Expressways, Market Access, and Industrial Development in China

Kecen Jing
School of Economics, Nankai University
jingkecen@gmail.com

Wen-Chi Liao
Department of Real Estate, National University of Singapore
wliao@nus.edu.sg

Abstract

Taking China as an example, this paper considers how the massive construction of national expressways improves county-level market access, and how the ameliorated market access influences industrial development from both the *extensive* and *intensive margin*. A practical means of estimating the critical parameter of market access is provided. To address the endogenous placement decision of expressways, a novel panel IV based on the minimal spanning tree (MST) method in graph theory is constructed to determine the actual expressway market access. The IV estimation results suggest that market access growth significantly bolsters firms' capital accumulation and the growth of firm quantities in counties. Capital-intensive manufacturing industries benefit more from the market access growth and are more inclined to concentrate in better-accessed locations; market access growth is more substantial in less-developed inland provinces and their urban cores. The findings shed light on the dramatic transitions in economic geography in China since the new millennium.

JEL Codes: O14; O18; R40

Key words: expressway, market access, firm performance, industrial development, location quotient

1. INTRODUCTION

Large-scale investment in transportation infrastructure in a large economy is an appealing economic experiment, as transportation plays a fundamental role in fostering regional growth, promoting industrial development, and shaping economic geography. However, such experimentation is difficult to observe in developed countries, like the U.S., that completed their transportation network construction years ago. As the largest developing country with massive ongoing transportation projects, China is an ideal place to study the impact of rapidly growing transport networks such as high-speed railways or expressways. (e.g., Faber, 2014; Qin, 2017; Baum-Snow et al., 2017; Lin, 2017)

This paper explores China's National Expressway Network, a mega-project that was initiated in the 1990s and is still under construction, as a large-scale natural experiment. Here, we investigate the National Expressway Network's impact on county-level market access, using the measurements provided by Donaldson and Hornbeck (2016), and scrutinize how the growth of market access contributes to industrial firms' development and geographic concentration patterns. As the national expressways were strategically planned in the early 1990s with the aim of connecting important city nodes, the market access level and growth of counties are correlated with local unobservable phenomena that also affect the outcome. To address this issue, we construct a hypothetical minimal spanning tree (MST) network as an instrument for the National Expressway Network, following Faber's (2014) novel instrument of a hypothetical road network, and use the calculated market access from the MST as the instrumental variable (IV). To solve the potential issue with Faber's (2014) method that arises from the endogenous selection of target cities in the MST construction procedure, we use walled cities' characteristics back in the 1820s to identify target cities.

In consideration of the extendibility of MST IV to a panel IV, we adopt Faber's (2014) minimal spanning tree (MST) IV rather than refer to historical road maps (e.g., Duranton & Turner, 2012). Despite its general application in economic literature (e.g., Groves et al., 1994; Levitt, 1997; Brandt

& Morrow, 2017), the panel IV is new in transportation research. To my knowledge, the only study that employs a panel transportation IV is Frye (2016), but it simply provides a dummy indicator for whether one county would have been connected to the inter-state highway in a specific year, rather than specifying an expanding route network as the present research does. The MST framework also allows for the construction of multiple IVs that facilitate the weak- and over-identification tests. This research purpose cannot be achieved with one static historical road map.

The result suggests a strong positive impact of market access growth on several industrial outcomes at the county level. First, there is higher industrial GDP growth. This impact is generable through two possible channels, namely the extensive margin (the change in total units) and the intensive margin (the change per unit), which are widely discussed in the trade and industrial organization literature (Chaney, 2008; Volpe Martincus & Carballo, 2008; Blundell et al., 2011; Byford & Gans, 2014; Santos Silva et al., 2014). From the extensive margin, the increased aggregate outcome is attributed to more firms being attracted to the counties. From the intensive margin, increased market access fosters the production of individual firms. Interestingly, the positive impact on production is made via the factor input of capital relative to labor and total factor productivity (TFP). Specifically, capital-intensive industries benefit more from locating themselves in well-accessed counties.

Heterogeneous tests across regions show that better market access helps industrial firms produce more in inland provinces and their regional cores. The results are in line with the core-periphery effect outlined in the new economic geography (NEG) theory (Krugman, 1991a; Krugman, 1991b; Fujita et al., 2001). Better market access also facilitates the concentration of capital-intensive manufacturing industries. These findings shed light on the well-reported transformation of economic geography in China, and indicate the importance of convenient transportation to the national mission of industrial upgrading and the enrichment of less-developed inland areas.

A substantial number of researchers, including Duranton et al. (2014), Coşar and Demir (2016), Storeygard (2016), and Donaldson (2018), have scrutinized how road infrastructure reduces transportation cost, increases trade integration and thus fosters regional growth. Another body of literature adopts the notion of market potential or market access while trying to appreciate transport networks from a holistic point of view (Harris, 1954; Redding & Venables, 2004; Hanson, 2005; Donaldson &Hornbeck, 2016; Fretz et al., 2017). This research is more closely related to studies on the impact of highways or high-speed railways in China (e.g., Faber, 2014; Qin, 2017; Baum-Snow et al., 2017). Faber (2014) focuses specifically on the National Expressway Network, as this thesis does, but mainly investigates the impact of being connected to the expressway on the county-level economic outcomes. However, the integrated notion of market access is not addressed in that study. Taking both highways and railways into account, Baum-Snow et al. (2017) illuminate the pattern of decentralization in Chinese prefectures since 1990 by pointing out that the growth of transport networks, measured either by the number of radical roads or ring roads, redistributes the central city's population and GDP to surrounding areas.

Of the relevant literature, this research is most closely related to but different in essence from that of Baum-Snow et al. (2018), who adopt the definition of market access from Donaldson and Hornbeck (2016) and who study the impact of market access gained via China's national highway network on the economic activities and populations of prefectures. The historical road network from 1962 is used to construct IV for the endogenous highway market access in 2010. They find small or negative effects on prefectures' GDP and populations on average, but the regional primaries achieve manufacturing and service growth from better regional market access, while the hinterland prefectures experience losses. That result is generally in accordance with ours, though our emphasis is on the industrial sector and change of economic geography at the county level. This paper also deviates from Baum-Snow et al. (2018) in terms of methodology: we are able to construct a panel IV at a more refined geographic level from the minimal-spanning-tree (MST) method, while their study, which uses a historical road map, cannot achieve a panel IV.

Despite the profusion of recent studies, little attention has been paid to the effect of market access on the manufacturing industrial development that forms the backbone of China's economic take-off in the last three decades. This research contributes to the understanding of extensive transportation infrastructure improvements in the following ways. First, it is among the earliest to examine the integrity of local market access improvements and the change of industrial landscapes, as well as to further analyze the mechanism behind such changes from an individual firm's perspective. Second, it constructs a novel panel instrumental variable from a hypothetical minimal spanning tree (MST) network. To the best of my knowledge, it is the first panel IV for market access. The IV from hypothetical spanning tree satisfies both the inclusionary and exclusionary restrictions of IV estimation. Third, the study explicitly estimates the power-decay parameter θ of market access in the domestic trade context of China, while most research (e.g., Baum-Snow et al., 2018) borrows the parameter value from previous literature.

The remainder of this paper is organized as follows. Section 2 briefly outlines the development of China's National Expressway Network since 1990s. Section 3 explains the empirical methodology adopted in this study, including the measurement of market access, the estimation strategy and a step-by-step description of the IV construction, as well as the data we use. The empirical results are presented and discussed in Section 4 and Section 5, while conclusions are offered in Section 6.

2. THE NATIONAL EXPRESSWAY NETWORK IN CHINA

China has experienced dramatic economic growth since its Open and Reform policy was proposed in 1978. This rapid economic development, however, has been coupled with an increasing restriction of transportation supply capacity that has become a bottleneck for further economic growth. Acknowledging the importance of a well-connected transportation network, the Chinese government has proposed a long-term planning blueprint for a National Expressway Network,[1]

---

[1] The national-level expressway system is often referred to the National Trunk Highway System (NTHS) in the literature (e.g., Faber, 2014), but to avoid the misleading use of the term "highway" in NTHS, this thesis simply refers to the NTHS

which has been in progress since the 1980s. The ambitiously planned National Expressway Network has three phases, namely the 7-5 network, the 7-9-18 network and the 7-11-18 network. The last phase is still under construction, and is targeted for completion by 2030. This paper focuses on the 7-9-18 network.

The first phase of the National Expressway Network was the 7-5 network approved by the China State Council in 1992. This network is named so because it is comprised of 7 vertical lines and 5 horizontal lines that constitute the main skeleton of the early road network. It is not an expressway network because lower-standard national highways[2] are also included in it. The planning objectives of the network were achieved ahead of schedule in 2007. However, the completed 7-5 network has been unable to cope with the burgeoning demand for road transport in the new millennium, despite the fact that the total expressway mileage of China ranked the second in the world by the end of 2001.

In 2004, a much more ambitious expressway plan, the 7-9-18 network, was unveiled. The 7-9-18 network comprises a grid of 7 radial expressways from Beijing, the capital of China; 9 north-south expressways; and 18 east-west expressways. It aims to connect cities with urban population larger than 200,000 and important city nodes such as ports, transportation hubs and boarder cities. In contrast to the 7-5 network, it consists exclusively of national-level expressways that satisfy a high technical standard: at least four-lane routes with limited access, comprehensive service systems, and communication monitoring equipment. National expressways are the highest level among all the national road types. Provincial expressways that connect to the 7-9-18 network are to be planned by local governments.

The third phase of China's national expressway construction, an extension of the 7-9-18 network,

as the National Expressway Network.

[2] National highways are generally of lower standard than the national expressways in terms of road quality, speed limits, service facilities and tolls.

was proposed by the Ministry of Transport in 2013. Two more north-south expressways will be added to the previous 7-9-18 network to make it a 7-11-18 network. The complete 7-11-18 network will connect all the newly-emerged prefecture capitals and cities with urban populations of more than 200,000.

The national expressway is centrally planned, and both central and local governments have made substantial investments in it. The Chinese government has made the decision to put effort into constructing such a superior network that its road quality, speed and tolls are of significant contrast to previously existing national and provincial highways. It is the reason we have chosen to study the effect of the national expressway while ignore other road layers[3].

This research looks into the period between 2000 and 2009, a period generally labeled as the rapid development of the expressway, during which the 7-9-18 network was constructed. Though in 2009 the expressway network was not yet as comprehensive as it is now, it had already connected major city nodes. Figure 1 depicts the expansion of the national expressway network from 2000 to 2009.

[Insert Figure 1 here]

3. EMPIRICAL METHODOLOGY AND DATA

Given the massive construction of expressway network in China, we turn to the question of how the large-scale infrastructure construction influences industrial development. We measure the county-level market access following the reduced form derived from Donaldson and Hornbeck (2016)'s general equilibrium model of domestic trade. Our main empirical strategy is an instrumental variable (IV) estimation. The instrumental variable is the market access calculated from a hypothetical minimal spanning tree (MST) network in China at various years. Thus, the MST method can generate a panel IV. In addition, the MST framework allows for the construction

[3] Technically speaking, China has roads almost everywhere. Counties had access to all other counties even back in the nineteenth century at a coarse resolution. In this regard, the access from normal roads will boil down to physical distance which can be captured by the county-fixed effect.

of more than one panel IV and facilitates testing the over-identification assumption in a two-stage least square (2SLS) estimation. Hence, my novel panel IV is preferable for analyzing the fast-growing expressway network than the conventional IV from a static historical road map.

3.1. Measuring County-Level Market Access

This research uses the concept of market access to examine the effect of the expansion of transport networks on local economies. The market access of one location is determined not by its local unobservable variables or its proximity to expressways, but by how conveniently it is connected to other regions and the economic performance of the regions it is connected to. Specifically, we adopt the first-order approximation of market access (MA) in Donaldson and Hornbeck's (2016) general equilibrium model, which takes the following form:

$$MA_{ct} = \sum_{c'}^{c' \neq c} T_{cc't}^{-\theta} \times pop_{c't} \tag{1}$$

The market access MA_{ct} of county c at year t is, intuitively interpreted, the weighted travel cost by the market size at destination county c'. t represents four periods from 2000 to 2009: 2000, 2003, 2006, and 2009.[4] Here, travel cost is measured by a power-decay function of travel time in minutes $T_{cc't}$ along the expressway network. θ is the power decay parameter.

