporation (Huang & Chen 1999).

rates. This new system was designed to gradually shrink the gap between official and floating exchange rate, while these two exchange rates were unified in 1994. In 1996, the RMB became convertible on the current account (Huang & Chen 1999). Tariffs were also extensively reduced much earlier than China's WTO accession. China reduced the import tariff of 265 commodities in 1991, 3,371 commodities in 1992, and 2,898 commodities in 1993. From 1994 to 1997, a series of tariff exemptions were applied for processing trade, productive equipment and technology import. These reforms reduced China's average tariff rate from about 40% to 17% by 1998 (Ianchovichina & Martin 2001), and built the foundation of the processing trade boom and technological equipment export exploration over the next ten years.

Table 5.3-1: Exporters Recorded in NBS and Custom Dataset

	2000	2001	2002	2003	2004	2005	2006
Custom Exporters	62,746	68,487	78,443	95,688	120,590	144,030	171,205
NBS Total Firms	162,885	171,256	181,557	196,222	278,980	271,835	301,961
NBS Export Firms	37,200	40,805	45,308	50,907	76,952	75,624	79,315
(in Custom Data)	(59.3%)	(59.6%)	(57.8%)	(53.2%)	(63.8%)	(52.5%)	(46.3%)
Matched Firms	20,944	24,135	28,499	34,042	51,294	53,412	58,892
(in NBS Exporters)	(56.3%)	(59.1%)	(62.9%)	(66.9%)	(66.7%)	(70.6%)	(74.3%)
(in Custom Data)	(33.4%)	(35.2%)	(36.3%)	(35.6%)	(42.5%)	(37.1%)	(34.4%)

Note: This table illustrates how many firm-level observations are left when custom data are merged with industrial firm dataset. Following the approach used by Ding, Jiang & Sun (2016), export firms in custom data are matched with NBS data according to their reported name, telephone number, address, and postcode. Those firms with more than one establishment are not included because their address cannot correctly reflect their locations.

During the observation period (2000 to 2006), China witnessed repaid export expansion, continuous entry of small and new exporters, and obvious export structure transition from processing trade to other trade modes. In the customs dataset, 62746 exporters are recorded for 2000, this number tripled to 171205 in 2006, implying a significant export exploration during the short seven years. However, more than half the exporters are small firms, because the NBS dataset only records 37200 and 79315 exporters (firms above designated size) during the same period, the proportion of small exporters increased from 40% to 55%, which are not included in the present study after data matching (Table 5.3-1).

Table 5.3-2: Export Structure across Different Trade Modes

	2000	2003	2006
NBS Total Firms	162,885	196,222	301,961
Export Transactions	5,199,738	9,241,319	16,174,046
Matched Transactions	1,143,406	2,336,315	4,798,011
	(22.0%)	(25.3%)	(29.7%)
Processing Trade:			
Imported materials	550,011	882,175	1,540,080
(in Matched Obs)	(48.1%)	(39.4%)	(32.1%)
Supplied materials	98,070	152,747	231,490
(in Matched Obs)	(8.6%)	(6.8%)	(4.8%)
Ordinary and Others	459,325	1,301,393	3,026,441
(in Matched Obs)	(40.2%)	(55.7%)	(63.1%)

Note: This table illustrates the export transaction variation, and export structure variation by different types of trade mode, including processing trade with imported materials, processing trade with supplied materials, ordinary and other trade modes.

In the transaction dimension, most export transactions are processing trade flows, which can be further classified into imported-material processing trade which import parts and materials as intermediate inputs, and provided-material processing trade which uses intermediate inputs directly provided by their trade partners. The proportion of imported-material flows gradually decreased from 50% to 30% from 2000 to 2006, while provided-material export also declined from 9% to 5%, implying that China's dramatic export exploration was also accompanied by endogenous transition or import substitution, from processing trade to ordinary and other trade modes (Table 5.3-2).

Egger & Falkinger (2003) claim that such kinds of trade structure transition can be affected by infrastructure development, because more a efficient road network can lower trade barriers through trade cost reduction effects, increase the quantity of intermediate producers, and make home countries become less dependent on foreign intermediate goods. Chuan (2007) also points out that outward processing trade in EU is more dependent on infrastructure, compared with inward processing trade; these facts suggest China's highway construction may play a similar important role in China's trade structure transition.

Table 5.3-3: Comparison of Trade Structure in 2006 and 2000

	Obs	Average	St Dev.	Min	5th	95th	Max
2006							
1. Original version:							
Quantity	4,798,011	97704.060	3648504	0	9	130061	1.23e+09
Value		111155.500	1521742	1	192	295786	5.31e+08
2. Aggregated by firm-product-destination-trade method groups:							
Quantity	390,365	760104.500	3.74e+07	0	13	569572	1.16e+10
Value		990871.800	1.69e+07	1	292	2162463	3.10e+09
3. St Dev. within each 8-digit HS code group:							
Quantity	6217	307837.700	4275803	0	10.167	532483.400	2.62e+08
Value		249002.600	1092851	0	4911.563	739909.800	2.74e+07
4. St Dev. within each firm-HS product group:							
Quantity	274711	50695.370	1436055	0	7.529	81616.790	3.13e+08
Value		59882.310	621536.800	0	178.572	176812.300	1.42e+08
5. St Dev. within each destination-HS product group:							
Quantity	151780	61008.090	1237563	0	3.500	116523.300	2.62e+08
Value		72266.560	550872.100	0	452.429	208123.300	5.02e+07
2000							
1. Original version:							
Quantity	1,143,406	118046.800	3518266	0	28	181051	1.23e+09
Value		84377.530	584008.400	1	240	277300	8.71e+07
2. Aggregated by firm-product-destination-trade method groups:							
Quantity	334,263	403800.100	2.06e+07	0	20	468868	9.81e+09
Value		288628.3	3106412	1	164	837112	5.82e+08
3. St Dev. within each 8-digit HS code group:							
Quantity	5025	282899.900	6655514	0	12.804	450025.200	4.62e+08
Value		160552.300	1012372	0	2251.666	458269.800	5.26e+07
4. St Dev. within each firm-HS product group:							
Quantity	81856	64260.510	1348034	0	23.6430	107812	3.01e+08
Value		51394.200	273269.100	0	234.0520	166909.300	1.83e+07
5. St Dev. within each destination-HS product group:							
Quantity	55304	73664.850	1526478	0	8.660254	132660.800	3.01e+08
Value		54079.870	292617.300	0	326.523	165516.700	1.83e+07

Note: Table 5.3-3 illustrates trade structure of matched transactions observations in 2006 and 2000. Line 1 shows summary statistics of matched export transactions. Line 2 aggregates monthly export quantity and value into annual observations, grouped by different firm, product, destination, and trade method. Line 3 shows the standard deviation of export quantity and value, within each 8-digit HS group. Line 4 provides the standard deviation of export quantity and value, within each firm and HS group. Line 5 provides the standard deviation of export quantity and value, within each destination and HS group.

Table 5.3-4: Summary Statistics of Different Scopes of Export Variety

statistics	All	2000	2003	2006
1.Destination by firm-sector				
Obs	233150	16583	24784	55789
Average	7.764	5.938	7.4174	8.846
Std. Dev.	(9.787)	(7.905)	(9.425)	(10.718)
2.Product by firm-sector				
Obs	233150	16583	24784	55789
Average	6.457	6.011	6.140	6.883
Std. Dev.	(10.897)	(11.464)	(10.028)	(11.323)
3.Destination-product by firm-sector				
Obs	233150	16583	24784	55789
Average	18.305	14.254	17.089	21.200
Std. Dev.	(39.869)	(33.993)	(35.826)	(45.611)
4.Product by firm-destination-sector				
Obs	1810212	98464	183832	493530
Average	2.358	2.401	2.304	2.396
Std. Dev.	(3.614)	(3.918)	(3.355)	(3.627)

Note: Table 5.3-4 shows export variety changes over the period from 2000 to 2006. Export variety are calculated in four different dimensions, i.e., variety of destination by firms within each 2-digit sector, variety of product for each firm within each 2-digit sector, variety of destination and product for each firm within each 2-digit sector, variety of product for each firm within each 2-digit sector to each destination.

China's export expansion is characterized by a dramatic aggregate export explosion on the one hand, and the decline of firm-level export intensity and the rise of export firm-level variety on the other. Similarly to Manova & Zhang (2012), Table 5.3-3 summarizes export flows within three dimensions, firm level, eight-digit product level, and destination level. Group one shows that the average transaction-level export quantity decreases from 118046.8 to 97704.1, while average transaction-level export value increases from 84377.5 to 111155.5, implying that China's export structure gradually changed from low-value and quality mode to high-price and quantity mode. By contrast, the firms' export of each good (eight-digit HS code) to each destination shows an opposite situation, i.e., their average export quantity increases from 430800.1 to 760104.5, while export value explodes from 288628.3 to 990871.8 (quality increase). These facts imply that firms' overall export volumes tend to increase, while more exporters tend to export more frequently with relative smaller export flows for each transaction. Groups three to five summarize the standard deviation of export quantity and value. Group three shows that the standard deviation within each eight-digit HS code group, whether they are estimated on the basis of quantity or value, tends to increase. By contrast, when the transaction flows are classified by firm and HS code, the standard deviation of export quantity decreases from 64260.51 in 2000 to 50695.37 in 2006; when the transaction flows are classified by firm, destination, and HS code, the standard deviation of export quantity decreases from 73664.85 in 2000 to 61008.09 in 2006. These facts suggest that the variation of overall export flows tend to increase, while the variation of firm-level export flows tend decline because of the smaller export volumes for each transaction. Consistent results can also be found in Table 5.3-4, it shows that firms' export variety by destination, product, and sector gradually increase during 2000 to 2006, implying the decline of firm-level export intensity is accompanied with the rise of export scopes.

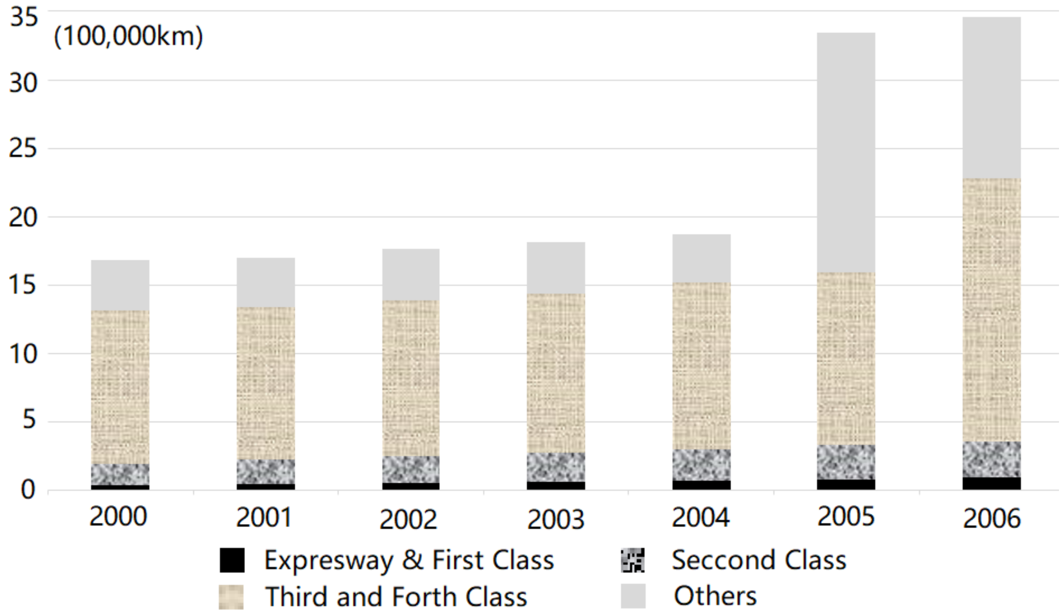
5.4 Methodology

5.4.1 Customs Data

This chapter mainly uses four main datasets. The manufacture firm dataset, regional dataset, and highway dataset, are the same as in the third and fourth chapters. Due to this chapter focusing on highways' export activities, a new customs transaction-level dataset is merged with these previous three datasets. This dataset contains a series of trade variables from 2000 to 2006, including trade price, quantity, origin, and destination countries within each 8-digit HS product code, which allows tracing the general direction of trade flows, and estimating transport costs according to road network data. The custom dataset is merged with industrial firm-level data according to the approach used by Ding, Jiang & Sun (2016). The firm's name is the primary choice to identify those firms appearing both in NBS and the custom datasets. The firm's name may change over time, however, name-changing errors are corrected by firms' telephone numbers, addresses, and postcodes. Due to NBS dataset only including state-owned enterprises (SOEs) and non-SOEs with more than five million RMB revenue, those small non-SOEs are dropped when custom data merged into the industrial firm dataset. In addition, the matched dataset in the present study is smaller than the dataset matched by Ding, Jiang & Sun (2016) because spatial analysis requires precise and unique records of firms' locations; therefore a number of large scale firms with multiple establishments in NBS dataset are dropped in the final merged dataset.

5.4.2 The Measurement on Domestic Transportation Cost

Figure 5.4-1: Total Length of Highway and Road from 2000 to 2006



From the bottom to the top, each bar illustrates the total length of four types of road, i.e., expressway and first class highway, second class highway, third and fourth class highway, and other roads.

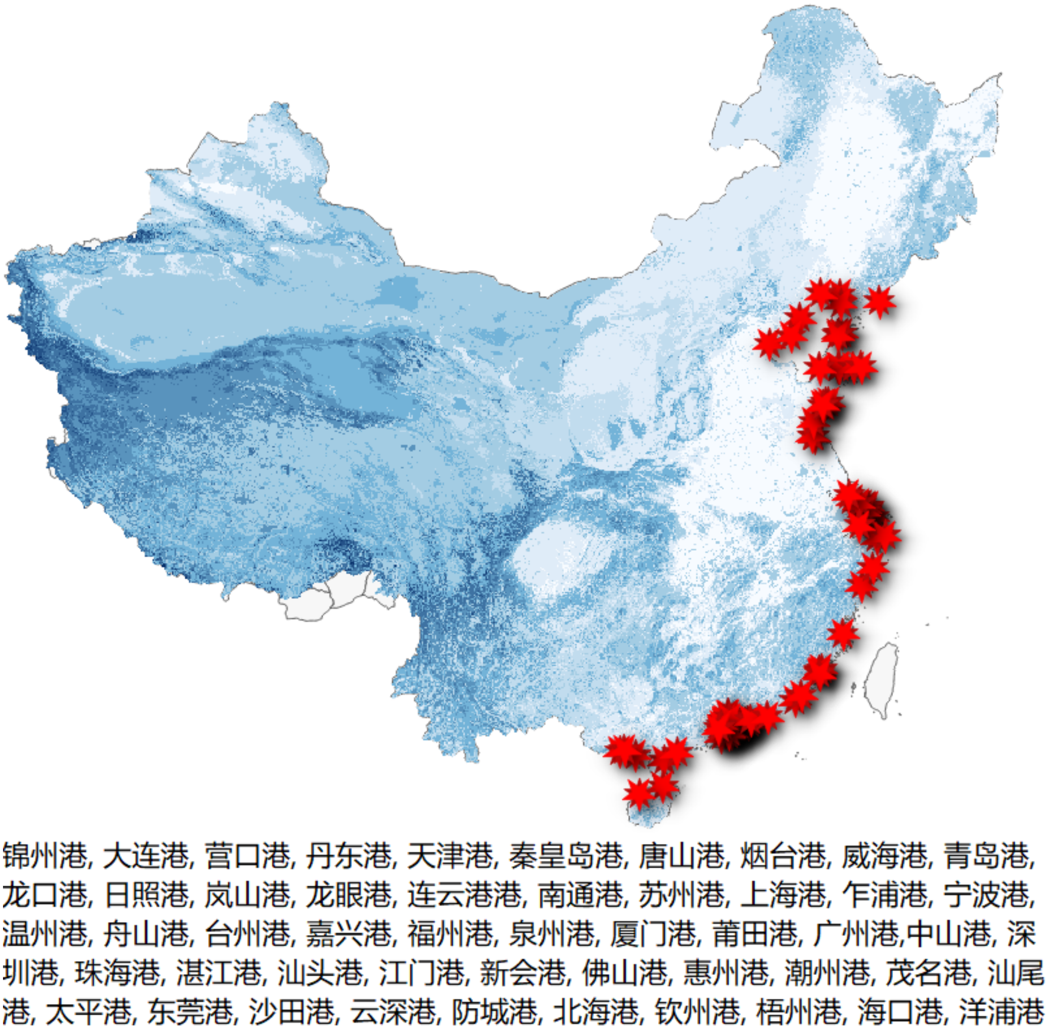
Firms choose to use new constructed highways because the potential benefits are larger than payments to use highways. Due to China' high-quality roads normally charging toll fees, while other roads with lower quality are usually free to use, when a firm can get better access to road networks, its transportation time will decrease while the toll fee payment may increase. On the one hand, toll fees increase firms' transport costs; on the other hand, new-constructed roads tend to have higher quality, which can save firms' time consumption, reduce fueling costs, increase firms' turnover rate and induce them to sell products more smoothly. Extant empirical studies normally agree that domestic integration and trade are largely promoted by infrastructure development. This implies that most firms tend to use highways even if they need to pay toll fees, as the benefits of using highways are larger than toll fee expenditures.

Table 5.4-5: Designed Speed of Different Types of Roads

Types of Roads	Expressway	First	Second	Third	Fourth
Designed Speed (km/h)	80-120	60-100	80-60	40-30	20

Data Source: China's Highway Engineering Technique Standard in 2003 (JTG B01-2003). The designed speed can vary according to terrain construction costs. For example, in some plain regions, expressways can reach their maximum designed speed (120km/h), while decreasing to only 60km/h in some rugged areas.

Figure 5.4-2: Seaports Reported by Chinese Statistical Yearbook from 2000 to 2006.



The Chinese name of the seaports are provided in this figure, collected form Chinese Statistical Yearbook.

Most previous trade literature acknowledges that transportation cost is an important component of trade cost, but transportation cost has not been precisely measured due to the limitation of spatial data. Manova et al. (2015) divide trade cost into fixed and variable trade costs. Fixed trade cost is associated with the cost of market investigation, preliminary investment, and customization of products; while the variable trade cost includes transportation, storage costs, and insurance payment. However, most existing studies use distance-based proxies to capture transportation costs or trade cost. For example, Manova et al. (2015) study the impacts of financial constraints on export activities, their trade cost is indicated by bilateral distance and administration cost.[12] Manova (2012) uses shipment data, and investigates the impacts of transportation cost on shipment weight, while the transportation cost is indicated by freight distance. Baldwin & Harrigan (2011) use bilateral distance to destinations to capture transportation cost. These measurements have two limitations. First, if the distance is

[12]The data is collected from the World Bank's 'Doing Business Report'.

calculated at an aggregated level, these measurements are only cross-section observations, they cannot capture the changes of transportation cost over time. Second, there are also some studies that use shipment data which can provide shipment distances to destinations, but such types of firm-level data are not always available in other countries.

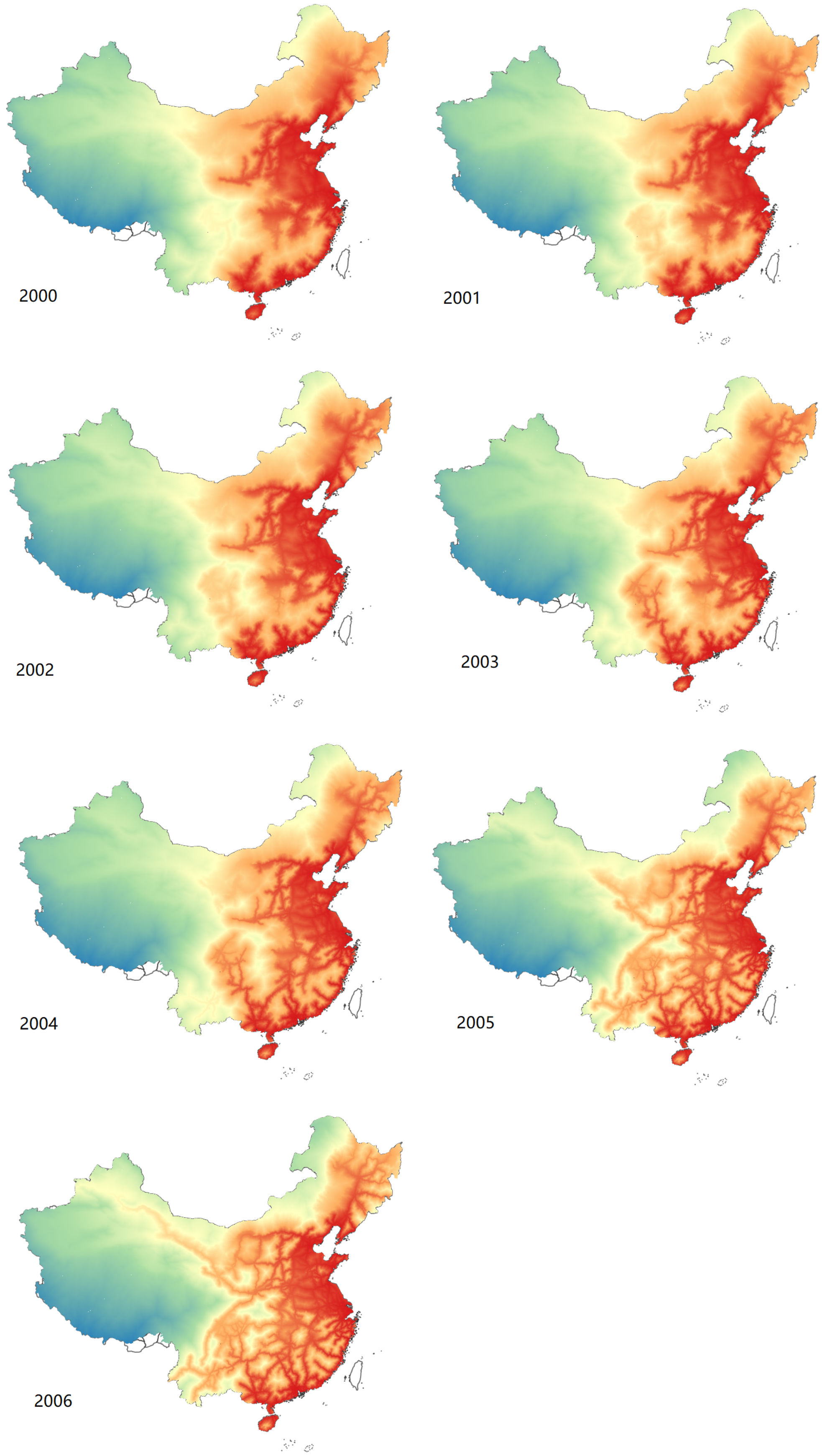
The present study estimates the accumulated domestic transportation costs to capture the impacts of highway development on trade activities. Some recent studies have proposed several different ways to estimate traffic accessibility or trade costs to capture the impacts of road connection, e.g., physical infrastructure stocks (Bougheas et al. 2000), distance to roads (Matthee & Naudé 2008, Faber 2014), distance to trade ports (Matthee & Naudé 2008), or the dummy to indicate whether a region is connected by traffic lines (Tang et al. 2019); however, these approaches are not specialized for international trade because the factors of trade ports are not considered. In this chapter, we want to estimate firms' trade costs to the nearest trade port along a given road network, which requires us to combine terrain data with road network data, then estimate how firms will select their optimal paths, conditional on the spatial distribution of roads and different terrain factors. The combination of highway and terrain data is called the travel cost map, defining the logistic costs of each unit of distance. Terrain data is directly the cost map used in the third and fourth chapters, composed by three parts: land ruggedness; elevation; and land cover maps. Road network data is the new ACASIAN data updated by published atlases (1998 to 2007).

To combine terrain and highway data, it is necessary to define their relative weight between terrain data and highway data. Our highway data can provide the information for high-quality highways (expressways and first-class highways), the land cost index in terrain data represents the overall trade cost on those lands without high-quality highways, and the transportation on these lands are mainly supported by low-class roads.[13] For example, Figure 5.4-1 and Table 5.4-5 show that the majority of China's roads are third or fourth class highways, or non-classified roads, which only allow vehicles going at speeds lower than 40km/h, implying that those lands without expressways have much higher speed limitation for goods transportation than expressways. Expressways have the highest quality and allow vehicles to reach a speed about 80-120km/h; while first-class roads allow the speed around 60-100km/h.[14]

[13]China's Highway Engineering Technique Standard in 2003 (JTG B01-2003) provide the definition of five types of roads, i.e., expressway, first, second, third, and fourth class roads. Due to our trade dataset covering the period from 2000 to 2006, we use China's Highway Engineering Technique Standard in 2003 (JTG B01-2003) to define highways' quality.

[14]First class roads in our dataset are mainly 'urban expressways', they were constructed to connect urban

Figure 5.4-3: Domestic Trade Cost along Highways to Trade Ports during 2000 to 2006



The light color means lower domestic trade cost.

---

roads with highway system, surrounding big cities, but are normally constructed with the technical standards of first class even though their name contains the term 'expressway'.

Besides the speed limits on different types of road, the average road length can also affect transportation speed. Low-quality road networks tend to have much shorter average road lengths and contain many more crossroads and signal lights than expressways, because the latter are normally designed to connect farther regions and big cities. These facts imply that the density of low-quality roads is very large in those regions without high-quality highways, and the relative speed between low- and high-quality roads can represent the relative trade cost in terrain data and highway data. We take the average speed on high-quality roads is 100km/h, while the average speed on land without high-quality roads should be lower than 20km/h, the relative transportation cost on terrain and highway dataset is 5:1; considering a series of transportation barriers on low-quality roads, this is still a conservative estimate.

The last step is to generate the map of accumulated travel cost (Figure 5.4-3), on the basis of seaport map and travel cost map. The seaport map can define the destinations that exporters want to arrive, while the travel cost map can reflect the travel costs that firms need to pay when they travel in China. The travel cost map is a raster map, and it divides Chinese land into many cells (approximately $0.82x0.82km^2$), while each cell can reflect the travel cost a firm travels across the cell. Consequently, the accumulated travel cost means that a firm chooses its optimum route from its address to the nearest seaport, we add up the travel cost of all cells to get the total travel cost.

The red colour in Figure 5.4-3 indicates the lower accumulated travel cost while blue indicates higher cost, the red areas tend to become larger with highway expansion. However, this measurement also has one limitation, i.e., it can only capture firms' traffic accessibility to trade ports, while there are also a small proportion of firms' exports on land, with countries sharing a border with China. [15]

5.4.3 Variable Selection and Model Specification

To investigate the impacts of infrastructure development on firms' export activities, the present study uses firms' product-destination-specific export quantities to capture the intensive margin and export scope to capture the extensive margin. The key explanatory variable is the domestic geographical transportation cost DGC, i.e., firms' accumulated transportation

[15] According to an announcement of Chinese government, there is less than 10% Chinese trade flows are transported on land, suggesting that we can consider more than 90% trade activities can be captured by our accumulated travel cost measurement.

cost along highways to the nearest trade ports. As robustness checks, this study also uses firms' distance to highways and highway density to capture domestic transportation costs. Besides the domestic transportation cost, we also consider the impacts of international geographical transportation cost $ForeignDist$, which is indicated by firms' spherical distance to destination countries (Table 5.4-6). To capture the agglomeration effects of local cities, this study uses prefecture-level population density (the same as in the third and fourth chapters) to capture agglomeration effects from local markets, and province-level fixed effects to capture urbanization effects.

Table 5.4-6: Summary Statistics of Transportation Cost Indicators

log	Obs	Full Sample Average (Std. Dev.)	Obs	2000 Average (Std. Dev.)	Obs	2003 Average (Std. Dev.)	Obs	2006 Average (Std. Dev.)
DGC	4,705,456	5.441	282,173	5.828	566,320	5.472	1,197,063	5.308
		(1.178)		(1.519)		(1.164)		(1.081)
(firm-level)								
full sample	264,550	5.628	20,186	5.937	33,269	5.668	57,736	5.493
		(1.283)		(1.559)		(1.281)		(1.170)
pre-2000 entry	98,379	5.559	20,186	5.937	13,434	5.420	10,632	5.194
		(1.355)		(1.559)		(1.228)		(1.148)
post-2000 entry	166,171	5.669	0	-	19,835	5.836	47,104	5.560
		(1.237)		(-)		(1.288)		(1.165)
Unchanged Address:								
full sample	187,357	5.655	20,186	5.937	19,797	5.694	40,296	5.506
		(1.304)		(1.559)		(1.296)		(1.172)
pre-2000 entry	58,253	5.637	20,186	5.937	5,649	5.349	3,894	5.071
		(1.397)		(1.559)		(1.248)		(1.176)
post-2000 entry	129,104	5.663	0	-	14,148	5.832	36,402	5.553
		(1.239)		(-)		(1.289)		(1.162)
Foreign Dist	4,734,187	16.023	284,491	16.035	570,000	16.021	1,203,745	16.023
		(.631)		(.620)		(.632)		(.632)
(firm-level)	235,554	16.027	17,317	16.042	24,783	16.027	55,786	16.023
		(.626)		(.621)		(.641)		(.626)

Note: The variable DGC is firms' accumulated geographical transportation cost along highways to the nearest trade ports, this index is constructed according to cost map, expressways, and trade ports (it dose not has unit.). $Foreign - Dist$ is country-level spherical distance from China to each destination country, each firm may have many destination countries, its unit is log meter.[16]

To investigate the impacts of firm entry and relocation, Table 5.4-6 divides exporters into several groups. First, exporters are divided into 'pre-2000 entry' and 'post-2000 entry' groups. to distinguish whether a firm has already been a exporter since 2000. The decrease of DGC in 'pre-2000 entry' group is faster than 'post-2000 entry' group, suggesting new exporters are normally in inland regions, far from trade ports. When we only keep those firms that

[16]The spherical distance is calculated according to longitude and latitude coordinates, i.e., $d_{ij} = R \times arccos[sin(latitude_i)sin(latitude_j) + cos(latitude_i)cos(latitude_j)cos(longitude_i - longitude_j)]$. R is average earth radius, i.e., 6371.004km.

have never changed their addresses, the decrease of DGC in 'pre-2000 entry' group is still faster than 'post-2000 entry' group, suggesting the firms' relocation is not the reason of new exporters emerging in inland regions. At the same time, the transaction-level DGC is smaller than firm-level DGC, suggesting although most new exporters are in inland regions, exporters closer to seaports tend to generate more transaction flow than inland exporters.

Figure 5.4-4: Kernel Density Estimate of DGC

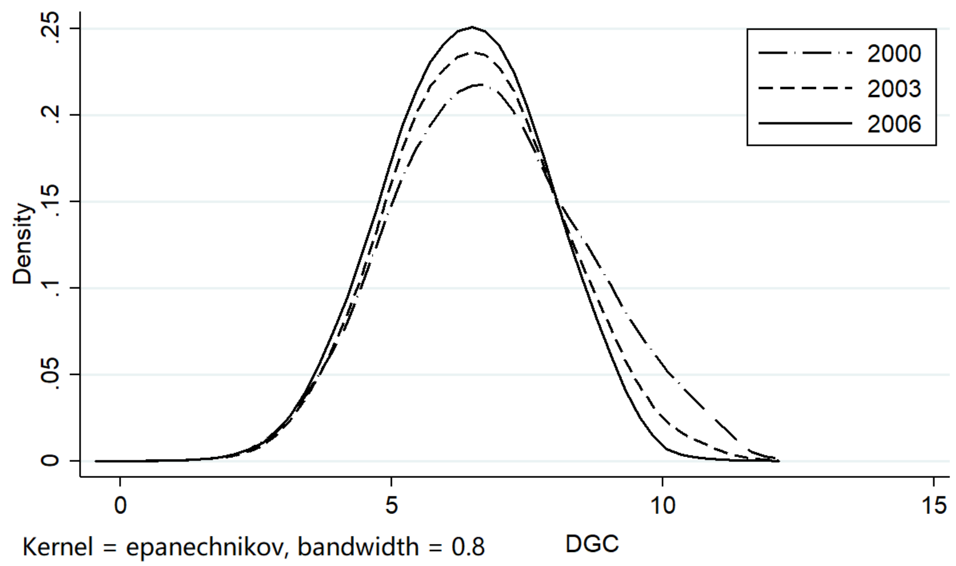


Figure 5.4-5: Kernel Density Estimate of Foreign Dist

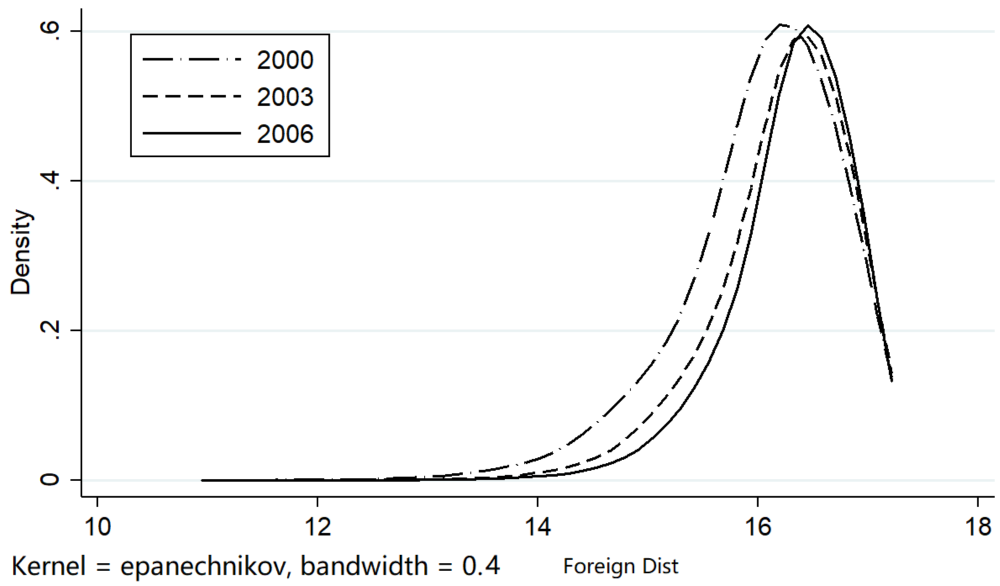


Figure 5.4-4 shows the transaction-level kernel density estimate of DGC; it can be found that the overall distribution moves from the right to left, its kurtosis tends to increase while the skewness gradually declines from 2000 to 2006, implying that domestic transportation tends

to decrease due to highway expansion. By contrast, Figure 5.4-5 illustrates the kernel density estimate on $Foreign - Dist$. The overall distribution moves from left to right, suggesting that more firms tend to export to further destinations with highway expansion.

