Reassessing the Firm Selection Hypothesis: New Evidence from Chinese Highways*

Jiawei Chen Yi Niu Matthew Shum

November 2025

Abstract

This paper re-examines the firm selection hypothesis—that larger markets eliminate more low-productivity firms—by extending the multi-market framework of Combes, Duranton, Gobillon, Puga, and Roux (2012) to allow for asymmetric trade costs. A city with both a larger local market and better accessibility to other cities has a larger overall market size and therefore stronger selection. Exploiting substantial heterogeneity in highway access across cities during China's massive highway expansion, we test this prediction using firm-level manufacturing data and find stronger selection in large cities with highway access than in small cities without it. The results remain robust to alternative specifications and to controls for the potential endogeneity of highway placement. We further show that less productive firms are more likely to exit after gaining highway access, consistent with the selection mechanism. These findings highlight transportation infrastructure as a key determinant of market integration and competitive selection.

Keywords: firm selection, market size, firm productivity, highway access.

1 Introduction

In the literature in trade and spatial economics, an important hypothesis linking firm productivity to market size is that larger markets induce tougher competition, thereby raising the productivity cutoff for firm survival and eliminating more least-productive firms (Syverson, 2004;

*Chen: University of California, Irvine (Irvine, CA, USA), jiaweic@uci.edu. Niu: Capital University of Economics and Business (Beijing, China), yiniu@cueb.edu.cn; corresponding author. Shum: Caltech (Pasadena, CA, USA), mshum@caltech.edu. We thank Qi Li for generously sharing the firm-level data. We thank Yongheng Deng, Jindong Pang, Matt Turner, Jian Wang, Jia Yan, Junfu Zhang, and participants at the 2019 International Symposium on the Frontier of International Trade and Regional Science (Wuhan), the Urban Economics Association 2020 virtual conference, the Virtual Workshops on Topics in Regional and Urban Economics (Jinan-Peking University), Brookings-Jinan China Microeconomic Policy Forum 2021, the 4th International Workshop on Market Studies and Spatial Economics (Université Libre de Bruxelles & ECARES), and the 1st Summer Meeting in Urban Economics, China (Peking University) for helpful suggestions. Qinwen Deng, Wanxin Meng and Sinan Ni provided excellent research assistance.

Melitz and Ottaviano, 2008). This firm selection hypothesis has broad implications for understanding trade, geography, competition and efficiency (Melitz and Redding, 2014; Behrens and Robert-Nicoud, 2015; Proost and Thisse, 2019).

To empirically assess this hypothesis, Combes, Duranton, Gobillon, Puga, and Roux (hereafter CDGPR, 2012) develop an influential quantile approach that exploits the entire productivity distributions to examine whether, and to what extent, larger markets more strongly left-truncate the distribution than smaller ones, while controlling for the potential right-shift and dilation of the distribution arising, among other factors, from agglomeration economies. Applying this approach to manufacturing firms in France, they find little evidence that the productivity distribution in larger cities is more left-truncated than in smaller ones, a result seemingly at odds with the firm selection hypothesis. However, it remains unclear whether this result constitutes a genuine falsification of the hypothesis or instead reflects the restrictive nature of certain modeling assumptions, such as symmetric trade costs.

In this paper, we revisit the firm selection hypothesis. We extend the multi-market framework of CDGPR (2012) by allowing for asymmetric trade costs, and propose a simple yet general strategy for comparing market sizes: a city with both a larger local market (e.g., higher population) and better accessibility than another city can be regarded as having a larger overall market size[1] and, consequently, a stronger selection effect. Our strategy yields a testable theoretical prediction without imposing strong distributional assumptions and integrates naturally with the quantile approach, which requires distinguishing between large and small markets but neither relies on a continuous measure of market size nor assumes any specific productivity distribution.