3.2. Estimation of the power-decay parameter $\boldsymbol{\theta}$

The key parameter θ determines how the market access decays with travel time along the expressway network. It is essential to the MA measurement and is specific to the research context. Estimation of θ is necessary, but extant literature often ignores the estimation and assigns a value to θ. (e.g., Baum-Snow et al., 2018).

[4] The reasons to choose four periods with three-year intervals are: 1) annual expressway data is not available; and 2) the impact of transportation infrastructure takes time to become apparent.

The power-decay parameter θ is also known as trade elasticity of travel impedance or travel cost (Burger et al., 2009; Coşar & Demir, 2016; Ewing & Cervero, 2010; Salon et al., 2012) developed from the gravity model of trade. Its estimation requires knowledge of bilateral domestic trade flow. Such data is unfortunately not available in China, so instead we exploit the microdata on county-to-port volumes of domestic freight transportation for export purposes, as reported by Chinese Customs in 2000 and 2003. Recording the trade volume from counties to ports essentially produces domestic trade data. The empirical model for estimating the trade elasticity θ takes the following linear function form in Equation (2). To the best of my knowledge, this study is the first one to estimate trade elasticity in a Chinese context.

$$ln\text{Volume}_{cpt} = \alpha - \theta lnT_{cpt} + \gamma PortSize_{pt} + \delta_c + \eta_t + \varepsilon_{cpt} \tag{2}$$

Volume_{cpt} is the trade volume in monetary value from county c to port p in year t for export purposes. The transportation cost is measured in travel time in minutes T_{cpt}. The estimated coefficient of travel time θ is the trade elasticity. The port size is controlled by its total export volume $PortSize_{pt}$. Time-fixed effect η_t and county-fixed effect δ_c control for general time trend and time-invariant county heterogeneity, respectively. ε_{cpt} is the error term.

The estimated value of parameter θ is 1.1 with a 95 percent confidence interval between 1.07 and 1.16. $\theta = 1.1$ is well in line with the trade elasticity values documented in the literature on regional and domestic transportation (e.g., Ewing & Cervero, 2010; Salon et al., 2012), though my estimated θ is smaller than those in international trade literature (e.g., Eaton & Kortum, 2002; Redding & Venables, 2004). The discrepancy between the two sets of literature is reasonable, because domestic trade is more efficient than its international counterpart.

3.3. Empirical Strategy

To answer the question of how the development of firms is influenced by the market access of the counties in which they are located, we regress the aggregated or average industrial firm outcome at county-year level Y_{ct} on market access MA_{ct}, while controlling for a vector of variable X_{ct} that includes the county population and the share of total numbers of firms for each industry in that county. County-fixed effect δ_c is included to absorb the time-invariant county-specific heterogeneities, such as distance to expressway or to targeted city nodes in Faber (2014). Province-year-fixed effect η_{rt} controls for the macroeconomic trend that applies to all the firms in the same province. ε_{icrt} is the error term.

$$lnY_{ct} = \alpha + \beta lnMA_{ct} + \gamma X_{ct} + \delta_c + \eta_{rt} + \varepsilon_{crt} \quad (3)$$

The reason we aggregate or average the firm outcome to county-year level is to accommodate the individual data for firm performance and county data for market access and instrumental variables. We keep the data structure consistent at the county-level and then assure the quality of two-stage least square (2SLS) estimation. The estimated coefficient β thus captures the effect on the performance of a representative firm in each county. A weight, the total number of firms at each county-year, is applied to the county-level regression in Equation (3). Without weights, the estimates could be driven by the extreme case in counties with few firms.

As counties not connected to the expressway system have zero value of MA_{ct}, we assign a small value, e^{-8}, to those counties to facilitate taking logs. In addition, a shifter is applied to $lnMA_{ct}$ by subtracting the MA value of non-connected counties (i.e., -18.42=ln(e^{-8})) and adding the minimal MA value of connected counties. This allows for the equation of the $lnMA_{ct}$ value of the connected county with least market access to that of the non-connected counties, thus avoiding a sudden jump in $lnMA_{ct}$ value when a county gets connected.

A first-difference model is also applied to investigate the effect of market access growth on the outcome growth of representative firms. The model is specified in Equation (4), where Δ is the operator of first difference between two consecutive years at the county level. County-fixed effect has been differenced out, and is therefore omitted from Equation (4). The control variable vector

X_{ct} is not differenced, in order to control county population and industry share in the starting year. Weights are also applied.

$$\Delta lnY_{ct} = \alpha + \beta \Delta lnMA_{ct} + \gamma X_{ct} + \eta_{rt} + \varepsilon_{crt} \quad (4)$$

The two estimation specifications, Equation (3) and Equation (4), are equivalent to large extent. They yield identical results if t equals two and X_{ct} is also differenced. While working with an unbalanced panel of the kind used in this research, the first-difference estimator may lose observations due to "gaps" in the time series. In that case, a fixed-effect model like Equation (3) is a better choice (Wooldridge, 2006). Accordingly, Equation (3) is the baseline regression model here.

An unbiased estimate of the coefficient of market access requires $lnMA_{ct}$ to be uncorrelated with the error term. Unfortunately, this is not the case in reality, due to the strategic placement of the expressway network by the government. As noted in Section 2, more economically important and larger cities are selected as connecting nodes. Chances are therefore high that these cities are more preferred under place-based policies or in agglomeration economies that are in favor of firm development. The local unobservable variables may affect firm performance and market access simultaneously.

Here, we adopt an instrumental variable (IV) estimation to address the endogeneity issue. We construct a hypothetical minimal spanning tree (MST) network to model the national expressway network, and use the calculated market access from the MST as the instrumental variable. Unlike the conventional way of exploiting a historical road map, the MST can generate a panel IV that enables mapping of the expansion of expressway network, while the historical road map is simply one cross-section. In addition, we am able to generate multiple IVs with the MST method that facilitate both the weak identification and over-identification tests.

3.4. A Panel IV for Expressway Network Expansion

A minimal spanning tree (MST) is a set of edges (lines) that have the minimal total cost of connecting all the target vertices (nodes) in a graph. The cost could be defined as total length, weight or any other cost measurement. MST is a method firstly introduced in the graph theory and now widely applied in geography and transportation research (e.g., Bunn et al., 2000; Faber, 2014). In the context of this research's, the MST is a hypothetical road network that connects all the pre-selected target cities and has a minimal total land cost based on the constructed cost surface of China, as shown in Figure 2.

[Insert Figure 2 here]

3.4.1. Construction of Minimal-Spanning-Tree (MST) Network

The construction of the MST takes four steps. The first step is to select the vertices (nodes) to be connected—that is, the targeted cites. We choose the targeted cities according to the proposal of National Expressway Network in 1990s. In descending order of importance, these include the capital city, Beijing; municipalities[5]; provincial capitals; special economic zones; ports; and the rest of cities with a population of more than one million in 1990. The constructed MST network is henceforth referred to as the Pop1990 MST.

The next step is to construct the optimal route graph between targeted cities using Dijkstra's (1959) algorithm. The optimal route is the least costly path between any pair of the targeted cities on the cost surface of China. The cost surface, presented in Figure 2, gives each 1km×1km land parcel a weighted land cost, which is calculated from China's land slope and remote sensing land cover data[6]. These data are given weights of 0.6 and 0.4, respectively. Lands with steeper slopes and less usable land, such as rivers or urban land, are assigned higher land costs and thus are less likely to have a through route. The minimal spanning tree network is generated based upon Kruskal's (1956)

[5] Municipalities include Beijing, Shanghai, Tianjin, and Chongqing.

[6] China's land slope is calculated from its elevation data from DIVA-GIS. The land cover remote sensing data is provided by the Cold and Arid Regions Science Data Center in China. Land cover data records the land use types for each land parcel in China. These include urban land, rivers, and forests.

algorithm in the third step. It is a subset of routes in the second step, which connects all the cities and has the least total land cost.

In the last step, we track the expansion of the Pop1990 MST by extracting the portion of MST that has the same length as the national expressway network for each time period t. The expansions of MST are assumed to follow two possible orders: from the most important target cities or from the least-cost route. The expansions of Pop1990 MST in different orders are plotted in Panel A of Figure 3. The blue routes denote road expansions in descending order of cities' importance, while the green routes denote expansions in ascending order of construction cost. The dynamic MST network allows for the calculation of a panel of hypothetical market access conditions to model the time-variant $lnMA_{ct}$. This research purpose cannot be achieved by using a cross-section historical road map.

[Insert Figure 3 here]

3.4.2. Wall-City MST versus Pop1990 MST

One concern with the use of the Pop1990 MST is the endogenous selection of target cities. Cities designated as important in 1990s are likely to have higher hypothetical market access and better industrial development, which is also a major critique of Fabler's (2014) MST method that directly uses important cities in 1990s as target cities.

To address the possible endogeneity in the selection of target cities, we propose a novel Wall-City MST to model the China expressway network by using the estimated population in 1990 from walled cities in the 1820, i.e., the late Qing Dynasty[7], to select target cities. Historical walled cities in the Qing Dynasty were built for the military purpose of homeland security—for example, to defend against the White Lotus Rebellion[8] or foreign invasion from the western frontiers. The

---

[7] The Qing Dynasty was the last imperial dynasty of China. It ruled China from 1644 to 1912.

[8] The White Lotus Rebellion was initiated by the members of the white Lotus Society, a secret religious society, in the

walled cities, therefore, are rarely likely to coincide with the important cities of current times. The exclusionary restriction of a Wall-City MST IV is therefore satisfied. On the other hand, these historical walled cities are linked to the current city ranks in the sense that the wall size distribution is positively correlated with city size distribution and has good predictive power for population density at the current time (Du & Zhang, 2017; Ioannides & Zhang, 2017). Thus, the Wall-City MST reasonably resembles the National Expressway Network.

Therefore, for the Wall-City MST, we use the characteristics of the historical city walls, e.g., the height or circumference of city walls, to predict the city population in 1990 and then select targeted cities with an estimated population larger than 500,000[9]. This makes the targeted walled cities less likely to be correlated with current industrial development than the target cities in the Pop1990 MST. The geographic distribution of targeted cities under both selection criteria are presented in Figure A1 of the Appendices. Targeted walled cities were densely located at the nation's boundary back in the Qing Dynasty and have a different distribution pattern from the current important cities.

The MST constructed based upon wall cites is henceforth referred to as the Wall-City MST. Analogously, we also define two means of expansion for the Wall-City MST: from the most important target cities or from the least-cost route. The expansion of the Wall-City MST is plotted in Figure 3, Panel B.

Given the Pop1990 MST and Wall-City MST, as well as their dynamic expansions in different orders, we can construct four hypothetical market access conditions, labeled as $MA\_pop90_{ct}^{order}$, $MA\_pop90_{ct}^{cost}$, $MA\_wallcity_{ct}^{order}$, $MA\_wallcity_{ct}^{cost}$, as the instrumental variables. MA_pop90 and $MA\_wallcity$ represent the market access constructed from the Pop1990 MST and Wall-City

---

mountainous areas such as Sichuan and Shaanxi provinces in the early Qing Dynasty.

[9] The number of cities with estimated city populations from historical wall cities is generally smaller than the real population in 1990. To make the number of selected cities comparable, we lower the selection criteria for estimated population from 1 million to 500,000. With this, the predicted targeted cities are different, but the total number is similar to that of the Pop1990 MST.

MST respectively, and the superscripts $order$ and $cost$ denote the expansion order of MST by target city importance order and by route construction cost. For better readability, $MA\_pop90_{ct}^{order}$ and $MA\_pop90_{ct}^{cost}$ will be referred as the Pop1990 IV and $MA\_wallcity_{ct}^{order}$ and $MA\_wallcity_{ct}^{cost}$ will be referred as the Wall-City IV.