The baseline regressions use Heckman two-stage regression to investigate firms' trade activities. The first stage describes how transportation costs affect firms' export decisions; the second stage describes the impacts on export intensive and extensive margins. We use this method because of self-selection issue, i.e., export firms are not randomly selected samples; if we directly regress the export quantity on transportation cost, the factors that have causal impacts on firms' export decision would not be considered and this will cause endogeneity issues. Besides Heckman two-stage approach, we also use instrument variables as robustness checks to address the endogeneity issues.

$$ExportDecision_{it} = \alpha_0 + \beta_1 DGC_{it} + \mu X_{it} + \epsilon_n + \epsilon_r + \epsilon_t + u_{it}. \tag{5.5}$$

For the first stage, the dependent variable indicates whether the firm i decides to export at time y. The coefficient of the key explanatory variable DGC_{it} on the right-hand side represents the influence of domestic transportation cost on firms' probability to start trading activities. The control variable vector X_{jt} includes firms' ownership, age, size, and prefecture-level population density. ϵ_n, ϵ_r, and ϵ_t represent 2-digit industry-level, province-level, and year fixed effects, u_{it} captures firm-level error terms.

$$Export_{ijdt} = \alpha_0 + \beta_1 DGC_{it} + \beta_2 Foreign - Dist_{idt} + \mu X_{idt} + \epsilon_n + \epsilon_r + \epsilon_t + u_{it}. \tag{5.6}$$

For the second stage, the variable $Export_{ijd,t}$ represents the export quantity of product j from firm i toward destination d at time y, or the export scope of firm i toward destination d at time y. Firm-level transportation cost is indicated by DGC_{it} and $Foreign - Dist_{idt}$. Unlike the first stage, the second-stage model contains the firm's destination factors because this can only be known after firms start export activities. X_{idt} includes firm-level ownership, age, and size effects, region-level urbanization, and infrastructure development effects; unlike the first stage, remoteness[17] and market scales (GDP) of destination countries are un-

[17] $Remoteness$ is defined as $Remoteness_d = \sum_o GDP_o distance_{od}$ (Manova & Zhang 2012).

der control, heterogeneity of export goods are also considered.[18], this information can only be known after firms start trading activities. ϵ_n, ϵ_r, and ϵ_t represent 2-digit industry-level, province-level, and year fixed effects, u_{it} captures the firm-level error term.

We decide these control variables on the basis of the existing trade literature, and their summary statistics are illustrated in Table H4. For example, Liu et al. (2017) include firm size, ownership, and productivity into consideration, while Tang et al. (2019) control firm size, age, and labour productivity in their control vector; because ownership can affect firms' trade decisions, larger and older firms tend to be more productive and more likely to be involved in international trade. To capture ownership effects, this study uses the same measurements as in the third and fourth chapters, i.e., the share of paid-in capital with different types of ownership. Firm size and age are also included to capture the scale economies and potential market influences of large and older firms.

To investigate the role of value-weight ratio across different goods, this study uses the four-digit industry-level measurement constructed by Syverson (2004).[19] Those industries with ratios larger than the median are defined as value intensive industries, while others are weight intensive industries ($Homo$). In addition, as an alternative proxy, this study also uses the HS-quoted names and export units to identify products' heterogeneity. There are three types of units in custom data: number, size, and weight (Table H3). The number-counted goods are normally finished devices and live animals, e.g., cars, microwave ovens, airplanes, shoes, pens, and pigs. Some of these number-counted products can be very large, while some others could be much smaller, i.e., large heterogeneity. The size-counted goods are normally intermediate goods, and their units include meters, square meters, liters, and cubic meters. Most of them have very low value-weight ratios, e.g., wood and glass; while some others are of much higher unit value, e.g., alcoholic beverages and cinematographic films. Due to goods' heterogeneity, the value-weight ratio of size-counted products cannot be directly calculated, an alternative measurement is value-size ratio. For liquid goods such as wine and soft drinks, their unit weight can be approximately represented by water density. The

[18] In order to identify homogeneous and heterogeneous goods, we use the product code provided by Rauch (1999) who classifies SITC 4-digit goods into three categories, i.e., goods traded in organized exchanges, goods with prices quoted in trade publications, and others. Goods in the first two categories are normally considered as homogeneous goods, which are widely used following trade studies.

[19] Some early studies, such as Pedersen & Gray (1998), to calculate the aggregated value-weight ratio in Norway. Syverson (2004) provides the 4-digit industry-level aggregated value-weight ratio, based on the data from American plant-level data, i.e., 1977 Census of Manufactures. Ong & Sou (2015) calculating value-weight ratio based on U.S. export and import data.

weight-counted goods are also normally intermediate products, their units include kilograms and grams; value-weight ratio can be directly calculated from these units.

In addition, this study also uses the reduced gravity model in robustness checks to capture the trade cost channel. Following McCallum (1995) and many empirical studies, trade costs t_{ij} from firm i to destination d at year y are assumed to follow a log-linear function, i.e., $t_{id,t} = Dist_{id}^{\gamma_1}\ Infra_{it}^{\gamma_2}\ e^{\gamma_0+\eta X_{id}}$. γ_1 is the elasticity of trade costs $t_{id,t}$ with respect to bilateral distance $Dist_j$, while γ_2 is the elasticity of trade cost with respect to infrastructure development $Infra_{it}$, where $TC_{it} = ln(Infra_{it})$. $X_{id,t}$ is a covariate term, it includes two factors to capture whether the trade partner has a shared border $border_{id}$ or a common language $language_{id}$.[20] To control agglomeration economies, this study uses region-level population density as indicator to control urbanization effects. Due to the fact that agglomeration economies can also be promoted by market accessibility. Similar to the extant literature, this study also calculates the market potential index based on the real road network. This index is calculated as the firm-level accumulated trade costs of core cities, while these core cities were defined by early highway construction plans in the early 1990s.

5.4.4 Instrument Variable Construction

Review of Instrument Variable Construction

Although this chapter constructs a new proxy based on the real road network to capture firms' traffic accessibility, domestic geographical transportation cost is potentially affected by the endogeneity issues, because when policy makers decide where to locate new roads their decisions can be affected by the spatial distribution of economic activities.[21]

To investigate the impacts of transportation infrastructure on trade flows, previous literature has proposed different proxies to capture transportation or overall trade cost. Most trade literature focuses on aggregate trade flow, so the infrastructure factors are also measured by aggregated proxies, e.g., Limao & Venables (2001) using country-level roads, railway density, and telephones per person to capture overall infrastructure density; Matthee & Naudé (2008), Duranton et al. (2014) calculate city-level distance to trade ports as the indicator of

[20]Substitute the trade cost function into the equation 5.5, generates the cross term with $Infra_{it}$, and let $\phi_1 = \gamma_0\beta_1$, $\phi_2 = \gamma_2\beta_1$, $\phi_3 = \gamma_1\beta_1$, $\phi_4 = \gamma_3\beta_1$, $\phi_5 = \gamma_4\beta_1$, get $Export_{id,t} = \alpha_0 + \phi_1 Infra_{it} + \phi_2 Infra_{it}^2 + \phi_3 Infra_{it} ln(Dist_{id}) + \phi_4 Infra_{it} language_{id} + \phi_5 Infra_{it} border_{id} + \mu X_{it} + \epsilon_n + \epsilon_d + \epsilon_r + \epsilon_t + u_{it}$.

[21]There are also several additional methods that can address the endogeneity issues, such as control-function-based estimation and 'special regressor' methods (Baum et al. 2012). IV approach is the most straightforward and widely used method for these approaches.

traffic accessibility; Poncet (2003) uses liberal distances between provincial capitals to capture domestic transportation cost; Coşar & Demir (2016) combine geographical locations and infrastructure factor to measure region-level transportation costs; Guo & Yang (2019) estimate region-level land costs to China's top-ten trade ports, by using charges data and spatial data of China's road and railway network. By contrast, there are also some studies that focus on firm-level trade activities, e.g., Liu et al. (2017) calculate firms' distance to highways and highway density, with a 20km or 30km radius for each firm as the proxies of traffic accessibility; Bougheas et al. (2000), Tang et al. (2019) use firms' location and regional data to capture transportation cost.

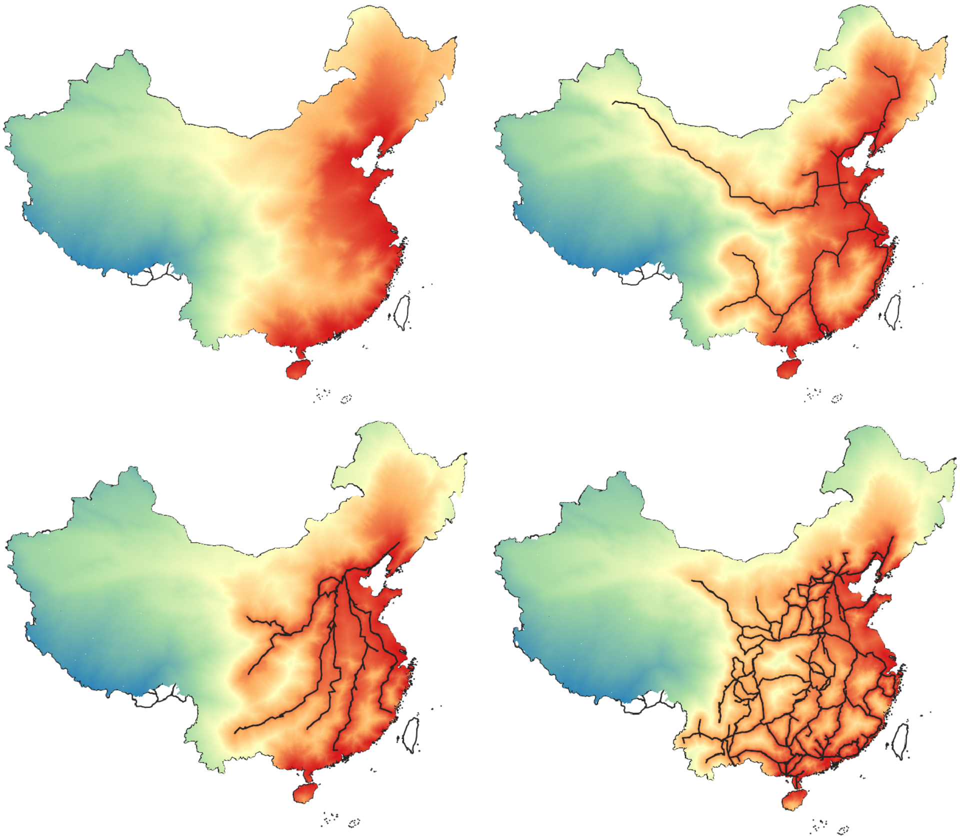
These studies have developed four types of instruments to address endogeneity issues, i.e., construction plan, historical roads, time-lagged terms, and counter-factual road approaches. For example, Martincus & Blyde (2013) use Peruvian historical roads as the instrument of new roads, Baum-Snow (2007), Michaels (2008) use U.S. highway construction plan as the instruments of current highways, while Duranton & Turner (2012) construct instruments basing on both construction plan and historical data. Coşar & Demir (2016) using time-lagged terms as the instrument of the key explanatory variables. Faber (2014), Liu et al. (2017), Tang et al. (2019) construct counter-factual roads as IVs based on geographical features.

Construction of Instrument Variables

Four different instruments (Terrain Surface, Ming Road, Qing Road, and Counter-factual Road) are constructed based on terrain data and road network data; they reflect firms' accumulated domestic transportation costs along historical roads or counter-factual roads to the nearest trade port. Historical roads are expected to have causal impacts but are exogenous to the spatial distribution of current roads. Counter-factual roads are constructed according to cost-minimized rule; they are also exogenous to the current road distribution.

Similar to previous chapters, counter-factual roads are generated by a least cost approach from cost map and node cities (similar as Faber (2014)), while cost map defines the road construction cost, and node cities are the points the policy maker wants to connect by highways. The historical roads include Ming dynasty roads (1363-1644) and Qing courier routes (1800-1900), provided by Fairbank Center for Chinese Studies of Harvard University and the Center for Historical Geographical Studies at Fudan University (CHGIS dataset). The node cities are defined by the '5-7' plan (5 longitudinal and 7 latitudinal highways), which was proposed in 1980s by the Ministry of Communications and approved by the State Council of China in 1993. These cities were selected to connect into a nationwide highway network, because of their higher level of urbanization, population density, and economic output in 1980s. The cost map is merged by land ruggedness map (34% share), elevation map (33% share) and land cover map (33% share). The land ruggedness map is provided by Nunn & Puga (2012), the elevation map is collected from USGS National Elevation Dataset (NED), while the cover map is collected from a dataset provided by Broxton et al. (2014).

Figure 5.4-6: Terrain Surface IV (upper left), Counter-factual Road IV (upper right), Qing Road IV (bottom left), and Ming Road IV (bottom right).



First, we hope to understand the individual impact of terrain data, the instrument Terrain Surface is firms' accumulated transportation cost on the surface without any road. Then, according to the given historical roads, two instruments are calculated from the firms' accumulated transportation costs along the roads from the Ming and Qing dynasties. The last instrument is based on Counter-factual Road, constructed from terrain data and counter-factual roads generated from the least cost approach. These instruments reflect how firms would select their optimal paths, conditional on the spatial distribution of the given road network and different terrain surfaces. Terrain data is directly the cost map used in the least cost approach, composed by land ruggedness, elevation, and land cover maps; then the terrain data is merged with road network data, with the weight of 5:1, the same as DGC.

Table 5.4-7: Summary Statistics of Instrument Variables

	statistics	All	2000	2003	2006
Terrain Surface	Obs	1,404,939	142,027	176,974	274,560
	Average	5933.930	7157.196	6149.108	5414.782
	Std. Dev.	(9876.882)	(10909.760)	(10203.24)	(9229.533)
Ming Road	Obs	1,404,939	142,027	176,974	274,560
	Average	1825.769	2197.341	1875.659	1675.530
	Std. Dev.	(3207.724)	(3956.905)	(3412.241)	(2686.619)
Qing Road	Obs	1,404,939	142,027	176,974	274,560
	Average	2744.203	3334.647	2809.604	2503.133
	Std. Dev.	(4704.200)	(5686.654)	(4944.703)	(4051.553)
Counter-factual Road	Obs	1,404,939	142,027	176,974	274,560
	Average	2169.923	2572.884	2226.958	2030.661
	Std. Dev.	(3351.596)	(3947.573)	(3504.807)	(2987.228)

Note: This table illustrates summary statics of four instrument variables (they do not have units). Two IVs are constructed according to historical roads from the Ming and Qing dynasties. The Counter-factual IV is constructed based on Counter-factual Roads, while the Terrain Surface IV is only based on geographical factors such as land ruggedness and land cover information.

Table 5.4-7 summarizes the four instruments, i.e., Terrain Surface IV, Qing Road IV, Ming Road IV, and Counter-factual Road IV. It is noteworthy that, even if the instruments are constructed as time-invariant surfaces, the firm-level observations still change over time, because of firms' entry and exits. If many new exporters emerged from custom data during 2000 to 2006, and their relative distances are closer to most trade ports, then the instruments will tend to decline. Similarly, if lots of firms gradually relocated to coastal areas, the IVs will also decrease over time. Table 5.4-7 shows the significant decline of IV values and the rise of sample size during 2000 to 2006, suggesting that the entry of new exporters plays an essential role.

5.5 Empirical Analysis

5.5.1 Baseline Results

This section investigates the impacts of highway construction on the pattern of China's export activities. At the first stage, highway development is expected to encourage more firms to export; at the second stage, the export pattern (export intensive and extensive margins) of those firms which have started trading activities will be further shaped by highway construction. At the same time, the key explanatory variable DGC can be affected by the endogeneity issues, because the spatial distribution of new roads can be affected by economic factors. Policy makers are motivated to locate new infrastructure in those regions with higher potential economic growth, while this advantage may remain stable until the new infrastructure has been completed. To solve the potential simultaneity issues, this study constructs several IVs based on historical roads and counter-factual roads. Although these IVs have many different combinations, we only present the results that can pass the Under-identification and Over-identification tests.

Table 5.5-8 investigates the impacts of geographical transportation costs on firms' export decisions. The fixed effects in columns (1) to (7) are controlled by two components, i.e., year fixed effects and time-invariant cross fixed effects. For the time-invariant cross fixed effects, column (1) controls time-invariant province-level and 4-digit industry fixed effects, column (2) controls time-invariant prefecture-level and 4-digit industry fixed effects, while column (3) controls time-invariant county-level and 4-digit industry fixed effects. Regional and industrial fixed effects can control persistent scale gaps across regions and industries, and the impacts of geographical factors such as natural resource endowment, overall terrain ruggedness level, and local climate; while year fixed effects can capture nationwide shocks on firm size in each year. Columns (4) to (7) further include firm-level fixed effects; column (4) only controls firm-level time-invariant fixed effects; column (5) controls time-invariant province-level, firm-level, and 4-digit industry fixed effects; column (6) controls time-invariant prefecture-level, firm-level, and 4-digit industry fixed effects; while column (7) controls time-invariant county-level, firm-level, and 4-digit industry fixed effects. Firm-level fixed effects can capture firm-level persistent scale gaps; the cross fixed effects between firm and region can control the relocation effect, which means that if a firm changes its address over the sample period, the cross fixed effects between firm and region can capture firm-level fixed effects in different regions separately; the cross fixed effects between firm and industry can control the case that a firm changes its 4-digit industry over sample period. Due to columns (5) to (7) control firm-level, region-level, and industry-level fixed effects simultaneously, the firm-level persistent size gaps, relocation effect, and industry transition effect can be captured simultaneously.

Table 5.5-8: The Impacts of Highways on Export Decision.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Decision	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.019***	-0.004*	-0.000	-0.003*	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Size	0.057***	0.059***	0.061***	0.031***	0.027***	0.027***	0.025***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Foreign-share	0.363***	0.331***	0.314***	0.033***	0.027***	0.027***	0.026***
	(0.007)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Rail-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Road-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TFP-ACF	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Collective-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	1435106	1428863	1405756	1331353	1264484	1264242	1249384
Group	10186	46856	116998	327712	343734	343702	345454
Under-identification test	400.875	1076.362	1262.561	7373.757	6583.139	6565.432	6130.146
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.0e+06	7.6e+05	4.6e+05	2.9e+05	2.5e+05	2.5e+05	2.4e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.251	0.060	0.320	1.225	0.853	1.081	0.290
P value	(0.616)	(0.806)	(0.572)	(0.268)	(0.356)	(0.298)	(0.590)
First-stage Results:							
Terrain-Surface	0.494***						
	(0.012)						
Counter-factual-Road	0.260***	0.491***	0.446***	0.478***	0.485***	0.485***	0.486***
	(0.013)	(0.007)	(0.009)	(0.006)	(0.007)	(0.007)	(0.007)
Ming-Road		0.370***	0.432***	0.502***	0.495***	0.495***	0.495***
		(0.008)	(0.009)	(0.006)	(0.007)	(0.007)	(0.007)
F	2744.440	1509.867	1985.517	8124.610	6745.826	6735.313	6310.122
R2	0.660	0.581	0.513	0.479	0.446	0.446	0.438
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I1. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

The key explanatory variable is domestic geographical transportation cost (DGC), it can reflect firms' accumulated travel cost to their nearest seaport. The variation of DGC can be affected by firms' location, as well as highways' expansion. The results in columns (1) to (3) show that, when the firm-level fixed effects are not included, a 10% decrease of DGC can increase 0.04% to 0.20% export probability. The results in columns (4) to (7) show that, when the firm-level fixed effects or firm-region-industry fixed effects are included, the coefficients of DGC become insignificant; suggesting that the impacts of highway construction on firms' export decisions are not robust. A possible explanatory is that the impacts of highway construction on firms' export decisions are absorbed by firm-region-industry fixed effects, because the value of DGC can also be affected by firms' location. At the same time, the two control variables provide consistent coefficients, i.e., larger and foreign-invested firms are more likely to export. A 10% increase of firm size is related to a 0.3% increase of export probability, a 10% increase of foreign-owned capital is related to about 0.26% increase of export probability. These results suggest that state-owned and collective-owned firms are more likely to focus on the domestic market rather than the overseas market, while foreign-owned firms tend to maintain both larger intensive and extensive margins, and are more affiliated to foreign markets, consistent with many recent Chinese studies such as Liu et al. (2017).

Table 5.5-9 investigates the impacts of geographical transportation cost on firm-level export values, controlled by the same fixed effects and control variables. The coefficient of DGC in column (1) is -0.062, suggesting that a 10% decrease of domestic transportation cost leads to about 0.8% increase of firm-level export value, consistent with Liu et al. (2017) and Tang et al. (2019). By contrast, the coefficients of DGC in other columns suggest that a 10% decrease of domestic transportation cost leads to about 0.2% to 0.4% decrease of firm-level export value, suggesting the location effects are absorbed by the cross fixed effects between firm and region. At the same time, Table 5.4-6 tells us new exporters are normally in inland regions (not very close to seaports). Therefore, the positive coefficients of DGC in columns (4) to (7) imply that, when the firm-level location effects are absorbed by region-firm-industry fixed effects, highway expansion has more significant impacts on the export values of those firms not very close to seaports.

Besides the key explanatory variable, there are also several control variables that provide consistent coefficients. The coefficients of TFP, GDP, and $Size$ are positive, a 10% increase of TFP is related to 1.4% increase of firm-level export value, a 10% increase of average GDP of firms' destination countries is related to 3.7% increase of firm-level export value, a 10% increase of asset size is related to 5.3% increase of firm-level export value, while a 10% decrease of firms' average $Remoteness$ is related to 2.8% increase of firm-level export value, consistent with Liu et al. (2017).

Table 5.5-9: The Impacts of Highways on Firm-level Export Value.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.077***	-0.021	0.040**	0.023*	0.035**	0.035**	0.035**
	(0.013)	(0.014)	(0.015)	(0.011)	(0.012)	(0.012)	(0.012)
TFP-ACF	0.087***	0.082***	0.110***	0.136***	0.139***	0.139***	0.138***
	(0.010)	(0.009)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
Average GDP	1.105***	1.006***	0.979***	0.389***	0.368***	0.367***	0.366***
	(0.083)	(0.068)	(0.065)	(0.054)	(0.057)	(0.057)	(0.057)
Average Foreign-Dist	0.164***	0.167***	0.157***	0.166***	0.172***	0.173***	0.175***
	(0.027)	(0.026)	(0.025)	(0.024)	(0.025)	(0.025)	(0.025)
Average Remoteness	-0.895***	-0.801***	-0.783***	-0.295***	-0.276***	-0.276***	-0.276***
	(0.080)	(0.066)	(0.064)	(0.054)	(0.057)	(0.057)	(0.057)
Size	0.612***	0.644***	0.679***	0.554***	0.530***	0.530***	0.529***
	(0.012)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)	(0.011)
Homo	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rail-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Road-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Collective-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Foreign-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	187185	184559	179183	170945	160049	160020	158273
Group	5235	13277	23471	48663	50957	50950	50998
Under-identification test	208.737	498.473	461.678	2774.507	2354.154	2351.269	2303.083
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.2e+05	1.1e+05	7.0e+04	6.1e+04	5.1e+04	5.1e+04	4.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	3.473	0.007	0.670	0.989	0.445	0.545	0.001
P value	(0.06)	(0.932)	(0.413)	(0.320)	(0.505)	(0.460)	(0.976)
First-stage Results:							
Counter-factual-Road	0.240***						
	(0.023)						
Qing-Road	0.512***	0.176***	-0.010	0.094*	0.132**	0.129**	0.128*
	(0.020)	(0.024)	(0.049)	(0.048)	(0.049)	(0.049)	(0.051)
Terrain-Surface		0.534***	0.729***	0.687***	0.640***	0.644***	0.643***
		(0.023)	(0.046)	(0.046)	(0.047)	(0.047)	(0.049)
F	1314.688	1457.317	1451.227	2188.465	1671.592	1672.099	1657.660
R2	0.646	0.608	0.576	0.619	0.584	0.584	0.586
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table J1. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

By contrast, a 10% increase of firms' average distance to destinations is related to 1.7% increase of firm-level export value, suggesting firms tend to have higher trade flow if a firm's average bilateral distance is larger. At the same time,some regional control variables such $RoadDensity$, $RailDensity$, $PopDensity$ can be affected by the potential multicollinearity issues. This issue could be induced by the inter-correlation across different types of transport routes and city scales, e.g., route density could be higher in coastal and higher income regions, route density could also be higher around some big cities or regional capitals. Due to the coefficients of these region-level variables are more likely to be biased, we do not provide further economic interpretation on these coefficients.

Table 5.5-10: The Impacts of Highways on Firm-level Export Share in Revenue.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Share	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.087***	-0.043**	0.032*	0.012	0.023*	0.022*	0.023*
	(0.013)	(0.014)	(0.015)	(0.010)	(0.011)	(0.011)	(0.011)
TFP-ACF	-0.293***	-0.280***	-0.243***	-0.113***	-0.108***	-0.108***	-0.108***
	(0.011)	(0.009)	(0.008)	(0.006)	(0.006)	(0.006)	(0.006)
Average GDP	0.839***	0.775***	0.779***	0.351***	0.336***	0.335***	0.337***
	(0.076)	(0.065)	(0.062)	(0.052)	(0.055)	(0.055)	(0.055)
Average Foreign-Dist	0.153***	0.150***	0.142***	0.158***	0.153***	0.154***	0.156***
	(0.027)	(0.025)	(0.025)	(0.023)	(0.024)	(0.024)	(0.024)
Average Remoteness	-0.632***	-0.574***	-0.590***	-0.259***	-0.247***	-0.246***	-0.249***
	(0.074)	(0.064)	(0.061)	(0.052)	(0.055)	(0.055)	(0.055)
Size	-0.086***	-0.061***	-0.032***	0.035***	0.036**	0.036**	0.037***
	(0.009)	(0.007)	(0.007)	(0.010)	(0.011)	(0.011)	(0.011)
Homo	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rail-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Road-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Collective-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Foreign-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	187185	184559	179183	170945	160049	160020	158273
Group	5235	13277	23471	48663	50957	50950	50998
Under-identification test	199.874	498.473	461.678	2774.507	2354.154	2351.269	2303.083
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.2e+05	1.1e+05	7.0e+04	6.1e+04	5.1e+04	5.1e+04	4.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	1.066	1.286	2.439	0.385	0.003	0.017	0.431
P value	(0.302)	(0.257)	(0.118)	(0.535)	(0.957)	(0.895)	(0.511)
First-stage Results:							
Ming-Road	0.064*						
	(0.032)						
Qing-Road	0.660***	0.176***	-0.010	0.094*	0.132**	0.129**	0.128*
	(0.026)	(0.024)	(0.049)	(0.048)	(0.049)	(0.049)	(0.051)
Terrain-Surface		0.534***	0.729***	0.687***	0.640***	0.644***	0.643***
		(0.023)	(0.046)	(0.046)	(0.047)	(0.047)	(0.049)
F	1264.284	1457.317	1451.227	2188.465	1671.592	1672.099	1657.660
R2	0.635	0.608	0.576	0.619	0.584	0.584	0.586
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I3. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 5.5-10 investigates the impacts of geographical transportation cost on firm-level export intensity, which is indicated by firms' exports-sales ratio. The fixed effects and control variables are the same as previous regressions. The coefficients of DGC in columns (1) and (2) show that a 10% decrease of domestic transportation cost can induce 0.4% to 0.9% increase of firm-level exports-sales ratios, consistent with Table 5.5-8. By contrast, the coefficients of DGC in other columns suggest that a 10% decrease of domestic transportation cost can induce 0.2% to 0.3% decrease of firm-level exports-sales ratios, suggesting the location effects are absorbed by firm-region-industry fixed effects. This pattern is consistent with Table 5.5-9, i.e., when the firm-level location effects are absorbed by region-firm-industry fixed effects, highway expansion has more significant impacts on the export values of inland firms. At the same time, there are also several control variables that provide consistent coefficients. The coefficients of $Size$, TFP and $Remoteness$ are negative, a 10% decrease of firm size is related to 0.4% increase of firm-level exports-sales ratios, a 10% increase of TFP is related to 1.1% decrease of firm-level exports-sales ratios, a 10% decrease of firms' average $Remoteness$is related to 2.5% increase of firm-level exports-sales ratios. The coefficients of GDP and $Foreign - Dist$ are positive, a 10% increase of GDP of destination country is related to 3.4% increase of firm-level exports-sales ratios, a 10% increase of firms' distance to destinations is related to 1.5% increase of firm-level export value. The singles of the coefficients $Foreign - Dist$, GDP and $Remoteness$ are consistent with Table 5.5-9, but the negative coefficients of TFP and $size$ suggest that productivity firms have lower exports-sales ratios and more diversified production structure.

Table 5.5-11 investigates the impacts of geographical transportation cost on firm-level export variety, which is indicated by the number of products. The fixed effects and control variables are the same as previous regressions. The coefficients of DGC in columns (1) and (2) show that a 10% decrease of domestic transportation cost can induce 0.2% to 0.5% increase of firm-level export variety, consistent with Liu et al. (2017). By contrast, the coefficients of DGC in other columns are insignificant, suggesting the impacts of highway expansion on firm-level export variety are not robust.

Table 5.5-11: The Impacts of Highways on Firm-level Export Variety (Product Variety).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.050***	-0.023**	0.010	-0.002	0.002	0.002	0.001
	(0.009)	(0.008)	(0.009)	(0.005)	(0.006)	(0.006)	(0.006)
TFP-ACF	0.012**	0.009*	0.010*	0.014***	0.017***	0.017***	0.017***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Average GDP	0.408***	0.414***	0.381***	0.180***	0.173***	0.173***	0.176***
	(0.049)	(0.034)	(0.032)	(0.022)	(0.023)	(0.023)	(0.023)
Average Foreign-Dist	0.069***	0.071***	0.060***	0.069***	0.069***	0.070***	0.070***
	(0.016)	(0.012)	(0.011)	(0.009)	(0.009)	(0.009)	(0.009)
Average Remoteness	-0.320***	-0.331***	-0.302***	-0.141***	-0.135***	-0.134***	-0.138***
	(0.048)	(0.034)	(0.032)	(0.022)	(0.023)	(0.023)	(0.023)
Size	0.146***	0.160***	0.174***	0.156***	0.151***	0.151***	0.150***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Homo	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rail-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Road-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Collective-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Foreign-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	187185	184559	179183	170945	160049	160020	158273
Group	5235	13277	23471	48663	50957	50950	50998
Under-identification test	242.984	498.473	461.678	2774.507	2354.154	2351.269	2303.083
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.3e+05	1.1e+05	7.0e+04	6.1e+04	5.1e+04	5.1e+04	4.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	3.035	1.915	0.480	0.093	0.501	0.349	0.363
P value	(0.082)	(0.166)	(0.488)	(0.760)	(0.479)	(0.555)	(0.547)
First-stage Results:							
Qing-Road	0.101***	0.176***	-0.010	0.094*	0.132**	0.129**	0.128*
	(0.018)	(0.024)	(0.049)	(0.048)	(0.049)	(0.049)	(0.051)
Terrain-Surface	0.592***	0.534***	0.729***	0.687***	0.640***	0.644***	0.643***
	(0.016)	(0.023)	(0.046)	(0.046)	(0.047)	(0.047)	(0.049)
F	1607.874	1457.317	1451.227	2188.465	1671.592	1672.099	1657.660
R2	0.653	0.608	0.576	0.619	0.584	0.584	0.586
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I4. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

At the same time, there are also several control variables that provide consistent coefficients. The coefficients of $Size$, TFP, GDP, and $Foreign - Dist$ are positive, a 10% increase of firm size is related to 1.5% increase of firm-level export variety of products, a 10% increase of TFP is related to 0.2% increase of firm-level export variety of products, a 10% increase of GDP of destination country is related to 1.7% increase of firm-level export variety of products, a 10% increase of firms' average distance to destination countries is related to 0.7% increase of firm-level export variety of products. By contrast, a 10% decrease of firms' average $Remoteness$is related to 1.4% increase of firm-level export variety of products. These results suggest that productive and larger firms have a higher variety of products, consistent with the heterogeneous firm trade models represented by Melitz (2003). However, the positive coefficients of $AverageForeign - Dist$ are unexpected results. A possible explanation is that larger and productive firms have more diversified trade relationships with more countries, so their average distances to destinations are also larger.