We apply this strategy using firm-level data on China's manufacturing sectors.[2] Because controlled-access highways (hereafter *highways*, analogous to interstate freeways in the United States) play a crucial role in determining trade costs of manufactured goods, as discussed in Section 2, and nearly 80 percent of domestically traded goods were shipped by road, we treat cities located near highways as having better accessibility than those located substantially farther away. Using the same quantile approach as in CDGPR (2012), we find clear evidence of stronger firm selection in large cities with highway access compared to small cities without: the productivity distribution of the small cities exhibits a left tail of low-productivity firms that is absent from that of the large cities. These findings are robust to using a traditional measure of market access

[1] This comparison remains valid under many variants of market access or market potential that are either ad hoc or built on additional assumptions, and can also be applied in other empirical contexts related to market size.

[2] Because the firm selection mechanism requires competitive market forces to operate effectively, we restrict our analyses to 3-digit manufacturing industries whose employment share of state-owned enterprises was below 50 percent in 1998.

that reflects accessibility through the entire road network, as we detail in Section 6.6. Moreover, after a city gains highway access, less productive firms are more likely to exit, consistent with the firm selection mechanism from a dynamic perspective.

Our study exploits the substantial heterogeneity in highway access across Chinese cities during China's massive highway expansion from virtually none to a nationwide network. The total length of highways was only 0.5 thousand kilometers (km) in 1990 but grew rapidly to 4.8 thousand km in 1997, 53.9 thousand km in 2007, and 190.7 thousand km in 2024. Our study period, 1998–2007, lies in the middle of this expansion, providing sufficient firm observations from both large cities with highway access and small cities without it. At the beginning of 1998, 9.5 percent of Chinese cities had highways within their boundaries, while 59.7 percent were located more than 50 km away from a highway; by the end of 2007, these shares were 41.0 percent and 14.0 percent, respectively. This unique setting provides a rare opportunity to revisit the firm selection hypothesis by comparing the productivity distributions of cities with and without highway access. Such an empirical environment is difficult to replicate in most developed countries, where major highway networks were completed long before detailed firm-level census data became available, or in smaller countries, where cross-city variation in market accessibility is relatively limited.

This paper bridges two strands of literature which have previously developed somewhat distinctly. First, the literature on firm selection, which largely overlooked the role of transportation costs, has reported mixed results. On the empirical front, scant evidence of firm selection from city size has been found for manufacturing sectors in France, Japan, and Italy during the 1990s and 2000s (CDGPR, 2012; Kondo, 2017; Accetturo et al. 2018), while Syverson (2004) reports confirmatory evidence for firm selection from market size in his study of the ready-mixed concrete industry. Backus (2020), also studying the ready-mixed concrete sector, finds less evidence for firm selection; rather, the competition in local markets (measured by firm density) improves firm productivity but does not eliminate low-productivity firms. Our paper provides an explanation which reconciles these divergent results, from the perspective of transportation costs: France, Japan and Italy all have mature and well-developed nationwide transportation infrastructures that are not explicitly modeled in the analyses, while transportation costs for concrete are prohibitive even for small distances, leading to extremely localized and geographically segmented markets.

Second, several papers have likewise exploited the growth in transportation infrastructure in China to empirically test economic theories, but they have not looked at the firm selection

hypothesis. Looking more historically, Banerjee et al. (2020) document the relation between recent economic growth in Chinese regions to proximity to imperial-era transport hubs, finding that highways and railroads promote economic growth. However, Faber (2014) shows that increasing highway connections have *reduced* GDP growth in counties adjacent to new highways. Importantly, he uses a novel approach based on "shortest path" distances between cities to construct exogenous instruments for highway construction, which we will also use in this paper. Baum-Snow et al. (2020) report spatially contrasting effects from highways. Specifically, regional highways increase various economic outcomes, such as GDP, population, and wage, for larger "regional primate" prefectures but reduce these outcomes for smaller "hinterland" prefectures. These papers have not examined the relation between highways and firm productivity, which is the primary focus in this paper.