Both the Pop1990 IV and Wall-City IV will be applied to the empirical estimations, but we attach greater importance to the Wall-City IV, as it has accounted for the endogenous selection of target cities and is more likely to satisfy the exclusionary restriction of IV estimation.

3.5. The Validity of Identifying Assumption

The validity of the minimal spanning tree (MST) IV relies on the assumption that it influences industrial firms' performance only through its correlation with a county's market access from national expressway network (inclusionary restriction) instead of through direct correlation with firm performance (exclusionary restriction). The MST method is my major methodological contribution and has the merit of satisfying the identifying assumptions.

The inclusionary restriction is intuitively valid. Conditional on connecting all the targeted cities, MST is constructed along the least-cost routes given the landscape and land use in China, which is also a practical rule for expressway construction. The market access developed from the MST network, though hypothetical, correlates with MA_{ct} to a large degree. The MST IV is less likely to be correlated with the dependent variable than the IV from a historical road map, as modern expressways may follow the path of historical roads and favor the historically important city nodes. In addition, the changing MST network allows for the construction of multiple panel IVs that are able to model the dynamic expansion of expressways and, above all, facilitate over-identification tests.

Another threat to the exclusionary restriction might result from the possibly endogenous hypothetical market access, as the evolution of the expressway network is more likely to be towards large cities where firms would produce more. This concern is addressed by using the Wall-City MST, in which the selection of target cites is based on predicted population from historical data and thus less likely to be correlated with current firms' performance.

We also include county-fixed effect to control for county-level attributes such as distance to prefecture centers or a county's historical characteristics, because these attributes may correlate with IV and attract better-performing firms (Faber, 2014) as well. The omission of county-level attributes would result in the violation of the exclusionary restriction.

Rigorous tests of the validity of IV are presented in the Results section.

3.6. Data

The geo-referenced national expressway data is obtained from ACASIAN database at Griffith University, Australia. The data is in a well-compiled shape-file format that facilitates further process in ArcGIS and has been used for several frequently-cited studies of China (e.g., World Bank, 2007; Faber, 2014). To calculate the travel time along the expressway network, we assign the national expressway a conventional speed of 100km/h.[10]

We obtain the geo-referenced county boundary map for 2000, which includes six-digit standard county codes, from the China Data Center at the University of Michigan, USA. The county allow me to match county-level geographic information like market access to the county-level socioeconomic characteristics such as population and area from the statistical yearbook.

The performance information of individual industrial firms from 2000 to 2009 is taken from the Annual Surveys of Industrial Firms (ASIF) conducted annually by the National Bureau of Statistics in China. Industrial firms are those in the mining, manufacturing and utilities industries[11]. As is

[10] According to the Road Traffic Safety Law of the People's Republic of China, expressways in China have a speed limit of 120 km/h, and a minimum speed of 70 km/h is also in force. 100 km/h is a reasonable value to represent the average speed of expressways in China.

[11] The utilities industry includes the production and supply of water, electricity, and gas.

reported by the National Bureau of Statistics, this survey covers enterprises above designated size, generally referred as "above-scale" firms, but the "above-scale" definition has changed several times. Before 2006, enterprise above designated size included all the industrial State-Owned Enterprises (SOEs) and industrial non-SOEs with annual revenue of 5 million yuan or more from their main business operations. From 2007 onward, the "above-scale" definition has covered all the industrial firms with annual revenue above 5 million yuan from main business operations, regardless of their ownership. However, we do not need to adjust the firm data to account for this standard change. According to the Economic Census in 2004 and 2008, the ratio of annual revenue from main business operations of above-scale enterprises to that of all industrial firms is 89.44 percent to 88.25 percent. It means that despite the change of standard, the representativeness of firms in the ASIF data remains almost the same. The change in the definition of "above-scale" is most likely due to inflation.

ASIF provides the most detailed and comprehensive information at the individual firm level. For instance, it provides identifying information like firm name, firm ID, address and six-digit county code, together with financial and production statements such as total asset, production value, employment, and revenue, etc.

Despite the rich information, the use of ASIF data is subject to several commonly acknowledged challenges (Brandt et al., 2014; Nie et al, 2012). As my research focus is on firms' production and We only use four years of ASIF data (2000, 2003, 2006, and 2009), most of the noted data issues are not relevant to this research. However, caveats remain. First, there is the fact the definition of the 4-digit Chinese Industry Classification (CIC) system changed in 2003. We adjust the firms' CIC code in 2000 data to the revised 2003 CIC code. Second, even given the extensive coverage of variables in the ASIF data, the set of variables included in each year are not consistent and there are missing variables. For example, the 2009 data does not have value added information. We calculate valued added by assuming the growth rate of value added from 2006 to 2009 is the mean of growth rate in 2000-2003 and 2003-2006. Third, the data processing largely depends on the six-digit county codes of firms to match to other county-level information, but the address and the six-digit county code in ASIF data are from a firm's registration address rather than its actual location. As stated in the Regulations of the People's Republic of China on Company Registration, there can only be one company address registered by the registration authority, and this address should be within the jurisdiction of the registration authority.[12] Therefore, only firms registered at the county level can definitely be assumed to have their registration addresses in the same counties as their actual locations. Accordingly, firms that are not registered at county level are dropped, and relocation firms are also dropped to preclude firm sorting that may challenge the identifying validity.

4. EMPIRICAL RESULTS

Executing the empirical estimations described above, we report and discuss the results with additional heterogeneous test by industries and locations in this section. Table 1 provides descriptive statistics for firm outcome averaged or aggregated to county-year level, the county-level market access from both expressway network and constructed MST, and the estimated population from historical wall cities and county population in 1990. Outcomes include the aggregated value added from firms at the county level, which is also interpreted as the GDP contributed by the industrial firms, and the total number of firms. At the county-mean level, they include firm production, employment, capital and TFP estimated from the Solow residual.

[Insert Table 1 here]

4.1. Growth of Market Access along the NTHS Network

Market access is calculated for all the expressway connected counties[13] for the four study years

[12] Source: http://home.saic.gov.cn/zw/zcfg/xzfg/201402/t20140227_215709.html.

[13] This research assumes a county is connected by the expressway if the county's center is within 20 km of any segment of the established expressway system.

(2000, 2003, 2006, and 2009) using the estimated θ value taken from China's domestic freight transportation data (i.e., θ=1.1). Naturally, counties without connections to expressways have minimal market access values after the shifter discussed in Section 3.3 is applied. Figure 4 displays the density of evolving market access from 2000 to 2009. [14]

[Insert Figure 4 here]

The pattern in Figure 4 clearly shows that the evolution of market access is reflected in both the magnitude of increases and geographic expansion. Market access increments are also more likely to increase around economically important or highly populated cities, indicating the influence of endogenous placement decisions in the expressway planning and justifying the necessity and usefulness of the instrumental variables.

4.2. Market Access and Performance of Firms

Before turning to the instrumental variable estimation, we run a couple of OLS regressions to establish the correlation between market access and firm performance at the county level. A two-way fixed-effect model (i.e. Equation [3]) and a first-difference model (i.e., Equation [4]) are both applied. The OLS results are presented in Table 2. The fixed-effect and first-difference regression results are quite similar in terms of coefficients and significance, and both can be interpreted as a growth-growth model (Woodridge, 2006). In the subsequent analysis, we will stick to the FE model, which is more efficient and able to make use of all periods' information in the unbalanced panel data.

[Insert Table 2 here]

We examine the correlation between county market access and a set of outcomes that has described in the summary statistics. The county-mean outcome can be interpreted as the performance of a

[14] For the purpose of better visualization, market access is plotted in smoothed density.

representative firm in that county. Incorporating both aggregated and averaged county outcomes helps us appreciate the contribution of market access to industrial development from both the extensive margin and the intensive margin. This phenomenon is not investigated by the extant literature, which simply looks at aggregated county-level or prefecture-level outcomes.

The OLS regression results show that counties with better market access improvements tend to have higher industrial GDP growth, while the performance of representative firms is not significantly correlated with market access, holding fixed the value of all control variables and fixed effects. The statistical insignificance and small coefficients may result from a downward bias arising from omitted local unobservable variables correlating with market access and firm performance, or from the selection of entry firms, which are usually smaller in size and production in well-linked-up counties.

[Insert Table 3 here]

To get an unbiased estimation of the effect of market access in both extensive and intensive margin, we adopt 2SLS estimation with the instrumental variables from the two sets of minimal spanning tree networks. The first-stage results are reported in Table 3. The four panel IVs are positively and significantly correlated with market access, indicating that they are not weak and are of reasonable size. Table 4 shows the second-stage regression with more rigorous tests of IV quality.

[Insert Table 4 here]

Panel A of Table 4 adopts the Pop1990 IV (i.e., $lnMA\_pop90^{order}$ and $lnMA\_pop90^{cost}$) and Panel B adopts the Wall-City IV (i.e., $lnMA\_wallcity^{order}$ and $lnMA\_wallcity^{cost}$). The simultaneous inclusion of two IVs in one regression allows for an over-identification test that is rarely taken in transportation literature. Both IV estimates from Panel A and B pass the weak identification and over-identification tests, suggesting the inclusionary restriction and exclusionary restriction for IV estimation are satisfied. Nevertheless, the Wall-City IV estimates are of higher F statistic value in the weak identification test and higher P-value in the over-identification test. The Wall-City IV is thus an even more valid IV for this research.

Further, the Wall-City IV estimates are slightly smaller than the Pop1990 IV estimates. This result is intuitively sound, because the Pop1990 MST connects more large cities that produce more due to agglomeration economies (e.g., Behrens et al., 2014; Glaeser & Gottlieb, 2009). It makes the Pop1990 IV estimates overestimated and generates a source of over-identification. On the other hand, the Wall-City MST is a more random road network and is less likely to correlate with firm outcome.

IV regression addresses the downward bias in OLS; thus, the IV estimates from Table 4 are larger and more significant than the OLS estimates. Specifically, Column 1 demonstrates the significantly positive impact of market access on industrial GDP, which is the aggregated outcome. The effect could be driven by either the increase of total number of firms within counties, i.e., the extensive margin (Column 2) or the better performance of individual firms, i.e., the intensive margin (Column 3-6). Column 2 shows that locations with better market access are appealing places for firms to locate, which is in line with the NEG story (e.g., Fujita et al., 2001). The total product of a representative firm increases when it is afforded better market access (Column 3). The channels of total production growth are explored through the examination of its contributing factors, which include employment, capital stock and TFP, as outlined in Columns 4–6. The production growth is mainly due to the significant increase in capital stock and marginally increase in TFP, while employment remain less responsive to market access growth.

These findings have plausible economics. Capital is tradable, with an elastic supply in the wake of the capital liberalization in China in the early 1990s. Labor supply, on the other hand, is inelastic given the nature of a local labor market. Thus, when demand for output rises, producers will substitute capital for labor, even if the production is constant. This trend explains the significant positive impact on capital usage and insignificant response of labor input at locations with better market access. Lastly, firms do not seem to meet the stronger demand through acquiring new technology for higher TFP on average, which is not surprising because R&D requires highly skilled labor that may be less available outside first-tier cities.

4.2.1. Parameter Sensitivity Test

The analysis above applies the θ value of 1.1. We also report second-stage results from using alternative value of θ equal to 1.07 and 1.16, the lower and upper boundary of the 95 percent confidence interval of θ esitmates. Table A1 in the Appendices presents the sensitivity test of the power-decay parameter θ. When θ is 1.07, suggesting transportation is less of an impediment, there is a higher marginal impact of market access on firm performance. On the other hand, the effect of market access is smaller when $\theta=1.16$. The parameter sensitivity test shows how much the estimation results depend on the choice of parameter value.

4.3. Heterogeneous Response to Market Access across Industries

Given that the demand for goods at one location with better market access is higher and labor supply is less elastic, this place would encourage the development of capital-intensive industries. Such industries are expected to achieve a higher production growth in places with better market access. Therefore, this subsection outlines tests of the heterogeneous impact of market access on the performance of different industries.