Table 5.5-12 also investigates the impacts of geographical transportation cost on firm-level export variety, but the variety is indicated by the number of destinations. The fixed effects and control variables are the same as previous regressions. The coefficients of DGC in columns (1) and (2) show that a 10% decrease of domestic transportation cost can induce 0.2% to 0.3% decrease of firm-level export variety of destinations, suggesting that those exporters close to seaports tend to focus on a smaller number of destination countries. By contrast, the coefficients of DGC in other columns are insignificant, suggesting the impacts of highway expansion on firm-level export variety are not robust, the same as Table 5.5-11.

For the control variables, the coefficients of $Size$, GDP, and $Foreign - Dist$ are positive, a 10% increase of firm size is related to 0.3% increase of firm-level export variety of destinations, a 10% increase of GDP of destination country is related to 3.4% increase of firm-level export variety of destinations, a 10% increase of firms' average distance to destination countries is related to 0.7% increase of firm-level export variety of destinations. By contrast, a 10% decrease of firms' average $Remoteness$is related to the 3% increase of firm-level export variety of destinations. These results are also consistent with the baseline results.

Table 5.5-12: The Impacts of Highways on Firm-level Export Variety (Destination Variety).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	0.031***	0.021***	0.023**	0.004	0.007	0.007	0.008
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
Average GDP	0.978***	0.908***	0.853***	0.357***	0.338***	0.339***	0.336***
	(0.058)	(0.037)	(0.034)	(0.023)	(0.025)	(0.025)	(0.025)
Average Foreign-Dist	0.054***	0.052***	0.047***	0.070***	0.072***	0.072***	0.072***
	(0.011)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Average Remoteness	-0.971***	-0.898***	-0.839***	-0.323***	-0.305***	-0.305***	-0.303***
	(0.057)	(0.037)	(0.034)	(0.024)	(0.025)	(0.025)	(0.025)
Size	0.040***	0.039***	0.039***	0.027***	0.028***	0.028***	0.027***
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.006)
Homo	Yes	Yes	Yes	Yes	Yes	Yes	Yes
TFP-ACF	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rail-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Road-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Collective-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Foreign-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	187185	184559	179183	170945	160049	160020	158273
Group	5235	13277	23471	48663	50957	50950	50998
Under-identification test	199.874	492.072	461.678	2774.507	2354.154	2351.269	2303.083
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.2e+05	1.0e+05	7.0e+04	6.1e+04	5.1e+04	5.1e+04	4.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.005	0.075	0.216	2.310	1.691	1.641	0.561
P value	(0.943)	(0.785)	(0.642)	(0.129)	(0.194)	(0.200)	(0.454)
First-stage Results:							
Ming-Road	0.064*	0.121***					
	(0.032)	(0.025)					
Qing-Road	0.660***	0.609***	-0.010	0.094*	0.132**	0.129**	0.128*
	(0.026)	(0.022)	(0.049)	(0.048)	(0.049)	(0.049)	(0.051)
Terrain-Surface			0.729***	0.687***	0.640***	0.644***	0.643***
			(0.046)	(0.046)	(0.047)	(0.047)	(0.049)
F	1264.256	1272.103	1451.210	2188.439	1671.558	1672.065	1657.624
R2	0.635	0.604	0.576	0.619	0.584	0.584	0.586
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I5. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 5.5-13 investigates the impacts of geographical transportation cost on firms' export variety in each destination, a new type of fixed effects (destination fixed effects) is added in all columns. The coefficients of DGC in columns (1) and (2) show that a 10% decrease of domestic transportation cost can induce about 0.1% increase of firms' export variety in each destination, suggesting that those exporters near to seaports tend to export more different products in each destination.

Table 5.5-13: The Impacts of Highways on Firm-level Export Variety in each Destination.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.021***	-0.012***	-0.005	0.006**	0.005*	0.005*	0.004
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
TFP-ACF	0.005***	0.005***	0.004***	0.005***	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Size	0.068***	0.078***	0.088***	0.084***	0.079***	0.079***	0.077***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Collective-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Remoteness	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Foreign-Dist	Yes	Yes	Yes	Yes	Yes	Yes	Yes
GDP	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rail-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Road-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Foreign-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Homo	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	1323028	1252179	1166083	1048896	963952	963841	952533
Group	115334	188739	247743	347052	345056	345028	343433
Under-identification test	5041.735	8445.844	6850.243	1.8e+04	1.2e+04	1.2e+04	1.3e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.8e+05	5.9e+05	3.9e+05	3.4e+05	2.5e+05	2.5e+05	2.7e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	8.115	3.276	0.794	0.317	3.098	3.101	0.399
P value	(0.004)	(0.070)	(0.373)	(0.573)	(0.078)	(0.078)	(0.528)
First-stage Results:							
Ming-Road	0.367***	0.182***	0.070***	0.182***	0.489***	0.489***	0.214***
	(0.003)	(0.007)	(0.010)	(0.008)	(0.005)	(0.005)	(0.008)
Counter-factual-Road	0.422***				0.464***	0.464***	
	(0.003)				(0.006)	(0.006)	
Qing-Road		0.555***	0.674***	0.670***			0.623***
		(0.006)	(0.008)	(0.006)			(0.007)
F	14199.671	12605.408	12276.596	12695.306	7798.121	7789.035	9500.496
R2	0.608	0.593	0.577	0.624	0.565	0.565	0.585
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Destination	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I8. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

By contrast, when the firm-level fixed effects are controlled, a 10% decrease of domestic transportation cost can induce 0.05% decrease of export variety, suggesting that when the firm-level location effects are absorbed by region-firm-industry fixed effects, highway expansion has slightly larger impacts on the export variety of those firms far from seaports, consistent with Table 5.5-8. For the control variables, a 10% increase of productivity is related to 0.06% increase of firms' export variety of destinations, a 10% increase of firm size is related to 0.8% increase of firms' export variety of destinations, consistent with the predictions of heterogeneous firms trade models. As robustness checks, Table I9 investigates the impacts of highway expansion on firm-level export value in each destination, but the coefficients of DGC are insignificant.

Table 5.5-14: The Impacts of Highways on Firms' Export Volume of Products.

Dependent Variable: Export Volume	Without Firm Fixed Effect (1)	(2)	(3)	With Firm Fixed Effect (4)	(5)	(6)	(7)
Main Results:							
DGC	0.131***	0.060***	0.053***	0.022**	0.026***	0.026**	0.023**
	(0.008)	(0.008)	(0.010)	(0.007)	(0.008)	(0.008)	(0.008)
Average GDP	0.830***	0.662***	0.645***	0.423***	0.408***	0.408***	0.402***
	(0.027)	(0.026)	(0.026)	(0.027)	(0.028)	(0.028)	(0.028)
Average Foreign-Dist	0.200***	0.195***	0.194***	0.198***	0.194***	0.194***	0.191***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)
Average Remoteness	-0.715***	-0.541***	-0.520***	-0.294***	-0.284***	-0.284***	-0.277***
	(0.027)	(0.026)	(0.026)	(0.027)	(0.028)	(0.028)	(0.028)
Foreign-share	0.192***	0.161***	0.175***	0.040**	0.047***	0.046***	0.045**
	(0.012)	(0.012)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
Size	0.196***	0.215***	0.244***	0.327***	0.313***	0.313***	0.311***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
TFP-ACF	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rail-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Road-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Collective-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Homo	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observation	991589	932422	866620	778814	721210	721005	712183
Group	123122	162112	193596	261495	258749	258694	257445
Under-identification test	2494.680	4110.599	4455.862	1.5e+04	1.2e+04	1.2e+04	1.2e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	4.6e+05	4.5e+05	3.4e+05	2.6e+05	2.3e+05	2.3e+05	2.2e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.051	1.762	2.518	1.794	0.358	0.352	0.744
P value	(0.821)	(0.184)	(0.113)	(0.180)	(0.550)	(0.553)	(0.389)
First-stage Results:							
Ming-Road	0.328***	0.345***		0.051***	0.072***	0.070***	0.058***
	(0.004)	(0.004)		(0.010)	(0.011)	(0.011)	(0.011)
Counter-factual-Road	0.439***	0.456***	0.226***				
	(0.003)	(0.004)	(0.007)				
Qing-Road			0.558***	0.746***	0.724***	0.727***	0.735***
			(0.006)	(0.008)	(0.008)	(0.008)	(0.008)
F	9431.627	6377.174	8497.859	9452.057	7588.718	7608.237	7573.425
R2	0.597	0.594	0.598	0.613	0.591	0.592	0.595
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
HS Code	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I10. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 5.5-14 investigating the impacts of geographical transportation cost on the export volume of products, a new type of fixed effects (6-digit HS product fixed effects) is added in all columns. When the firm-level fixed effects are not included, a 10% decrease of domestic transportation cost can induce 0.5% to 1.3% decrease of export volume of products. Similarly, when the firm-level fixed effects are included, a 10% decrease of domestic transportation cost can induce 0.2% decrease of the export volume of products, suggesting that those exporters near to seaports tend to export less for each product.

For the control variables, the coefficients of $Size$, GDP, and $Foreign - Dist$ are positive, a 10% increase of firm size is related to 3.1% increase of export volume of products, a 10% increase of average GDP of destination country is related to 4% increase of export volume of products, a 10% increase of firms' average distance to destination countries is related to 1.9% increase of export volume of products, while a 10% decrease of firms' average $Remoteness$is related to 2.8% increase of export volume of products. At the same time, the coefficients of $Foreign - Capital$ show that a 10% increase of foreign capital share is related to 0.5% increase of export volume of products. These results are consistent with the fact that foreign firms are more likely to be involved in international trade. As robustness checks, Table I11 uses the export value to substitute volumes, but the coefficients of DGC are insignificant.

Table 5.5-15 investigates the impacts of geographical transportation cost on export volume for each 6-digit HS product in each destination, the firm-region-industry-product-destination fixed effects are controlled in these regressions. The coefficients of DGC are consistent across all columns, i.e., a 10% decrease of DGC can induce 0.2% decrease of the export volume of products in each destination. These results are consistent with the previous regressions of export variety, i.e., firms closer to seaports (or coastal firms) tend to have higher firm-level export variety, and firms' export variety in each destination, so the export volume of products in each destination will be smaller for firms closer to seaports. For the control variables, a 10% increase of firms' average distance to destination countries is related to 0.3% to 0.7% increase of export volume of products in each destination, a 10% decrease of $Remoteness$ is related to 3% to 5% increase of export volume of products in each destination, a 10% increase of productivity is related to 0.3% to 0.5% increase of export volume of products in each destination. As robustness checks, Table I7 uses export value rather than volumes to indicate firms' export volume, but the coefficients of DGC are insignificant.

Table 5.5-15: The Impacts on Export Volume of Products by Destination.

Dependent Variable:	Without Region and Industry Fixed Effect				With Region and Industry Fixed Effect			
Export Volume	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
DGC	0.013	0.013*	0.020**	0.020***	0.017	0.020*	0.021**	0.019***
	(0.011)	(0.007)	(0.007)	(0.005)	(0.013)	(0.008)	(0.007)	(0.005)
Foreign-Dist	0.042***	0.071***	0.020	0.031*	0.041***	0.029	0.069***	0.050***
	(0.005)	(0.003)	(0.019)	(0.012)	(0.004)	(0.023)	(0.003)	(0.014)
TFP-ACF	0.034***	0.031***	0.044***	0.054***	0.034***	0.039***	0.035***	0.054***
	(0.006)	(0.004)	(0.004)	(0.003)	(0.006)	(0.004)	(0.004)	(0.003)
Remoteness	0.017	0.315***	0.500***	0.334***	0.014	0.499***	0.310***	0.315***
	(0.016)	(0.011)	(0.070)	(0.047)	(0.015)	(0.079)	(0.010)	(0.052)
Size	0.102***	0.156***	0.138***	0.271***	0.107***	0.140***	0.153***	0.262***
	(0.009)	(0.006)	(0.006)	(0.004)	(0.010)	(0.007)	(0.006)	(0.005)
GDP	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rail-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Road-Density	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Pop	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Collective-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Foreign-share	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Homo	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
Observation	3185913	2898105	2936770	1992389	3149756	2853872	2829325	1799670
Group	62987	325901	439865	706578	74561	468816	344727	682088
Under-identification test	1025.925	3878.997	6579.468	3.9e+04	788.573	4482.375	2933.159	2.8e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.5e+06	1.2e+06	1.2e+06	6.3e+05	1.4e+06	1.1e+06	1.1e+06	5.2e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.036	0.696	0.077	0.045	0.284	0.114	0.033	2.890
P value	(0.850)	(0.404)	(0.782)	(0.833)	(0.594)	(0.736)	(0.856)	(0.089)
First-stage Results:								
Qing-Road	0.680***	0.665***	0.687***	0.681***	0.670***	0.686***	0.662***	0.683***
	(0.0462)	(0.0155)	(0.0110)	(0.00442)	(0.0293)	(0.0112)	(0.0149)	(0.00485)
Ming-Road	0.157*	0.172***	0.144***	0.143***	0.146***	0.122***	0.149***	0.117***
	(0.0643)	(0.0207)	(0.0148)	(0.00586)	(0.0388)	(0.0147)	(0.0197)	(0.00644)
F	633.027	2786.014	5144.459	26198.236	480.399	3878.587	2223.090	19688.038
R2	0.627	0.623	0.627	0.626	0.592	0.594	0.589	0.596
Fixed Effects:								
Region					county	county	county	county
Industry					4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6-digit HS		Yes		Yes		Yes		Yes
Destination			Yes	Yes			Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I6. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

The previous studies using region-level data mostly conclude that aggregate trade volumes tend to increase after infrastructure construction, e.g., Limao & Venables (2001), Huang & Xu (2012), Donaldson (2018), Li et al. (2018), and Tang et al. (2019)[22] Due to data constraints, there are a limited number of studies that have investigated firm-level spatial impacts

[22]Limao & Venables (2001) find that aggregate stocks of transportation and communication infrastructure can affect trade flow through the trade costs reduction effect; Huang & Xu (2012) find that aggregated transportation cost has insignificant impacts on export intensive margin; Donaldson (2018) find that railroad construction during the British Raj in India can promote the inter-regional trade; Li et al. (2018) find that railway connection can promote exports from China to Europe and Central Asia; Tang et al. (2019) find that firms in those cities connected by high-speed railways witness export growth by 12.7%, but this effect only remains significant when newly-constructed railway stations are located within 30km distance from cities.

generated by infrastructure development. Our firm-level studies provide more details at the firm and transaction levels.

Results show that firms closer to seaports (or coastal firms) are more likely to involve in international trade, which have higher firm-level export values and higher exports-sales ratios. When we consider the export intensive and extensive margins, coastal firms normally export more types of different products, but focus on a smaller number of destinations. Consequently, firms closer to seaports tend to have lower export intensive margins, e.g., export volume of products, or export volume of products in each destination. These results are consistent with firm-level studies such as Liu et al. (2017), who show that the increase of traffic accessibility can promote firms' exports and increase firms' import and export scope. At the same time, our studies also show that highway expansion also has heterogeneous impacts on coastal and inland firms. Our key explanatory variable DGC can not only reflect the impacts of highways' expansion, but also firms' location effects, while the location effects can be absorbed by the cross fixed effects between firm and region. When we control the firm-region-industry fixed effects, the results suggest that highway expansion can promote the export of firms far from seaports. These results are also consistent with our summary statistics that new exporters tend to emerge in inland regions. This results provide a new explanation on the basis of Tang et al. (2019), while they find that high-speed railway connections can also increase firms' export scope and that this impact is stronger for the eastern region, capital- and technology-intensive sectors.

5.5.2 Robustness Checks

Table 5.5-16 uses two province-level proxies to measure transportation costs; these measurements are province-level line distances to China's 14 main trade ports in 2002, provided by Huang & Xu (2012). That study finds the province-level transportation cost has significant impacts on firm-level export value and variety of products. Columns (1), (2), (5) and (6) control 4-digit industry, province, and year fixed effects, and their results show that a 10% increase of transportation cost is related to 0.5% to 1% increase of export value, and 0.5% increase of export variety respectively. These results are consistent with our baseline results in the previous section, and existing region-level studies such as Huang & Xu (2012), e.g., when we only control region and industry fixed effects, firms closer to trade ports have more significant trade activities. Columns (3), (4), (7) and (8) control firm and year fixed effects, the coefficients of province-level transportation cost are insignificant, which are the expected results because region-level factors will be absorbed by firm-level fixed effects, the same as the baseline results in the previous section.

Table 5.5-16: Robustness Checks on Alternative Measurement on Trade Cost.

Dependent Variable:	Firm-level Export Value				Firm-level Export Variety			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
Codist1	-0.048***		0.054		-0.045***		-0.084	
	(0.010)		(0.142)		(0.006)		(0.118)	
Codist2		-0.101***		0.113		-0.065***		-0.050
		(0.016)		(0.148)		(0.010)		(0.135)
R2	0.220	0.221	0.158	0.158	0.0789	0.0787	0.0619	0.0619
Observation	169571	169571	171294	171294	169571	169571	171294	171294
Group	4966	4966	60031	60031	4966	4966	60031	60031
F	219.0	239.8	558.7	558.6	101.8	101.7	195.8	195.8
Fixed Effects:								
Region	province	province			province	province		
Industry	4-digit	4-digit			4-digit	4-digit		
Firm			Yes	Yes			Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I18. Columns (1), (2), (5) and (6) control 4-digit industry, province, and year fixed effects; columns (3), (4), (7) and (8) control firm and year fixed effects. The cluster standard error is always applied in each column, according their fixed-effect groups. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

As alternative robustness checks, Table 5.5-17 uses the reduced gravity model to test the role of transportation cost channel, i.e., how DGC affects the implied trade cost, then the implied trade cost affects export activities. On the basis of previous trade literature such as Jacks et al. (2011) and Chen & Novy (2018), our implied trade cost function is decided by the following variables, i.e., the present study includes DGC, $Foreign-Dist$, the dummy variable of common language and border to capture cost channel.[23]

The dependent variable is firm-level export value. For columns (1) to (4), the coefficients of $Foreign-Dist$ are positive and consistent with the baseline results, suggesting that a 10% increase of firms' average distance to the destination is related to 1.8% increase of export value. The coefficients of common language and borders are negative, suggesting export flows toward destinations with common language or borders have lower export values. Columns (5) to (8) test the combined effects, i.e., how DGC affects the trade cost function, then changes the export activities. The coefficients of $DGC * Foreign-Dist$ are posi-

[23]There economies have common language with mainland China, i.e., Taiwan, Hong Kong, Macao, and Singapore. The common border economies include China's neighbouring countries plus Hong Kong and Macao

tive, while the coefficients of $DGC * Language$ and $DGC * Border$ are negative, implying that there is a positive relationship between DGC and the implied trade cost. At the first stage, the decrease of DGC is also positively related to the decrease of implied trade cost. At the second stage, the negative coefficients of $DGC * DGC$ suggest the lower implied transportation cost is related to the higher export value. Table J2 uses firm-level export variety of products as the dependent variables to substitute firm-level export value, the positive relationship between DGC and implied trade cost still exists.

Table 5.5-17: Channel Study: The Impacts of Implied Trade Cost on Export Value.

Dependent Variable: Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
DGC	-0.017*	-0.013	-0.013	-0.013	-0.427***	-0.451***	-0.451***	-0.451***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.074)	(0.078)	(0.078)	(0.078)
Foreign-Dist	0.178***	0.185***	0.185***	0.185***				
	(0.024)	(0.025)	(0.025)	(0.025)				
Language	-0.076	-0.052	-0.052	-0.052				
	(0.042)	(0.045)	(0.045)	(0.045)				
Border	-0.107**	-0.109*	-0.109*	-0.109*				
	(0.041)	(0.044)	(0.044)	(0.044)				
DGC*DGC					-0.008**	-0.007*	-0.007*	-0.007*
					(0.003)	(0.003)	(0.003)	(0.003)
DGC*Foreign-Dist					0.032***	0.033***	0.033***	0.033***
					(0.004)	(0.004)	(0.004)	(0.004)
DGC*Language					-0.010	-0.006	-0.006	-0.006
					(0.007)	(0.008)	(0.008)	(0.008)
DGC*Border					-0.015*	-0.016*	-0.016*	-0.016*
					(0.007)	(0.008)	(0.008)	(0.008)
R2	0.153	0.138	0.138	0.138	0.154	0.138	0.138	0.138
Observation	189868	188049	188049	188049	189868	188049	188049	188049
Group	67586	80774	80774	80774	67586	80774	80774	80774
F	541.1	446.5	446.5	446.5	517.2	426.9	426.9	426.9
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry		4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table J5. Columns (1) and (5) control firm and year fixed effects, columns (2) and (6) control 4-digit industry, province, and year fixed effects; columns (3) and (7) control 4-digit industry, prefecture, and year fixed effects; columns (4) and (8) control 4-digit industry, county, and year fixed effects. The cluster standard error is always applied in each column, according their fixed-effect groups. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

The previous summary statistics Table 5.4-6 tells us new exporters are normally in inland regions (not very close to seaports). At the same time, our baseline results also show that, when the region-firm-industry fixed effects are controlled, highway expansion has more significant impacts on the export value of those firms not very close to seaports, suggesting the entry of new exporters may affect firms' export value and variety. The dependent variable in Table 5.5-18 is firm-level export variety. It splits the full sample into two groups, one is the exporter started their export earlier than 2000, while another group contains firms who started their exports after 2000. The coefficients of DGC in columns (4) to (6) show that, for the exporters entered after 2000, a 10% decrease of DGC can induce 0.4% to 0.6% increase of the export variety. This impacts are stronger for firms entered before 2000, suggesting that new entry exporters closer to seaports have higher export variety. At the same time, new entry exporters far from seaports also have much lower export variety than other firms. Table 5.5-18 does not present the regressions with firm-level fixed effects, because we found that once we control firm-level fixed effects, the coefficients of DGC will become insignificant, suggesting that the results in Table 5.5-18 mainly reflect firm-level heterogeneity. Table I13 uses firm-level export value to substitute firm-level export variety, but the coefficients of DGC are same in the two groups, suggesting the highways' impacts on export value are consistent across pre-2000 and post-2000 entry exporters.

Table 5.5-18: The Impacts on Firm-level Product Variety.

Dependent Variable:	Without Entry			Post-2000 Entry		
Product Variety	(1)	(2)	(3)	(4)	(5)	(6)
Main Results:						
DGC	-0.031**	-0.007	0.015	-0.063***	-0.042***	-0.003
	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)	(0.013)
Observation	65381	64248	62404	121142	118692	114318
Group	3158	6287	9543	4600	10953	18128
Under-identification test	147.876	387.808	363.908	218.569	376.792	380.044
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	4.3e+04	3.4e+04	2.6e+04	7.9e+04	6.8e+04	3.4e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	1.424	0.483	0.060	0.356	1.064	0.072
P value	(0.233)	(0.487)	(0.806)	(0.551)	(0.302)	(0.788)
First-stage Results:						
Qing-Road	-0.034	-0.037	-0.264***	0.126***	0.369***	0.301***
	(0.038)	(0.055)	(0.075)	(0.018)	(0.015)	(0.053)
Terrain-Surface	0.746***	0.762***	0.996***	0.546***		0.381***
	(0.035)	(0.053)	(0.071)	(0.014)		(0.051)
Counter-factual-Road					0.379***	
					(0.014)	
F	1017.157	1323.309	1207.325	1067.945	350.036	377.677
R2	0.695	0.639	0.636	0.616	0.576	0.454
Fixed Effects:						
Region	province	prefecture	county	province	prefecture	county
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I12. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Similarly, Table 5.5-19 investigates the impacts of DGC on firms' export volume of products in each destination. It only includes firms started exporting earlier than 2000. The coefficients of DGC are consistent, a 10% decrease of DGC can induce 0.2% decrease of export volume, consistent with baseline regressions. By contrast, Table I15 that only includes firms started exporting after 2000, but the coefficients of DGC are insignificant. These results suggest that, for those firms started exporting before 2000, those exporters far from seaports have a higher export volume of products in each destination.

Table 5.5-19: The Impacts on Export Volume of Products in each Destination (without entry).

Dependent Variable:	Without Region and Industry Fixed Effect				With Region and Industry Fixed Effect			
Export Volume	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
DGC	0.009	0.017*	0.016	0.022***	0.012	0.016	0.024**	0.020**
	(0.014)	(0.008)	(0.008)	(0.006)	(0.016)	(0.009)	(0.009)	(0.006)
Observation	1318544	1223246	1256291	936419	1289589	1205642	1177681	828991
Group	16026	119026	144790	294057	21854	163181	130688	288502
Under-identification test	652.476	2651.824	4788.989	2.7e+04	538.924	3620.733	2145.528	2.1e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.5e+05	5.4e+05	5.6e+05	3.2e+05	6.1e+05	5.1e+05	5.0e+05	2.6e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.024	0.701	0.515	0.966	0.183	0.058	0.193	0.910
P value	(0.877)	(0.403)	(0.473)	(0.326)	(0.669)	(0.810)	(0.661)	(0.340)
First-stage Results:								
Qing-Road	0.569***	0.570***	0.576***	0.574***	0.571***	0.579***	0.571***	0.577***
	(0.036)	(0.018)	(0.014)	(0.007)	(0.040)	(0.016)	(0.020)	(0.007)
Ming-Road	0.197***	0.191***	0.181***	0.177***	0.199***	0.183***	0.191***	0.177***
	(0.046)	(0.024)	(0.019)	(0.009)	(0.051)	(0.021)	(0.027)	(0.010)
F	223.693	1024.415	1375.524	7739.823	221.203	1148.802	931.534	6422.417
R2	0.482	0.482	0.483	0.488	0.460	0.464	0.461	0.470
Fixed Effects:								
Region					county	county	county	county
Industry					4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6-digit HS		Yes		Yes		Yes		Yes
Destination			Yes	Yes			Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I14. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Similarly, Table 5.5-20 investigates the impacts of DGC on firms' export volume of products. It only includes firms started exporting earlier than 2000. The coefficients of DGC are consistent, a 10% decrease of DGC can induce 0.2% to 0.3% decrease of export volume, consistent with baseline regressions. By contrast, Table I17 that only includes firms started exporting after 2000, but the coefficients of DGC are largely insignificant. These results suggest that, for those firms started exporting before 2000, those exporters far from seaports have a higher export volume of products.

Table 5.5-20: The Impacts on Firms' Export Volume of Products (without entry).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Share	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	0.092***	0.032*	0.026*	0.019*	0.028**	0.027**	0.020*
	(0.012)	(0.013)	(0.013)	(0.009)	(0.010)	(0.010)	(0.010)
Observation	394923	375068	353336	354658	324162	323986	319303
Group	58392	73080	84277	102912	104364	104322	104136
Under-identification test	2106.163	3859.983	3934.435	9894.387	8589.122	8572.756	8371.008
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.5e+05	1.6e+05	1.4e+05	1.3e+05	1.1e+05	1.1e+05	1.1e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.145	0.066	3.824	0.000	0.000	0.000	0.167
P value	(0.703)	(0.797)	(0.051)	(0.985)	(0.997)	(0.999)	(0.683)
First-stage Results:							
Ming-Road	0.434***	0.383***		0.041**	0.050***	0.047***	0.032*
	(0.007)	(0.006)		(0.013)	(0.014)	(0.014)	(0.014)
Counter-factual-Road	0.366***	0.445***	0.222***				
	(0.006)	(0.006)	(0.009)				
Qing-Road			0.586***	0.768***	0.755***	0.759***	0.769***
			(0.007)	(0.010)	(0.011)	(0.011)	(0.011)
F	4303.630	5199.974	5975.387	7298.846	5599.841	5627.408	5620.669
R2	0.610	0.606	0.640	0.653	0.630	0.630	0.634
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
HS Code	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I16. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

This section provides more details on the basis of baseline regression. First, when we use region-level measurements of transportation cost (provided by Huang & Xu (2012)), the results show that firms closer to seaports (or coastal firms) tend to have higher export values and variety, consistent with our baseline results. Second, we use the implied transportation cost function to investigate the relationship between DGC and implied transportation, and find this correlation is positive, suggesting the importance of the trade cost channel. Third, the baseline results show that firms closer to seaports (or coastal firms) normally export more types of different products, we find the new entry exporters play an essential role in this process. Forth, the baseline results also show that when we control the firm-region-industry fixed effects, highway expansion can promote the export of firms far from seaports. This section further reveals that if an firm has already started exporting before 2000, highway expansion has more significant impacts on the export volume of these exporters. A possible explanation is that those exporters entered after 2000 need a period of learning process to expand their export volume, then those inland exporters existing before 2000 are more sensitive to highway expansion.

5.5.3 Channel Studies and Heterogeneity Studies

Infrastructure development can affect firms' processing or outsourcing decisions, through both trade costs and market selection channels. Processing trade accounts for about 50% of China' imports during the early years of China's WTO accession, while previous summary statistics in Table 5.3-2 show that the proposition of processing trade in all transactions tends to decrease from 2000 to 2006. On the one hand, processing trade is more likely to be affected by infrastructure development, because both input and output goods need to be transported through traffic lines from or toward foreign markets; on the other hand, infrastructure development can induce the entry of intermediate producers in the long run; then these processing firms will become gradually more dependent on domestic rather than foreign suppliers. Due to processing firms tend to be less productive than non-processing firms (Yu 2015, Dai et al. 2016), they tend to be replaced by ordinary exporters, because they tend to be more productive and capital-intensive, hire more skilled employees, and earn higher revenues than non-exporters (Bernard et al. 2012). This process can induce a transition from a processing-oriented trade mode to non-processing or ordinary-trade dominated trade mode, accompanied by the increase of economic complexity and income.

Table 5.5-21 investigates the impacts of DGC on firms' probability of whether a firm is processing an exporter. The coefficients of DGC are consistent and significant in columns (1) and (7), i.e., a 10% decrease of DGC can induce 0.06% to 0.1% decrease of firms' processing probability. Column (7) controls firm-industry-region fixed effects and year fixed effects, so it provides strong evidence that new entry processing exporters tend to choose addresses not very close to seaports. At the same time, Table 5.4-6 also shows that new exporters normally choose the addresses not very close to seaports, consistent with Table

5.5-21. A possible explanation is that those potential processing exporters close to seaports have already started processing exporting because of its lower entry barriers than ordinary exports.

Table 5.5-21: The Impacts on Processing Export Decision.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect				
Export Decision	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Main Results:								
DGC	0.011*	0.002	0.005	0.004	0.005	0.005	0.006*	
	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	
Observation	194318	191681	186164	177720	166412	166383	164566	
Group	5320	13630	24242	50521	52901	52894	52944	
Under-identification test	211.158	518.302	593.974	2837.417	2374.376	2371.487	2326.982	
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Weak identification test	1.3e+05	1.1e+05	6.4e+04	6.2e+04	5.2e+04	5.2e+04	5.1e+04	
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Over-identification test	1.545	0.852	3.012	0.598	0.061	0.082	0.254	
P value	(0.214)	(0.356)	(0.083)	(0.439)	(0.804)	(0.774)	(0.614)	
First-stage Results:								
Qing-Road	0.508***			0.725***	0.680***	0.681***	0.689***	
	(0.020)			(0.016)	(0.017)	(0.017)	(0.017)	
Counter-factual-Road	0.247***	0.479***	0.455***					
	(0.023)	(0.012)	(0.014)					
Ming-Road		0.323***	0.350***	0.099***	0.147***	0.146***	0.134***	
		(0.013)	(0.014)	(0.020)	(0.022)	(0.022)	(0.022)	
F	1585.793	861.887	947.043	2560.868	1936.808	1936.281	1921.338	
R2	0.646	0.599	0.548	0.612	0.579	0.579	0.581	
Fixed Effects:								
Region	province	prefecture	county		province	prefecture	county	
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit	
Firm				Yes	Yes	Yes	Yes	
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table J7. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 5.5-22 investigates the impacts of DGC on firms' processing share in total export value. When the firm-level fixed effects are not included, a 10% decrease of DGC can reduce 0.4% to 1% firms' processing share, consistent with Table 5.5-21 and Table 5.4-6. The new entry processing exporters are normally in regions not very close to seaports (or inland regions), so the processing share is higher in inland regions. When the firm-industry-region fixed effects are included, a 10% decrease of DGC can increase 0.3% firms' processing share in total export value. These results suggest that, when we strictly control firm-industry-region time-invariant heterogeneity, the expansion of highways can induce the processing exporters to transform into ordinary or other exporters. This explanation is also consistent with the study of Egger & Falkinger (2003), they show that more efficient road networks can lower the trade barriers and promote trade structure transition from processing-oriented modes to a more diversified structure.

Table 5.5-22: The Impacts on Firm-level Process Share in Export Value.