Three studies apply CDGPR's quantile approach and find certain evidence for the selection effect. Ding and Niu (2019) report stronger selection effect in larger provinces of China, Accetturo et al. (2018) find stronger selection in Italian cities and provinces with greater market potential, and Arimoto et al. (2014) document stronger selection in Japan's industrial clusters relative to other areas. Our study differs from them in several respects. First, we introduce asymmetric trade costs into the multi-market model of CDGPR (2012) without imposing additional strong assumptions, and derive a directly testable prediction. For comparison, Ding and Niu (2019) assume symmetric trade costs between provinces, while Accetturo et al. (2018) allow for asymmetric trade costs but restrict their analysis to a three-city model with identical city populations, making the prediction harder to test. Second, we focus on access to transportation infrastructure—a key determinant of domestic trade costs and market size—that all three studies overlook. Third, the trade barriers examined by Ding and Niu (2019) and Accetturo et al. (2018) are either unobservable (informal inter-provincial barriers in Ding and Niu) or fixed over time (Euclidean distance in Accetturo et al.), preventing them from examining the dynamics of firm selection as trade costs evolve. In contrast, we exploit variation from the expansion of China's highway network, which directly affects trade costs and firm exit over time. Finally, the selection effect found by Arimoto et al. (2014) is driven by competition in the input market, whereas our paper focuses on competition in the product market.

The remainder of the paper proceeds as follows. Section 2 describes the dramatic expansion of China's highway network and its impact on transportation costs. Section 3 develops our theoretical framework. Section 4 outlines our quantile-based empirical approach. Section 5 describes the data and variables. Section 6 presents productivity distributions and estimated

results of the quantile approach. Section 7 provides panel evidence on how highway access affects firm exit and productivity. Section 8 concludes.

2 Controlled-Access Highways and Transportation Costs

Controlled-access highways are vitally important for domestic trade within China; the share of domestically traded goods delivered via trucks has surpassed 70 percent since 1985, and reached 78 percent in 1998, the beginning of our study period.[3] Highway infrastructure in China has grown phenomenally within the past three decades. From 1990 to 2019, the length of all types of classified roads increased over six-fold, from 0.74 to 4.70 million km, with the fastest growth being in controlled-access highways, which are the largest roads in terms of both lanes and traffic.[4] From 1990 to 2019, the length of controlled-access highways increased around 300 times, from 0.5 thousand to 149.6 thousand km, with an annual growth rate of 21.7 percent. The increase in controlled-access highways was especially pronounced during our study period, growing from 4.8 thousand km in 1997 to 53.9 thousand km in 2007, a growth rate of 27.4 percent per year.[5]

Controlled-access highways differ from other roads mainly because they are designed exclusively for high-speed, unhindered vehicular traffic. In particular, they have no traffic signals, intersections or property access, do not allow pedestrians, non-motorized vehicles or farm machinery, have high-quality road surfaces, and allow vehicles to travel at high speeds up to 120 kilometers per hour (kph).[6] Not surprisingly, controlled-access highways have dramatically reduced travel time within China: during the 1980s, for instance, driving from Beijing to Tianjin (about 160 km) took approximately 7 hours, and it took almost one day to travel the 300 km to Shijiazhuang, a nearby provincial capital city. Two new controlled-access highways built in 1993 and 1994 cut the travel time to 1.5 hours and 3 hours, respectively. Similarly, the new controlled-access highways reduced driving times between Shanghai and Nanjing (1996; from 6-7 hours to

[3]Source: *China Statistical Yearbook*, various years.

[4]Classified roads in China, known as *dengji gonglu*, are divided into five classes, controlled-access highways and Classes 1-4. Controlled-access highways have divided lanes and full access control, and are restricted to automobile traffic. The designed average daily traffic (ADT) is 25,000 – 55,000 vehicles for four-lane controlled-access highways, 45,000 to 80,000 for the six-lane, and 60,000 to 100,000 for the eight-lane. First-class highways have divided lanes but not necessarily access control. Their designed ADT is 15,000 – 30,000 vehicles for four-lane highways, and 25,000 – 55,000 for six-lane highways. Second-class highways have two lanes and designed ADT of 3,000 – 7,500. Third-class highways also have two lanes but a lower ADT of 1,000 – 4,000. Fourth-class highways have two lanes or a single lane, with a designed ADT below 1,500 for the two-lane and below 200 for the single-lane.