To this end, we stratify individual firms into 39 industries according to the 2-digit CIC code. We then run the regression model in Column 1 of Panel B in Table 4 for each industry and obtain the point IV estimate.[15] Figure 5 is a scatterplot that plots the coefficients of manufacturing industries against the industry's capital-labor ratio.[16] The pattern is quite clear: while some industries report

[15] The reason we do not run regression with interaction between $lnMA$ and industry dummy is to avoid including too many instrumental variables' interactions with the industry dummy.

[16] We drop the mining and utilities industries (e.g., the production of water, electricity and gas) from the data, as these are more dependent on natural resources than market access; in addition, mining and utilities account for a much smaller share of industrial firms than manufacturing does.

a negative coefficient, industries with a higher degree of capital-intensive ratio tend to have more positive coefficients of market access. The fact that capital-intensive industries benefit more from locating in well-accessed counties is quite consistent with the findings in Table 4 and the NEG story discussed above.

[Insert Figure 5 here]

We further classify manufacturing industries into emerging and traditional industries. An emerging industry is defined as one with a production growth rate in the top 50^{th} percentile among all the industries from 2000 to 2009. In general, emerging industries are capital-intensive and more dependent on market access. This finding may shed light on the remarkable industrial transformation happening in China. That is, market access brings about disproportionate capital increases for manufacturing firms and boosts the industry transformation from labor-intensive to capital-intensive. This transformation is the purpose that expressway planners intend to achieve (The Ministry of Transport, 2004).

4.4. Heterogeneous Effect across Geographic Locations

Another interesting observation emerging from the distinct results between OLS and 2SLS estimations is that the local unobservable variables seem to affect market access and industrial development to a great extent. That is, there is considerable heterogeneity across locations. Table 5 examines the heterogeneous effect of market access in the geographic dimension. To this end, we include the interaction term of $lnMA$ and various location indicators for well-acknowledged developed locations, such as the southeast coastal provinces[17] or first-tier cities[18], to examine whether increasing market access are more favorable for firms in those locations. The instrumental variable, that is the Wall-City IV, is accordingly combined with location dummies. The same

[17] The southeast coastal provinces include Jiangsu, Zhejiang, Fujian, and Guangdong. To distinguish the impact from that of the first-tier cities, Shanghai, Guangzhou and Shenzhen are excluded from the coastal provinces.

[18] The first-tier cities of China are widely acknowledged as Beijing, Shanghai, Guangzhou, and Shenzhen.

dependent variables as in Table 4 are adopted.

[Insert Table 5 here]

Panel A presents the effect of market access in the southeast coastal provinces where the manufacturing industries are most developed. The results suggest the traditional bellwethers of manufacturing are losing their lead with the development of expressways. Compared to the inland provinces, southeast coastal provinces have experienced much less marginal impact on total industrial GDP from market access growth, which is mainly due to the loss of capital per firm (intensive margin) and reduction in firm quantities (extensive margin). No significant pattern for the first-tier cities is identified from the data presented in Panel B, though there is a slight decrease in TFP. Panels A and B support an overall "moving inland" narrative that market access growth benefits the less-developed inland provinces more.

Panels C and D present investigations the heterogeneous impact of market access within inland provinces. Inland provincial capitals make better use of market access growth in terms of total industrial GDP and capital per firm. As provincial capitals are the cores of provinces, Panel C shows a core-periphery pattern, as proposed in NEG model. Here, the evolving concentration of production, which is from periphery to core, rises with the reduction in transportation costs between asymmetric locations. Taking one step further, Panel D presents a more refined geographic scope and demonstrates how the inland urban districts benefit from better market access compared to the rural counties within one prefecture[19]. The results consistently and astonishingly suggest that market access growth contributes to significantly higher industrial GDP growth in urban districts, and that it is mainly due to a larger number of firms, higher capital growth and even larger-scale employment in urban districts. The findings presented in Panel D lend strong support to the

[19] A typical prefecture-level city in China consists of county-level urban districts and other counties on their outskirts that are referred to as rural counties.

concentration of production in urban area or, in short, urbanization.

My result does not contradict the decentralization narrative put forward by Baum-Snow et al. (2017). They emphasize regional decentralization, while this research tracks urbanization at the national level. In addition, Baum-Snow et al. (2017) focus on the decentralization of total population and GDP, while we consider industrial development.

To sum up, Table 5 shows a very interesting pattern—that is, the concurrent trends of industrial urbanization and inland migration of manufacturers, which are well-supported by anecdotal evidence and literature.[20] This research further contributes to the "moving inland" narrative by specifically emphasizing the role of transportation.

5. GEOGRAPHIC CONCENTRATION OF INDUSTRIAL DEVELOPMENT

The previous section separately assesses the heterogeneous impact of market access by industry and by location. In this section, we integrate the two dimensions and look into the geographic concentration of manufacturing industries.

5.1. The Location Quotient

Table 6 reports the effect of market access on location quotient and how this effect varies across industries and locations. A location quotient is a widely-applied ratio that measures geographic concentrations of economic activity. For the purposes of this research, the LQ of industry i in county c is defined as:

$$LQ_{ic} = \frac{Y_{ic}/Y_c}{Y_i/Y} \tag{5}$$

where Y_{ic}/Y_c is county c's share of total production of industry i, and Y_i/Y is industry i's share of

[20] For example, https://www.reuters.com/article/us-china-manufacturing/special-report-worlds-workshop-heads-to-inland-china-idUSTRE67O19420100825.

total industrial production in China. Thus, LQ measures the relative concentration of production in a county compared to the nation. The county-level and industry-level production values in Equation (5) are aggregated from the ASIF data. The IV estimates in Table 6 illustrate how a county's market access influences the concentration or dispersion of industries.

[Insert Table 6 here]

Despite the overall positive effect of market access on LQ, manufacturing industries are much more concentrated at locations with better market access via expressway than mining and utilities industries, because freight transportation is more relevant to manufacturers. It facilitates the access to upstream industries and downstream consumers. This finding is in accordance with Chinese reality. For instance, Foxconn, the world's largest contract electronics manufacturer, has established a new plant in Zhengzhou, an inland but transportation-hub city with good market access, where half of the world's iPhones are produced.[21] Market access also has a larger impact on capital-intensive or emerging industries' concentrations.[22] Similarly, the inland provinces have gained higher concentration of manufacturing than the southeast coastal area.

The analysis above provides solid evidence for the role of market access in helping the less-developed inland provinces accumulate capital and develop capital-intensive industries. China used to be the "world's factory," thanks to its massive input of cheap labor. This has led to the success of manufacturing industries in the southeast coastal provinces. Today, however, the inland provinces do not necessarily stick to the old labor-intensive path. A good transportation network helps industries to make a leap forward. That is indeed the industrial development for which the

---

[21] Source: http://en.people.cn/n3/2017/0918/c90000-9270552.html.

[22] To further investigate the concentration pattern across industries, we plot the coefficient for each manufacturing industry with respect to the industry's capital-labor ration in Figure A2 of Appendices, similar to what we did in Figure 5. Figure A2 backs up the estimation results shown in Table 6, which indicate that capital-intensive and emerging industries are more likely to concentrate in locations with better market access. It implies that market access helps counties to develop the fast-growing emerging industries as their base sectors, which, in turn, fosters regional development in the long run.

Chinese government strives.

5.2. Geographic Concentration Pattern of Typical Industries

Two classical industries that highlight the early and mature phases of industrial development are highlighted in this section. They are the textile and telecommunication and computer manufacturing industry, respectively. We plot their initial LQ in 2000 and LQ growth from 2000 to 2009 in Figure 6. The textile industry is representative of traditional industry and is the largest industrial sector in China in the sample period. Its geographic concentration pattern, shown in the left graph of Panel A, demonstrates that textile production is widely distributed all over the country, except for in the Tibet area, where the road access was quite limited in 2000. The right graph in Panel A displays the change of LQ in 2000-2009 for the textile industry. The blue dots denote the production concentrations at the county level. The emergence of textile concentration is more likely to be witnessed in central China, which has an extensive expressway network and thus better market access. Historically developed regions such as Beijing-Tianjin-Hebei area and the pearl-river delta in Guangdong province are losing the traditional textile industry.

[Insert Figure 6 here]

On the other hand, the sizzling hot telecommunications and computer manufacturing industry, the representative for emerging industries, was majorly concentrated in the southern China in 2000, but it has since experienced a spurt in well-accessed central and east coastal regions like Wuhan and Shanghai. The capital-intensive industry does favor regions with good access to large markets.

Notwithstanding the various concentration patterns across manufacturing industries, one common feature in the remarkable transformation of economic geography is that the concentration has risen in regions with good market access, whether they are inland or coastal, and whether they include first-tier cities or not. It indicates that the market access has a crucial role in the geographic concentration of manufacturing industries that are more dependent on freight transportation along expressway network.

6. CONCLUSION

Of all the massive transport-infrastructure projects around the globe in recent years, the National Expressway Network in China is the mega-project that has attracted public and academic attention due to its ambitious scale, enormous investment and, most importantly, sophisticated and complete network. However, its network effect on industrial development has rarely been investigated in literature. This research makes an initial attempt to study how the construction of the expressway network improves county-level market access, a deliverable identified in Donaldson and Hornbeck's (2016) general equilibrium trade model. It also considers how ameliorated market access influences industrial development from both the extensive and the intensive margin. In addition, this study provides a practical way of estimating the power-decay parameter of market access in the Chinese context.

The major caveats to the empirical estimation are the endogenous placement decisions regarding expressways. We construct a novel panel IV with reference to the minimal spanning tree method in graph theory to model the actual expressway market access, and address the endogenous selection of target cities in MST by utilizing historical information about walled cities in the 1820s. The resulting IV estimates show that increased market access significantly fosters the capital accumulation of individual firms and the growth of firm numbers in counties. Capital-intensive manufacturing industries benefit more from incremental increases in market access.

The finding implies that an expressway is a critical infrastructure element that helps manufacturing industries, especially in accessing the market. It promotes the industrial transformation to capital-insensitive industries. This research also sheds lights on the radical change of economic geography in China since the new millennium. With the improvement of market access, the inland provinces and regional cores, such as provincial capitals or urban districts within prefectures, receive larger amounts of industrial production and numbers of firms. This finding is consistent with the anecdotal “concurrent trends of industrial urbanization and inland migration of manufacturers” narrative.

REFERENCE

Alder, S. (2014). *Chinese roads in India: The effect of transport infrastructure on economic development*. University of North Carolina at Chapel Hill.

Allen, T., & Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, *129*(3), 1085–1140.

Angrist, J. D., & Pischke, J. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton: Princeton University Press.

Banerjee, A., Duflo, E., & Qian, N. (2012). *On the road: Access to transportation infrastructure and economic growth in China* (No. w17897). Cambridge: National Bureau of Economic Research.

Baum-Snow, N. (2007). Suburbanization and transportation in the monocentric model. *Journal of Urban Economics*, *62*(3), 405–423.

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

Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2018). Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*, https://doi.org/10.1016/j.jue.2018.05.001.

Byford, M. C., & Gans, J. S. (2014). Collusion at the extensive margin. *International Journal of Industrial Organization*, *37*, 75–83.

Behrens, K., Duranton, G., & Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, *122*(3), 507–553.

Blundell, R., Bozio, A., & Laroque, G. (2011). Labor supply and the extensive margin. *The American Economic Review*, *101*(3), 482–486.

Brandt, L., & Morrow, P. M. (2017). Tariffs and the organization of trade in China. *Journal of International Economics*, *104*, 85–103.

Brandt, L., Van Biesebroeck, J., & Zhang, Y. (2014). Challenges of working with the Chinese NBS firm-level data. *China Economic Review*, *30*, 339–352.

Bunn, A. G., Urban, D. L., & Keitt, T. H. (2000). Landscape connectivity: a conservation application of graph theory. *Journal of Environmental Management*, *59*(4), 265–278.

Burger, M., Oort, F. van, & Linders, G.-J. (2009). On the specification of the gravity model of trade: Zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis*, 4(2), 167–190.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *The American Economic Review*, *98*(4), 1707–1721.