Dependent Variable: Processing Share	Without Firm Fixed Effect (1)	(2)	(3)	With Firm Fixed Effect (4)	(5)	(6)	(7)
Main Results:							
DGC	0.105***	0.039*	0.010	-0.009	-0.033**	-0.033**	-0.030*
	(0.020)	(0.018)	(0.021)	(0.011)	(0.011)	(0.011)	(0.012)
Observation	92963	91530	88935	85687	79614	79595	78818
Group	2995	6836	11329	23401	24862	24856	24914
Under-identification test	115.604	344.171	385.678	1817.523	1483.068	1480.780	1478.621
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	4.2e+04	4.3e+04	2.9e+04	2.9e+04	2.5e+04	2.5e+04	2.4e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	1.298	4.028	0.576	0.512	0.141	0.133	0.233
P value	(0.255)	(0.045)	(0.448)	(0.474)	(0.708)	(0.716)	(0.629)
First-stage Results:							
Ming-Road	0.350***	0.325***	0.351***	0.472***	0.482***	0.483***	0.478***
	(0.010)	(0.013)	(0.015)	(0.013)	(0.014)	(0.014)	(0.014)
Counter-factual-Road	0.443***	0.476***	0.453***	0.469***	0.464***	0.464***	0.471***
	(0.010)	(0.012)	(0.015)	(0.013)	(0.014)	(0.014)	(0.014)
F	1290.597	746.500	766.183	1567.594	1137.600	1135.452	1113.531
R2	0.615	0.598	0.549	0.589	0.560	0.559	0.561
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table J6. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

The previous regressions focus on highways' export-side effects; at the same time, highway construction can also affect input cost, while the lower costs of sourcing inputs from abroad may also promote firms' export volume or export variety. This channel is also supported by previous studies such as Egger & Falkinger (2003), who show that infrastructure development can lower trade barriers, increase the quantity of intermediate producers. Table 5.5-23 investigates the impacts of DGC on firm-level import value. When the firm-level fixed effects are not controlled, a 10% decrease of DGC can increase 0.5% to 1.3% firm-level import value. The coefficient of column (4) suggests that when we only control firm-level and year fixed effects, highway expansion tends to increase the import value of those firms not very close to seaports (or inland firms). However, when the firm-industry-region fixed effects are controlled, the coefficients of $Dist$ become insignificant, suggesting the impacts of highway expansion tend to be absorbed by firm-level firm-industry-region time-invariant heterogeneity.

Table 5.5-23: The Impacts of Highways on Firm-level Import Value.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Variety by Destination	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.132***	-0.049**	0.007	0.028*	0.017	0.016	0.023
	(0.017)	(0.018)	(0.019)	(0.012)	(0.013)	(0.013)	(0.013)
Observation	131176	128660	123723	116459	107712	107694	106244
Group	4688	10299	17583	33229	34716	34711	34722
Under-identification test	232.039	477.367	396.463	2355.830	1939.321	1936.699	1925.363
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	8.6e+04	7.2e+04	5.0e+04	4.6e+04	3.8e+04	3.8e+04	3.7e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	11.839	0.001	2.491	1.174	1.209	0.949	1.268
P value	(0.001)	(0.981)	(0.115)	(0.279)	(0.272)	(0.330)	(0.260)
First-stage Results:							
Qing-Road	-0.006		-0.277***	-0.059	-0.057	-0.065	-0.062
	(0.023)		(0.059)	(0.052)	(0.059)	(0.059)	(0.061)
Terrain-Surface	0.690***	0.677***	0.973***	0.823***	0.812***	0.820***	0.814***
	(0.021)	(0.020)	(0.056)	(0.049)	(0.056)	(0.056)	(0.057)
Ming-Road		0.019					
		(0.024)					
F	1866.897	1697.186	1604.835	2014.155	1476.207	1477.897	1471.587
R2	0.653	0.620	0.604	0.655	0.620	0.620	0.624
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table J1. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 5.5-24 investigates the impacts of DGC on firm-level import variety of products. When the firm-level fixed effects are not controlled, a 10% decrease of DGC can increase 0.4% to 1.1% firm-level import variety, suggesting those firms near to seaports tend to import more diversified products. The same as Table 5.5-23, when the firm-industry-region fixed effects are controlled, the coefficients of $Dist$ become insignificant. These results suggest that, when we strictly control firm-industry-region time-invariant heterogeneity, the impacts of highway expansion tend to be absorbed by firm-level heterogeneity.

Table 5.5-24: The Impacts of Highways on Firm-level Import Variety.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Variety by Destination	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.110***	-0.040***	-0.010	0.001	0.007	0.007	0.009
	(0.011)	(0.010)	(0.010)	(0.006)	(0.007)	(0.007)	(0.007)
Observation	131176	128660	123723	116459	107712	107694	106244
Group	4688	10299	17583	33229	34716	34711	34722
Under-identification test	232.039	479.977	396.463	2355.830	1950.705	1947.931	1957.897
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	8.6e+04	7.2e+04	5.0e+04	4.6e+04	3.8e+04	3.8e+04	3.7e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	4.692	0.090	1.170	4.052	2.398	2.461	2.708
P value	(0.030)	(0.764)	(0.280)	(0.044)	(0.122)	(0.117)	(0.100)
First-stage Results:							
Terrain-Surface	0.690***	0.687***	0.973***	0.823***			
	(0.021)	(0.035)	(0.056)	(0.049)			
Qing-Road	-0.006	0.006	-0.277***	-0.059	0.616***	0.617***	0.619***
	(0.023)	(0.036)	(0.059)	(0.052)	(0.012)	(0.012)	(0.012)
Counter-factual-Road					0.243***	0.242***	0.238***
					(0.017)	(0.017)	(0.018)
F	1866.845	1727.036	1604.801	2014.105	1323.162	1322.620	1312.674
R2	0.653	0.620	0.604	0.655	0.620	0.620	0.624
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I4. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 5.5-25 uses firms' import value as control variables, the dependent variables in columns (1) and (2) are firms' export values in columns (3) and (4) are firms' export scope, in columns (5) and (6) are firms' export share in sales. The fixed effects in columns (1), (3), (5) are firm fixed effects and year fixed effects, while fixed effects in columns (2), (4), (6) are firm-county-industry fixed effects and year fixed effects. The regressions are OLS so the coefficients can only reflect correlation rather causality impacts. The coefficients of firm-level import value suggest that a 10% decrease of import value is related to 2.2% to 2.8% increase of firm-level export value, 0.6% to 1.1% increase of firm-level export variety of products, 1.9% to 2.4% increase of firm-level export share in sales.

Table 5.5-25: Import Decision as Control Variable.

Dependent Variable:	Export Value		Product Scope		Export Share in Sales	
	(1)	(2)	(3)	(4)	(5)	(6)
Main Results:						
Import	0.288***	0.225***	0.112***	0.060***	0.236***	0.189***
	(0.019)	(0.015)	(0.010)	(0.007)	(0.018)	(0.014)
Observation	188049	188049	188049	188049	188049	188049
Group	32337	80774	32337	80774	32337	80774
Fixed Effects:						
Region	county	county	county	county	county	county
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Firm		Yes		Yes		Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table J3. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Previous trade literature supports that infrastructure development tends to benefit some certain sectors such as high-tech and textile industries. According to Li et al. (2018), railway connection can promote exports from China to Europe and Central Asia, while China's export expansion mainly concentrates in equipment and manufactured industries, which are also normally defined or included in the high-tech sector by previous trade literature. Similarly, Amiti & Freund (2008) shows that China's export structure has transformed from a low-tech to a high-tech oriented mode from 2000 to 2006; Khandelwal et al. (2013) find Chinese textile and apparel industries also expanded very fast after China's WTO accession. Table 5.5-28 high-tech industries expanded much more significantly than other sectors.

Following Feng et al. (2017), Table 5.5-26 uses dummy variables to indicate high-tech industries (HS84-85, 90-92), textile and apparel (HS50-63) industries. The results across columns (1) to (4) show that DGC has insignificant impacts on the export volume of high-tech industries. Results across columns (5) to (8) show that, when firm-level fixed effects are controlled, a 10% decrease of DGC can decrease 0.3% firm-level export volume in textile industries. These results suggest that high-tech industries are not sensitive to highway expansion, while the impacts of highways on textile industries are consistent with baseline regressions, i.e., firms are not very close to seaports tend to benefit from highway expansion.

Table 5.5-26: The Impacts on Export Volume of Textile and High-tech Sectors Respectively.

Dependent Variable: Export Volume	High-tech Sector (1)	(2)	(3)	(4)	Textile Sector (5)	(6)	(7)	(8)
Main Results:								
DGC	0.038	0.013	0.013	-0.003	0.014	0.027*	0.027*	0.027*
	(0.030)	(0.020)	(0.020)	(0.020)	(0.018)	(0.012)	(0.012)	(0.012)
Observation	227771	147180	147171	145129	416802	278420	278278	274851
Group	20211	53534	53530	53227	17099	97436	97406	97017
Under-identification test	492.632	2096.530	2096.333	2031.176	355.990	5343.266	5325.812	5187.120
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.0e+05	4.4e+04	4.4e+04	4.3e+04	2.1e+05	9.3e+04	9.4e+04	9.2e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.591	1.580	1.536	0.595	0.280	0.003	0.004	0.008
P value	(0.442)	(0.209)	(0.215)	(0.441)	(0.597)	(0.954)	(0.953)	(0.928)
First-stage Results:								
Ming-Road	-0.035	0.038	0.037	0.030	-0.021	-0.010	-0.014	-0.032
	(0.046)	(0.020)	(0.020)	(0.020)	(0.070)	(0.019)	(0.019)	(0.019)
Qing-Road	0.832***	0.779***	0.779***	0.787***	0.811***	0.782***	0.788***	0.800***
	(0.034)	(0.015)	(0.015)	(0.015)	(0.056)	(0.015)	(0.015)	(0.014)
F	305.283	1353.082	1353.121	1336.426	356.462	3602.463	3622.361	3625.949
R2	0.614	0.562	0.562	0.566	0.633	0.617	0.617	0.620
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry		4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table J8. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 5.5-27 investigates the impacts of DGC on firms' export decisions in high-tech and textile and apparel industries respectively. The dependent variables are dummy variables to indicate whether a new entry exporter is in high-tech or textile industries. The results across columns (1) to (4) show that, when firm-level fixed effects are included, a 10% decrease of DGC can increase 0.02% firms' export probability in high-tech industries, consistent with the baseline results. Results across columns (5) to (8) show that, when firm-level fixed effects are included, a 10% decrease of DGC can decrease 0.05% to 0.07% firms' export probability in textile industries, the absolute values are larger than the baseline results. These results confirm that high-tech industries are not sensitive to highway expansion, while the impacts of highways on textile industries are stronger than other sectors.

Table 5.5-27: The Heterogeneous Impacts on Export Decision across Textile and High-tech Sectors.

Dependent Variable:	High-tech Sector				Textile Sector			
Entry Decision	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
DGC	-0.002	-0.000	-0.001	-0.002*	-0.007***	-0.005*	0.002	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
Observation	1031690	1025018	1002571	928114	1031690	1025018	1002571	928114
Group	9867	42732	100925	261537	9867	42732	100925	261537
Under-identification test	387.226	1009.079	1076.551	6721.549	387.226	999.046	1064.764	6834.871
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	7.2e+05	6.0e+05	3.7e+05	2.3e+05	7.2e+05	5.8e+05	3.5e+05	2.2e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.846	2.705	1.048	0.003	3.677	0.320	0.686	2.637
P value	(0.358)	(0.100)	(0.306)	(0.958)	(0.055)	(0.572)	(0.407)	(0.104)
First-stage Results:								
Counter-factual-Road	0.478***	0.377***	0.306***	0.332***	0.478***	0.494***	0.452***	0.475***
	(0.007)	(0.009)	(0.011)	(0.007)	(0.007)	(0.008)	(0.010)	(0.007)
Ming-Road	0.347***				0.347***	0.353***	0.403***	0.476***
	(0.008)				(0.008)	(0.008)	(0.010)	(0.007)
Terrain-Surface		0.432***	0.516***	0.549***				
		(0.010)	(0.010)	(0.006)				
F	2432.217	1864.512	2320.952	8651.339	2432.217	1388.425	1688.516	7542.666
R2	0.650	0.597	0.537	0.526	0.650	0.591	0.523	0.512
Region	province	prefecture	county		province	prefecture	county	
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit	
Firm				Yes				Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	

Note: The full regression results are provided in Table J9. The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table 5.5-28: Export Structure across High-tech and Textile Industries.

	2000	2003	2006
Total Transactions	1,143,406	2,336,315	4,798,011
Total Value (Billion Yuan)	96.48	206.94	533.33
High-tech Industries	301,877	688,175	1,506,697
Transactions (Proportion)	(26.4%)	(27.2%)	(31.4%)
Value (Billion Yuan)	45.51	115.23	320.07
(Proportion)	(47.2%)	(55.7%)	(60.0%)
Textile Industries	311,440	553,821	1,035,339
Transactions (Proportion)	(27.2%)	(23.7%)	(21.6%)
Value (Billion Yuan)	17.47	28.33	53.51
(Proportion)	(18.1%)	(13.7%)	(10.0%)

Note: This table illustrates the export transaction variation, and export structure variation across high-tech and textile industries.

To interpret these results, Table 5.5-28 summarizes the trade structure transitions across high-tech and textile industries from 2000 to 2006. It can be found that both the export value and quantity of these two sectors increase very dramatically during this period, but the expansion of high-tech is very fast, which induces its export proportion to increase very quickly, and it dominates China's export products (more than 50% export value); on the other side, the export proportion of textile sector tends to decrease even though its absolute export volume tends to increase. These results are consistent with Amiti & Freund (2008), i.e., significant high-tech booming after trade liberalization and facilitation shocks, while the importance of the labour-intensive sector, such as the textile industry, gradually deceases over the same period. At the same time, these results provide more details to explain why high-tech sectors are not sensitive to highway expansion, i.e., the development of high-tech sectors are so fast, highways only have minicule and insignificant impacts on firms' entry and export value increase in these sectors.

5.6 Conclusions

Results show that, from 2000 to 2006, firms closer to seaports (or coastal firms) are more likely to involve in international trade, having higher firm-level export values and higher exports-sales ratios. A 10% decrease of travel cost can induce 0.04% to 0.2% increase of firms' export probability, up to 0.8% increase of firm-level export value and export-sale ratio. When we consider the export intensive and extensive margins, coastal firms normally export more types of different products, but focus on a smaller number of destinations. A 10% decrease of travel cost can induce 0.8% increase of firm-level export variety of products, and 0.8% decrease of firm-level export variety of destinations. Consequently, firms closer to seaports tend to have lower export intensive margins. A 10% decrease of travel cost can induce up to 0.2% to 1.3% decrease of firms' export volume of products, and 0.2% decrease of firms' export volume of products in each destination. At the first stage, larger and foreign firms are more likely to be involved in international trade. For the second stage, larger and productive exporters have higher firm-level export value, export export variety, and higher export volume of products in each destination, consistent with Liu et al. (2017). At the same time, larger and productive exporters have lower export exports-sales ratio, suggesting they have more diversified production structures than other firms, both foreign and domestic markets are important for these larger and productive firms. These results are consistent with firm-level studies such as Liu et al. (2017), who show that the increase of traffic accessibility can promote firms' exports and increase firms' import and export scope. This fact is also partly consistent with the prediction of Krugman-style love-of-variety models rather than with Armington models, i.e., if an economy doubles its export value with its economic growth, its export extensive margin is expected to increase faster than its intensive margin (Hummels & Klenow 2005). The decline of transportation costs can strengthen the connection across up- and down-stream industries, encourage incumbent or new entry firms to produce differentiated goods, then motivate the export market to evolve toward a more diversified structure, which is also referred to as economic complexity in aggregate studies.[24]

At the same time, our studies also show that highway expansion also has heterogeneous impacts on coastal and inland firms. First, our key explanatory variable DGC can not only reflect the impacts of highways' expansion, but also firms' location effects, while the location effects can be absorbed by the cross fixed effects between firm and region. When we control the firm-region-industry fixed effects, highway expansions tend to promote export value and export intensive margin of firms not very close to seaports. A possible explanation is that those potential exporters close to seaports have already started exporting. These results are also consistent with our summary statistics that new exporters tend to emerge in inland regions. Second, a 10% increase of travel cost to seaports can induce 0.4% to 0.6% decrease of firms' export variety for those firms starting exporting after 2000, this impacts are stronger than baseline regressions, suggesting that these new exporters tend to have smaller export scopes. These results provide a new explanation on the basis of Tang et al. (2019). Third, highway expansion has more significant impacts on the export volume of those firms that have already started exporting before 2000. A possible explanation is that those exporters entered after 2000 need a period of learning process to expand their export volume, then those inland exporters existing before 2000 are more sensitive to highway expansion.

Infrastructure development can affect firms' processing decisions through both trade costs

[24]Cross-country evidence shows that big countries tend to have more diversified economies and higher level of economic complexity; they not only have larger export and import value, but also have higher demand and production capacity of larger category of heterogeneous products (extensive margin), compared with smaller economies. When firms produce heterogeneous goods, each type of goods faces less competition than homogeneous goods. At the same time, the production of complex goods poses higher requirements on the quality and variety of intermediate inputs, andthese firms face higher technological frontiers than homogeneous producers. There are many trade models that specify the relationship across quality, diversification strategy, and trade cost; see (Baldwin & Harrigan 2011, Johnson 2012, Fan et al. 2015, Feng et al. 2016, Fan et al. 2018).

and market selection channels. Processing trade accounts for about 50% of China' imports during the early years of China's WTO accession, while the proposition of processing trade in all transactions tends to decrease from 2000 to 2006. On the one hand, processing trade is more likely to be affected by infrastructure development, because both input and output goods need to be transported through traffic lines from or toward foreign markets; on the other hand, infrastructure development can induce the entry of intermediate producers in the long run; then these processing firms will become gradually more dependent on domestic rather than foreign suppliers. Due to processing firms tend to be less productive than non-processing firms (Yu 2015, Dai et al. 2016), they tend to be replaced by ordinary exporters, because they tend to be more productive and capital-intensive, hire more skilled employees, and earn higher revenues than non-exporters (Bernard et al. 2012). This process can induce a transition from a processing-oriented trade mode to non-processing or ordinary-trade dominated trade mode, accompanied by the increase of economic complexity and income. Our results show that, different from baseline regressions, new entry processing exporters tend to choose addresses not very close to seaports, a 10% decrease of travel cost can induce 0.06% to 0.1% decrease of firms' processing probability. A possible explanation is that those potential processing exporters close to seaports have already started processing exporting because of its lower entry barriers than ordinary exports. A 10% decrease of travel cost can increase 0.3% firms' processing share in the total export values. These results suggest that, when we strictly control firm-industry-region time-invariant heterogeneity, the decrease of travel cost to seaports can reduce firms' processing share in export value, suggesting that the expansion of highways can induce the processing exporters to transform into ordinary or other exporters. This explanation is also consistent with the study of Egger & Falkinger (2003), they show that more efficient road networks can lower the trade barriers and promote trade structure transition from processing-oriented mode to more diversified structure.

Previous trade literature supports that infrastructure development tends to benefit some certain sectors such as high-tech and textile industries. According to Li et al. (2018), railway connection can promote exports from China to Europe and Central Asia, while China's export expansion mainly concentrates in equipment and manufactured industries, which are also normally defined or included in the high-tech sector by previous trade literature. Similarly, Amiti & Freund (2008) show that China's export structure has transformed from a low-tech to a high-tech oriented mode from 2000 to 2006; Khandelwal et al. (2013) find Chinese textile and apparel industries also expanded very fast after China's WTO accession.

Table 5.5-28 high-tech industries expanded much more significantly than other sectors. Our data show that the export value and quantity of high-tech and textile industries increase very dramatically from 2000 to 2006, but the expansion of high-tech is very fast, which induces its export proportion to increase very quickly, and it dominates China's export products (more than 50% export value); on the other side, the export proportion of textile sector tends to decrease even though its absolute export volume tends to increase. At the same time, the regressions results show that high-tech industries are not sensitive to highway expansion, highways' impacts on textile industries are stronger than other industries.

.1 Appendix H

Additional Statistical Results

Table H1: Summary Statistics of Export Volume

	statistics	All	2000	2001	2002	2003	2004	2005	2006
Quantity	Obs	4873126	291075	354945	441393	586590	897789	1059744	1241529
	Average	346403.7	382802.1	360243	322894.2	361073.3	343715.2	329758.5	351492.9
	Std. Dev.	(1.80e+07)	(1.10e+07)	(9259015)	(1.12e+07)	(1.26e+07)	(2.02e+07)	(2.00e+07)	(2.16e+07)
Value	Obs	4873126	291075	354945	441393	586590	897789	1059744	1241529
	Average	368143.5	318830.6	333397.6	292745.3	345465.4	359591.3	382063.8	421461.5
	Std. Dev.	(7455175)	(3132374)	(3980442)	(4541084)	(6275909)	(7140283)	(8080363)	(9632030)
Value (firm level)	Obs	264630	20195	23518	27170	33280	50267	52449	57751
	Average	6779333	4595429	5031798	4756461	6089204	6422587	7719686	9060565
	Std. Dev.	(7.85e+07)	(2.18e+07)	(3.06e+07)	(3.85e+07)	(6.01e+07)	(6.80e+07)	(9.29e+07)	(1.14e+08)

Note: This table provides summary statistics of firms' export quantity. For comparison, export value is also illustrated even through it does not appear in regressions. Variables $Quantity$ and $Value$ are firms' annual export quantity and value for each product (8-digit HS code) and destination. Variables $Value(firm)$ is firms' total export value for all products.

Table H2: Transaction Structure of Transportation Mode

	2000	2001	2002	2003	2004	2005	2006
Matched Obs	1,143,406	1,522,200	1,736,124	2,336,315	3,475,469	4,024,127	4,798,011
River&Sea	661,773	961,111	1,156,725	1,572,186	2,381,722	2,780,636	3,376,584
	(57.9%)	(63.1%)	(66.6%)	(67.3%)	(68.5%)	(69.1%)	(70.4%)
Air	154,053	205,237	239,173	346,733	562,050	676,095	802,026
	(13.5%)	(13.5%)	(13.4%)	(14.8%)	(16.2%)	(16.8%)	(16.7%)
Road	317,138	342,539	328,346	402,117	512,083	545,857	596,019
	(27.7%)	(22.5%)	(18.9%)	(17.2%)	(14.7%)	(13.6%)	(12.4%)
Railway	5,608	7,170	5,885	7,413	9,016	9,439	12,340
	(0.5%)	(0.5%)	(0.3%)	(0.3%)	(0.3%)	(0.2%)	(0.3%)
Mail	3,554	4,867	5,045	6,855	8,987	9,710	7,956
	(0.3%)	(0.3%)	(0.3%)	(0.3%)	(0.3%)	(0.2%)	(0.2%)
Others	1,280	1,237	949	1,011	1,611	1,293	3,086

Note: Export products are transported by six kinds of transport modes, i.e., motor transport, rail transport, air transport, river&sea transport, mail transport, and others, while water transport is the most widely used transport mode. For example, 70% transactions in 2006 are transported by river&sea mode, 12,7% by motor&rail mode, while 16.7% by air mode. However, some records are obviously wrong because they state some trade products are transported to U.S. and Canada through highways.

Table H3: Unit of Different Types of Goods (8-digit HS code)

Unit	Total	Number	Weight	Size	Unknown
Types of Goods	8153	2317	5373	457	-

Table H4: Summary Statistics of Other Control Variables

	Obs	Full Sample Average (Std. Dev.)	Obs	2000 Average (Std. Dev.)	Obs	2003 Average (Std. Dev.)	Obs	2006 Average (Std. Dev.)
GDP (National)	1270	2.80e+11	178	2.62e+11	181	2.73e+11	184	2.99e+11
		(1.15e+12)		(1.09e+12)		(1.14e+12)		(1.24e+12)
(merged)	3981643	2.52e+12	232139	2.42e+12	389349	2.37e+12	1094141	2.60e+12
		(4.09e+12)		(3.73e+12)		(3.89e+12)		(4.29e+12)
Homo	7243133	.5468509	445023	.5452123	766453	.6178409	1572778	.6256922
		(.4978002)		(.4979522)		(.4859155)		(.4839438)
Remoteness	5059362	5.13e+18	311009	5.14e+18	599960	5.05e+18	1261015	5.15e+18
		(8.19e+18)		(7.58e+18)		(7.90e+18)		(8.60e+18)
State	6354553	.0524994	387013	.1311388	574537	.061449	1501739	.0180052
		(.2061058)		(.3126862)		(.2203693)		(.1143561)
Collective	6354628	.0562484	387003	.1313725	574535	.0510545	1501738	.0226491
		(.2080249)		(.3007775)		(.195416)		(.1319823)
Foreign	6354517	.4487034	387022	.4462003	574496	.4607077	1501739	.4871491
		(.4552889)		(.4418918)		(.4523583)		(.4618148)
Size	5945396	10.55083	351880	10.39968	539793	10.62352	1430058	10.79085
		(1.694309)		(1.650287)		(1.643019)		(1.720748)
Age	6405172	9.607131	393006	11.56487	579050	9.261041	1508890	8.589213
		(36.7952)		(33.89717)		(8.957782)		(9.176639)
Homo	7243133	.5468509	445023	.5452123	766453	.6178409	1572778	.6256922
		(.4978002)		(.4979522)		(.4859155)		(.4839438)

Note: GDP and GDP per capita of destination countries are included to capture market scales and individual purchasing power. $Homo$ is the dummy variable of homogeneous goods, i.e., Rauch (1999) base on SITC 4-digit code, goods traded in organized exchanges, or with prices quoted in trade publications are defined as homogeneous goods. $Remoteness$ is defined as $Remoteness_d = \sum_o GDP_o distance_{od}$ (Manova & Zhang 2012). $Distexpressway$, $Roaddensity$, $State$, $Collective$, $Foreign$, $Size$, and Age capture firms distance to expressways, provincial highway density, ownership effects, size and age effects.