[5]For comparison, in 2018, the length of controlled-access highways is 77,960 km in US and 78,097 km in EU.

[6]For comparison, the speed limit is 60-100 kph for Class 1 highways, 40-80 kph for Class 2, 30-60 kph for Class 3, and 20-60 kph for Class 4.

less than 3 hours), Shenyang and Panjin (2000; from 3 to 1 hour), and Changsha and Yongzhou (2003; from 6 to 3 hours).[7]

Controlled-access highways significantly lower the costs of shipping goods for a number of reasons. First, the higher driving speeds reduce time costs and hence wages paid to drivers. Second, the higher speed limit and fewer turns and brakes together increase vehicular fuel efficiency and reduce fuel costs per kilometer, as well as repair and maintenance costs. Third, the improved surfaces reduce transit damage (particularly for fragile items) and permit heavier loading of trucks which leads to larger scale economies in transportation. Fourth, accident rates are lowered largely due to the access control.

For instance, in a 1998 survey on the controlled-access highway connecting Shijiazhuang with Beijing, Wu (2005) finds that it reduced the costs of repair and maintenance by 77 percent and drivers' wages by 72 percent, as compared to a second-class highway (National Highway #107) which ran parallel to the controlled-access highway. Jia et al. (2004) find that, for a medium-sized truck running at the speed of 100 km/h, the Changping Highway in Liaoning province reduces fuel consumption by 30 percent compared to a parallel national highway #102.

3 Theory

In this section, we develop the theoretical background showing that a city with a larger population and better accessibility than another is expected to exhibit a larger market size and, consequently, a stronger selection effect. The analysis builds on the frameworks of Melitz and Ottaviano (2008) and CDGPR (2012).

There are I cities, and the population of city i is denoted by L_i. An individual consumer's utility is given by

$$U = q_0 + \alpha \int_{k\in\Omega} q^k dk - \frac{1}{2}\gamma \int_{k\in\Omega} \left(q^k\right)^2 dk - \frac{1}{2}\eta \left(\int_{k\in\Omega} q^k dk\right)^2 \tag{1}$$

where the individual's consumption of a homogeneous numeraire good, q_0, is assumed positive, q^k is the consumption of a variety k of differentiated goods from a set Ω, and α, γ and η are positive parameters representing the consumer's taste. Under utility maximization with budget

[7] Sources are as follows. Beijing-Tianjin: http://www.xinhuanet.com/politics/2018-08/06/c_1123270958.htm; Beijing-Shijiazhuang: http://m.xinhuanet.com/2018-07/13/c_1123121492.htm; Shanghai-Nanjing: http://58.213.139.243:8088/imgpath/zz3/1996-4/D1/D1.html; Shenyang-Panjin: http://news.sina.com.cn/china/2000-09-15/127175.html; Changsha-Yongzhou: http://news.sina.com.cn/o/2003-12-27/08271443738s.shtml; all accessed on November 1, 2020.

constraint, the consumer's demand for any variety k is

$$q^k = \begin{cases} \frac{1}{\gamma+\eta N}\left(\alpha + \frac{\eta}{\gamma}NP\right) - \frac{1}{\gamma}p^k, & \text{if } p^k \leq \overline{p} \equiv \frac{\gamma}{\gamma+\eta N}\left(\alpha + \frac{\eta}{\gamma}NP\right), \\ 0, & \text{if } p^k > \overline{p}. \end{cases} \tag{2}$$

where p^k denotes the price of the variety k, N denotes the measure of the set of varieties with $q^k > 0$ in equilibrium (Ω^*), $P = \frac{1}{N}\int_{k\in\Omega^*} p^k dk$ denotes the average price of these consumed varieties, and $\overline{p}$ is a choke price such that any variety more expensive than $\overline{p}$ is not purchased by any consumers.