Coşar, A. K., & Demir, B. (2016). Domestic road infrastructure and international trade: Evidence from Turkey. *Journal of Development Economics*, *118*, 232–244.

Coşar, A. K., & Fajgelbaum, P. D. (2016). Internal geography, international trade, and regional specialization. *American Economic Journal: Microeconomics, 8*(1), 24–56.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, *1*(1), 269–271.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review, 108*(4–5), 899–934.

Donaldson, D., & Hornbeck, R. (2016). Railroads and American economic growth: A "market access" approach. *Quarterly Journal of Economics, 131*(2), 799-858.

Du, R., & Zhang, J. (2017). *Walled Cities and Urban Density in China*. Clark University.

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. (2015). *Urban land use*. Handbook of Regional and Urban Economics (Vol. 5). Elsevier.

Duranton, G., & Turner, M. A. (2012). Urban growth and transportation. *The Review of Economic Studies*, *79*(4), 1407–1440.

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, *70*(5), 1741-1779.

Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. *Journal of the American Planning Association*, *76*(3), 265–294.

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

Fretz, S., Parchet, R., & Robert-Nicoud, F. (2017). *Highways, Market Access, and Spatial Sorting* (No. DP12437). CESifo Group Munich.

Frye, D. (2016). *Transportation Networks and the Geographic Concentration of Employment*. Vassar College.

Fujita, M., Krugman, P. R., & Venables, A. J. (2001). *The spatial economy: Cities, regions, and international trade*. Cambridge, MA: MIT Press.

Glaeser, E. L., & Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature*, *47*(4), 983–1028.

Groves, T., Hong, Y., McMillan, J., & Naughton, B. (1994). Autonomy and incentives in Chinese state enterprises. *The Quarterly Journal of Economics*, *109*(1), 183–209.

Hanson, G. H. (2005). Market potential, increasing returns and geographic concentration. *Journal of International Economics*, *67*(1), 1–24.

Harris, C. D. (1954). The market as a factor in the localization of industry in the United States. *Annals of the Association of American Geographers, 44*, 315.

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

Ioannides, Y. M., & Zhang, J. (2017). Walled cities in late imperial China. *Journal of Urban Economics*, *97*, 71–88.

Jing, K., & Liao, W.-C. (2018). *The butterfly effect of transport network connectivity*. Manuscript.

Krugman, P. (1991a). Increasing returns and economic geography. *Journal of political economy*, *99*(3), 483–499.

Krugman, P. (1991b). History and industry location: The case of the manufacturing belt. *The American Economic Review, 81*(2), 80–83.

Kruskal, J. B. (1956). On the shortest spanning subtree of a graph and the traveling salesman problem. *Proceedings of the American Mathematical Society*, *7*(1), 48–50.

Levitt, S. D. (1997). Using electoral cycles in police hiring to estimate the effect of police on crime. *The American Economic Review, 87*(3), 270–290.

Li, D., Lu, Y., & Wu, M. (2012). Industrial agglomeration and firm size: Evidence from China. *Regional Science and Urban Economics*, *42*(1–2), 135–143.

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

Lin, Y. (2017). Travel costs and urban specialization patterns: Evidence from China's high-speed railway system. *Journal of Urban Economics*, *98*, 98–123.

Melo, P. C., Graham, D. J., & Brage-Ardao, R. (2013). The productivity of transport infrastructure investment: A meta-analysis of empirical evidence. *Regional Science and Urban Economics*, *43*(5), 695–706.

Nie, H., Jiang, T., & Yang, R. (2012). A review and reflection on the use and abuse of Chinese industrial enterprises database. *World Economy*, (5), 142–158.

Qin, Y. (2017). "No county left behind?": The distributional impact of high-speed rail upgrades in China. *Journal of Economic Geography*, *17*(3), 489–520.

Redding, S., & Venables, A. J. (2004). Economic geography and international inequality. *Journal of international Economics*, *62*(1), 53–82.

Salon, D., Boarnet, M. G., Handy, S., Spears, S., & Tal, G. (2012). How do local actions affect VMT? A critical review of the empirical evidence. *Transportation Research Part D: Transport and Environment*, *17*(7), 495–508.

Santos Silva, J. M. C., Tenreyro, S., & Wei, K. (2014). Estimating the extensive margin of trade. *Journal of International Economics*, *93*(1), 67–75.

Storeygard, A. (2016). Farther on down the road: transport costs, trade and urban growth in Sub-Saharan Africa. *The Review of Economic Studies*, *83*(3), 1263–1295.

The Ministry of Transport. (2004). *National Trunk Highway System Planning*. China: The Ministry of Transport.

The Ministry of Transport. (2013). *National Highway Network Planning (2013–2030).* China: The Ministry of Transport.

Volpe Martincus, C., & Carballo, J. (2008). Is export promotion effective in developing countries? Firm-level evidence on the intensive and the extensive margins of exports. *Journal of International Economics*, *76*(1), 89–106.

Woodridge, J. M. (2005). Instrumental variables estimation with panel data. *Econometric Theory, 21*(4), 865–869.

Woodridge, J. M. (2006). *Introductory Econometrics: A modern Approach*. Third Edition. Thomson South-Western.

World Bank. (2007). *Domestic trade impacts of the expansion of the national expressway network* *in China* (East Asia Region Transport Sector working paper: No. 14). Washington, DC: World Bank.

World Bank. (2014). *Regional Economic Impact Analysis of High-Speed Rail in China* (No: ACS9734). Washington, DC: World Bank.

World Bank. (2017). *Global Mobility Report*. Washington, DC: World Bank.

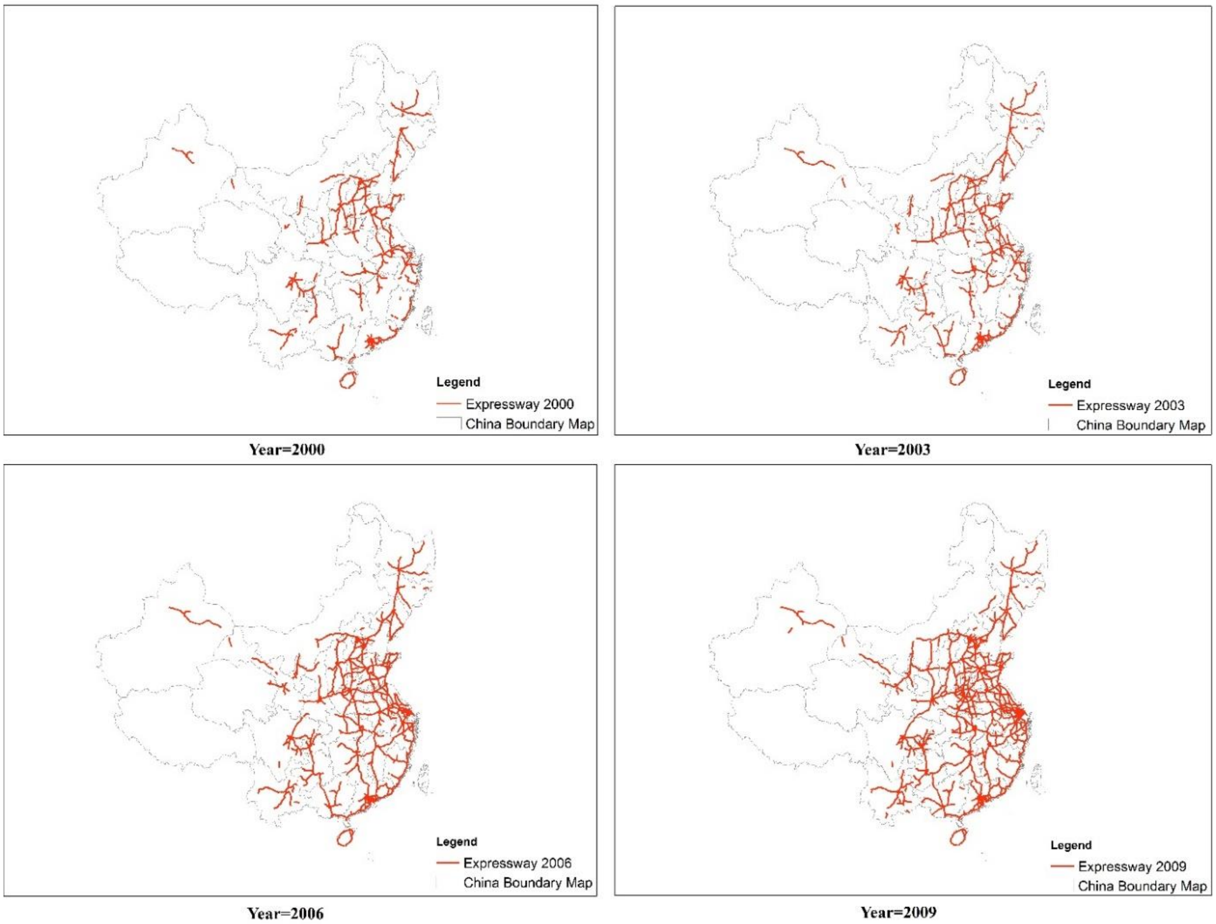


Figure 1. Expressway Network in China

Note: The red lines represents the national expressways. We additionally apply a typology check for the network to ensure road segments are correctly connected at the nodes.

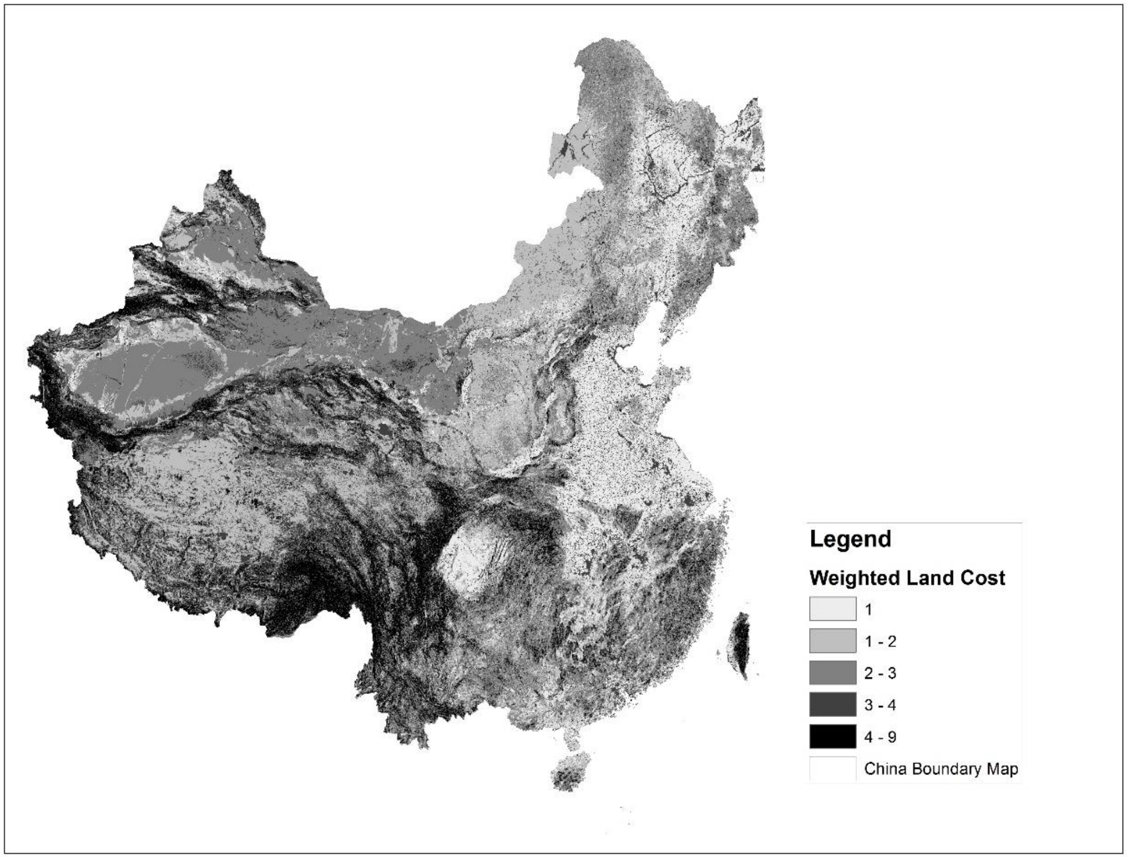
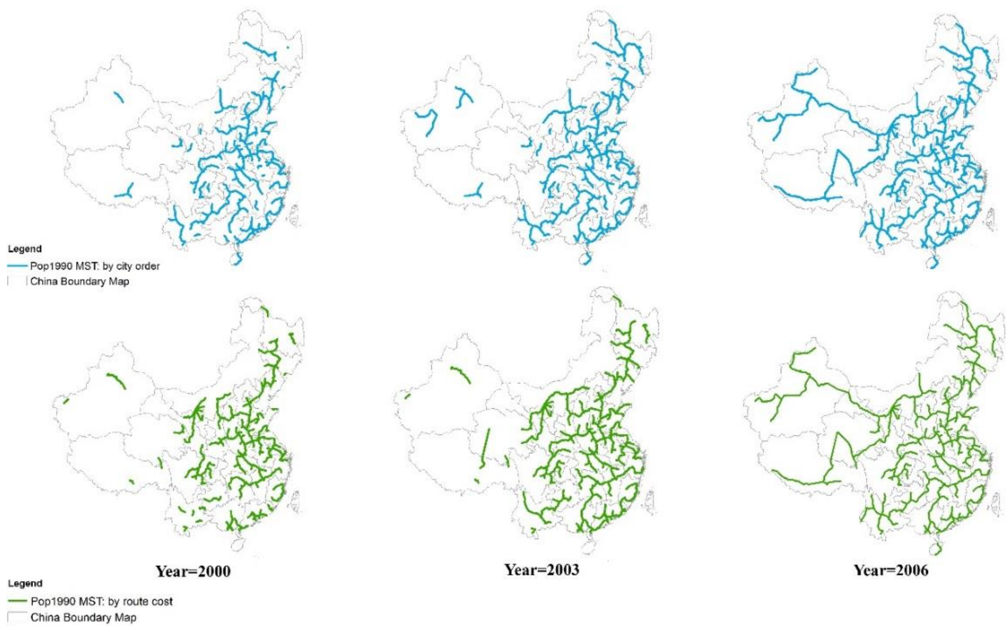