.2 Appendix I

Robustness Checks on Baseline Regressions

Table I1: The Impacts of Highways on Export Decision.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Decision	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.019***	-0.004*	-0.000	-0.003*	-0.002	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Rail-Density	-0.037***	-0.036***	-0.035***	-0.011***	-0.007***	-0.007***	-0.006**
	(0.007)	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Road-Density	-0.066***	-0.058***	-0.054***	-0.010***	-0.006***	-0.006***	-0.004**
	(0.004)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
TFP-ACF	-0.006***	-0.005***	-0.003***	0.006***	0.006***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Pop	-0.013***	-0.028***	-0.028***	-0.006**	-0.000	-0.000	-0.000
	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)	(0.002)
State-share	-0.011***	-0.008**	-0.006*	0.001	0.002	0.002	0.003
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Collective-share	-0.015***	-0.013***	-0.012***	0.001	0.001	0.000	0.001
	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
Foreign-share	0.363***	0.331***	0.314***	0.033***	0.027***	0.027***	0.026***
	(0.007)	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Age	0.000***	0.000***	0.000***	0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.057***	0.059***	0.061***	0.031***	0.027***	0.027***	0.025***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observation	1435106	1428863	1405756	1331353	1264484	1264242	1249384
Group	10186	46856	116998	327712	343734	343702	345454
Under-identification test	400.875	1076.362	1262.561	7373.757	6583.139	6565.432	6130.146
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.0e+06	7.6e+05	4.6e+05	2.9e+05	2.5e+05	2.5e+05	2.4e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.251	0.060	0.320	1.225	0.853	1.081	0.290
P value	(0.616)	(0.806)	(0.572)	(0.268)	(0.356)	(0.298)	(0.590)
First-stage Results:							
Terrain-Surface	0.494***						
	(0.012)						
Counter-factual-Road	0.260***	0.491***	0.446***	0.478***	0.485***	0.485***	0.486***
	(0.013)	(0.007)	(0.009)	(0.006)	(0.007)	(0.007)	(0.007)
Ming-Road		0.370***	0.432***	0.502***	0.495***	0.495***	0.495***
		(0.008)	(0.009)	(0.006)	(0.007)	(0.007)	(0.007)
F	2744.440	1509.867	1985.517	8124.610	6745.826	6735.313	6310.122
R2	0.660	0.581	0.513	0.479	0.446	0.446	0.438
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I2: The Impacts of Highways on Firm-level Export Value.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.077***	-0.021	0.040**	0.023*	0.035**	0.035**	0.035**
	(0.013)	(0.014)	(0.015)	(0.011)	(0.012)	(0.012)	(0.012)
TFP-ACF	0.087***	0.082***	0.110***	0.136***	0.139***	0.139***	0.138***
	(0.010)	(0.009)	(0.009)	(0.006)	(0.006)	(0.006)	(0.006)
GDP	1.105***	1.006***	0.979***	0.389***	0.368***	0.367***	0.366***
	(0.083)	(0.068)	(0.065)	(0.054)	(0.057)	(0.057)	(0.057)
Foreign-Dist	0.164***	0.167***	0.157***	0.166***	0.172***	0.173***	0.175***
	(0.027)	(0.026)	(0.025)	(0.024)	(0.025)	(0.025)	(0.025)
Remoteness	-0.895***	-0.801***	-0.783***	-0.295***	-0.276***	-0.276***	-0.276***
	(0.080)	(0.066)	(0.064)	(0.054)	(0.057)	(0.057)	(0.057)
Rail-Density	-0.284***	-0.227***	-0.218***	-0.230***	-0.218***	-0.218***	-0.218***
	(0.048)	(0.041)	(0.040)	(0.030)	(0.033)	(0.033)	(0.034)
Road-Density	0.100*	0.067	0.090**	0.212***	0.176***	0.176***	0.169***
	(0.050)	(0.037)	(0.033)	(0.026)	(0.026)	(0.026)	(0.026)
Pop	-0.020	-0.190***	-0.134***	-0.135***	-0.091***	-0.091***	-0.093***
	(0.020)	(0.027)	(0.027)	(0.014)	(0.014)	(0.014)	(0.014)
State-share	-0.501***	-0.440***	-0.313***	-0.022	-0.025	-0.025	-0.022
	(0.051)	(0.053)	(0.056)	(0.043)	(0.045)	(0.045)	(0.046)
Collective-share	-0.118**	-0.082*	-0.081*	0.006	-0.012	-0.013	-0.013
	(0.037)	(0.036)	(0.036)	(0.027)	(0.028)	(0.028)	(0.028)
Foreign-share	0.744***	0.753***	0.698***	0.013	0.023	0.023	0.025
	(0.029)	(0.022)	(0.022)	(0.019)	(0.020)	(0.020)	(0.020)
Age	-0.001*	-0.001*	-0.001*	-0.001*	-0.001*	-0.001*	-0.001*
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.612***	0.644***	0.679***	0.554***	0.530***	0.530***	0.529***
	(0.012)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)	(0.011)
Homo	0.001	0.001	0.016	-0.064***	-0.060***	-0.060***	-0.059***
	(0.023)	(0.017)	(0.017)	(0.010)	(0.011)	(0.011)	(0.011)
Observation	187185	184559	179183	170945	160049	160020	158273
Group	5235	13277	23471	48663	50957	50950	50998
Under-identification test	208.737	498.473	461.678	2774.507	2354.154	2351.269	2303.083
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.2e+05	1.1e+05	7.0e+04	6.1e+04	5.1e+04	5.1e+04	4.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	3.473	0.007	0.670	0.989	0.445	0.545	0.001
P value	(0.060)	(0.932)	(0.413)	(0.320)	(0.505)	(0.460)	(0.976)
First-stage Results:							
Counter-factual-Road	0.240***						
	(0.023)						
Qing-Road	0.512***	0.176***	-0.010	0.094*	0.132**	0.129**	0.128*
	(0.020)	(0.024)	(0.049)	(0.048)	(0.049)	(0.049)	(0.051)
Terrain-Surface		0.534***	0.729***	0.687***	0.640***	0.644***	0.643***
		(0.023)	(0.046)	(0.046)	(0.047)	(0.047)	(0.049)
F	1314.688	1457.317	1451.227	2188.465	1671.592	1672.099	1657.660
R2	0.646	0.608	0.576	0.619	0.584	0.584	0.586
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I3: The Impacts of Highways on Firm-level Export Share in Revenue.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Share	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.087***	-0.043**	0.032*	0.012	0.023*	0.022*	0.023*
	(0.013)	(0.014)	(0.015)	(0.010)	(0.011)	(0.011)	(0.011)
TFP-ACF	-0.293***	-0.280***	-0.243***	-0.113***	-0.108***	-0.108***	-0.108***
	(0.011)	(0.009)	(0.008)	(0.006)	(0.006)	(0.006)	(0.006)
GDP	0.839***	0.775***	0.779***	0.351***	0.336***	0.335***	0.337***
	(0.076)	(0.065)	(0.062)	(0.052)	(0.055)	(0.055)	(0.055)
Foreign-Dist	0.153***	0.150***	0.142***	0.158***	0.153***	0.154***	0.156***
	(0.027)	(0.025)	(0.025)	(0.023)	(0.024)	(0.024)	(0.024)
Remoteness	-0.632***	-0.574***	-0.590***	-0.259***	-0.247***	-0.246***	-0.249***
	(0.074)	(0.064)	(0.061)	(0.052)	(0.055)	(0.055)	(0.055)
Rail-Density	-0.154**	-0.109**	-0.113**	-0.153***	-0.143***	-0.143***	-0.147***
	(0.050)	(0.040)	(0.039)	(0.029)	(0.032)	(0.032)	(0.033)
Road-Density	-0.041	-0.070	-0.060	0.024	-0.004	-0.004	-0.012
	(0.047)	(0.036)	(0.032)	(0.025)	(0.026)	(0.026)	(0.026)
Pop	0.014	-0.171***	-0.108***	-0.106***	-0.073***	-0.073***	-0.075***
	(0.019)	(0.025)	(0.025)	(0.013)	(0.013)	(0.013)	(0.013)
State-share	-0.205***	-0.155**	-0.098	-0.019	-0.015	-0.015	-0.015
	(0.047)	(0.050)	(0.054)	(0.042)	(0.043)	(0.043)	(0.045)
Collective-share	-0.088*	-0.047	-0.035	0.010	-0.005	-0.005	-0.004
	(0.038)	(0.036)	(0.036)	(0.027)	(0.028)	(0.028)	(0.028)
Foreign-share	0.810***	0.798***	0.724***	-0.000	0.011	0.011	0.013
	(0.029)	(0.021)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)
Age	-0.001	-0.001	-0.001	-0.001*	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Size	-0.086***	-0.061***	-0.032***	0.035***	0.036**	0.036**	0.037***
	(0.009)	(0.007)	(0.007)	(0.010)	(0.011)	(0.011)	(0.011)
Homo	0.026	0.024	0.038*	-0.054***	-0.051***	-0.052***	-0.051***
	(0.022)	(0.016)	(0.016)	(0.010)	(0.010)	(0.010)	(0.010)
Observation	187185	184559	179183	170945	160049	160020	158273
Group	5235	13277	23471	48663	50957	50950	50998
Under-identification test	199.874	498.473	461.678	2774.507	2354.154	2351.269	2303.083
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.2e+05	1.1e+05	7.0e+04	6.1e+04	5.1e+04	5.1e+04	4.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	1.066	1.286	2.439	0.385	0.003	0.017	0.431
P value	(0.302)	(0.257)	(0.118)	(0.535)	(0.957)	(0.895)	(0.511)
First-stage Results:							
Ming-Road	0.064*						
	(0.032)						
Qing-Road	0.660***	0.176***	-0.010	0.094*	0.132**	0.129**	0.128*
	(0.026)	(0.024)	(0.049)	(0.048)	(0.049)	(0.049)	(0.051)
Terrain-Surface		0.534***	0.729***	0.687***	0.640***	0.644***	0.643***
		(0.023)	(0.046)	(0.046)	(0.047)	(0.047)	(0.049)
F	1264.284	1457.317	1451.227	2188.465	1671.592	1672.099	1657.660
R2	0.635	0.608	0.576	0.619	0.584	0.584	0.586
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table I4: The Impacts of Highways on Firm-level Export Variety (Product Variety).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.050***	-0.023**	0.010	-0.002	0.002	0.002	0.001
	(0.009)	(0.008)	(0.009)	(0.005)	(0.006)	(0.006)	(0.006)
TFP-ACF	0.012**	0.009*	0.010*	0.014***	0.017***	0.017***	0.017***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
GDP	0.408***	0.414***	0.381***	0.180***	0.173***	0.173***	0.176***
	(0.049)	(0.034)	(0.032)	(0.022)	(0.023)	(0.023)	(0.023)
Foreign-Dist	0.069***	0.071***	0.060***	0.069***	0.069***	0.070***	0.070***
	(0.016)	(0.012)	(0.011)	(0.009)	(0.009)	(0.009)	(0.009)
Remoteness	-0.320***	-0.331***	-0.302***	-0.141***	-0.135***	-0.134***	-0.138***
	(0.048)	(0.034)	(0.032)	(0.022)	(0.023)	(0.023)	(0.023)
Rail-Density	-0.060**	-0.046*	-0.041*	-0.063***	-0.063***	-0.062***	-0.063***
	(0.022)	(0.020)	(0.019)	(0.014)	(0.016)	(0.016)	(0.016)
Road-Density	0.048	0.027	0.030	0.059***	0.043**	0.042**	0.038**
	(0.025)	(0.019)	(0.017)	(0.013)	(0.013)	(0.013)	(0.013)
Pop	0.026*	-0.040**	-0.016	-0.017*	-0.007	-0.008	-0.008
	(0.012)	(0.013)	(0.012)	(0.007)	(0.006)	(0.006)	(0.006)
State-share	-0.016	-0.030	-0.008	0.028	0.025	0.025	0.025
	(0.026)	(0.027)	(0.028)	(0.020)	(0.021)	(0.021)	(0.021)
Collective-share	-0.035	-0.028	-0.026	0.009	0.010	0.009	0.011
	(0.019)	(0.018)	(0.019)	(0.013)	(0.014)	(0.014)	(0.014)
Foreign-share	0.209***	0.205***	0.182***	-0.000	0.000	0.001	0.001
	(0.014)	(0.011)	(0.011)	(0.009)	(0.010)	(0.010)	(0.010)
Age	0.000	0.000	0.000	-0.000*	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.146***	0.160***	0.174***	0.156***	0.151***	0.151***	0.150***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)
Homo	-0.055*	-0.058***	-0.059***	-0.083***	-0.078***	-0.078***	-0.078***
	(0.022)	(0.013)	(0.012)	(0.006)	(0.006)	(0.006)	(0.006)
Observation	187185	184559	179183	170945	160049	160020	158273
Group	5235	13277	23471	48663	50957	50950	50998
Under-identification test	242.984	498.473	461.678	2774.507	2354.154	2351.269	2303.083
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.3e+05	1.1e+05	7.0e+04	6.1e+04	5.1e+04	5.1e+04	4.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	3.035	1.915	0.480	0.093	0.501	0.349	0.363
P value	(0.082)	(0.166)	(0.488)	(0.760)	(0.479)	(0.555)	(0.547)
First-stage Results:							
Qing-Road	0.101***	0.176***	-0.010	0.094*	0.132**	0.129**	0.128*
	(0.018)	(0.024)	(0.049)	(0.048)	(0.049)	(0.049)	(0.051)
Terrain-Surface	0.592***	0.534***	0.729***	0.687***	0.640***	0.644***	0.643***
	(0.016)	(0.023)	(0.046)	(0.046)	(0.047)	(0.047)	(0.049)
F	1607.874	1457.317	1451.227	2188.465	1671.592	1672.099	1657.660
R2	0.653	0.608	0.576	0.619	0.584	0.584	0.586
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I5: The Impacts of Highways on Firm-level Export Variety (Destination Variety).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	0.031***	0.021***	0.023**	0.004	0.007	0.007	0.008
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)
TFP-ACF	-0.002	0.001	0.005	0.007*	0.008*	0.008*	0.008*
	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
GDP	0.978***	0.908***	0.853***	0.357***	0.338***	0.339***	0.336***
	(0.058)	(0.037)	(0.034)	(0.023)	(0.025)	(0.025)	(0.025)
Foreign-Dist	0.054***	0.052***	0.047***	0.070***	0.072***	0.072***	0.072***
	(0.011)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Remoteness	-0.971***	-0.898***	-0.839***	-0.323***	-0.305***	-0.305***	-0.303***
	(0.057)	(0.037)	(0.034)	(0.024)	(0.025)	(0.025)	(0.025)
Rail-Density	0.008	-0.001	-0.009	-0.019	-0.030	-0.031	-0.027
	(0.021)	(0.018)	(0.018)	(0.017)	(0.018)	(0.018)	(0.019)
Road-Density	-0.035*	-0.025	-0.018	0.005	0.000	0.001	0.003
	(0.018)	(0.016)	(0.016)	(0.016)	(0.017)	(0.017)	(0.017)
Pop	0.007	0.006	0.010	-0.019	-0.010	-0.009	-0.006
	(0.010)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
State-share	-0.058**	-0.036	-0.036	0.012	0.006	0.008	0.007
	(0.019)	(0.019)	(0.020)	(0.022)	(0.024)	(0.024)	(0.025)
Collective-share	-0.021	-0.009	-0.004	-0.006	-0.003	-0.003	-0.003
	(0.015)	(0.015)	(0.015)	(0.017)	(0.018)	(0.018)	(0.018)
Foreign-share	-0.011	-0.004	0.001	-0.001	0.005	0.004	0.007
	(0.010)	(0.009)	(0.009)	(0.012)	(0.013)	(0.013)	(0.013)
Age	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.040***	0.039***	0.039***	0.027***	0.028***	0.028***	0.027***
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)	(0.006)
Homo	0.117***	0.114***	0.121***	0.144***	0.143***	0.143***	0.143***
	(0.018)	(0.012)	(0.011)	(0.008)	(0.009)	(0.009)	(0.009)
Observation	187185	184559	179183	170945	160049	160020	158273
Group	5235	13277	23471	48663	50957	50950	50998
Under-identification test	199.874	492.072	461.678	2774.507	2354.154	2351.269	2303.083
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.2e+05	1.0e+05	7.0e+04	6.1e+04	5.1e+04	5.1e+04	4.9e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.005	0.075	0.216	2.310	1.691	1.641	0.561
P value	(0.943)	(0.785)	(0.642)	(0.129)	(0.194)	(0.200)	(0.454)
First-stage Results:							
Ming-Road	0.064*	0.121***					
	(0.032)	(0.025)					
Qing-Road	0.660***	0.609***	-0.010	0.094*	0.132**	0.129**	0.128*
	(0.026)	(0.022)	(0.049)	(0.048)	(0.049)	(0.049)	(0.051)
Terrain-Surface			0.729***	0.687***	0.640***	0.644***	0.643***
			(0.046)	(0.046)	(0.047)	(0.047)	(0.049)
F	1264.256	1272.103	1451.210	2188.439	1671.558	1672.065	1657.624
R2	0.635	0.604	0.576	0.619	0.584	0.584	0.586
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I6: The Impacts on Export Volume of Products in each Destination.

Dependent Variable:	Without Region and Industry Fixed Effect				With Region and Industry Fixed Effect			
Export Volume	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
DGC	0.013	0.013*	0.020**	0.020***	0.017	0.020*	0.021**	0.019***
	(0.011)	(0.007)	(0.007)	(0.005)	(0.013)	(0.008)	(0.007)	(0.005)
Foreign-Dist	0.042***	0.071***	0.020	0.031*	0.041***	0.029	0.069***	0.050***
	(0.005)	(0.003)	(0.019)	(0.012)	(0.004)	(0.023)	(0.003)	(0.014)
TFP-ACF	0.034***	0.031***	0.044***	0.054***	0.034***	0.039***	0.035***	0.054***
	(0.006)	(0.004)	(0.004)	(0.003)	(0.006)	(0.004)	(0.004)	(0.003)
Remoteness	0.017	0.315***	0.500***	0.334***	0.014	0.499***	0.310***	0.315***
	(0.016)	(0.011)	(0.070)	(0.047)	(0.015)	(0.079)	(0.010)	(0.052)
GDP	0.165***	-0.024*	0.000	0.000	0.168***	0.000	-0.018	0.000
	(0.016)	(0.011)	(.)	(.)	(0.015)	(.)	(0.010)	(.)
Rail-Density	-0.073*	-0.069***	-0.114***	-0.164***	-0.062	-0.113***	-0.055**	-0.154***
	(0.031)	(0.019)	(0.021)	(0.014)	(0.034)	(0.023)	(0.020)	(0.016)
Road-Density	0.048	0.073***	0.077***	0.107***	0.046	0.081***	0.051***	0.085***
	(0.026)	(0.015)	(0.017)	(0.012)	(0.027)	(0.017)	(0.015)	(0.012)
Pop	-0.036**	-0.019*	-0.037***	-0.034***	-0.030*	-0.030***	-0.019*	-0.028***
	(0.013)	(0.009)	(0.008)	(0.006)	(0.013)	(0.008)	(0.008)	(0.006)
State-share	-0.015	-0.009	-0.010	0.028	-0.017	0.003	-0.004	0.040*
	(0.043)	(0.024)	(0.024)	(0.018)	(0.048)	(0.026)	(0.026)	(0.019)
Collective-share	-0.006	0.024	-0.016	0.033**	-0.028	-0.035	0.005	0.016
	(0.027)	(0.018)	(0.017)	(0.012)	(0.027)	(0.018)	(0.017)	(0.013)
Foreign-share	0.008	0.023*	0.005	0.006	0.018	0.012	0.025*	0.007
	(0.022)	(0.011)	(0.012)	(0.008)	(0.023)	(0.013)	(0.012)	(0.009)
Age	-0.001**	-0.001***	-0.001***	-0.001***	-0.001**	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.102***	0.156***	0.138***	0.271***	0.107***	0.140***	0.153***	0.262***
	(0.009)	(0.006)	(0.006)	(0.004)	(0.010)	(0.007)	(0.006)	(0.005)
Homo	-0.007	0.000	0.055***	0.000	-0.009	0.054***	0.000	0.000
	(0.022)	(.)	(0.009)	(.)	(0.020)	(0.009)	(.)	(.)
Observation	3185913	2898105	2936770	1992389	3149756	2853872	2829325	1799670
Group	62987	325901	439865	706578	74561	468816	344727	682088
Under-identification test	1025.925	3878.997	6579.468	3.9e+04	788.573	4482.375	2933.159	2.8e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.5e+06	1.2e+06	1.2e+06	6.3e+05	1.4e+06	1.1e+06	1.1e+06	5.2e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.036	0.696	0.077	0.045	0.284	0.114	0.033	2.890
P value	(0.850)	(0.404)	(0.782)	(0.833)	(0.594)	(0.736)	(0.856)	(0.089)
First-stage Results:								
Qing-Road	0.680***	0.665***	0.687***	0.681***	0.670***	0.686***	0.662***	0.683***
	(0.0462)	(0.0155)	(0.0110)	(0.00442)	(0.0293)	(0.0112)	(0.0149)	(0.00485)
Ming-Road	0.157*	0.172***	0.144***	0.143***	0.146***	0.122***	0.149***	0.117***
	(0.0643)	(0.0207)	(0.0148)	(0.00586)	(0.0388)	(0.0147)	(0.0197)	(0.00644)
F	633.027	2786.014	5144.459	26198.236	480.399	3878.587	2223.090	19688.038
R2	0.627	0.623	0.627	0.626	0.592	0.594	0.589	0.596
Fixed Effects:								
Region					county	county	county	county
Industry					4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6-digit HS		Yes		Yes		Yes		Yes
Destination			Yes	Yes			Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I7: The Impacts of Highways on Export Value of Products in each Destination.

Dependent Variable:	Without Region and Industry Fixed Effect				With Region and Industry Fixed Effect			
Export Volume	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
DGC	0.003	-0.005	0.008	0.003	0.002	0.005	-0.001	0.000
	(0.010)	(0.007)	(0.007)	(0.005)	(0.011)	(0.008)	(0.008)	(0.005)
Foreign-Dist	0.041***	0.069***	0.010	0.017	0.040***	0.013	0.068***	0.033*
	(0.004)	(0.003)	(0.018)	(0.012)	(0.004)	(0.021)	(0.003)	(0.014)
TFP-ACF	0.065***	0.062***	0.074***	0.083***	0.062***	0.069***	0.063***	0.081***
	(0.006)	(0.004)	(0.004)	(0.003)	(0.007)	(0.004)	(0.004)	(0.003)
Remoteness	-0.080***	0.281***	0.356***	0.286***	-0.081***	0.396***	0.277***	0.283***
	(0.015)	(0.010)	(0.066)	(0.046)	(0.015)	(0.074)	(0.010)	(0.051)
GDP	0.253***	0.016	0.000	0.000	0.255***	0.000	0.021*	0.000
	(0.015)	(0.010)	(.)	(.)	(0.015)	(.)	(0.010)	(.)
Rail-Density	0.015	0.001	-0.026	-0.094***	0.013	-0.036	0.009	-0.088***
	(0.030)	(0.018)	(0.021)	(0.014)	(0.034)	(0.024)	(0.020)	(0.016)
Road-Density	-0.026	0.028	-0.006	0.067***	-0.021	0.004	0.008	0.046***
	(0.027)	(0.014)	(0.016)	(0.011)	(0.028)	(0.017)	(0.015)	(0.012)
Pop	-0.041**	-0.031***	-0.036***	-0.040***	-0.042**	-0.035***	-0.033***	-0.039***
	(0.013)	(0.009)	(0.008)	(0.006)	(0.013)	(0.008)	(0.009)	(0.006)
State-share	-0.049	0.014	-0.013	0.054**	-0.044	-0.001	0.012	0.054**
	(0.035)	(0.021)	(0.025)	(0.018)	(0.038)	(0.026)	(0.023)	(0.019)
Collective-share	-0.032	0.023	-0.043**	0.041***	-0.040	-0.048**	0.015	0.028*
	(0.026)	(0.016)	(0.016)	(0.012)	(0.027)	(0.017)	(0.016)	(0.012)
Foreign-share	-0.006	0.013	-0.013	-0.001	0.004	-0.007	0.016	-0.000
	(0.020)	(0.011)	(0.012)	(0.008)	(0.020)	(0.013)	(0.011)	(0.009)
Age	-0.001**	-0.001***	-0.001***	-0.001***	-0.001**	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.108***	0.162***	0.135***	0.273***	0.109***	0.135***	0.158***	0.265***
	(0.009)	(0.006)	(0.006)	(0.004)	(0.010)	(0.007)	(0.006)	(0.005)
Homo	0.363***	0.000	0.401***	0.000	0.369***	0.408***	0.000	0.000
	(0.022)	(.)	(0.009)	(.)	(0.020)	(0.008)	(.)	(.)
Observation	3194840	2905921	2945619	1997779	3158567	2862511	2836917	1804532
Group	63003	326839	440318	708697	74582	469310	345671	684094
Under-identification test	1023.368	3888.350	6567.974	3.9e+04	784.566	4474.120	2939.366	2.8e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.5e+06	1.2e+06	1.2e+06	6.3e+05	1.4e+06	1.1e+06	1.1e+06	5.2e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.005	1.624	1.178	1.002	0.184	1.997	0.049	0.129
P value	(0.945)	(0.202)	(0.278)	(0.317)	(0.668)	(0.158)	(0.825)	(0.720)
First-stage Results:								
Qing-Road	0.680***	0.665***	0.687***	0.681***	0.670***	0.686***	0.662***	0.683***
	(0.0462)	(0.0155)	(0.0110)	(0.00442)	(0.0293)	(0.0112)	(0.0149)	(0.00485)
Ming-Road	0.157*	0.172***	0.144***	0.143***	0.146***	0.122***	0.149***	0.117***
	(0.0643)	(0.0207)	(0.0148)	(0.00586)	(0.0388)	(0.0147)	(0.0197)	(0.00644)
F	633.027	2786.014	5144.459	26198.236	480.399	3878.587	2223.090	19688.038
R2	0.627	0.623	0.627	0.626	0.592	0.594	0.589	0.596
Fixed Effects:								
Region					county	county	county	county
Industry					4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6-digit HS		Yes		Yes		Yes		Yes
Destination			Yes	Yes			Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I8: The Impacts of Highways on Firm-level Export Variety in each Destination.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.021***	-0.012***	-0.005	0.006**	0.005*	0.005*	0.004
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Foreign-Dist	0.000	-0.016**	-0.015	-0.003	-0.002	-0.002	-0.001
	(0.004)	(0.006)	(0.009)	(0.005)	(0.006)	(0.006)	(0.006)
TFP-ACF	0.005***	0.005***	0.004***	0.005***	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Remoteness	0.017	-0.005	-0.030	-0.134***	-0.118***	-0.117***	-0.118***
	(0.023)	(0.022)	(0.022)	(0.019)	(0.021)	(0.021)	(0.021)
GDP	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Rail-Density	0.005	0.018**	0.004	0.002	0.003	0.003	0.005
	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
Road-Density	0.028***	0.021***	0.016***	0.018***	0.008	0.008	0.008
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Pop	0.002	-0.010**	-0.007*	-0.001	0.003	0.003	0.001
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
State-share	0.015*	0.013	0.028**	0.038***	0.040***	0.040***	0.038***
	(0.007)	(0.008)	(0.009)	(0.007)	(0.008)	(0.008)	(0.008)
Collective-share	-0.012**	-0.007	-0.015**	0.022***	0.018***	0.018***	0.012*
	(0.004)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Foreign-share	0.070***	0.075***	0.071***	-0.001	0.000	0.000	0.001
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)
Age	0.000*	0.000**	0.000**	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.068***	0.078***	0.088***	0.084***	0.079***	0.079***	0.077***
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Homo	0.008	0.006	0.002	-0.040***	-0.037***	-0.037***	-0.040***
	(0.005)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Observation	1323028	1252179	1166083	1048896	963952	963841	952533
Group	115334	188739	247743	347052	345056	345028	343433
Under-identification test	5041.735	8445.844	6850.243	1.8e+04	1.2e+04	1.2e+04	1.3e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.8e+05	5.9e+05	3.9e+05	3.4e+05	2.5e+05	2.5e+05	2.7e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	8.115	3.276	0.794	0.317	3.098	3.101	0.399
P value	(0.004)	(0.070)	(0.373)	(0.573)	(0.078)	(0.078)	(0.528)
First-stage Results:							
Ming-Road	0.367***	0.182***	0.070***	0.182***	0.489***	0.489***	0.214***
	(0.003)	(0.007)	(0.010)	(0.008)	(0.005)	(0.005)	(0.008)
Counter-factual-Road	0.422***				0.464***	0.464***	
	(0.003)				(0.006)	(0.006)	
Qing-Road		0.555***	0.674***	0.670***			0.623***
		(0.006)	(0.008)	(0.006)			(0.007)
F	14199.671	12605.408	12276.596	12695.306	7798.121	7789.035	9500.496
R2	0.608	0.593	0.577	0.624	0.565	0.565	0.585
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Destination	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I9: The Impacts of Highways on Firm-level Export Value in each Destination.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.020***	-0.003	-0.008	0.014*	0.011	0.011	0.006
	(0.005)	(0.006)	(0.008)	(0.006)	(0.007)	(0.007)	(0.007)
Foreign-Dist	0.008	-0.020	-0.032	-0.023	-0.015	-0.016	-0.011
	(0.009)	(0.017)	(0.023)	(0.015)	(0.018)	(0.018)	(0.018)
TFP-ACF	0.110***	0.098***	0.110***	0.095***	0.095***	0.095***	0.094***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Remoteness	0.616***	0.593***	0.496***	0.149*	0.227***	0.229***	0.223***
	(0.075)	(0.068)	(0.069)	(0.058)	(0.064)	(0.064)	(0.065)
GDP	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Rail-Density	-0.024	-0.015	-0.064**	-0.088***	-0.072***	-0.072***	-0.072***
	(0.022)	(0.022)	(0.021)	(0.018)	(0.020)	(0.020)	(0.020)
Road-Density	0.007	0.026	0.026	0.089***	0.074***	0.073***	0.075***
	(0.017)	(0.016)	(0.016)	(0.014)	(0.014)	(0.014)	(0.014)
Pop	0.002	-0.078***	-0.086***	-0.051***	-0.038***	-0.038***	-0.042***
	(0.009)	(0.011)	(0.011)	(0.008)	(0.008)	(0.008)	(0.008)
State-share	-0.187***	-0.170***	-0.097***	0.046*	0.050*	0.050*	0.048*
	(0.021)	(0.024)	(0.025)	(0.022)	(0.023)	(0.023)	(0.024)
Collective-share	-0.063***	-0.048**	-0.052**	0.022	0.011	0.011	-0.000
	(0.015)	(0.015)	(0.016)	(0.014)	(0.015)	(0.015)	(0.015)
Foreign-share	0.116***	0.169***	0.170***	-0.001	0.001	0.001	0.002
	(0.009)	(0.009)	(0.010)	(0.011)	(0.011)	(0.011)	(0.011)
Age	-0.002***	-0.002***	-0.001***	-0.001**	-0.001**	-0.001**	-0.001*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.355***	0.389***	0.411***	0.361***	0.348***	0.348***	0.349***
	(0.004)	(0.004)	(0.004)	(0.005)	(0.006)	(0.006)	(0.006)
Homo	0.001	0.015	0.028**	0.010	0.020*	0.021*	0.014
	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Observation	1323028	1252179	1166083	1048896	963952	963841	952533
Group	115334	188739	247743	347052	345056	345028	343433
Under-identification test	5041.735	8653.055	6850.243	1.8e+04	1.2e+04	1.2e+04	1.3e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.8e+05	5.8e+05	3.9e+05	3.4e+05	2.5e+05	2.5e+05	2.7e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.344	0.053	2.381	0.268	0.412	0.431	0.047
P value	(0.558)	(0.819)	(0.123)	(0.605)	(0.521)	(0.512)	(0.828)
First-stage Results:							
Ming-Road	0.365***	0.330***	0.0860	0.191***	0.491***	0.491***	0.226***
	(0.0138)	(0.0144)	(0.0483)	(0.0328)	(0.0196)	(0.0196)	(0.0342)
Counter-factual-Road	0.425***	0.478***			0.463***	0.463***	
	(0.0126)	(0.0147)			(0.0203)	(0.0203)	
Qing-Road			0.656***	0.666***			0.617***
			(0.0412)	(0.0254)			(0.0266)
F	674.571	425.776	740.705	852.078	564.245	563.527	682.930
R2	0.613	0.591	0.565	0.622	0.563	0.563	0.581
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Destination	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I10: The Impacts of Highways on Firms' Export Volume of Products.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Volume	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	0.131***	0.060***	0.053***	0.022**	0.026***	0.026**	0.023**
	(0.008)	(0.008)	(0.010)	(0.007)	(0.008)	(0.008)	(0.008)
TFP-ACF	-0.032***	-0.005	0.032***	0.069***	0.073***	0.072***	0.074***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.005)
GDP	0.830***	0.662***	0.645***	0.423***	0.408***	0.408***	0.402***
	(0.027)	(0.026)	(0.026)	(0.027)	(0.028)	(0.028)	(0.028)
Foreign-Dist	0.200***	0.195***	0.194***	0.198***	0.194***	0.194***	0.191***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.010)
Remoteness	-0.715***	-0.541***	-0.520***	-0.294***	-0.284***	-0.284***	-0.277***
	(0.027)	(0.026)	(0.026)	(0.027)	(0.028)	(0.028)	(0.028)
Rail-Density	-0.122***	-0.149***	-0.187***	-0.208***	-0.216***	-0.216***	-0.214***
	(0.027)	(0.026)	(0.025)	(0.022)	(0.023)	(0.023)	(0.023)
Road-Density	0.063**	0.076***	0.101***	0.172***	0.132***	0.131***	0.131***
	(0.024)	(0.022)	(0.021)	(0.019)	(0.020)	(0.020)	(0.020)
Pop	0.002	-0.031*	-0.041**	-0.084***	-0.052***	-0.051***	-0.053***
	(0.013)	(0.015)	(0.015)	(0.010)	(0.010)	(0.010)	(0.010)
State-share	-0.192***	-0.174***	-0.070*	0.047	0.046	0.042	0.038
	(0.029)	(0.031)	(0.034)	(0.029)	(0.031)	(0.031)	(0.031)
Collective-share	-0.063**	-0.016	-0.003	0.050**	0.042*	0.042*	0.035
	(0.020)	(0.021)	(0.022)	(0.019)	(0.020)	(0.020)	(0.020)
Foreign-share	0.192***	0.161***	0.175***	0.040**	0.047***	0.046***	0.045**
	(0.012)	(0.012)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
Age	-0.001**	-0.001**	-0.001**	-0.001**	-0.001**	-0.001**	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.196***	0.215***	0.244***	0.327***	0.313***	0.313***	0.311***
	(0.005)	(0.005)	(0.005)	(0.007)	(0.007)	(0.007)	(0.007)
Homo	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Observation	991589	932422	866620	778814	721210	721005	712183
Group	123122	162112	193596	261495	258749	258694	257445
Under-identification test	2494.680	4110.599	4455.862	1.5e+04	1.2e+04	1.2e+04	1.2e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	4.6e+05	4.5e+05	3.4e+05	2.6e+05	2.3e+05	2.3e+05	2.2e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.051	1.762	2.518	1.794	0.358	0.352	0.744
P value	(0.821)	(0.184)	(0.113)	(0.180)	(0.550)	(0.553)	(0.389)
First-stage Results:							
Ming-Road	0.328***	0.345***		0.051***	0.072***	0.070***	0.058***
	(0.004)	(0.004)		(0.010)	(0.011)	(0.011)	(0.011)
Counter-factual-Road	0.439***	0.456***	0.226***				
	(0.003)	(0.004)	(0.007)				
Qing-Road			0.558***	0.746***	0.724***	0.727***	0.735***
			(0.006)	(0.008)	(0.008)	(0.008)	(0.008)
F	9431.627	6377.174	8497.859	9452.057	7588.718	7608.237	7573.425
R2	0.597	0.594	0.598	0.613	0.591	0.592	0.595
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
HS Code	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I11: The Impacts of Highways on Firms' Export Value by Product.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	0.037***	0.012	0.011	-0.013	-0.007	-0.007	-0.008
	(0.007)	(0.007)	(0.009)	(0.009)	(0.010)	(0.010)	(0.011)
Foreign-Dist	0.030***	0.025***	0.025***	0.073***	0.072***	0.072***	0.070***
	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
TFP-ACF	0.066***	0.070***	0.076***	0.061***	0.063***	0.063***	0.064***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)
Remoteness	0.303***	0.317***	0.268***	0.166***	0.159***	0.159***	0.165***
	(0.020)	(0.020)	(0.020)	(0.023)	(0.024)	(0.024)	(0.024)
GDP	-0.192***	-0.196***	-0.144***	-0.034	-0.030	-0.030	-0.035
	(0.022)	(0.020)	(0.021)	(0.023)	(0.024)	(0.024)	(0.024)
Rail-Density	-0.035	-0.031	-0.021	-0.046	-0.030	-0.029	-0.029
	(0.027)	(0.027)	(0.027)	(0.025)	(0.028)	(0.028)	(0.028)
Road-Density	-0.074**	-0.061**	-0.026	0.046*	0.035	0.033	0.036
	(0.024)	(0.022)	(0.023)	(0.023)	(0.023)	(0.024)	(0.024)
Pop	0.049***	-0.088***	-0.086***	-0.084***	-0.069***	-0.069***	-0.069***
	(0.011)	(0.016)	(0.016)	(0.013)	(0.014)	(0.014)	(0.014)
State-share	-0.135***	-0.142***	-0.123***	-0.045	-0.050	-0.054	-0.052
	(0.026)	(0.029)	(0.032)	(0.034)	(0.036)	(0.036)	(0.036)
Collective-share	-0.009	0.024	0.022	-0.004	-0.000	-0.000	0.005
	(0.018)	(0.019)	(0.021)	(0.023)	(0.024)	(0.024)	(0.024)
Foreign-share	0.090***	0.104***	0.116***	0.018	0.023	0.022	0.027
	(0.010)	(0.010)	(0.011)	(0.016)	(0.017)	(0.017)	(0.018)
Age	-0.002***	-0.002***	-0.002***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.049***	0.055***	0.075***	0.138***	0.140***	0.140***	0.141***
	(0.004)	(0.004)	(0.004)	(0.008)	(0.009)	(0.009)	(0.009)
Homo	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Observation	950170	891247	825835	738062	682394	682194	673648
Group	120186	157470	186966	250609	247162	247108	245719
Under-identification test	2412.075	4019.426	4310.061	1.2e+04	9725.742	9708.670	9377.079
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	4.3e+05	4.2e+05	3.2e+05	2.1e+05	1.8e+05	1.8e+05	1.8e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.338	1.091	2.543	0.417	1.890	1.960	2.397
P value	(0.561)	(0.296)	(0.111)	(0.518)	(0.169)	(0.162)	(0.122)
First-stage Results:							
Ming-Road	0.328***	0.349***		0.476***	0.488***	0.489***	0.483***
	(0.004)	(0.004)		(0.006)	(0.007)	(0.007)	(0.007)
Counter-factual-Road	0.437***	0.452***	0.218***	0.450***	0.439***	0.437***	0.445***
	(0.003)	(0.004)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)
Qing-Road			0.565***				
			(0.006)				
F	9351.842	6346.955	8289.935	6662.031	5057.741	5046.668	4973.568
R2	0.595	0.593	0.599	0.586	0.564	0.564	0.567
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
HS Code	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I12: The Impacts on Firm-level Product Variety.

Dependent Variable:	Without Entry			Post-2000 Entry		
Product Variety	(1)	(2)	(3)	(4)	(5)	(6)
Main Results:						
DGC	-0.031**	-0.007	0.015	-0.063***	-0.042***	-0.003
	(0.010)	(0.011)	(0.010)	(0.011)	(0.011)	(0.013)
TFP-ACF	0.015	0.016*	0.009	0.007	0.003	0.005
	(0.008)	(0.008)	(0.007)	(0.005)	(0.005)	(0.005)
GDP	0.585***	0.578***	0.476***	0.330***	0.348***	0.324***
	(0.075)	(0.060)	(0.061)	(0.047)	(0.036)	(0.034)
Foreign-Dist	0.021	0.031	0.032	0.091***	0.095***	0.075***
	(0.023)	(0.021)	(0.021)	(0.016)	(0.013)	(0.013)
Remoteness	-0.496***	-0.503***	-0.409***	-0.246***	-0.266***	-0.246***
	(0.075)	(0.060)	(0.060)	(0.046)	(0.035)	(0.034)
Rail-Density	-0.145***	-0.100***	-0.077***	0.197**	0.124	0.212**
	(0.024)	(0.021)	(0.020)	(0.073)	(0.070)	(0.068)
Road-Density	0.123***	0.104***	0.099***	-0.056	-0.058**	-0.050*
	(0.033)	(0.028)	(0.028)	(0.030)	(0.022)	(0.020)
Pop	0.007	-0.006	0.024	0.043***	-0.043**	-0.031*
	(0.017)	(0.018)	(0.016)	(0.013)	(0.016)	(0.015)
State-share	-0.037	-0.049	-0.036	-0.016	-0.071*	-0.052
	(0.037)	(0.043)	(0.040)	(0.031)	(0.034)	(0.038)
Collective-share	-0.073**	-0.055	-0.048	-0.042*	-0.034	-0.010
	(0.028)	(0.029)	(0.029)	(0.021)	(0.022)	(0.023)
Foreign-share	0.159***	0.158***	0.110***	0.197***	0.203***	0.194***
	(0.024)	(0.024)	(0.023)	(0.016)	(0.012)	(0.013)
Age	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.171***	0.178***	0.186***	0.118***	0.134***	0.152***
	(0.007)	(0.007)	(0.008)	(0.004)	(0.004)	(0.005)
Homo	-0.074*	-0.065***	-0.057**	-0.045*	-0.051***	-0.056***
	(0.030)	(0.019)	(0.018)	(0.022)	(0.015)	(0.013)
Observation	65381	64248	62404	121142	118692	114318
Group	3158	6287	9543	4600	10953	18128
Under-identification test	147.876	387.808	363.908	218.569	376.792	380.044
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	4.3e+04	3.4e+04	2.6e+04	7.9e+04	6.8e+04	3.4e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	1.424	0.483	0.060	0.356	1.064	0.072
P value	(0.233)	(0.487)	(0.806)	(0.551)	(0.302)	(0.788)
First-stage Results:						
Qing-Road	-0.034	-0.037	-0.264***	0.126***	0.369***	0.301***
	(0.038)	(0.055)	(0.075)	(0.018)	(0.015)	(0.053)
Terrain-Surface	0.746***	0.762***	0.996***	0.546***		0.381***
	(0.035)	(0.053)	(0.071)	(0.014)		(0.051)
Counter-factual-Road					0.379***	
					(0.014)	
F	1017.157	1323.309	1207.325	1067.945	350.036	377.677
R2	0.695	0.639	0.636	0.616	0.576	0.454
Fixed Effects:						
Region	province	prefecture	county	province	prefecture	county
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I13: The Impacts on Firm-level Export Value.