The only input in the production is labor. The numeraire is produced competitively with the marginal cost of one and without any sunk costs. Shipping the numeraire anywhere is free. As a result, the cost of one unit of labor is unity. Shipping differentiated goods inside a city is also free.

Shipping differentiated goods between any two cities incurs iceberg trade costs. If a firm in city i delivers τ_{ij} (> 1) units of its products to city j, only one unit will arrive. The trade costs are pairwise symmetric, i.e., $\tau_{ij} = \tau_{ji}$ for all i and j, and they are allowed to differ across any pairs of cities.

As discussed in Section 2, highway access significantly lowers transportation costs. We model this by assuming a hierarchy of iceberg trade costs. Define Ψ as the subset of cities with highway access, and Φ as the subset of cities without. We assume the trade costs between any two cities that both have highway access are lower than the trade costs between any city with highway access and any city without highway access ($\tau_{ab} < \tau_{cd}$, $\forall a, b \in \Psi$, $\forall c \in \Psi$, $\forall d \in \Phi$), which are in turn lower than the trade costs between any two cities that both lack highway access ($\tau_{cd} < \tau_{ef}$, $\forall c \in \Psi$, $\forall d \in \Phi$, $\forall e, f \in \Phi$).

To produce one variety of the differentiated goods, a firm must make an irreversible investment f_E, and then randomly draws the marginal cost c from a common distribution with probability density function $g(c)$ and cumulative density function (CDF) $G(c)$. Therefore, the number of varieties the firm produces is one. If the marginal cost is greater than the choke price, $c > \overline{p}$, the firm cannot earn any operational profits and has to exit the market. If $c < \overline{p}$, the firm survives and produces. Therefore, the choke price $\overline{p}$ also serves as the marginal cost cutoff.

Suppose a survived firm located in city i with marginal cost c sells its products in city j under the price $p_{ij}(c)$, the demand from city j for its products is, according to equation (2), $Q_{ij}(c) = \frac{L_j}{\gamma}[\overline{p}_j - p_{ij}(c)]$, and the profit is $\pi_{ij}(c) = Q_{ij}(c)[p_{ij}(c) - \tau_{ij}c]$. The first-order condition for profit maximization yields the firm's optimal price for city j, $p^*_{ij}(c) = \frac{1}{2}\left(\overline{p}_j + \tau_{ij}c\right)$. At this optimal

price, the firm's operational profits from selling to city j are $\pi_{ij}^*(c) = \frac{L_j}{\gamma}[\overline{p}_j - p_{ij}^*(c)][p_{ij}^*(c) - \tau_{ij}c] = \frac{L_j}{4\gamma}[\overline{p}_j - \tau_{ij}c]^2$.

The free entry of new firms drives the ex-ante expected profits from producing differentiated goods in any city i to zero. Therefore, in equilibrium

$$\frac{L_i}{4\gamma}\int_0^{\overline{p}_i}(\overline{p}_i - c)^2 g(c)\,dc + \sum_{j\neq i}\frac{L_j}{4\gamma}\int_0^{\frac{\overline{p}_j}{\tau_{ij}}}\left(\overline{p}_j - \tau_{ij}c\right)^2 g(c)\,dc = f_E \tag{3}$$

where the two components on the left-hand side of the equation represent operational profits the firm earns from selling in the local city i, and from selling to other cities, respectively.

The equilibrium condition (3) endogenously determines the marginal cost cutoff in any city i, $\overline{p}_i$, which subsequently defines the exit rate of entrants $S_i = 1 - G(\overline{p}_i)$, or the strength of firm selection. It implies that, ceteris paribus, the strength of firm selection in a city depends on the sizes of cities and trade costs between all pairs of cities. We express the main theoretical result in the following proposition.