Figure 2: The Weighted Cost Surface of China

Note: This figure displays the weighted cost for every1km×1km land parcel in China. The weighted cost is calculated using China's slope and land use layer. The weight for these two layer is 6:4 respectively.

Panel A. Pop1990 Minimal Spanning Tree



Panel B. Wall-City Minimal Spanning Tree

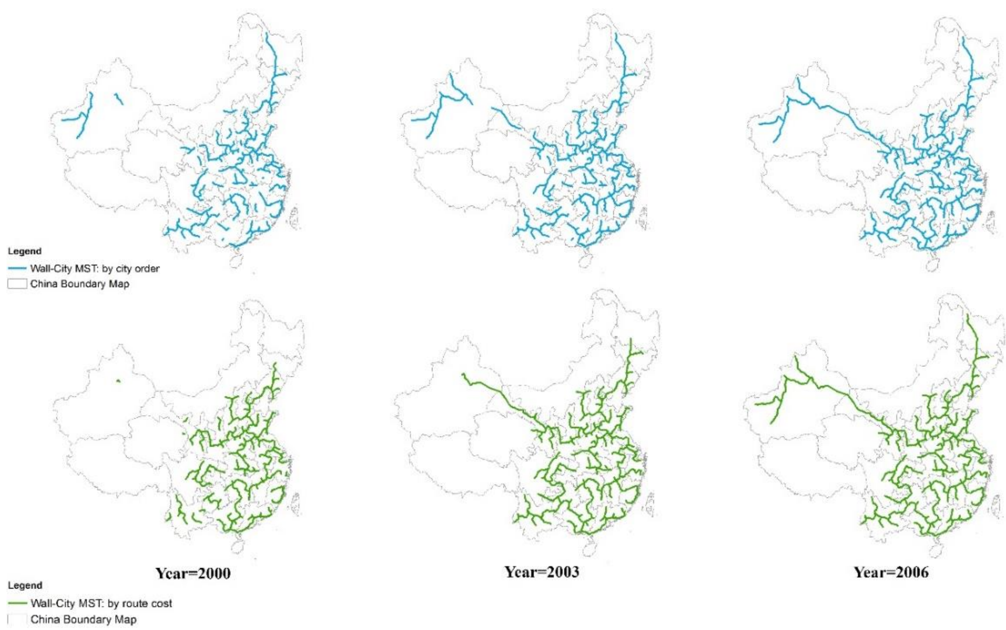


Figure 3. Pop1990 MST and Wall-City MST

Note: Panel A shows the dynamic Pop1990 MST from 2000 to 2006 and panel B depicts the Wall-City MST expansion. The blue routes denote the road expansion in the descending order of city's importance, while the green routes denotes the expansion in ascending order of construction cost. We obtain the subset of MST for each year based on the actual length of expressway network. As the constructed MST is shorter than the real network, thus length of the full MST network is only comparable to the length of real network in 2006. There will be no change of MST network from 2006 to 2009.

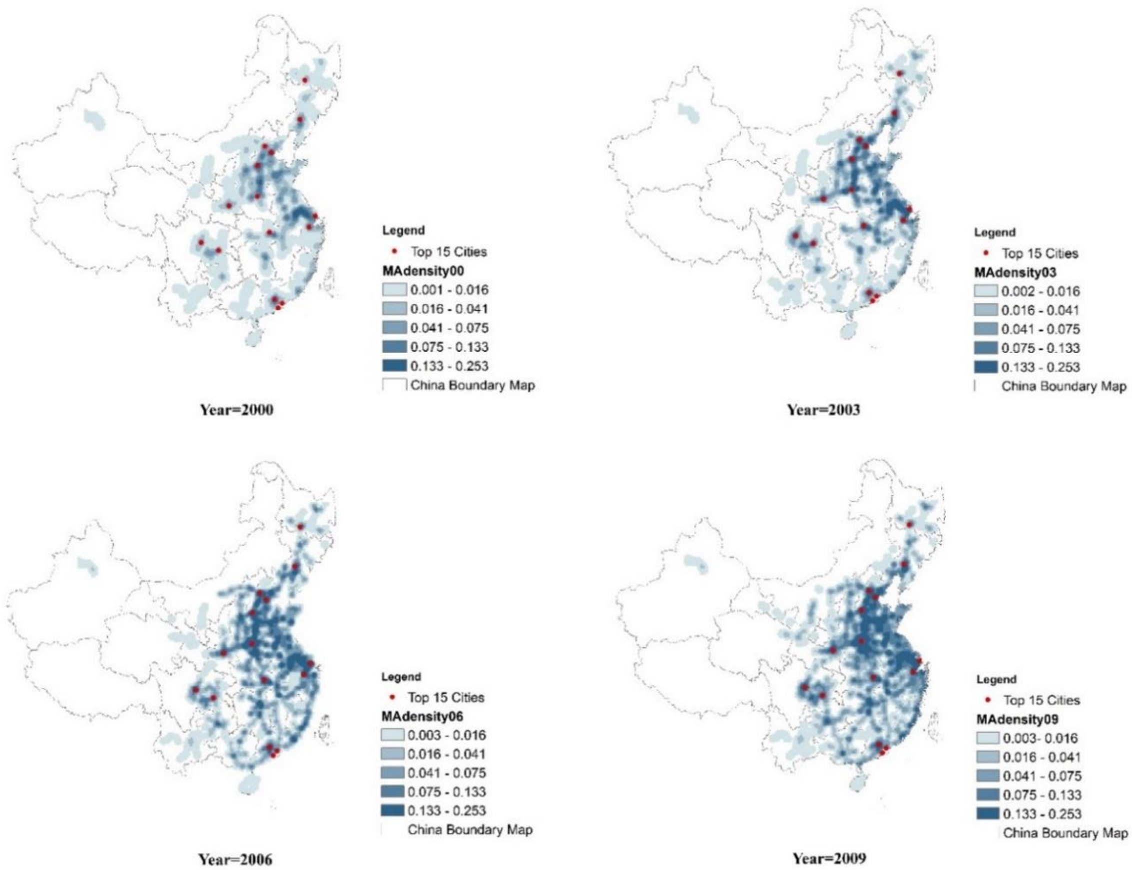


Figure 4: Market Access from 2000-2009

Note: The blue area denotes the density of market access. The darker the color, the higher the market access density. Market access values in each subgraph have been categorized to 5 groups by the same threshold value for the sake of compression.

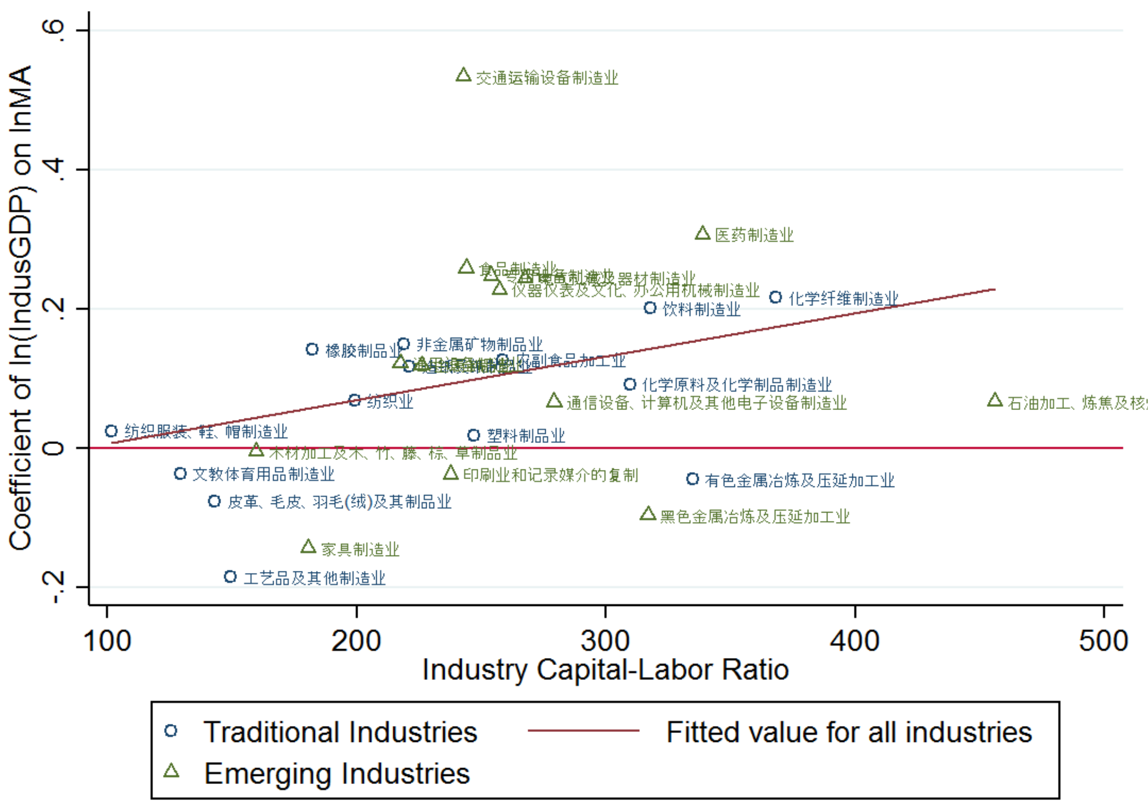
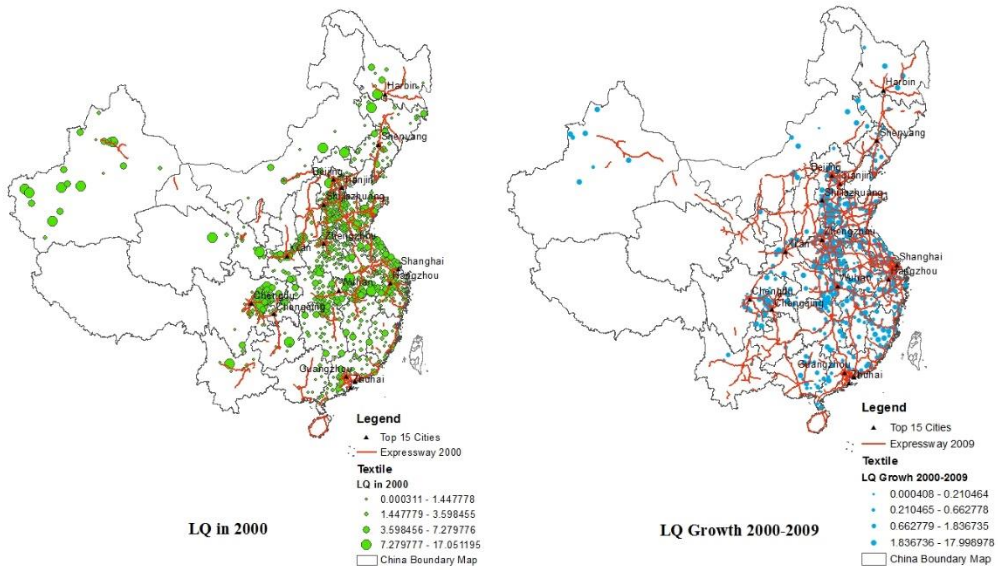


Figure 5: Heterogeneous Response to Market Access across Industries

Note: Each dot (circle or triangle) represents an estimated impact of market access growth on industrial GDP growth for one manufacturing industry.