Dependent Variable:	Without Entry			Post-2000 Entry		
Product Variety	(1)	(2)	(3)	(4)	(5)	(6)
Main Results:						
DGC	-0.075***	-0.000	0.044*	-0.075***	-0.038	0.014
	(0.017)	(0.020)	(0.019)	(0.016)	(0.020)	(0.025)
TFP-ACF	0.143***	0.145***	0.149***	0.048***	0.040***	0.080***
	(0.015)	(0.014)	(0.015)	(0.012)	(0.011)	(0.010)
GDP	1.218***	1.113***	0.990***	1.042***	0.929***	0.923***
	(0.127)	(0.112)	(0.113)	(0.084)	(0.075)	(0.075)
Foreign-Dist	0.073	0.142**	0.146**	0.202***	0.189***	0.171***
	(0.044)	(0.046)	(0.045)	(0.029)	(0.029)	(0.029)
Remoteness	-0.991***	-0.912***	-0.807***	-0.853***	-0.740***	-0.736***
	(0.124)	(0.111)	(0.112)	(0.082)	(0.073)	(0.073)
Rail-Density	-0.391***	-0.283***	-0.209***	0.087	-0.022	-0.025
	(0.056)	(0.043)	(0.041)	(0.129)	(0.129)	(0.141)
Road-Density	0.201**	0.139*	0.146**	-0.029	-0.027	0.010
	(0.065)	(0.055)	(0.050)	(0.060)	(0.047)	(0.043)
Pop	-0.070*	-0.147***	-0.073*	0.040	-0.126***	-0.102**
	(0.031)	(0.037)	(0.036)	(0.022)	(0.032)	(0.032)
State-share	-0.562***	-0.485***	-0.388***	-0.517***	-0.498***	-0.297***
	(0.074)	(0.078)	(0.079)	(0.073)	(0.081)	(0.087)
Collective-share	-0.192***	-0.137*	-0.155**	-0.130**	-0.088	-0.030
	(0.054)	(0.055)	(0.054)	(0.049)	(0.049)	(0.049)
Foreign-share	0.667***	0.626***	0.489***	0.675***	0.705***	0.681***
	(0.043)	(0.042)	(0.042)	(0.030)	(0.025)	(0.025)
Age	-0.002	-0.001*	-0.001	-0.004***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Size	0.649***	0.672***	0.696***	0.553***	0.586***	0.633***
	(0.019)	(0.015)	(0.016)	(0.013)	(0.011)	(0.012)
Homo	0.014	0.021	0.013	-0.008	-0.006	0.009
	(0.032)	(0.026)	(0.026)	(0.025)	(0.021)	(0.021)
Observation	65381	64248	62404	121142	118692	114318
Group	3158	6287	9543	4600	10953	18128
Under-identification test	184.291	387.808	401.393	191.200	385.644	425.467
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	4.3e+04	3.4e+04	2.6e+04	8.3e+04	6.3e+04	3.7e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	1.115	2.311	1.549	2.985	0.287	0.103
P value	(0.291)	(0.129)	(0.213)	(0.084)	(0.592)	(0.748)
First-stage Results:						
Terrain-Surface	0.695***	0.762***	0.603***	0.427***	0.427***	0.396***
	(0.026)	(0.053)	(0.020)	(0.014)	(0.023)	(0.016)
Counter-factual-Road	0.026		0.182***	0.277***		0.340***
	(0.033)		(0.026)	(0.018)		(0.019)
Qing-Road		-0.037			0.259***	
		(0.055)			(0.024)	
F	955.711	1323.273	1114.429	1036.498	395.678	374.521
R2	0.695	0.639	0.639	0.629	0.555	0.470
Fixed Effects:						
Region	province	prefecture	county	province	prefecture	county
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I14: The Impacts on Export Volume by Firm, Destination and Product (without entry).

Dependent Variable:	Without Region and Industry Fixed Effect				With Region and Industry Fixed Effect			
Export Volume	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
DGC	0.009	0.017*	0.016	0.022***	0.012	0.016	0.024**	0.020**
	(0.014)	(0.008)	(0.008)	(0.006)	(0.016)	(0.009)	(0.009)	(0.006)
Foreign-Dist	0.053***	0.082***	0.046*	0.050***	0.051***	0.062*	0.079***	0.075***
	(0.009)	(0.005)	(0.022)	(0.014)	(0.008)	(0.026)	(0.005)	(0.016)
TFP-ACF	0.027**	0.018**	0.035***	0.035***	0.030**	0.031***	0.026***	0.037***
	(0.010)	(0.006)	(0.006)	(0.004)	(0.011)	(0.007)	(0.006)	(0.005)
Remoteness	0.059*	0.494***	0.529***	0.490***	0.053*	0.485***	0.488***	0.465***
	(0.028)	(0.018)	(0.092)	(0.062)	(0.026)	(0.106)	(0.017)	(0.070)
GDP	0.139***	-0.173***	0.000	0.000	0.144***	0.000	-0.165***	0.000
	(0.028)	(0.018)	(.)	(.)	(0.026)	(.)	(0.018)	(.)
Rail-Density	-0.076*	-0.085***	-0.120***	-0.181***	-0.068	-0.118***	-0.068**	-0.171***
	(0.033)	(0.020)	(0.022)	(0.015)	(0.037)	(0.025)	(0.022)	(0.017)
Road-Density	0.008	0.074**	0.021	0.094***	0.014	0.027	0.047	0.048*
	(0.044)	(0.025)	(0.027)	(0.018)	(0.046)	(0.029)	(0.026)	(0.019)
Pop	-0.045*	-0.005	-0.043***	-0.022**	-0.044*	-0.037**	-0.013	-0.026**
	(0.020)	(0.013)	(0.012)	(0.008)	(0.020)	(0.013)	(0.013)	(0.009)
State-share	-0.034	-0.016	-0.036	0.022	-0.040	-0.020	-0.006	0.044
	(0.056)	(0.031)	(0.031)	(0.023)	(0.062)	(0.033)	(0.034)	(0.025)
Collective-share	-0.002	0.037	-0.017	0.054**	-0.025	-0.039	0.000	0.030
	(0.040)	(0.028)	(0.024)	(0.017)	(0.039)	(0.026)	(0.024)	(0.018)
Foreign-share	0.042	0.028	0.034	-0.000	0.045	0.036	0.015	-0.015
	(0.039)	(0.018)	(0.019)	(0.013)	(0.043)	(0.021)	(0.018)	(0.014)
Age	-0.001*	-0.001***	-0.001***	-0.001***	-0.001*	-0.001**	-0.001***	-0.001**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.086***	0.136***	0.115***	0.242***	0.095***	0.121***	0.132***	0.227***
	(0.016)	(0.010)	(0.009)	(0.007)	(0.018)	(0.010)	(0.010)	(0.007)
Homo	0.126**	0.000	0.179***	0.000	0.124***	0.179***	0.000	0.000
	(0.038)	(.)	(0.015)	(.)	(0.035)	(0.015)	(.)	(.)
Observation	1318544	1223246	1256291	936419	1289589	1205642	1177681	828991
Group	16026	119026	144790	294057	21854	163181	130688	288502
Under-identification test	652.476	2651.824	4788.989	2.7e+04	538.924	3620.733	2145.528	2.1e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.5e+05	5.4e+05	5.6e+05	3.2e+05	6.1e+05	5.1e+05	5.0e+05	2.6e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.024	0.701	0.515	0.966	0.183	0.058	0.193	0.910
P value	(0.877)	(0.403)	(0.473)	(0.326)	(0.669)	(0.810)	(0.661)	(0.340)
First-stage Results:								
Qing-Road	0.569***	0.570***	0.576***	0.574***	0.571***	0.579***	0.571***	0.577***
	(0.036)	(0.018)	(0.014)	(0.007)	(0.040)	(0.016)	(0.020)	(0.007)
Ming-Road	0.197***	0.191***	0.181***	0.177***	0.199***	0.183***	0.191***	0.177***
	(0.046)	(0.024)	(0.019)	(0.009)	(0.051)	(0.021)	(0.027)	(0.010)
F	223.693	1024.415	1375.524	7739.823	221.203	1148.802	931.534	6422.417
R2	0.482	0.482	0.483	0.488	0.460	0.464	0.461	0.470
Fixed Effects:								
Region					county	county	county	county
Industry					4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6-digit HS		Yes		Yes		Yes		Yes
Destination			Yes	Yes			Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I15: The Impacts on Export Intensity by Firm, Destination and Product (post-2000 entry).

Dependent Variable:	Without Region and Industry Fixed Effect				With Region and Industry Fixed Effect			
Export Volume	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
DGC	0.010	0.006	0.020	0.007	0.013	0.015	0.004	0.003
	(0.018)	(0.013)	(0.014)	(0.010)	(0.020)	(0.015)	(0.014)	(0.011)
Foreign-Dist	0.036***	0.064***	-0.061	-0.024	0.035***	-0.068	0.064***	-0.019
	(0.005)	(0.003)	(0.035)	(0.023)	(0.005)	(0.041)	(0.003)	(0.026)
TFP-ACF	0.037***	0.039***	0.048***	0.066***	0.035***	0.043***	0.038***	0.062***
	(0.007)	(0.005)	(0.005)	(0.004)	(0.008)	(0.005)	(0.005)	(0.004)
Remoteness	-0.010	0.189***	0.426***	0.055	-0.011	0.489***	0.189***	0.057
	(0.018)	(0.012)	(0.103)	(0.069)	(0.017)	(0.114)	(0.012)	(0.076)
GDP	0.181***	0.080***	0.000	0.000	0.183***	0.000	0.082***	0.000
	(0.018)	(0.013)	(.)	(.)	(0.018)	(.)	(0.012)	(.)
Rail-Density	-0.092	0.087	-0.126	0.036	-0.024	-0.105	0.101	0.029
	(0.090)	(0.063)	(0.072)	(0.055)	(0.104)	(0.084)	(0.073)	(0.063)
Road-Density	0.057	0.037	0.096***	0.065***	0.044	0.091***	0.020	0.059***
	(0.032)	(0.019)	(0.021)	(0.015)	(0.033)	(0.021)	(0.019)	(0.015)
Pop	-0.016	-0.021*	-0.023*	-0.030***	-0.010	-0.017	-0.013	-0.019*
	(0.017)	(0.011)	(0.011)	(0.008)	(0.017)	(0.011)	(0.010)	(0.008)
State-share	0.024	0.007	0.046	0.040	0.031	0.054	-0.002	0.031
	(0.066)	(0.035)	(0.039)	(0.028)	(0.075)	(0.042)	(0.038)	(0.031)
Collective-share	-0.013	0.005	-0.020	-0.005	-0.033	-0.032	0.008	-0.012
	(0.037)	(0.021)	(0.024)	(0.017)	(0.039)	(0.025)	(0.022)	(0.018)
Foreign-share	-0.019	0.021	-0.019	0.015	-0.002	-0.006	0.036*	0.031**
	(0.025)	(0.014)	(0.016)	(0.011)	(0.024)	(0.016)	(0.015)	(0.012)
Age	-0.002**	-0.001***	-0.002*	-0.001***	-0.002**	-0.002*	-0.001***	-0.001**
	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Size	0.106***	0.161***	0.145***	0.277***	0.104***	0.140***	0.154***	0.265***
	(0.011)	(0.007)	(0.008)	(0.006)	(0.012)	(0.008)	(0.008)	(0.006)
Homo	-0.107***	0.000	-0.043***	0.000	-0.106***	-0.042***	0.000	0.000
	(0.026)	(.)	(0.011)	(.)	(0.024)	(0.010)	(.)	(.)
Observation	1867369	1674859	1680479	1055970	1860167	1648230	1651644	970679
Group	46961	206875	295075	412521	52707	305635	214039	393586
Under-identification test	347.017	1057.271	1159.810	8460.768	261.671	766.190	871.767	6390.742
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	6.3e+05	5.1e+05	4.8e+05	2.3e+05	5.8e+05	4.4e+05	4.6e+05	1.9e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.145	0.001	0.023	4.114	0.089	0.107	0.583	1.131
P value	(0.703)	(0.973)	(0.881)	(0.043)	(0.765)	(0.744)	(0.445)	(0.288)
First-stage Results:								
Qing-Road	0.689***	0.669***	0.695***	0.691***	0.692***	0.709***	0.681***	0.707***
	(0.065)	(0.021)	(0.015)	(0.006)	(0.040)	(0.015)	(0.020)	(0.006)
Ming-Road	0.167	0.188***	0.155***	0.154***	0.136**	0.110***	0.144***	0.106***
	(0.089)	(0.027)	(0.019)	(0.007)	(0.052)	(0.019)	(0.026)	(0.008)
F	518.011	2156.692	4222.114	20753.227	371.731	3037.571	1662.248	15055.188
R2	0.672	0.666	0.669	0.667	0.639	0.639	0.635	0.638
Fixed Effects:								
Region					county	county	county	county
Industry					4-digit	4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
6-digit HS		Yes		Yes		Yes		Yes
Destination			Yes	Yes			Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I16: The Impacts on Firms' Export Volume of Products (without entry).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Share	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	0.092***	0.032*	0.026*	0.019*	0.028**	0.027**	0.020*
	(0.012)	(0.013)	(0.013)	(0.009)	(0.010)	(0.010)	(0.010)
TFP-ACF	-0.010	0.017	0.071***	0.056***	0.062***	0.062***	0.066***
	(0.009)	(0.009)	(0.009)	(0.007)	(0.007)	(0.007)	(0.007)
GDP	0.988***	0.772***	0.745***	0.503***	0.479***	0.479***	0.472***
	(0.046)	(0.046)	(0.047)	(0.043)	(0.045)	(0.045)	(0.046)
Foreign-Dist	0.215***	0.224***	0.218***	0.206***	0.198***	0.198***	0.192***
	(0.013)	(0.015)	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)
Remoteness	-0.874***	-0.653***	-0.620***	-0.382***	-0.367***	-0.368***	-0.359***
	(0.045)	(0.046)	(0.047)	(0.044)	(0.046)	(0.046)	(0.046)
Rail-Density	-0.126***	-0.150***	-0.196***	-0.245***	-0.254***	-0.255***	-0.251***
	(0.030)	(0.027)	(0.027)	(0.023)	(0.024)	(0.024)	(0.024)
Road-Density	0.004	0.031	0.059	0.127***	0.073*	0.069*	0.072*
	(0.035)	(0.033)	(0.033)	(0.030)	(0.031)	(0.031)	(0.031)
Pop	-0.040*	-0.023	-0.035	-0.088***	-0.052***	-0.051***	-0.057***
	(0.020)	(0.021)	(0.021)	(0.015)	(0.015)	(0.015)	(0.015)
State-share	-0.219***	-0.212***	-0.095*	0.022	0.028	0.029	0.026
	(0.040)	(0.043)	(0.045)	(0.037)	(0.039)	(0.039)	(0.039)
Collective-share	-0.053	-0.033	0.022	0.069**	0.051	0.051	0.040
	(0.029)	(0.030)	(0.031)	(0.026)	(0.027)	(0.027)	(0.027)
Foreign-share	0.126***	0.087***	0.083***	0.050*	0.047*	0.047*	0.045*
	(0.021)	(0.022)	(0.024)	(0.020)	(0.021)	(0.021)	(0.021)
Age	-0.001	-0.001	-0.000	-0.000*	-0.000	-0.000	-0.000
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.192***	0.201***	0.224***	0.280***	0.255***	0.255***	0.249***
	(0.008)	(0.008)	(0.009)	(0.011)	(0.011)	(0.011)	(0.011)
Homo	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Observation	394923	375068	353336	354658	324162	323986	319303
Group	58392	73080	84277	102912	104364	104322	104136
Under-identification test	2106.163	3859.983	3934.435	9894.387	8589.122	8572.756	8371.008
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.5e+05	1.6e+05	1.4e+05	1.3e+05	1.1e+05	1.1e+05	1.1e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.145	0.066	3.824	0.000	0.000	0.000	0.167
P value	(0.703)	(0.797)	(0.051)	(0.985)	(0.997)	(0.999)	(0.683)
First-stage Results:							
Ming-Road	0.434***	0.383***		0.041**	0.050***	0.047***	0.032*
	(0.007)	(0.006)		(0.013)	(0.014)	(0.014)	(0.014)
Counter-factual-Road	0.366***	0.445***	0.222***				
	(0.006)	(0.006)	(0.009)				
Qing-Road			0.586***	0.768***	0.755***	0.759***	0.769***
			(0.007)	(0.010)	(0.011)	(0.011)	(0.011)
F	4303.630	5199.974	5975.387	7298.846	5599.841	5627.408	5620.669
R2	0.610	0.606	0.640	0.653	0.630	0.630	0.634
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
HS Code	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e. valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I17: The Impacts on Firms' Export Volume of Products (post-2000 entry).

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Export Share	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	0.152***	0.072***	0.083***	0.002	0.004	0.004	0.009
	(0.010)	(0.011)	(0.015)	(0.015)	(0.016)	(0.016)	(0.016)
Foreign-Dist	-0.061***	-0.079***	-0.073***	-0.054***	-0.050***	-0.050***	-0.051***
	(0.008)	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)
TFP-ACF	-0.039***	-0.012	0.027***	0.073***	0.075***	0.074***	0.075***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
Remoteness	-0.214***	-0.147***	-0.154***	-0.073***	-0.065**	-0.065**	-0.059**
	(0.026)	(0.025)	(0.025)	(0.022)	(0.022)	(0.022)	(0.022)
GDP	0.106***	0.043	0.062*	0.015	0.009	0.009	0.003
	(0.027)	(0.025)	(0.025)	(0.022)	(0.023)	(0.023)	(0.023)
Rail-Density	0.225*	0.138	0.093	0.175	0.183	0.183	0.162
	(0.091)	(0.093)	(0.092)	(0.091)	(0.100)	(0.100)	(0.101)
Road-Density	0.058	0.089**	0.076*	0.157***	0.126***	0.126***	0.124***
	(0.032)	(0.031)	(0.030)	(0.027)	(0.027)	(0.027)	(0.027)
Pop	0.084***	-0.037*	-0.032	-0.050***	-0.029*	-0.029*	-0.028*
	(0.016)	(0.018)	(0.018)	(0.013)	(0.013)	(0.013)	(0.013)
State-share	-0.143**	-0.093	0.023	0.083	0.083	0.073	0.062
	(0.049)	(0.055)	(0.057)	(0.049)	(0.052)	(0.052)	(0.053)
Collective-share	-0.055	0.016	0.039	0.023	0.027	0.026	0.027
	(0.029)	(0.031)	(0.032)	(0.029)	(0.030)	(0.030)	(0.031)
Foreign-share	0.244***	0.197***	0.205***	0.029	0.046*	0.046*	0.046*
	(0.014)	(0.014)	(0.016)	(0.018)	(0.019)	(0.019)	(0.019)
Age	-0.002***	-0.003***	-0.002***	-0.002***	-0.002***	-0.002***	-0.002***
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.224***	0.246***	0.278***	0.344***	0.334***	0.334***	0.333***
	(0.006)	(0.006)	(0.007)	(0.010)	(0.010)	(0.010)	(0.010)
Homo	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Observation	542771	502097	460838	398945	373090	373066	369125
Group	80205	102585	118612	150581	146239	146228	145144
Under-identification test	1837.644	2245.227	2119.421	3094.747	2476.267	2476.227	2357.940
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	2.8e+05	2.5e+05	1.6e+05	9.3e+04	8.4e+04	8.4e+04	8.2e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.283	2.955	0.981	0.286	2.902	2.865	3.981
P value	(0.595)	(0.086)	(0.322)	(0.593)	(0.089)	(0.091)	(0.046)
First-stage Results:							
Ming-Road	0.233***	0.296***		0.361***	0.380***	0.380***	0.375***
	(0.005)	(0.006)		(0.010)	(0.011)	(0.011)	(0.012)
Counter-factual-Road	0.504***	0.483***	0.341***	0.501***	0.492***	0.492***	0.503***
	(0.004)	(0.006)	(0.009)	(0.010)	(0.011)	(0.011)	(0.011)
Qing-Road			0.418***				
			(0.008)				
F	5939.626	2468.707	2747.975	2111.709	1809.608	1809.485	1800.261
R2	0.591	0.571	0.507	0.489	0.481	0.481	0.482
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
HS Code	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table I18: Robustness Checks on Alternative Measurement on Trade Cost.

Dependent Variable:	Firm-level Export Value				Firm-level Export Variety			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
Codist1	-0.048***		0.054		-0.045***		-0.084	
	(0.010)		(0.142)		(0.006)		(0.118)	
Codist2		-0.101***		0.113		-0.065***		-0.050
		(0.016)		(0.148)		(0.010)		(0.135)
Foreign-Dist	0.144***	0.149***	0.167***	0.167***	0.060**	0.064***	0.072***	0.072***
	(0.029)	(0.028)	(0.024)	(0.024)	(0.018)	(0.018)	(0.009)	(0.009)
TFP-ACF	0.108***	0.109***	0.139***	0.139***	0.012*	0.014**	0.015***	0.015***
	(0.011)	(0.011)	(0.006)	(0.006)	(0.005)	(0.005)	(0.003)	(0.003)
GDP	1.165***	1.173***	0.406***	0.406***	0.395***	0.399***	0.166***	0.166***
	(0.090)	(0.088)	(0.056)	(0.056)	(0.053)	(0.052)	(0.023)	(0.023)
Remoteness	-0.953***	-0.962***	-0.309***	-0.309***	-0.305***	-0.309***	-0.128***	-0.128***
	(0.086)	(0.085)	(0.056)	(0.056)	(0.052)	(0.051)	(0.023)	(0.023)
Rail-Density	-0.273***	-0.275***	-0.259***	-0.259***	-0.038	-0.042	-0.068***	-0.068***
	(0.049)	(0.049)	(0.030)	(0.030)	(0.023)	(0.023)	(0.015)	(0.015)
Road-Density	0.109*	0.109*	0.236***	0.236***	0.050	0.053*	0.069***	0.070***
	(0.055)	(0.055)	(0.027)	(0.027)	(0.026)	(0.026)	(0.014)	(0.014)
Pop	0.006	0.004	-0.147***	-0.147***	0.031**	0.041***	-0.016*	-0.016*
	(0.019)	(0.018)	(0.013)	(0.013)	(0.010)	(0.010)	(0.006)	(0.006)
State-share	-0.464***	-0.463***	-0.013	-0.013	0.036	0.035	0.038	0.038
	(0.054)	(0.054)	(0.046)	(0.046)	(0.028)	(0.028)	(0.021)	(0.021)
Collective-share	-0.160***	-0.159***	0.010	0.010	-0.031	-0.032	0.019	0.019
	(0.041)	(0.042)	(0.031)	(0.031)	(0.021)	(0.021)	(0.015)	(0.015)
Foreign-share	0.756***	0.744***	0.007	0.007	0.208***	0.203***	-0.009	-0.009
	(0.032)	(0.032)	(0.019)	(0.019)	(0.015)	(0.015)	(0.010)	(0.010)
Age	-0.001*	-0.001*	-0.001*	-0.001*	0.000	0.000	-0.000*	-0.000*
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.609***	0.609***	0.555***	0.555***	0.144***	0.145***	0.156***	0.156***
	(0.013)	(0.013)	(0.011)	(0.011)	(0.005)	(0.005)	(0.005)	(0.005)
Homo	0.008	0.007	-0.071***	-0.071***	-0.043	-0.043	-0.085***	-0.085***
	(0.025)	(0.025)	(0.011)	(0.011)	(0.024)	(0.025)	(0.007)	(0.007)
R2	0.220	0.221	0.158	0.158	0.0789	0.0787	0.0619	0.0619
Observation	169571	169571	171294	171294	169571	169571	171294	171294
Group	4966	4966	60031	60031	4966	4966	60031	60031
F	219.0	239.8	558.7	558.6	101.8	101.7	195.8	195.8
Fixed Effects:								
Region	province	province			province	province		
Industry	4-digit	4-digit			4-digit	4-digit		
Firm			Yes	Yes			Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The full regression results are provided in Table I18. Columns (1), (2), (5) and (6) control 4-digit industry, province, and year fixed effects; columns (3), (4), (7) and (8) control firm and year fixed effects. The cluster standard error is always applied in each column, according their fixed-effect groups. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

.3 Appendix J

Channel Studies and Heterogeneity Studies

Table J1: The Impacts of Highways on Firm-level Import Value.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Variety by Destination	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.132***	-0.049**	0.007	0.028*	0.017	0.016	0.023
	(0.017)	(0.018)	(0.019)	(0.012)	(0.013)	(0.013)	(0.013)
TFP-ACF	0.163***	0.167***	0.156***	0.105***	0.104***	0.104***	0.106***
	(0.011)	(0.011)	(0.011)	(0.008)	(0.009)	(0.009)	(0.009)
GDP	-0.895***	-0.736***	-0.567***	0.016	0.013	0.016	0.017
	(0.106)	(0.085)	(0.086)	(0.080)	(0.084)	(0.084)	(0.084)
Foreign-Dist	0.272***	0.239***	0.207***	0.198***	0.189***	0.189***	0.190***
	(0.038)	(0.038)	(0.034)	(0.030)	(0.032)	(0.032)	(0.032)
Remoteness	0.948***	0.791***	0.632***	0.066	0.070	0.069	0.068
	(0.105)	(0.085)	(0.086)	(0.079)	(0.083)	(0.083)	(0.083)
Rail-Density	-0.305***	-0.256***	-0.139**	-0.117**	-0.131**	-0.132**	-0.115**
	(0.059)	(0.053)	(0.049)	(0.038)	(0.042)	(0.042)	(0.042)
Road-Density	0.244*	0.125	0.106	0.166***	0.160***	0.161***	0.160***
	(0.108)	(0.069)	(0.058)	(0.044)	(0.047)	(0.047)	(0.047)
Pop	-0.081**	-0.128***	-0.065**	-0.044**	-0.036*	-0.038*	-0.031*
	(0.030)	(0.025)	(0.024)	(0.015)	(0.015)	(0.015)	(0.015)
State-share	-0.524***	-0.509***	-0.392***	-0.064	-0.086	-0.088	-0.060
	(0.071)	(0.074)	(0.079)	(0.070)	(0.076)	(0.076)	(0.076)
Collective-share	-0.091	-0.078	-0.054	-0.008	-0.012	-0.012	-0.008
	(0.066)	(0.061)	(0.063)	(0.050)	(0.053)	(0.053)	(0.054)
Foreign-share	1.568***	1.402***	1.185***	-0.010	-0.019	-0.019	-0.018
	(0.052)	(0.039)	(0.035)	(0.028)	(0.029)	(0.029)	(0.029)
Age	-0.003**	-0.003**	-0.002**	-0.000	-0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.848***	0.856***	0.856***	0.635***	0.620***	0.621***	0.623***
	(0.016)	(0.011)	(0.011)	(0.016)	(0.017)	(0.017)	(0.017)
Homo	-0.603***	-0.521***	-0.430***	-0.169***	-0.167***	-0.166***	-0.166***
	(0.030)	(0.020)	(0.018)	(0.014)	(0.015)	(0.015)	(0.015)
Observation	131176	128660	123723	116459	107712	107694	106244
Group	4688	10299	17583	33229	34716	34711	34722
Under-identification test	232.039	477.367	396.463	2355.830	1939.321	1936.699	1925.363
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	8.6e+04	7.2e+04	5.0e+04	4.6e+04	3.8e+04	3.8e+04	3.7e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	11.839	0.001	2.491	1.174	1.209	0.949	1.268
P value	(0.001)	(0.981)	(0.115)	(0.279)	(0.272)	(0.330)	(0.260)
First-stage Results:							
Qing-Road	-0.006		-0.277***	-0.059	-0.057	-0.065	-0.062
	(0.023)		(0.059)	(0.052)	(0.059)	(0.059)	(0.061)
Terrain-Surface	0.690***	0.677***	0.973***	0.823***	0.812***	0.820***	0.814***
	(0.021)	(0.020)	(0.056)	(0.049)	(0.056)	(0.056)	(0.057)
Ming-Road		0.019					
		(0.024)					
F	1866.897	1697.186	1604.835	2014.155	1476.207	1477.897	1471.587
R2	0.653	0.620	0.604	0.655	0.620	0.620	0.624
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table J2: The Impacts of Highways on Firm-level Import Variety.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Variety by Destination	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	-0.110***	-0.040***	-0.010	0.001	0.007	0.007	0.009
	(0.011)	(0.010)	(0.010)	(0.006)	(0.007)	(0.007)	(0.007)
TFP-ACF	0.034***	0.030***	0.022***	0.012**	0.014***	0.014***	0.014***
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)
GDP	-0.183***	-0.105**	0.008	0.159***	0.176***	0.176***	0.174***
	(0.052)	(0.041)	(0.040)	(0.031)	(0.033)	(0.033)	(0.033)
Foreign-Dist	0.165***	0.156***	0.147***	0.112***	0.103***	0.103***	0.102***
	(0.016)	(0.016)	(0.015)	(0.011)	(0.012)	(0.012)	(0.012)
Remoteness	0.286***	0.197***	0.084*	-0.095**	-0.111***	-0.111***	-0.109***
	(0.052)	(0.040)	(0.039)	(0.031)	(0.033)	(0.033)	(0.033)
Rail-Density	-0.285***	-0.232***	-0.193***	-0.159***	-0.141***	-0.141***	-0.143***
	(0.032)	(0.025)	(0.024)	(0.017)	(0.019)	(0.019)	(0.019)
Road-Density	0.205***	0.130***	0.113***	0.092***	0.061**	0.060**	0.049*
	(0.032)	(0.027)	(0.026)	(0.020)	(0.021)	(0.021)	(0.021)
Pop	-0.058***	-0.094***	-0.064***	-0.033***	-0.034***	-0.034***	-0.033***
	(0.016)	(0.016)	(0.015)	(0.008)	(0.008)	(0.008)	(0.008)
State-share	-0.200***	-0.243***	-0.216***	-0.043	-0.020	-0.021	-0.015
	(0.035)	(0.035)	(0.038)	(0.029)	(0.032)	(0.032)	(0.033)
Collective-share	-0.061	-0.063*	-0.065*	-0.010	-0.001	-0.001	-0.004
	(0.038)	(0.031)	(0.030)	(0.021)	(0.022)	(0.022)	(0.023)
Foreign-share	1.055***	0.929***	0.772***	0.009	-0.000	-0.000	-0.001
	(0.030)	(0.021)	(0.018)	(0.013)	(0.013)	(0.013)	(0.013)
Age	-0.001*	-0.001*	-0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.360***	0.369***	0.364***	0.236***	0.224***	0.224***	0.225***
	(0.009)	(0.006)	(0.006)	(0.007)	(0.008)	(0.008)	(0.008)
Homo	-0.244***	-0.219***	-0.193***	-0.098***	-0.099***	-0.099***	-0.100***
	(0.015)	(0.011)	(0.010)	(0.007)	(0.007)	(0.007)	(0.007)
Observation	131176	128660	123723	116459	107712	107694	106244
Group	4688	10299	17583	33229	34716	34711	34722
Under-identification test	232.039	479.977	396.463	2355.830	1950.705	1947.931	1957.897
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	8.6e+04	7.2e+04	5.0e+04	4.6e+04	3.8e+04	3.8e+04	3.7e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	4.692	0.090	1.170	4.052	2.398	2.461	2.708
P value	(0.030)	(0.764)	(0.280)	(0.044)	(0.122)	(0.117)	(0.100)
First-stage Results:							
Terrain-Surface	0.690***	0.687***	0.973***	0.823***			
	(0.021)	(0.035)	(0.056)	(0.049)			
Qing-Road	-0.006	0.006	-0.277***	-0.059	0.616***	0.617***	0.619***
	(0.023)	(0.036)	(0.059)	(0.052)	(0.012)	(0.012)	(0.012)
Counter-factual-Road					0.243***	0.242***	0.238***
					(0.017)	(0.017)	(0.018)
F	1866.845	1727.036	1604.801	2014.105	1323.162	1322.620	1312.674
R2	0.653	0.620	0.604	0.655	0.620	0.620	0.624
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table J3: Import Decision as Control Variable.

Dependent Variable:	Export Value		Product Scope		Export Share in Sales	
	(1)	(2)	(3)	(4)	(5)	(6)
Main Results:						
Imp	0.288***	0.225***	0.112***	0.060***	0.236***	0.189***
	(0.019)	(0.015)	(0.010)	(0.007)	(0.018)	(0.014)
DGC	0.014	-0.024**	-0.001	-0.011**	0.011	-0.019*
	(0.010)	(0.009)	(0.006)	(0.004)	(0.010)	(0.008)
TFP-ACF	0.108***	0.138***	0.009*	0.017***	-0.245***	-0.109***
	(0.009)	(0.006)	(0.004)	(0.003)	(0.008)	(0.006)
GDP	0.986***	0.364***	0.383***	0.175***	0.785***	0.335***
	(0.065)	(0.057)	(0.032)	(0.023)	(0.062)	(0.055)
Foreign-Dist	0.156***	0.174***	0.059***	0.069***	0.141***	0.156***
	(0.025)	(0.025)	(0.011)	(0.009)	(0.025)	(0.024)
Remoteness	-0.792***	-0.273***	-0.305***	-0.138***	-0.597***	-0.247***
	(0.064)	(0.057)	(0.032)	(0.023)	(0.061)	(0.055)
Rail-Density	-0.170***	-0.214***	-0.023	-0.061***	-0.074	-0.142***
	(0.039)	(0.034)	(0.019)	(0.016)	(0.039)	(0.033)
Road-Density	0.079*	0.168***	0.026	0.037**	-0.069*	-0.014
	(0.033)	(0.026)	(0.017)	(0.013)	(0.032)	(0.026)
Pop	-0.107***	-0.092***	-0.006	-0.006	-0.086***	-0.071***
	(0.025)	(0.013)	(0.012)	(0.006)	(0.023)	(0.012)
State-share	-0.316***	-0.018	-0.009	0.026	-0.100	-0.012
	(0.056)	(0.046)	(0.028)	(0.021)	(0.054)	(0.044)
Collective-share	-0.082*	-0.009	-0.027	0.012	-0.036	-0.001
	(0.036)	(0.028)	(0.019)	(0.014)	(0.036)	(0.028)
Foreign-share	0.689***	0.026	0.179***	0.001	0.716***	0.014
	(0.022)	(0.020)	(0.011)	(0.010)	(0.020)	(0.019)
Age	-0.001**	-0.001	0.000	-0.000**	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.671***	0.523***	0.171***	0.149***	-0.039***	0.033**
	(0.009)	(0.011)	(0.004)	(0.005)	(0.008)	(0.011)
Homo	0.016	-0.058***	-0.059***	-0.078***	0.038*	-0.051***
	(0.017)	(0.011)	(0.012)	(0.006)	(0.016)	(0.010)
Observation	188049	188049	188049	188049	188049	188049
Group	32337	80774	32337	80774	32337	80774
Fixed Effects:						
Region	county	county	county	county	county	county
Industry	4-digit	4-digit	4-digit	4-digit	4-digit	4-digit
Firm		Yes		Yes		Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table J4: Channel Study: The Impacts of Implied Trade Cost on Export Value.

Dependent Variable: Export Value	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
DGC	-0.017*	-0.013	-0.013	-0.013	-0.427***	-0.451***	-0.451***	-0.451***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.074)	(0.078)	(0.078)	(0.078)
Foreign-Dist	0.178***	0.185***	0.185***	0.185***				
	(0.024)	(0.025)	(0.025)	(0.025)				
Language	-0.076	-0.052	-0.052	-0.052				
	(0.042)	(0.045)	(0.045)	(0.045)				
Border	-0.107**	-0.109*	-0.109*	-0.109*				
	(0.041)	(0.044)	(0.044)	(0.044)				
DGC*DGC					-0.008**	-0.007*	-0.007*	-0.007*
					(0.003)	(0.003)	(0.003)	(0.003)
DGC*Foreign-Dist					0.032***	0.033***	0.033***	0.033***
					(0.004)	(0.004)	(0.004)	(0.004)
DGC*Language					-0.010	-0.006	-0.006	-0.006
					(0.007)	(0.008)	(0.008)	(0.008)
DGC*Border					-0.015*	-0.016*	-0.016*	-0.016*
					(0.007)	(0.008)	(0.008)	(0.008)
TFP-ACF	0.135***	0.138***	0.138***	0.138***	0.135***	0.138***	0.138***	0.138***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
GDP	0.326***	0.314***	0.314***	0.314***	0.345***	0.330***	0.330***	0.330***
	(0.056)	(0.059)	(0.059)	(0.059)	(0.055)	(0.058)	(0.058)	(0.058)
Remoteness	-0.247***	-0.236***	-0.236***	-0.236***	-0.261***	-0.249***	-0.249***	-0.249***
	(0.055)	(0.058)	(0.058)	(0.058)	(0.055)	(0.058)	(0.058)	(0.058)
Rail-Density	-0.242***	-0.236***	-0.236***	-0.236***	-0.249***	-0.244***	-0.244***	-0.244***
	(0.030)	(0.034)	(0.034)	(0.034)	(0.031)	(0.034)	(0.034)	(0.034)
Road-Density	0.219***	0.177***	0.177***	0.177***	0.222***	0.178***	0.178***	0.178***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Pop	-0.164***	-0.119***	-0.119***	-0.119***	-0.169***	-0.122***	-0.122***	-0.122***
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.014)	(0.014)	(0.014)
State-share	-0.022	-0.021	-0.021	-0.021	-0.019	-0.019	-0.019	-0.019
	(0.043)	(0.046)	(0.046)	(0.046)	(0.043)	(0.046)	(0.046)	(0.046)
Collective-share	0.006	-0.013	-0.013	-0.013	0.005	-0.014	-0.014	-0.014
	(0.027)	(0.028)	(0.028)	(0.028)	(0.027)	(0.028)	(0.028)	(0.028)
Foreign-share	0.013	0.024	0.024	0.024	0.013	0.024	0.024	0.024
	(0.019)	(0.020)	(0.020)	(0.020)	(0.019)	(0.020)	(0.020)	(0.020)
Age	-0.001*	-0.001*	-0.001*	-0.001*	-0.001*	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.553***	0.527***	0.527***	0.527***	0.553***	0.527***	0.527***	0.527***
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)
Homo	-0.065***	-0.059***	-0.059***	-0.059***	-0.065***	-0.059***	-0.059***	-0.059***
	(0.010)	(0.011)	(0.011)	(0.011)	(0.010)	(0.011)	(0.011)	(0.011)
R2	0.153	0.138	0.138	0.138	0.154	0.138	0.138	0.138
Observation	189868	188049	188049	188049	189868	188049	188049	188049
Group	67586	80774	80774	80774	67586	80774	80774	80774
F	541.1	446.5	446.5	446.5	517.2	426.9	426.9	426.9
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry		4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Columns (1) and (5) control firm and year fixed effects, columns (2) and (6) control 4-digit industry, province, and year fixed effects; columns (3) and (7) control 4-digit industry, prefecture, and year fixed effects; columns (4) and (8) control 4-digit industry, county, and year fixed effects. The cluster standard error is always applied in each column, according their fixed-effect groups. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table J5: Channel Study: The Impacts of Implied Trade Cost on Export Scope.

Dependent Variable: Export Variety	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
DGC	-0.008*	-0.007	-0.007	-0.008	-0.261***	-0.265***	-0.267***	-0.268***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.030)	(0.032)	(0.032)	(0.032)
Foreign-Dist	0.073***	0.074***	0.074***	0.074***				
	(0.009)	(0.009)	(0.009)	(0.009)				
Language	0.004	0.000	0.000	0.007				
	(0.016)	(0.017)	(0.017)	(0.017)				
Border	-0.085***	-0.083***	-0.083***	-0.089***				
	(0.016)	(0.017)	(0.017)	(0.017)				
DGC*DGC					0.003*	0.004**	0.004**	0.004**
					(0.001)	(0.002)	(0.002)	(0.002)
DGC*Foreign-Dist					0.013***	0.013***	0.013***	0.013***
					(0.001)	(0.002)	(0.002)	(0.002)
DGC*Language					0.003	0.001	0.001	0.003
					(0.003)	(0.003)	(0.003)	(0.003)
DGC*Border					-0.014***	-0.014***	-0.014***	-0.015***
					(0.003)	(0.003)	(0.003)	(0.003)
TFP-ACF	0.014***	0.017***	0.017***	0.017***	0.014***	0.017***	0.017***	0.017***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
GDP	0.158***	0.150***	0.150***	0.154***	0.167***	0.157***	0.157***	0.162***
	(0.023)	(0.024)	(0.024)	(0.024)	(0.023)	(0.024)	(0.024)	(0.024)
Remoteness	-0.126***	-0.118***	-0.118***	-0.123***	-0.133***	-0.124***	-0.123***	-0.129***
	(0.023)	(0.024)	(0.024)	(0.024)	(0.023)	(0.024)	(0.024)	(0.024)
Rail-Density	-0.066***	-0.066***	-0.066***	-0.067***	-0.063***	-0.062***	-0.061***	-0.063***
	(0.014)	(0.016)	(0.016)	(0.016)	(0.015)	(0.016)	(0.016)	(0.016)
Road-Density	0.061***	0.044***	0.044***	0.040**	0.060***	0.043**	0.043**	0.039**
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Pop	-0.022***	-0.012*	-0.013*	-0.014*	-0.019**	-0.010	-0.010	-0.012
	(0.007)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)
State-share	0.028	0.025	0.025	0.025	0.028	0.025	0.025	0.025
	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)	(0.021)	(0.021)	(0.021)
Collective-share	0.009	0.009	0.009	0.011	0.009	0.010	0.009	0.011
	(0.013)	(0.014)	(0.014)	(0.014)	(0.013)	(0.014)	(0.014)	(0.014)
Foreign-share	-0.000	0.000	0.001	0.001	-0.000	0.000	0.001	0.001
	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)
Age	-0.000*	-0.000**	-0.000**	-0.000**	-0.000*	-0.000**	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.156***	0.150***	0.150***	0.149***	0.156***	0.150***	0.150***	0.149***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Homo	-0.083***	-0.078***	-0.078***	-0.078***	-0.083***	-0.078***	-0.078***	-0.078***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
R2	0.0608	0.0529	0.0530	0.0527	0.0609	0.0531	0.0531	0.0528
Observation	189868	188049	188049	188049	189868	188049	188049	188049
Group	67586	78957	78979	80774	67586	78957	78979	80774
F	194.9	160.1	160.1	157.4	186.8	153.3	153.4	150.9
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry		4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Columns (1) and (5) control firm and year fixed effects, columns (2) and (6) control 4-digit industry, province, and year fixed effects; columns (3) and (7) control 4-digit industry, prefecture, and year fixed effects; columns (4) and (8) control 4-digit industry, county, and year fixed effects. The cluster standard error is always applied in each column, according their fixed-effect groups. Significant level: * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table J6: The Impacts on Process Share in Term of Export Value.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect			
Processing Share	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Main Results:							
DGC	0.105***	0.039*	0.010	-0.009	-0.033**	-0.033**	-0.030*
	(0.020)	(0.018)	(0.021)	(0.011)	(0.011)	(0.011)	(0.012)
TFP-ACF	-0.237***	-0.211***	-0.170***	-0.029***	-0.025***	-0.025***	-0.025***
	(0.015)	(0.013)	(0.012)	(0.006)	(0.006)	(0.006)	(0.006)
GDP	-0.488***	-0.443***	-0.323**	-0.125	-0.107	-0.106	-0.074
	(0.125)	(0.115)	(0.105)	(0.066)	(0.067)	(0.067)	(0.066)
Foreign-Dist	0.037	0.075*	0.096**	-0.011	-0.003	-0.004	-0.001
	(0.040)	(0.037)	(0.037)	(0.024)	(0.025)	(0.025)	(0.025)
Remoteness	0.459***	0.404***	0.285**	0.107	0.089	0.088	0.055
	(0.123)	(0.116)	(0.105)	(0.067)	(0.068)	(0.068)	(0.067)
Rail-Density	0.028	-0.009	-0.025	-0.031	-0.016	-0.017	-0.016
	(0.055)	(0.050)	(0.045)	(0.032)	(0.033)	(0.033)	(0.033)
Road-Density	-0.215***	-0.143**	-0.074	0.022	0.027	0.028	0.019
	(0.065)	(0.052)	(0.050)	(0.038)	(0.038)	(0.038)	(0.038)
Pop	0.079*	0.125***	0.092***	0.045***	0.030**	0.030**	0.033**
	(0.031)	(0.023)	(0.023)	(0.012)	(0.010)	(0.010)	(0.011)
State-share	-0.075	-0.095	-0.013	0.021	-0.004	-0.000	-0.039
	(0.093)	(0.089)	(0.090)	(0.078)	(0.083)	(0.083)	(0.077)
Collective-share	0.153**	0.144*	0.167**	-0.035	-0.028	-0.027	-0.043
	(0.057)	(0.056)	(0.054)	(0.043)	(0.044)	(0.044)	(0.044)
Foreign-share	0.514***	0.374***	0.337***	-0.008	0.001	0.001	0.001
	(0.039)	(0.033)	(0.033)	(0.019)	(0.020)	(0.020)	(0.020)
Age	0.003	0.003	0.003	-0.000	-0.000	-0.000	0.000
	(0.002)	(0.001)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Size	-0.064***	-0.063***	-0.080***	-0.029*	-0.032**	-0.032**	-0.034**
	(0.015)	(0.013)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)
Homo	0.019	0.025	0.036	0.010	0.008	0.008	0.010
	(0.061)	(0.037)	(0.034)	(0.014)	(0.014)	(0.014)	(0.014)
Observation	92963	91530	88935	85687	79614	79595	78818
Group	2995	6836	11329	23401	24862	24856	24914
Under-identification test	115.604	344.171	385.678	1817.523	1483.068	1480.780	1478.621
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	4.2e+04	4.3e+04	2.9e+04	2.9e+04	2.5e+04	2.5e+04	2.4e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	1.298	4.028	0.576	0.512	0.141	0.133	0.233
P value	(0.255)	(0.045)	(0.448)	(0.474)	(0.708)	(0.716)	(0.629)
First-stage Results:							
Ming-Road	0.350***	0.325***	0.351***	0.472***	0.482***	0.483***	0.478***
	(0.010)	(0.013)	(0.015)	(0.013)	(0.014)	(0.014)	(0.014)
Counter-factual-Road	0.443***	0.476***	0.453***	0.469***	0.464***	0.464***	0.471***
	(0.010)	(0.012)	(0.015)	(0.013)	(0.014)	(0.014)	(0.014)
F	1290.597	746.500	766.183	1567.594	1137.600	1135.452	1113.531
R2	0.615	0.598	0.549	0.589	0.560	0.559	0.561
Fixed Effects:							
Region	province	prefecture	county		province	prefecture	county
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm				Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table J7: The Impacts on Processing Export Decision.

Dependent Variable:	Without Firm Fixed Effect			With Firm Fixed Effect				
Export Decision	(1)	(2)	(3)	(4)	(5)	(6)	(7)	
Main Results:								
DGC	0.011*	0.002	0.005	0.004	0.005	0.005	0.006*	
	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	
Rail-Density	-0.043***	-0.051***	-0.038***	-0.006	-0.000	0.000	0.002	
	(0.010)	(0.010)	(0.010)	(0.008)	(0.008)	(0.008)	(0.009)	
Road-Density	0.031*	0.038***	0.037***	0.020**	0.020**	0.020**	0.020**	
	(0.013)	(0.009)	(0.008)	(0.007)	(0.007)	(0.007)	(0.007)	
TFP-ACF	-0.029***	-0.020***	-0.017***	-0.002	-0.003	-0.003*	-0.003	
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	
Pop	-0.036***	-0.004	-0.002	0.005	0.002	0.002	0.003	
	(0.008)	(0.005)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	
State-share	-0.017*	-0.012	-0.012	-0.002	-0.001	-0.001	-0.000	
	(0.009)	(0.008)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	
Collective-share	0.026***	0.031***	0.037***	0.003	0.005	0.005	0.004	
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	
Foreign-share	0.220***	0.168***	0.152***	0.001	0.005	0.005	0.005	
	(0.009)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)	
Age	0.000	0.000**	0.000**	0.000	-0.000	-0.000	-0.000	
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Size	0.014***	0.007***	0.004**	-0.003	-0.005*	-0.005*	-0.005	
	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	
Observation	194318	191681	186164	177720	166412	166383	164566	
Group	5320	13630	24242	50521	52901	52894	52944	
Under-identification test	211.158	518.302	593.974	2837.417	2374.376	2371.487	2326.982	
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Weak identification test	1.3e+05	1.1e+05	6.4e+04	6.2e+04	5.2e+04	5.2e+04	5.1e+04	
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Over-identification test	1.545	0.852	3.012	0.598	0.061	0.082	0.254	
P value	(0.214)	(0.356)	(0.083)	(0.439)	(0.804)	(0.774)	(0.614)	
First-stage Results:								
Qing-Road	0.508***			0.725***	0.680***	0.681***	0.689***	
	(0.020)			(0.016)	(0.017)	(0.017)	(0.017)	
Counter-factual-Road	0.247***	0.479***	0.455***					
	(0.023)	(0.012)	(0.014)					
Ming-Road		0.323***	0.350***	0.099***	0.147***	0.146***	0.134***	
		(0.013)	(0.014)	(0.020)	(0.022)	(0.022)	(0.022)	
F	1585.793	861.887	947.043	2560.868	1936.808	1936.281	1921.338	
R2	0.646	0.599	0.548	0.612	0.579	0.579	0.581	
Fixed Effects:								
Region	province	prefecture	county		province	prefecture	county	
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit	
Firm				Yes	Yes	Yes	Yes	
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table J8: The Heterogeneous Impacts on Export Volume across Textile and High-tech Sectors.

Dependent Variable: Export Volume	High-tech Sector (1)	(2)	(3)	(4)	Textile Sector (5)	(6)	(7)	(8)
Main Results:								
DGC	0.038	0.013	0.013	-0.003	0.014	0.027*	0.027*	0.027*
	(0.030)	(0.020)	(0.020)	(0.020)	(0.018)	(0.012)	(0.012)	(0.012)
TFP-ACF	0.070***	0.069***	0.069***	0.069***	0.049***	0.083***	0.083***	0.085***
	(0.018)	(0.010)	(0.010)	(0.010)	(0.010)	(0.007)	(0.007)	(0.007)
GDP	1.503***	0.484***	0.484***	0.484***	0.640***	0.451***	0.451***	0.443***
	(0.096)	(0.066)	(0.066)	(0.067)	(0.051)	(0.039)	(0.039)	(0.040)
Foreign-Dist	0.463***	0.215***	0.215***	0.210***	0.291***	0.186***	0.186***	0.184***
	(0.042)	(0.024)	(0.024)	(0.025)	(0.019)	(0.013)	(0.013)	(0.013)
Remoteness	-1.112***	-0.298***	-0.299***	-0.299***	-0.460***	-0.358***	-0.358***	-0.349***
	(0.096)	(0.066)	(0.066)	(0.067)	(0.051)	(0.040)	(0.040)	(0.040)
Rail-Density	0.021	-0.103	-0.103	-0.116	-0.068	-0.218***	-0.216***	-0.213***
	(0.093)	(0.063)	(0.063)	(0.064)	(0.050)	(0.032)	(0.032)	(0.033)
Road-Density	0.191*	0.248***	0.248***	0.265***	0.083	0.099***	0.092**	0.081**
	(0.092)	(0.050)	(0.050)	(0.050)	(0.043)	(0.030)	(0.030)	(0.030)
Pop	-0.032	-0.051**	-0.051**	-0.057**	-0.120***	-0.082***	-0.080***	-0.083***
	(0.033)	(0.019)	(0.019)	(0.019)	(0.029)	(0.018)	(0.018)	(0.018)
State-share	-0.117	-0.076	-0.076	-0.063	0.117	0.160**	0.159**	0.165**
	(0.104)	(0.069)	(0.069)	(0.071)	(0.067)	(0.050)	(0.050)	(0.051)
Collective-share	0.063	0.063	0.063	0.040	0.011	0.022	0.022	0.020
	(0.100)	(0.052)	(0.052)	(0.052)	(0.049)	(0.029)	(0.029)	(0.029)
Foreign-share	0.069	0.148***	0.148***	0.153***	0.031	0.029	0.029	0.027
	(0.052)	(0.032)	(0.032)	(0.032)	(0.030)	(0.021)	(0.021)	(0.021)
Age	-0.001	-0.000	-0.000	-0.000	-0.001	-0.001	-0.001	-0.001
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Size	0.179***	0.375***	0.375***	0.364***	0.089***	0.230***	0.229***	0.227***
	(0.029)	(0.018)	(0.018)	(0.018)	(0.016)	(0.011)	(0.011)	(0.011)
Homo	-0.191***	0.000	0.000	0.000	0.203***	0.000	0.000	0.000
	(0.033)	(.)	(.)	(.)	(0.035)	(.)	(.)	(.)
Observation	227771	147180	147171	145129	416802	278420	278278	274851
Group	20211	53534	53530	53227	17099	97436	97406	97017
Under-identification test	492.632	2096.530	2096.333	2031.176	355.990	5343.266	5325.812	5187.120
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	1.0e+05	4.4e+04	4.4e+04	4.3e+04	2.1e+05	9.3e+04	9.4e+04	9.2e+04
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.591	1.580	1.536	0.595	0.280	0.003	0.004	0.008
P value	(0.442)	(0.209)	(0.215)	(0.441)	(0.597)	(0.954)	(0.953)	(0.928)
First-stage Results:								
Ming-Road	-0.035	0.038	0.037	0.030	-0.021	-0.010	-0.014	-0.032
	(0.046)	(0.020)	(0.020)	(0.020)	(0.070)	(0.019)	(0.019)	(0.019)
Qing-Road	0.832***	0.779***	0.779***	0.787***	0.811***	0.782***	0.788***	0.800***
	(0.034)	(0.015)	(0.015)	(0.015)	(0.056)	(0.015)	(0.015)	(0.014)
F	305.283	1353.082	1353.121	1336.426	356.462	3602.463	3622.361	3625.949
R2	0.614	0.562	0.562	0.566	0.633	0.617	0.617	0.620
Fixed Effects:								
Region		province	prefecture	county		province	prefecture	county
Industry		4-digit	4-digit	4-digit		4-digit	4-digit	4-digit
Firm	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Table J9: The Heterogeneous Impacts on Export Decision across Textile and High-tech Sectors.

Dependent Variable:	High-tech Sector				Textile Sector			
Entry Decision	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Results:								
DGC	-0.002	-0.000	-0.001	-0.002*	-0.007***	-0.005*	0.002	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
Rail-Density	0.009**	0.007**	0.006**	0.002	-0.007	-0.004	-0.003	-0.003*
	(0.003)	(0.002)	(0.002)	(0.002)	(0.004)	(0.003)	(0.002)	(0.002)
Road-Density	0.002	0.001	-0.001	-0.000	0.005*	0.003	0.001	0.004**
	(0.003)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.002)	(0.001)
TFP-ACF	-0.002**	-0.002***	-0.002***	-0.000	-0.004***	-0.004***	-0.004***	0.000
	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.000)
Pop	0.002	-0.004	-0.004	0.001	-0.002	-0.009	-0.005	0.001
	(0.001)	(0.004)	(0.004)	(0.002)	(0.002)	(0.004)	(0.004)	(0.001)
State-share	-0.011***	-0.010***	-0.010***	0.001	0.001	0.000	-0.002	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Collective-share	-0.012***	-0.011***	-0.011***	0.002**	-0.002	-0.001	-0.003*	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Foreign-share	0.172***	0.161***	0.148***	0.004*	0.123***	0.123***	0.125***	0.002
	(0.010)	(0.005)	(0.004)	(0.002)	(0.016)	(0.008)	(0.006)	(0.002)
Age	-0.000*	-0.000**	-0.000	0.000***	0.000**	0.000***	0.000**	0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Size	0.022***	0.023***	0.025***	0.012***	0.015***	0.016***	0.019***	0.009***
	(0.001)	(0.001)	(0.001)	(0.000)	(0.002)	(0.001)	(0.001)	(0.000)
R2	0.0902	0.0766	0.0660	0.00655	0.0539	0.0502	0.0491	0.00354
Observation	1031690	1025018	1002571	928114	1031690	1025018	1002571	928114
Group	9867	42732	100925	261537	9867	42732	100925	261537
F	50.84	109.9	172.5	119.1	10.75	28.55	58.77	62.87
Under-identification test	387.226	1009.079	1076.551	6721.549	387.226	999.046	1064.764	6834.871
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Weak identification test	7.2e+05	6.0e+05	3.7e+05	2.3e+05	7.2e+05	5.8e+05	3.5e+05	2.2e+05
P value	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Over-identification test	0.846	2.705	1.048	0.003	3.677	0.320	0.686	2.637
P value	(0.358)	(0.100)	(0.306)	(0.958)	(0.055)	(0.572)	(0.407)	(0.104)
First-stage Results:								
Counter-factual-Road	0.478***	0.377***	0.306***	0.332***	0.478***	0.494***	0.452***	0.475***
	(0.007)	(0.009)	(0.011)	(0.007)	(0.007)	(0.008)	(0.010)	(0.007)
Ming-Road	0.347***				0.347***	0.353***	0.403***	0.476***
	(0.008)				(0.008)	(0.008)	(0.010)	(0.007)
Terrain-Surface		0.432***	0.516***	0.549***				
		(0.010)	(0.010)	(0.006)				
F	2432.217	1864.512	2320.952	8651.339	2432.217	1388.425	1688.516	7542.666
R2	0.650	0.597	0.537	0.526	0.650	0.591	0.523	0.512
Region	province	prefecture	county		province	prefecture	county	
Industry	4-digit	4-digit	4-digit		4-digit	4-digit	4-digit	
Firm				Yes				Yes
Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	

Note: The Under-identification test reports Kleibergen-Paap Wald rk F statistic and its p-value, with null hypothesis that the equation is underidentified; the Weak identification test reports Cragg-Donald F statistic and its p-value, with null hypothesis that the equation is weakly identified by the instruments; the Over-identification test reports Sargan-Hansen statistic and its p-value, with null hypothesis that the instruments are uncorrelated with error term, i.e., valid instruments. Significant level: * p<0.05, ** p<0.01, *** p<0.001.

Chapter 6

Conclusion and Discussion

On the basis of the existing economic geographical literature, this dissertation provides more details to explain the spillover effects of Chinese highway expansion. Existing studies on China do not always support that there are positive effects of infrastructure improvement on economic growth, but more significant relocation effects (Faber 2014, Baum-Snow et al. 2017); while studies for western developed countries tend to conclude that road construction has positive effects on economic efficiency (Garcia-Mila et al. 1996, Holl 2016, Gibbons et al. 2016). Faber (2014) finds the improvement of transportation infrastructure opens up competition between rich and rural regions, it promotes population relocation and the development of rich regions, but also accelerates the outflow of population in rural regions; similarly, (Baum-Snow et al. 2017) can also find road and railroad construction promote population decentralization, from city central areas toward suburban areas.

This study investigates the impacts of highway construction on firm-level productivity. To address potential endogeneity issues, we construct IVs based on historical roads and counterfactual roads approaches. The baseline regressions show that firms closer to newly constructed highways have productivity premiums, a 10% decrease of distance to highways can increase firm-level productivity by 0.2%-0.3%. The decrease of firms' distance to highways is induced by two mechanisms: on the one side, highway development can reduce $Dist$ of address-unchanged firms; on the other side, highway development can also promote firms' entry and relocation, then reduce the overall $Dist$. Highways have lagged impacts on productivity growth, if we replace $Dist$ as one-year-lagged $Dist$, a 10% decrease of lagged $Dist$ is related to a 0.8% or 1.0% increase of firm-level productivity. These results are robust across different types of transport modes (road, railway, waterway), different productivity measurements (OLS, OP, LP, and ACF productivity), and distance measurements (entrance distances). Channel studies show that firms closer to highways have higher inventory level, while higher inventory level is related to larger firm size and higher productivity; by contrast, firms closer to highways have higher outsourcing level, while higher outsourcing level is related to smaller firm size and lower productivity, but their coefficients are smaller. The highways' impacts on inventory and outsourcing activities seem to conflict with each other, considering the very large and significant coefficients of firm size, a possible explanation is that the firm size increase is much faster even if the outsourcing promotion effect can reduce the average firm size. The increase of firm size is accompanied with productivity increases because large firms can normally reap more benefits, a potential consequence is that firm size dispersion will be increased during this process. This firm size mechanism is further studied in the next chapter. Heterogeneity studies show that highway construction effects in coastal provinces are stronger than in inland provinces, firms' productivity in coastal regions shows a more significant increase after highway construction than inland regions.

This second topic finds that highway expansion tends to increase firm size, but this impact varies from large to small firms, i.e., most large and medium enterprises tend to expand their scale when traffic accessibility increases, especially for large firms; a few small firms located at the far left of the size distribution tend to decrease their scale when traffic accessibility increase, implying large enterprises tend to get more benefits form highway expansion. Due to the heterogeneous growth of firm scales, the overall size dispersion tends to increase with traffic accessibility increase, suggesting that road expansion generates very significant industrial agglomeration effects. A 10% decrease of firms' distance to high-class highways can increase the firm size dispersion by 0.4% to 1.3%, increase the market concentration by 1.4% to 2.0%. The increase of size dispersion and market concentration are motivated by the rapid expansion of large firms, outsourcing activities, and the establishment of new and small firms. Channel studies show that firms closer to expressways are more likely to outsource their intermediate inputs and have higher inventory levels, but new entry firms tend to choose addresses around a certain brand of distance from highways, i.e., not very close and not very far from highways. Higher inventory levels are correlated with larger firm size, while new entry and outsourcing firms tend to be smaller. These results are consistent with Cabral & Mata (2003), Huber et al. (2013), Alfaro & Chari (2014), they find that deregulation and trade openness policies tend to reduce transaction costs, then induce the market share of medium-size firms to be replace by large firms and new entry firms, so the size dispersion tends to increase.

The first and second topics reveal that firms will be attracted to move or establish around new constructed expressways, consistent with the findings of Faber (2014). At the same time, new entry firms tend to choose the addresses around a certain brand of distance from highways, those addresses very close to highways are not attractive for new entry firms. At the same time, even for those regions that are linked by new expressways, it is not certain that they will benefit from infrastructure development. Due to the increase of traffic accessibility encouraging outsourcing of activities, and upstream- and downstream-linkages, the comparative advantages of industrial clusters will become much more significant than in the past. However, the nature of agglomeration economies induces industrial clusters to emerge in those regions with the best endowments of resources and factors, then the concentration of population and economic activities will widen the gaps between rich and rural regions. During this process, large firms tend to benefit more from infrastructure development and transaction cost decline, which presents an issue for policy makers, i.e., to regulate the market appropriately, and make sure the decline of transaction costs will encourage productivity-motivated expansion rather than the expansion fueled by market influence.

The third topic shows that, from 2000 to 2006, firms closer to seaports (or coastal firms) are more likely to involve in international trade, having higher firm-level export values and higher exports-sales ratios. When we consider the export intensive and extensive margins, coastal firms normally export more types of different products, but focus on a smaller number of destinations, these results stay robust even if the proxy of transportation cost is substituted by the proxies used by Huang & Xu (2012). Consequently, firms closer to seaports tend to have a lower export intensive margin, i.e., export volume of products, export volume of products in each destination. This fact is also partly consistent with the prediction of Krugman-style love-of-variety models rather than with Armington models, i.e., if an economy doubles its export value with its economic growth, its export extensive margin is expected to increase faster than its intensive margin (Hummels & Klenow 2005). Second, our data also show that new exporters tend to emerge in inland regions. When we control the firm-region-industry fixed effects, the results suggest that highway expansion can promote the export value and export intensive margin of those firms not very close to seaports. A possible explanation is that those potential exporters close to seaports have already started exporting, while the new exporters in inland regions normally have smaller export scope and scales. Third, new entry processing exporters tend to choose addresses not very close to seaports. A possible explanation is that those potential processing exporters close to seaports have already started processing exporting because of its lower entry barriers than ordinary exports. When we strictly control firm-industry-region time-invariant heterogeneity, the decrease of travel cost to seaports can reduce firms' processing share in export value, suggesting that highway expansion can induce the processing exporters to transform into ordinary or other exporters. This explanation is also consistent with the study of Egger & Falkinger (2003), they show that more efficient road networks can lower the trade barriers and promote trade structure transition from processing-oriented modes to more diversified structures. Forth, high-tech industries expanded much more significantly than other sectors and dominated China's export products (more than 50% export value) at the end of our sample period. At the same time, the regressions results show that high-tech industries are not sensitive to highway expansion, highways' impacts on textile industries are stronger than other industries.

These facts confirm that China's infrastructure expansion is accompanied with a significant increase of economic complexity, which is mainly motivated by transportation cost reduction, the entry of new exporters, and population agglomeration effect. Cross-country evidence shows that big countries tend to have more diversified economies and higher levels of economic complexity; they not only have larger export and import value, but also have higher demand and production capacity of larger categories of heterogeneous products (extensive margin), compared with smaller economies. Similarly, when firms produce heterogeneous goods, each type of goods faces less competition than homogeneous goods.On the other side, the production of complex goods proposes higher requirements on quality and variety of intermediate inputs, so these firms face higher technological frontiers than homogeneous producers. The decline of transportation costs induced by highway construction can strengthen the connection across up- and down-stream industries, encourage incumbent or new entry firms to produce differentiated goods, then motivate the export market to evolve toward a more diversified structure, accompanied with increased economic complexity.