Panel A: Traditional Industry: Textile



Panel B: Emerging Industry: Telecommunication, Computer and Other Devices

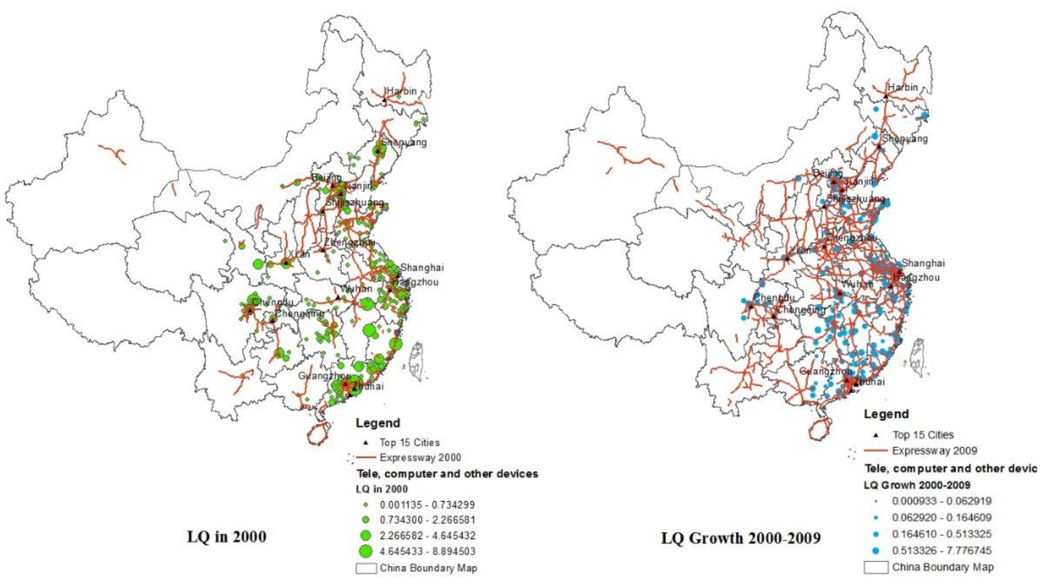


Figure 6: Initial LQ and LQ Growth of Typical Industries

Note: The left graph (green bubbles) in each panel exhibits the initial LQ in 2000, while the right graph (blue dots) shows the LQ growth from 2000 to 2009. For better visualization, only counties with positive LQ growth are plotted. The blue dot suggests an increase in LQ from 2000, and equivalently the industry is more concentrated in a county.

Table 1: Summary Statistics

County Characteristics	2000			2003			2006			2009		
Industrial Outcome	**Obs**	**Mean**	**S.d.**	**Obs**	**Mean**	**S.d.**	**Obs**	**Mean**	**S.d.**	**Obs**	**Mean**	**S.d.**
County sum: GDP from Industrial Firm[a] ('000,000)	1,876	624.83	1,700.00	1,892	1,273.00	3,341.00	1,814	1,459.00	3,668.00	1,822	2,272.00	5,298.00
County sum: Firm Number	1,876	43.71	78.97	1,892	60.49	125.00	1,814	73.48	137.90	1,822	69.77	128.50
County mean: Employment	1,876	229.80	102.80	1,892	209.50	82.77	1,814	147.40	58.64	1,822	170.80	75.62
County mean: Product Value ('000,000)	1,876	195.23	108.07	1,892	277.19	141.17	1,814	376.32	214.78	1,822	756.40	429.07
County mean: Capital ('000,000)	1,876	274.54	154.86	1,892	318.21	177.34	1,814	270.52	157.80	1,822	506.40	333.78
County mean: TFP	1,876	1.04	0.14	1,892	1.04	0.16	1,814	1.05	0.17	1,822	1.20	0.27
County-Level Market Access												
MA	1,016	44.15	28.41	1,167	76.17	41.98	1,402	104.96	47.55	1,562	117.30	48.70
MA_pop90^{order}	983	111.70	81.74	1,106	114.70	80.34	1,193	13 4.10	79.44	1,193	135.80	81.23
MA_pop90^{cost}	960	117.40	84.85	1,101	127.20	77.52	1,193	133.20	79.24	1,193	134.80	81.04
$MA\_wallcity^{order}$	865	38.58	20.68	946	52.90	28.01	981	68.05	23.55	981	69.25	24.01
$MA\_wallcity^{cost}$	894	57.29	24.82	963	67.52	22.40	981	67.92	23.42	981	69.11	23.87
Population		**1990**										
Population ('000)	2,224	417.30	318.00									
Estimated Population ('000)	1,780	462.10	111.40									

Note: [a] GDP from industrial firms is calculated from the aggregated value added of all the industrial firms in the county.

Table 2: OLS Regression

	(1)	(2)	(3)	(4)	(5)	(6)
Data Structure	Aggregated	Aggregate	Mean	Mean	Mean	Mean
	Panel A: Two-way Fixed-Effect Model					
D.V.	$\ln IndusGDP$	$\ln Firm\#$	$\ln Product$	$\ln Emply$	$\ln Capital$	$\ln TFP$
$lnMA$	0.0179***	0.0105*	0.0022	0.0019	0.0058	0.0001
	(0.0066)	(0.0063)	(0.0037)	(0.0035)	(0.0041)	(0.0021)
Observations	7,006	7,016	7,016	7,016	7,016	7,016
R-squared	0.9495	0.8858	0.8864	0.8396	0.8327	0.6000
Other Controls[a]	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES
Province-Year FE	YES	YES	YES	YES	YES	YES
	Panel B: First-Difference Model					
D.V.	$\Delta\ln IndusGDP$	$\Delta\ln Firm\#$	$\Delta \ln Product$	$\Delta\ln Emply$	$\Delta\ln Capital$	$\Delta\ln TFP$
$\Delta lnMA$	0.0196**	0.0117	0.0034	0.0051	0.0059	-0.0013
	(0.0089)	(0.0095)	(0.0047)	(0.0040)	(0.0047)	(0.0025)
Observations	5,156	4,973	5,167	5,167	5,167	5,167
R-squared	0.2886	0.4113	0.4494	0.4659	0.5458	0.3769
Other Controls	YES	YES	YES	YES	YES	YES
County FE	NO	NO	NO	NO	NO	NO
Province-Year FE	YES	YES	YES	YES	YES	YES

Note: [a] Other controls include county population and the share of total number of firms for each industry in that county. Same control variables are applied to the rest of regression tables unless specified otherwise. Standard errors are clustered at city-year level and are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Model (1) and (3)-(6) are weighted by the number of firms in the county.

Table 3: First-stage Regression

	(1)	(2)	(3)	(4)	(5)	(6)
D.V.	$lnMA$	$lnMA$	$lnMA$	$lnMA$	$lnMA$	$lnMA$
$lnMA\_pop90^{order}$	0.1333***		0.0833*			
	(0.0318)		(0.0426)			
$lnMA\_pop90^{cost}$		0.0666***	0.0405*			
		(0.0157)	(0.0211)			
$lnMA\_wallcity^{order}$				0.1045***		0.1182***
				(0.0270)		(0.0274)
$lnMA\_wallcity^{cost}$					0.1879**	0.2430***
					(0.0895)	(0.0897)
Observations	7,016	7,016	7,016	7,016	7,016	7,016
R-squared	0.8548	0.8548	0.8551	0.8545	0.8541	0.8550
Other Controls	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES
Province-Year FE	YES	YES	YES	YES	YES	YES

Note: Standard errors are clustered at city-year level and are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Regressions are weighted by the number of firms in the county.

Table 4: Second-stage Regression

	(1)	(2)	(3)	(4)	(5)	(6)
D.V.	$\ln IndusGDP$	$\ln Firm\#$	$\ln Product$	$\ln Emply$	$\ln Capital$	$\ln TFP$
Data Structure	Aggregated	Aggregated	Mean	Mean	Mean	Mean
	Panel A: Instruments from the Pop1990 MST					
$lnMA$	0.2831***	0.5055***	0.0907***	0.0206	0.0735***	0.0221**
	(0.0589)	(0.1369)	(0.0253)	(0.0183)	(0.0259)	(0.0109)
R-squared	0.9223	0.7328	0.8627	0.8372	0.8128	0.5785
Weak Identification Test [a]	16.56	11.64	16.60	16.60	16.60	16.60
Underidentification Test [b] P-value	0.000	0.000	0.000	0.000	0.000	0.000
Over-identification Test [c] P-value	0.176	0.747	0.944	0.000	0.102	0.181
	Panel B: Instruments from the Wall-City MST					
$lnMA$	0.1634***	0.4550***	0.0471**	0.0134	0.0583**	0.0094
	(0.0501)	(0.1314)	(0.0189)	(0.0172)	(0.0238)	(0.0092)
R-squared	0.9413	0.7625	0.8803	0.8387	0.8207	0.5962
Weak Identification Test	20.12	11.13	20.15	20.15	20.15	20.15
Underidentification Test P-value	0.000	0.000	0.000	0.000	0.000	0.000
Over-identification Test P-value	0.930	0.067	0.312	0.496	0.808	0.914
Observations	7,006	7,016	7,016	7,016	7,016	7,016
Other Controls	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES
Province-Year FE	YES	YES	YES	YES	YES	YES

Note: [a] The weak-identification test reports the correspondingly-robust Kleibergen-Paap robust rk Wald F statistic when clustered standard error is applied. The critical value to pass the weak-identification test is 11.59. [b] The null hypothesis for the underidentification LM test is that the equation is underidentified which means the excluded instruments are irrelevant. [c] In the case of clustered standard error, the Hansen J Statistic is report for the Sargan-Hansen test of overidentifying restrictions, of which the null hypothesis is the instruments are uncorrelated with the error term and exclusionary restriction is valid. Standard errors are clustered at city-year level and are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column (1) and (3)-(6) are weighted by the number of firms in the county.