Bibliography

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

Ackerberg, D., Benkard, C. L., Berry, S. & Pakes, A. (2007), 'Econometric tools for analyzing market outcomes', *Handbook of econometrics* **6**, 4171–4276.

Alfaro, L. & Chari, A. (2014), 'Deregulation, misallocation, and size: Evidence from india', *The Journal of Law and Economics* **57**(4), 897–936.

Alonso, W. et al. (1964), 'Location and land use'.

Amador, J. & Opromolla, L. D. (2013), 'Product and destination mix in export markets', *Review of World Economics* **149**(1), 23–53.

Amiti, M. & Freund, C. (2008), *The anatomy of China's export growth*, The World Bank.

Amiti, M. & Konings, J. (2007), 'Trade liberalization, intermediate inputs, and productivity: Evidence from indonesia', *American Economic Review* **97**(5), 1611–1638.

Anderson, J. E. & Van Wincoop, E. (2003), 'Gravity with gravitas: a solution to the border puzzle', *American economic review* **93**(1), 170–192.

Anderson, J. E. & Van Wincoop, E. (2004), 'Trade costs', *Journal of Economic literature* **42**(3), 691–751.

Angelini, P. & Generale, A. (2008), 'On the evolution of firm size distributions', *American Economic Review* **98**(1), 426–38.

Arrow, K. J. (1971), The economic implications of learning by doing, *in* 'Readings in the Theory of Growth', Springer, pp. 131–149.

Asturias, J., García-Santana, M. & Ramos, R. (2019), 'Competition and the welfare gains from transportation infrastructure: Evidence from the golden quadrilateral of india', *Journal of the European Economic Association* **17**(6), 1881–1940.

Aw, B. Y., Chung, S. & Roberts, M. J. (2000), ‘Productivity and turnover in the export market: micro-level evidence from the republic of korea and taiwan (china)’, *The World Bank Economic Review* **14**(1), 65–90.

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

Baldwin, R. & Harrigan, J. (2011), ‘Zeros, quality, and space: Trade theory and trade evidence’, *American Economic Journal: Microeconomics* **3**(2), 60–88.

Banerjee, A., Duflo, E. & Qian, N. (2012), On the road: Access to transportation infrastructure and economic growth in china, Technical report, National Bureau of Economic Research.

Bartelsman, E. J., Gautier, P. A. & De Wind, J. (2011), ‘Employment protection, technology choice, and worker allocation’.

Bartelsman, E. J., Haskel, J. & Martin, R. (2008), ‘Distance to which frontier? evidence on productivity convergence from international firm-level data’.

Baum, C. F., Dong, Y., Lewbel, A., Yang, T. et al. (2012), Binary choice models with endogenous regressors, *in* ‘Stata conference’.

Baum-Snow, N. (2007), ‘Did highways cause suburbanization?’, *The Quarterly Journal of Economics* **122**(2), 775–805.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A. & Zhang, Q. (2012), Roads, railroads and decentralization of chinese cities, Technical report, Working paper.

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.

Bee, M., Riccaboni, M. & Schiavo, S. (2017), ‘Where gibrat meets zipf: Scale and scope of french firms’, *Physica A: Statistical Mechanics and its Applications* **481**, 265–275.

Behar, A. & Venables, A. J. (2011), ‘Transport costs and international trade’, *Handbook of transport economics* pp. 97–115.

Benkard, C. L. (1999), Learning and forgetting: The dynamics of aircraft production, Technical report, National bureau of economic research.

Bernard, A. B., Dhyne, E., Magerman, G., Manova, K. & Moxnes, A. (2019), The origins of firm heterogeneity: a production network approach, Technical report, National Bureau of Economic Research.

Bernard, A. B. & Jensen, J. B. (1999), 'Exceptional exporter performance: cause, effect, or both?', *Journal of international economics* **47**(1), 1–25.

Bernard, A. B., Jensen, J. B., Redding, S. J. & Schott, P. K. (2012), 'The empirics of firm heterogeneity and international trade', *Annu. Rev. Econ.* **4**(1), 283–313.

Bernard, A. B., Jensen, J. B. & Schott, P. K. (2005), Importers, exporters, and multinationals: A portrait of firms in the us that trade goods, Technical report, National Bureau of Economic Research.

Bernard, A. B., Redding, S. J. & Schott, P. K. (2009), 'Products and productivity', *The Scandinavian Journal of Economics* **111**(4), 681–709.

Bernard, A. B., Redding, S. J. & Schott, P. K. (2010), 'Multiple-product firms and product switching', *The American Economic Review* **100**(1), 70–97.

Bernard, A. B., Redding, S. J. & Schott, P. K. (2011), 'Multiproduct firms and trade liberalization', *The Quarterly journal of economics* **126**(3), 1271–1318.

Bernard, A. B. & Wagner, J. (1997), 'Exports and success in german manufacturing', *Weltwirtschaftliches Archiv* **133**(1), 134–157.

Bertrand, M. & Schoar, A. (2003), 'Managing with style: The effect of managers on firm policies', *The Quarterly Journal of Economics* **118**(4), 1169–1208.

Bloom, N., Draca, M. & Van Reenen, J. (2009), 'Trade induced technical change? the impact of chinese imports on innovation and information technology', *CEP Discussion Papers* (1000).

Bloom, N., Genakos, C., Sadun, R. & Van Reenen, J. (2009), 'Does management matter? new empirics and old theories', *Unpublished manuscript, London School Econ./Stanford Univ* .

Bloom, N., Schankerman, M. & Van Reenen, J. (2013), 'Identifying technology spillovers and product market rivalry', *Econometrica* **81**(4), 1347–1393.

Bloom, N. & Van Reenen, J. (2007), 'Measuring and explaining management practices across firms and countries forthcoming quarterly journal of economics'.

Bloom, N. & Van Reenen, J. (2010), 'Why do management practices differ across firms and countries?', *The Journal of Economic Perspectives* **24**(1), 203–224.

Blundell, R. & Bond, S. (2000), 'Gmm estimation with persistent panel data: an application to production functions', *Econometric reviews* **19**(3), 321–340.

Bonfatti, R. & Poelhekke, S. (2017), 'From mine to coast: Transport infrastructure and the direction of trade in developing countries', *Journal of Development Economics* **127**, 91–108.

Bougheas, S., Demetriades, P. O. & Mamuneas, T. P. (2000), 'Infrastructure, specialization, and economic growth', *Canadian Journal of Economics/Revue canadienne d'économique* **33**(2), 506–522.

Braguinsky, S., Branstetter, L. G. & Regateiro, A. (2011), The incredible shrinking portuguese firm, Technical report, National Bureau of Economic Research.

Brandt, L., Van Biesebroeck, J. & Zhang, Y. (2012), 'Creative accounting or creative destruction? firm-level productivity growth in chinese manufacturing', *Journal of development economics* **97**(2), 339–351.

Broxton, P. D., Zeng, X., Sulla-Menashe, D. & Troch, P. A. (2014), 'A global land cover climatology using modis data', *Journal of Applied Meteorology and Climatology* **53**(6), 1593–1605.

Brun, J.-F., Carrère, C., Guillaumont, P. & De Melo, J. (2005), 'Has distance died? evidence from a panel gravity model', *The World Bank Economic Review* **19**(1), 99–120.

Cabral, L. & Mata, J. (2003), 'On the evolution of the firm size distribution: Facts and theory', *American economic review* **93**(4), 1075–1090.

Cavallo, E. & Daude, C. (2011), 'Public investment in developing countries: A blessing or a curse?', *Journal of Comparative Economics* **39**(1), 65–81.

Cavallo, E., Galiani, S., Noy, I. & Pantano, J. (2013), 'Catastrophic natural disasters and economic growth', *Review of Economics and Statistics* **95**(5), 1549–1561.

Chandler, A. D., Hikino, T. & Chandler, A. D. (2009), *Scale and scope: The dynamics of industrial capitalism*, Harvard University Press.

Chaney, T. (2008), 'Distorted gravity: the intensive and extensive margins of international trade', *American Economic Review* **98**(4), 1707–21.

Chen, N. & Novy, D. (2018), 'Currency unions, trade, and heterogeneity'.

Chuan, L. (2007), 'Infrastructure development in china', *Economic Research Project Report* (2).

Ciccone, A. (2002), 'Agglomeration effects in europe', *European economic review* **46**(2), 213–227.

Ciccone, A. & Hall, R. E. (1996), Productivity and the density of economic activity, Technical report.

Clerides, S. K., Lach, S. & Tybout, J. R. (1998), 'Is learning by exporting important? micro-dynamic evidence from colombia, mexico, and morocco', *The quarterly journal of economics* **113**(3), 903–947.

Coles, M. G. & Mortensen, D. T. (2016), 'Equilibrium labor turnover, firm growth, and unemployment', *Econometrica* **84**(1), 347–363.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D. & Roux, S. (2012), 'The productivity advantages of large cities: Distinguishing agglomeration from firm selection', *Econometrica* **80**(6), 2543–2594.

Coşar, A. K. & Demir, B. (2016), 'Domestic road infrastructure and international trade: Evidence from turkey', *Journal of Development Economics* **118**, 232–244.

Costa-Campi, M. T. & Viladecans-Marsal, E. (1999), 'The district effect and the competitiveness of manufacturing companies in local productive systems', *Urban Studies* **36**(12), 2085–2098.

Crespi, G., Criscuolo, C., Haskel, J. E. & Slaughter, M. (2008), Productivity growth, knowledge flows, and spillovers, Technical report, National Bureau of Economic Research.

Cummins, J. G. & Violante, G. L. (2002), 'Investment-specific technical change in the united states (1947–2000): Measurement and macroeconomic consequences', *Review of Economic dynamics* **5**(2), 243–284.

Dai, M., Maitra, M. & Yu, M. (2016), 'Unexceptional exporter performance in china? the role of processing trade', *Journal of Development Economics* **121**, 177–189.

Datta, S. (2012), 'The impact of improved highways on indian firms', *Journal of Development Economics* **99**(1), 46–57.

De Loecker, J. (2007*a*), 'Do exports generate higher productivity? evidence from slovenia', *Journal of international economics* **73**(1), 69–98.

De Loecker, J. (2007*b*), 'Product differentiation, multi-product firms and estimating the impact of trade liberalization on productivity'.

De Loecker, J., Goldberg, P. K., Khandelwal, A. K. & Pavcnik, N. (2016), 'Prices, markups, and trade reform', *Econometrica* **84**(2), 445–510.

De Loecker, J. & Warzynski, F. (2012), 'Markups and firm-level export status', *American economic review* **102**(6), 2437–71.

Debaere, P. & Mostashari, S. (2010), 'Do tariffs matter for the extensive margin of international trade? an empirical analysis', *Journal of International Economics* **81**(2), 163–169.

Del Gatto, M., Ottaviano, G. I. & Pagnini, M. (2008), 'Openness to trade and industry cost dispersion: Evidence from a panel of italian firms', *Journal of Regional Science* **48**(1), 97–129.

Di Giovanni, J., Levchenko, A. A. & Mejean, I. (2014), 'Firms, destinations, and aggregate fluctuations', *Econometrica* **82**(4), 1303–1340.

Ding, S., Guariglia, A. & Harris, R. (2016), 'The determinants of productivity in chinese large and medium-sized industrial firms, 1998–2007', *Journal of Productivity Analysis* **45**(2), 131–155.

Ding, S., Guariglia, A. & Knight, J. (2013), 'Investment and financing constraints in china: does working capital management make a difference?', *Journal of Banking & Finance* **37**(5), 1490–1507.

Ding, S., Jiang, W. & Sun, P. (2016), 'Import competition, dynamic resource allocation and productivity dispersion: micro-level evidence from china', *Oxford Economic Papers* **68**(4), 994–1015.

Ding, S., Sun, P. & Jiang, W. (2016), 'The effect of foreign entry regulation on downstream productivity: Microeconomic evidence from china', *The Scandinavian Journal of Economics* .

Dixit, A. K. & Stiglitz, J. E. (1977), 'Monopolistic competition and optimum product diversity', *The American Economic Review* **67**(3), 297–308.

Donaldson, D. (2018), 'Railroads of the raj: Estimating the impact of transportation infrastructure', *American Economic Review* **108**(4-5), 899–934.

Doraszelski, U. & Jaumandreu, J. (2013), 'R&d and productivity: Estimating endogenous productivity', *The Review of Economic Studies* **80**(4), 1338–1383.

Dunne, T., Roberts, M. J. & Samuelson, L. (1988), 'Patterns of firm entry and exit in us manufacturing industries', *The RAND journal of Economics* pp. 495–515.

Duranton, G. (2015), 'Roads and trade in colombia', *Economics of Transportation* **4**(1-2), 16–36.

Duranton, G., Henderson, V. & Strange, W. (2015), *Handbook of regional and urban economics*, Elsevier.

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

Duranton, G. & Puga, D. (2004), Micro-foundations of urban agglomeration economies, *in* 'Handbook of regional and urban economics', Vol. 4, Elsevier, pp. 2063–2117.

Duranton, G. & Turner, M. A. (2012), 'Urban growth and transportation', *Review of Economic Studies* **79**(4), 1407–1440.

Eaton, J. & Kortum, S. (2002), 'Technology, geography, and trade', *Econometrica* **70**(5), 1741–1779.

Eaton, J., Kortum, S. & Kramarz, F. (2004), 'Dissecting trade: Firms, industries, and export destinations', *American Economic Review* **94**(2), 150–154.

Eaton, J., Kortum, S. & Kramarz, F. (2011), 'An anatomy of international trade: Evidence from french firms', *Econometrica* **79**(5), 1453–1498.

Egger, H. & Falkinger, J. (2003), 'The role of public infrastructure for firm location and international outsourcing'.

Faber, B. (2014), 'Trade integration, market size, and industrialization: evidence from china's national trunk highway system', *The Review of Economic Studies* p. rdu010.

Fan, H., Li, Y. A. & Yeaple, S. R. (2015), 'Trade liberalization, quality, and export prices', *Review of Economics and Statistics* **97**(5), 1033–1051.

Fan, H., Li, Y. A. & Yeaple, S. R. (2018), 'On the relationship between quality and productivity: Evidence from china's accession to the wto', *Journal of International Economics* **110**, 28–49.

Feng, L., Li, Z. & Swenson, D. L. (2016), 'The connection between imported intermediate inputs and exports: Evidence from chinese firms', *Journal of International Economics* **101**, 86–101.

Feng, L., Li, Z. & Swenson, D. L. (2017), 'Trade policy uncertainty and exports: Evidence from china's wto accession', *Journal of International Economics* **106**, 20–36.

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

Forbes, S. J. & Lederman, M. (2010), 'Does vertical integration affect firm performance? evidence from the airline industry', *The RAND Journal of Economics* **41**(4), 765–790.

Foster, L., Haltiwanger, J. C. & Krizan, C. J. (2001), Aggregate productivity growth: lessons from microeconomic evidence, *in* 'New developments in productivity analysis', University of Chicago Press, pp. 303–372.

Foster, L., Haltiwanger, J. & Syverson, C. (2008), 'Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?', *The American economic review* **98**(1), 394–425.

Foster, L., Haltiwanger, J. & Syverson, C. (2016), 'The slow growth of new plants: Learning about demand?', *Economica* **83**(329), 91–129.

Fox, J. T. & Smeets, V. (2011), 'Does input quality drive measured differences in firm productivity?', *International Economic Review* **52**(4), 961–989.

Freund, C. & Pierola, M. D. (2015), 'Export superstars', *Review of Economics and Statistics* **97**(5), 1023–1032.

Gabaix, X. (2011), 'The granular origins of aggregate fluctuations', *Econometrica* **79**(3), 733–772.

Galindo-Rueda, F. & Haskel, J. (2005), *Skills, workforce characteristics and firm-level productivity in England*, Citeseer.

Gao, B., Chan, W. K. V. & Li, H. (2015), 'On the increasing inequality in size distribution of china's listed companies', *China Economic Review* **36**, 25–41.

Garcia-Mila, T., McGuire, T. J. & Porter, R. H. (1996), 'The effect of public capital in state-level production functions reconsidered', *The review of economics and statistics* pp. 177–180.

Gibbons, S., Lyytikäinen, T., Overman, H. G. & Sanchis-Guarner, R. (2016), 'New road infrastructure: the effects on firms'.

Gordon, R. J. (2000), 'Does the" new economy" measure up to the great inventions of the past?', *Journal of economic perspectives* **14**(4), 49–74.

Graham, D. J. (2007), 'Variable returns to agglomeration and the effect of road traffic congestion', *Journal of Urban Economics* **62**(1), 103–120.

Guo, J., Xu, Q., Chen, Q. & Wang, Y. (2013), 'Firm size distribution and mobility of the top 500 firms in china, the united states and the world', *Physica A: Statistical Mechanics and its Applications* **392**(13), 2903–2914.

Guo, L. & Yang, Z. (2019), 'Relationship between shipping accessibility and maritime transport demand: the case of mainland china', *Networks and Spatial Economics* **19**(1), 149–175.

Hart, P. E. & Prais, S. J. (1956), 'The analysis of business concentration: a statistical approach', *Journal of the Royal Statistical Society. Series A (General)* **119**(2), 150–191.

Hashiguchi, Y. & Tanaka, K. (2015), 'Agglomeration and firm-level productivity: Ab ayesian spatial approach', *Papers in Regional Science* **94**, S95–S114.

Helpman, E., Melitz, M. & Rubinstein, Y. (2008), 'Estimating trade flows: Trading partners and trading volumes', *The quarterly journal of economics* **123**(2), 441–487.

Henderson, J. (1986), 'Efficiency of resource usage and city size', *Journal of Urban economics* **19**, 47–70.

Henderson, J. V. (2003), 'Marshall's scale economies', *Journal of urban economics* **53**(1), 1–28.

Henderson, V., Kuncoro, A. & Turner, M. (1995), 'Industrial development in cities', *Journal of political economy* **103**(5), 1067–1090.

Holl, A. (2011), 'Market potential and firm-level productivity in spain', *Journal of Economic Geography* **12**(6), 1191–1215.

Holl, A. (2016), 'Highways and productivity in manufacturing firms', *Journal of Urban Economics* **93**, 131–151.

Holmes, T. J., Hsu, W.-T. & Lee, S. (2014), 'Allocative efficiency, mark-ups, and the welfare gains from trade', *Journal of International Economics* **94**(2), 195–206.

Holmes, T. J. & Mitchell, M. F. (2008), 'A theory of factor allocation and plant size', *The RAND Journal of Economics* **39**(2), 329–351.

Holtz-Eakin, D. & Schwartz, A. E. (1995), 'Spatial productivity spillovers from public infrastructure: evidence from state highways', *International Tax and Public Finance* **2**(3), 459–468.

Hornung, E. (2015), 'Railroads and growth in prussia', *Journal of the European Economic Association* **13**(4), 699–736.

Hortaçsu, A. & Syverson, C. (2007), 'Cementing relationships: Vertical integration, foreclosure, productivity, and prices', *Journal of political economy* **115**(2), 250–301.

Hottman, C. J., Redding, S. J. & Weinstein, D. E. (2016), 'Quantifying the sources of firm heterogeneity', *The Quarterly Journal of Economics* **131**(3), 1291–1364.

Hu, A. G., Jefferson, G. H. & Jinchang, Q. (2005), 'R&d and technology transfer: firm-level evidence from chinese industry', *Review of Economics and Statistics* **87**(4), 780–786.

Huang, J. & Chen, C. (1999), Effects of trade liberalization on agriculture in china: institutional and structural aspects, Technical report.

Huang, J. & Xu, W. (2012), 'Domestic transportation cost and chinese export pattern', *The Journal of World Economy* **1**, 58–77.

Huber, P., Oberhofer, H. & Pfaffermayr, M. (2013), 'Job creation and the intra-distribution dynamics of the firm size distribution', *Industrial and Corporate Change* **23**(1), 171–197.

Hummels, D. & Klenow, P. J. (2005), 'The variety and quality of a nation's exports', *American Economic Review* **95**(3), 704–723.

Hutchinson, J. & Persyn, D. (2012), 'Globalisation, concentration and footloose firms: in search of the main cause of the declining labour share', *Review of World Economics* **148**(1), 17–43.

Ianchovichina, E. & Martin, W. (2001), *Trade liberalization in China's accession to the World Trade Organization*, The World Bank.

Ijiri, Y. & Simon, H. A. (1964), 'Business firm growth and size', *The American Economic Review* **54**(2), 77–89.

Ilmakunnas, P., Maliranta, M. & Vainiomäki, J. (2004), 'The roles of employer and employee characteristics for plant productivity', *Journal of productivity analysis* **21**(3), 249–276.

Imbs, J. & Wacziarg, R. (2003), 'Stages of diversification', *American Economic Review* **93**(1), 63–86.

Jacks, D. S., Meissner, C. M. & Novy, D. (2010), 'Trade costs in the first wave of globalization', *Explorations in Economic History* **47**(2), 127–141.

Jacks, D. S., Meissner, C. M. & Novy, D. (2011), 'Trade booms, trade busts, and trade costs', *Journal of International Economics* **83**(2), 185–201.

Jefferson, G. H., Huamao, B., Xiaojing, G. & Xiaoyun, Y. (2006), 'R&d performance in chinese industry', *Economics of Innovation and New Technology* **15**(4-5), 345–366.

Johnson, R. C. (2012), 'Trade and prices with heterogeneous firms', *Journal of International Economics* **86**(1), 43–56.

Jordaan, J. A. & Rodriguez-Oreggia, E. (2012), 'Regional growth in mexico under trade liberalisation: how important are agglomeration and fdi?', *The Annals of Regional Science* **48**(1), 179–202.

Jorgenson, D. W., Ho, M. S. & Stiroh, K. J. (2008), 'A retrospective look at the us productivity growth resurgence', *The Journal of Economic Perspectives* **22**(1), 3–24.

Khandelwal, A. K., Schott, P. K. & Wei, S.-J. (2013), 'Trade liberalization and embedded institutional reform: Evidence from chinese exporters', *American Economic Review* **103**(6), 2169–95.

Konings, J., Cayseele, P. V. & Warzynski, F. (2005), ‘The effects of privatization and competitive pressure on firms’ price-cost margins: Micro evidence from emerging economies’, *Review of Economics and Statistics* **87**(1), 124–134.

Krugman, P. (1980), ‘Scale economies, product differentiation, and the pattern of trade’, *The American Economic Review* **70**(5), 950–959.

Krugman, P. (1991), ‘Increasing returns and economic geography’, *Journal of political economy* **99**(3), 483–499.

Krugman, P. & Venables, A. J. (1995), ‘Globalization and the inequality of nations’, *The quarterly journal of economics* **110**(4), 857–880.

Lall, S., Deichmann, U. & Shalizi, Z. (1999), *Agglomeration economies and productivity in Indian industry*, The World Bank.

Lardy, N. R. (1993), *Foreign trade and economic reform in China*, Cambridge University Press.

Lentz, R. & Mortensen, D. T. (2008), ‘An empirical model of growth through product innovation’, *Econometrica* **76**(6), 1317–1373.

Levinsohn, J. & Petrin, A. (2000), Estimating production functions using inputs to control for unobservables, Technical report, National bureau of economic research.

Levitt, S. D., List, J. A. & Syverson, C. (2011), How does learning by doing happen?, Technical report, Working Paper, University of Chicago, Chicago.

Li, H. & Li, Z. (2013), ‘Road investments and inventory reduction: Firm level evidence from china’, *Journal of Urban Economics* **76**, 43–52.

Li, Y., Bolton, K. & Westphal, T. (2018), ‘The effect of the new silk road railways on aggregate trade volumes between china and europe’, *Journal of Chinese Economic and Business Studies* **16**(3), 275–292.

Limao, N. & Venables, A. J. (2001), ‘Infrastructure, geographical disadvantage, transport costs, and trade’, *The World Bank Economic Review* **15**(3), 451–479.

Liu, D., Sheng, L. & Yu, M. (2017), ‘Highways and firms’ exports: Evidence from china’, *Peking University* .

Lotti, F. & Santarelli, E. (2004), 'Industry dynamics and the distribution of firm sizes: a nonparametric approach', *Southern Economic Journal* pp. 443–466.

Lu, Y. & Yu, L. (2015), 'Trade liberalization and markup dispersion: evidence from china's wto accession', *American Economic Journal: Applied Economics* **7**(4), 221–53.

Luo, X. (2004), *The Role of Infrastructure Investment Location in China's Western Development*, The World Bank.

Luttmer, E. G. (2007), 'Selection, growth, and the size distribution of firms', *The Quarterly Journal of Economics* **122**(3), 1103–1144.

Magerman, G., De Bruyne, K., Dhyne, E. & Van Hove, J. (2016), Heterogeneous firms and the micro origins of aggregate fluctuations, Technical report, Working Paper Research.

Manova, K. (2012), 'Credit constraints, heterogeneous firms, and international trade', *Review of Economic Studies* **80**(2), 711–744.

Manova, K., Wei, S.-J. & Zhang, Z. (2015), 'Firm exports and multinational activity under credit constraints', *Review of Economics and Statistics* **97**(3), 574–588.

Manova, K. & Zhang, Z. (2009), China's exporters and importers: Firms, products and trade partners, Technical report, National Bureau of Economic Research.

Manova, K. & Zhang, Z. (2012), 'Export prices across firms and destinations', *The Quarterly Journal of Economics* **127**(1), 379–436.

Mansfield, E. (1962), 'Entry, gibrat's law, innovation, and the growth of firms', *The American economic review* **52**(5), 1023–1051.

Martin, R. (2008), *Productivity dispersion, competition and productivity measurement*, number 692, Centre for Economic Performance, London School of Economics and Political ....

Martincus, C. V. & Blyde, J. (2013), 'Shaky roads and trembling exports: Assessing the trade effects of domestic infrastructure using a natural experiment', *Journal of International Economics* **90**(1), 148–161.

Matthee, M. & Naudé, W. (2008), 'The determinants of regional manufactured exports from a developing country', *International Regional Science Review* **31**(4), 343–358.

McCallum, J. (1995), ‘National borders matter: Canada-us regional trade patterns’, *The American Economic Review* **85**(3), 615–623.

Melitz, M. J. (2003), ‘The impact of trade on intra-industry reallocations and aggregate industry productivity’, *Econometrica* **71**(6), 1695–1725.

Melitz, M. J. & Ottaviano, G. I. (2008), ‘Market size, trade, and productivity’, *The review of economic studies* **75**(1), 295–316.

Michaels, G. (2008), ‘The effect of trade on the demand for skill: Evidence from the interstate highway system’, *The Review of Economics and Statistics* **90**(4), 683–701.

Mollisi, V. & Rovigatti, G. (2017), ‘Theory and practice of tfp estimation: the control function approach using stata’.

Moretti, E. (2004), ‘Workers’ education, spillovers, and productivity: evidence from plant-level production functions’, *The American Economic Review* **94**(3), 656–690.

Nakamura, R. (1985), ‘Agglomeration economies in urban manufacturing industries: a case of japanese cities’, *Journal of Urban economics* **17**(1), 108–124.

Novy, D. (2013), ‘Gravity redux: measuring international trade costs with panel data’, *Economic inquiry* **51**(1), 101–121.

Nunn, N. & Puga, D. (2012), ‘Ruggedness: The blessing of bad geography in africa’, *Review of Economics and Statistics* **94**(1), 20–36.

Okubo, T. (2009), ‘Trade liberalisation and agglomeration with firm heterogeneity: Forward and backward linkages’, *Regional Science and Urban Economics* **39**(5), 530–541.

Oliner, S. D., Sichel, D. E. & Stiroh, K. J. (2008), ‘Explaining a productive decade’, *Journal of policy Modeling* **30**(4), 633–673.

Olley, G. S. & Pakes, A. (1992), The dynamics of productivity in the telecommunications equipment industry, Technical report, National Bureau of Economic Research.

Ong, G. P. & Sou, W. S. (2015), ‘Modeling commodity value–weight trends between the united states and its trading partners’, *Transportation Research Record* **2477**(1), 93–105.

Ottaviano, G., Tabuchi, T. & Thisse, J.-F. (2002), ‘Agglomeration and trade revisited’, *International Economic Review* **43**(2), 409–435.

Pagano, P. & Schivardi, F. (2003), ‘Firm size distribution and growth’, *Scandinavian Journal of Economics* **105**(2), 255–274.

Pavcnik, N. (2000), Trade liberalization, exit, and productivity improvements: Evidence from chilean plants, Technical report, National bureau of economic research.

Pedersen, E. L. & Gray, R. (1998), ‘The transport selection criteria of norwegian exporters’, *International Journal of Physical Distribution & Logistics Management* .

Poncet, S. (2003), ‘Measuring chinese domestic and international integration’, *China Economic Review* **14**(1), 1–21.

Rauch, J. E. (1999), ‘Networks versus markets in international trade’, *Journal of international Economics* **48**(1), 7–35.

Roller, L.-H. & Waverman, L. (2001), ‘Telecommunications infrastructure and economic development: A simultaneous approach’, *American economic review* **91**(4), 909–923.

Rosenthal, S. S. & Strange, W. C. (2003), ‘Geography, industrial organization, and agglomeration’, *review of Economics and Statistics* **85**(2), 377–393.

Sakellaris, P. & Wilson, D. J. (2004), ‘Quantifying embodied technological change’, *Review of Economic Dynamics* **7**(1), 1–26.

Segarra, A. & Teruel, M. (2012), ‘An appraisal of firm size distribution: Does sample size matter?’, *Journal of Economic Behavior & Organization* **82**(1), 314–328.

Shefer, D. (1973), ‘Localization economies in smsa’s: a production function analysis’, *Journal of Regional Science* **13**(1), 55–64.

Silva, J. S. & Tenreyro, S. (2006), ‘The log of gravity’, *The Review of Economics and statistics* **88**(4), 641–658.

Simon, H. A. & Bonini, C. P. (1958), ‘The size distribution of business firms’, *The American economic review* **48**(4), 607–617.

Sun, P., Zhang, T. & Yao, S. (2019), ‘Tariff transmission,domestic transport costs and retail prices’, *China Economic Review* (3), 135–149.

Syverson, C. (2004), ‘Product substitutability and productivity dispersion’, *Review of Economics and Statistics* **86**(2), 534–550.

Syverson, C. (2011), ‘What determines productivity?’, *Journal of Economic literature* **49**(2), 326–365.

Tabuchi, T. (1986), ‘Urban agglomeration, capital augmenting technology, and labor market equilibrium’, *Journal of Urban Economics* **20**(2), 211–228.

Tang, Y., Yu, F., Lin, F. & Zhang, M. (2019), ‘China’s high-speed railway, trade cost and firm export’, *China Economic Review* (7), 158–173.

Thornton, R. A. & Thompson, P. (2001), ‘Learning from experience and learning from others: An exploration of learning and spillovers in wartime shipbuilding’, *American Economic Review* pp. 1350–1368.

Tiwari, P., Itoh, H. & Doi, M. (2003), ‘Shippers’ port and carrier selection behaviour in china: a discrete choice analysis’, *Maritime Economics & Logistics* **5**(1), 23–39.

Van Ark, B., O’Mahony, M. & Timmer, M. P. (2008), ‘The productivity gap between europe and the united states: trends and causes’, *The Journal of Economic Perspectives* **22**(1), 25–44.

Van Beveren, I. (2012), ‘Total Factor productivity Estimation: A PracticalReview (CTAN)’, *Journal of Economic Surveys* **26**(1), 98–128.

Van Biesebroeck, J. (2003), ‘Productivity dynamics with technology choice: An application to automobile assembly’, *The Review of Economic Studies* **70**(1), 167–198.

Van Biesebroeck, J. (2005), ‘Exporting raises productivity in sub-saharan african manufacturing firms’, *Journal of International economics* **67**(2), 373–391.

Venables, A. J. (1994), ‘Economic integration and industrial agglomeration’, *Economic and Social Review* **26**(1).

Venables, A. J. (1996), ‘Equilibrium locations of vertically linked industries’, *International economic review* pp. 341–359.

Vijil, M. & Wagner, L. (2012), ‘Does aid for trade enhance export performance? investigating the infrastructure channel’, *The World Economy* **35**(7), 838–868.

Wen, T. (2013), ‘The eight crises: the true experience from china’, *DongFang Press* .

Wooldridge, J. M. (2009), ‘On estimating firm-level production functions using proxy variables to control for unobservables’, *Economics Letters* **104**(3), 112–114.

Xu, M. (2017), ‘Riding on the new silk road: Quantifying the welfare gains from high-speed railways (job market paper)’.

Yang, X. & Ng, Y.-K. (2015), *Specialization and economic organization: A new classical microeconomic framework*, Vol. 215, Elsevier.

Yu, M. (2015), ‘Processing trade, tariff reductions and firm productivity: evidence from chinese firms’, *The Economic Journal* **125**(585), 943–988.

Yuan, Q., Zhao, Y., Shang, H., Zhang, W. & Umar, Z. (2016), ‘Financing constraints on the size distribution of industrial firms: the chinese experience’, *Applied Economics* **48**(41), 3899–3911.

Zhang, J., Chen, Q. & Wang, Y. (2009), ‘Zipf distribution in top chinese firms and an economic explanation’, *Physica A: statistical Mechanics and its applications* **388**(10), 2020–2024.