Table 5: Market Access and Geographic Heterogeneity

	(1)	(2)	(3)	(4)	(5)	(6)
D.V.	$\ln IndusGDP$	$\ln Firm\#$	$\ln Product$	$\ln Emply$	$\ln Capital$	$\ln TFP$
Data Structure	Aggregated	Aggregated	Mean	Mean	Mean	Mean
IV	Wall-City	Wall-City	Wall-City	Wall-City	Wall-City	Wall-City
	Panel A: South-East Coastal Provinces					
$lnMA$	0.4178***	0.6474***	0.0882*	0.0515	0.1411***	0.0177
	(0.1307)	(0.2133)	(0.0474)	(0.0338)	(0.0520)	(0.0204)
$lnMA \times$ South-East Coastal	-0.3686***	-0.6668***	-0.0564	-0.0499	-0.1166**	-0.0101
	(0.1319)	(0.2193)	(0.0496)	(0.0386)	(0.0551)	(0.0216)
R-squared	0.9175	0.6838	0.8732	0.8326	0.7936	0.5905
	Panel B: First-Tier Cities					
$lnMA$	0.1633***	0.4252***	0.0483**	0.0158	0.0595**	0.0093
	(0.0501)	(0.1284)	(0.0189)	(0.0171)	(0.0239)	(0.0093)
$lnMA \times$ First-Tier City	0.1257	1.6106	-0.0437	-0.0121	0.1715	-0.0381*
	(0.2164)	(1.4026)	(0.0664)	(0.0583)	(0.1160)	(0.0197)
R-squared	0.9413	0.7589	0.8801	0.8383	0.8178	0.5967
	Panel C: Inland Provincial Capitals					
$lnMA$	0.1310**	0.3553**	0.0496**	0.0132	0.0478*	0.0077
	(0.0512)	(0.1612)	(0.0205)	(0.0182)	(0.0246)	(0.0094)
$lnMA \times$ Inland Provincial Capital [a]	0.3258**	0.3941	-0.0261	0.0033	0.1093*	0.0152
	(0.1534)	(0.2856)	(0.0523)	(0.0407)	(0.0571)	(0.0279)
R-squared	0.9385	0.7919	0.8800	0.8387	0.8176	0.5956
	Panel D: Inland Urban Districts					
$lnMA$	-0.0322	-0.1989	0.0253	-0.0077	0.0151	0.0052
	(0.0424)	(0.1714)	(0.0197)	(0.0183)	(0.0224)	(0.0102)
$lnMA \times$ Inland Urban District [b]	1.0675***	1.7607***	0.1308**	0.1298**	0.2560***	0.0214
	(0.2240)	(0.5518)	(0.0578)	(0.0517)	(0.0746)	(0.0235)
R-squared	0.9161	0.6856	0.8796	0.8295	0.8092	0.5973
Observations	7,006	7,016	7,016	7,016	7,016	7,016
Other Controls	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES
Province-Year FE	YES	YES	YES	YES	YES	YES

Note: [a] Inland provinces are provinces excluding the south-east coastal provinces and first-tier cities. [b] Inland urban districts are the urban-district counties in inland provinces. Standard errors are clustered at city-year level and are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column (1) and (3)-(6) are weighted by the number of firms in the county.

Table 6: Market Access and Location Quotient

	(1)	(2)	(3)	(4)	(5)	(6)
D.V.	$lnLQ$	$lnLQ$	$lnLQ$	$lnLQ$	$lnLQ$	$lnLQ$
IV	Wall-City	Wall-City	Wall-City	Wall-City	Wall-City	Wall-City
Sample	All Industries	Manufacturing	Mining and Utilities	Manufacturing	Manufacturing	Manufacturing
$lnMA$	0.1128***	0.1225***	0.0498	0.1119***	0.0671*	0.1001***
	(0.0331)	(0.0350)	(0.0851)	(0.0355)	(0.0355)	(0.0231)
$lnMA$ × High K/L Industry				0.0228*		
				(0.0127)		
$lnMA$ × Emerging Industry					0.1175***	
					(0.0128)	
$lnMA$ × Inland Provinces						0.0395*
						(0.0237)
Observations	111,733	94,156	17,468	94,156	94,156	115,282
R-squared	0.2322	0.2067	0.5057	0.2064	0.2051	0.2018
Other Controls	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES
Industry-Year FE	YES	YES	YES	YES	YES	YES

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Instrumental variable is the Wall-City IV.

APPENDICES

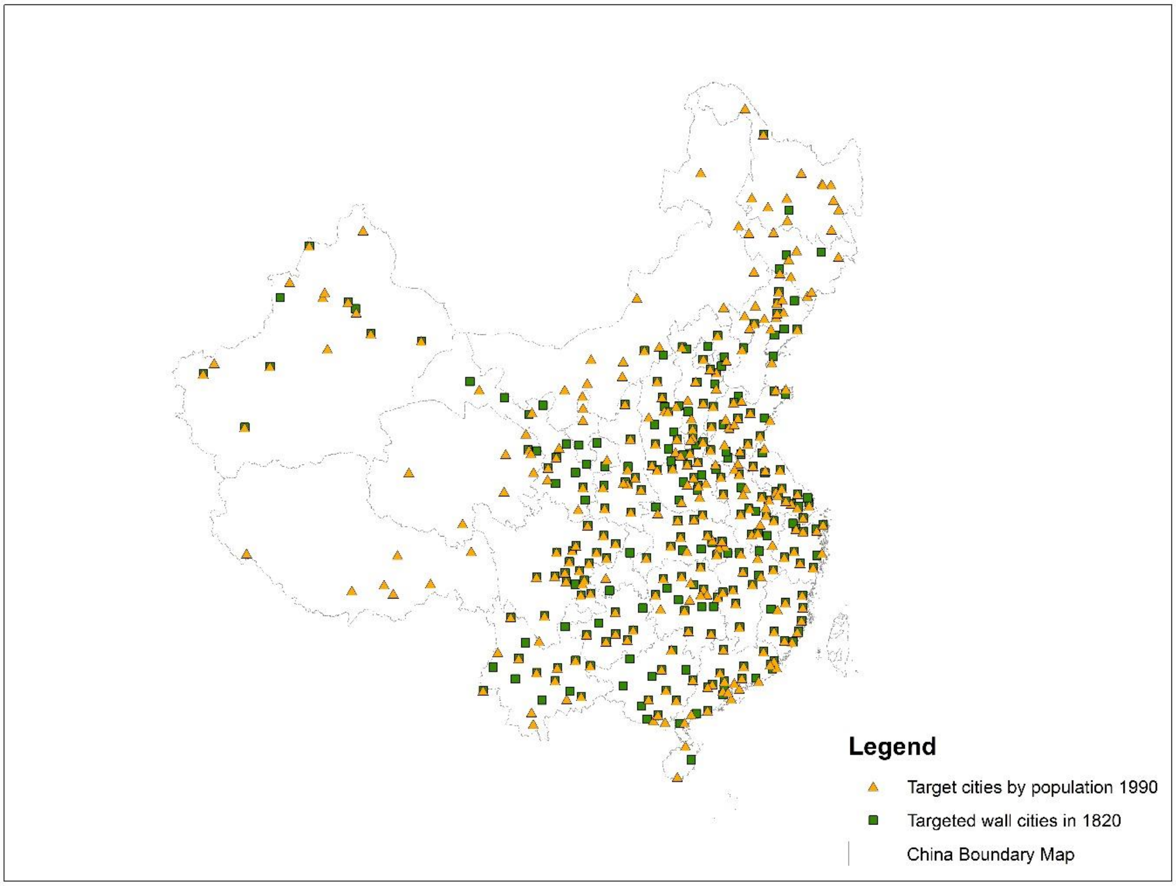


Figure A1. Target Cities for the MST Construction

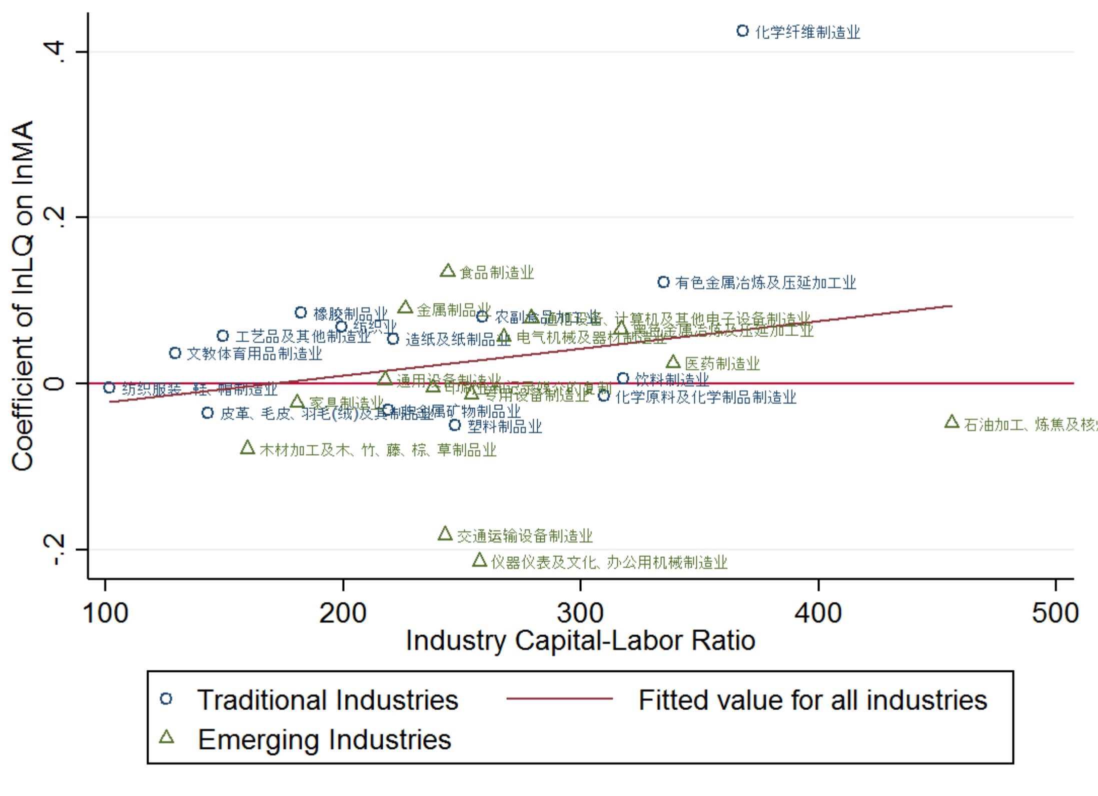


Figure A2: Geographic Concentration of Manufacturing Industries

Note: This figures plots the coefficients by regression lnLQ on lnMA for all the manufacturing industries.

Table A1: Sensitivity Test of the Power-Decay Parameter θ

	(1)	(2)	(3)	(4)	(5)	(6)
D.V.	$\ln IndusGDP$	$\ln Firm\#$	$\ln Product$	$\ln Emply$	$\ln Capital$	$\ln TFP$
Data Structure	Aggregated	Aggregated	Mean	Mean	Mean	Mean
	Panel A: Instruments from the Pop1990 MST					
$lnMA$ ($\theta = 1.07$)	0.2907***	0.5547***	0.0935***	0.0213	0.0759***	0.0227**
	(0.0614)	(0.1598)	(0.0262)	(0.0188)	(0.0269)	(0.0112)
R-squared	0.9205	0.6999	0.8610	0.8370	0.8112	0.5773
$lnMA$ ($\theta = 1.16$)	0.2691***	0.4311***	0.0856***	0.0194	0.0690***	0.0210**
	(0.0545)	(0.1063)	(0.0237)	(0.0174)	(0.0243)	(0.0104)
R-squared	0.9254	0.7762	0.8655	0.8376	0.8155	0.5807
	Panel B: Instruments from the Wall-City MST					
$lnMA$ ($\theta = 1.07$)	0.1681***	0.4876***	0.0486**	0.0139	0.0598**	0.0096
	(0.0516)	(0.1481)	(0.0194)	(0.0177)	(0.0245)	(0.0094)
R-squared	0.9407	0.7429	0.8798	0.8386	0.8199	0.5960
$lnMA$ ($\theta = 1.16$)	0.1546***	0.4070***	0.0440**	0.0125	0.0554**	0.0090
	(0.0475)	(0.1086)	(0.0180)	(0.0164)	(0.0225)	(0.0089)
R-squared	0.9425	0.7884	0.8813	0.8389	0.8221	0.5966
Observations	7,006	7,016	7,016	7,016	7,016	7,016
Other Controls	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES
Province-Year FE	YES	YES	YES	YES	YES	YES

Note: Standard errors are clustered at city-year level and are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Column (1) and (3)-(6) are weighted by the number of firms in the county.