PROPOSITION 1: Consider any two cities b and d such that b has both a larger population ($L_b > L_d$) and better accessibility ($b \in \Psi$, $d \in \Phi$, so that $\tau_{bk} < \tau_{dk}$, $\forall k$). City b has a lower marginal cost cutoff ($\overline{p}_b < \overline{p}_d$) and therefore exhibits stronger firm selection ($S_b > S_d$).

See Online Appendix A for the proof.

4 Empirical Approach

Our primary methodology for testing the firm selection hypothesis (Proposition 1) is to compare the distributions of firm productivity in city b versus city d as in the proposition, and to examine whether the distribution in b exhibits stronger left-truncation than that in d. A credible test of firm selection should account for agglomeration economies arising from city size, which may also influence the productivity distribution. To address this concern, we employ the quantile approach developed by CDGPR (2012), and we briefly describe below.

Denote a firm's log productivity as $\phi = \log(1/c)$. Given the underlying CDF of marginal cost $G(c)$, the underlying CDF of log productivity is $\tilde{F}(\phi) = 1 - G(e^{-\phi})$. As discussed in CDGPR (2012), if city size generates agglomeration economies—such as knowledge spillovers among workers—that increase the productivity of all active firms in the city, the distribution of firm productivity shifts to the right; if more productive firms benefit more from agglomeration, the distribution also becomes more dilated. We can express the observed CDF of log productivity in

city i as

$$F_i(\phi) = \max\left\{0,\ \frac{\tilde{F}\left(\frac{\phi - A_i}{D_i}\right) - S_i}{1 - S_i}\right\} \tag{4}$$

where A_i represents the extent that agglomeration right-shifts the distribution; D_i is a dilation term which indicates whether more productive firms benefit more, the same, or less from agglomeration, depending on whether D_i is greater than, equal to, or less than one.

Figure 1 illustrates the combined effects of selection and agglomeration on the productivity distribution of active firms in a large city with highway access and a small city without highway access. In Panel A, firm productivities are observed without noise. Relative to the underlying distribution (dotted line), both city distributions are left-truncated due to firm selection and right-shifted due to agglomeration economies, although the magnitudes differ. The large city with highway access (solid line), according to Proposition 1, exhibits stronger selection and thus greater left truncation than the small city without highway access (dashed line). At the same time, stronger agglomeration in the large city produces a larger rightward shift. In Panel B, firm productivities are observed with noise, as is typical in empirical studies. While this obscures the left truncation, the large city still shows noticeably lower weights at the low-productivity end compared to the small city.

[Figure 1 here]

Since the underlying distribution $\tilde{F}(\phi)$ is unobservable, it is hard to estimate the absolute strength of firm selection—that is, the share of firm entrants eliminated by selection in city i, denoted S_i in equation (4). Nevertheless, under the assumption that $\tilde{F}(\phi)$ is identical across cities, we can estimate the relative strength of firm selection, that is, whether, and to what extent, city b eliminates more entrants than city d, as shown below.

Consider cities b and d as in Proposition 1, where $L_b > L_d$, $b \in \Psi$, $d \in \Phi$, and thereby $S_b > S_d$. The CDFs of firm productivity in the two cities are linked by the following transformation:

$$F_b(\phi) = \max\left\{0, \frac{F_d\left(\frac{\phi - A}{D}\right) - S}{1 - S}\right\} \tag{5}$$

where $S = (S_b - S_d)/(1 - S_d)$, $A = A_b - DA_d$, and $D = D_b/D_d$. These three parameters capture, respectively, the relative strength of selection, agglomeration and dilation in city b compared to city d. It suggests that dilating F_d by D, right-shifting it by A and left-truncating it by S will generate F_b.

To test Proposition 1, we will fit equation (5) to the estimated firm productivity distributions in large cities with highways vs. small cities without highways. We follow CDGPR (2012) and Gobillon and Roux (2010) to estimate the parameters of S, A and D, and obtain their standard errors using the bootstrap method.[8] We expect a stronger selection effect in large cities with highways, i.e., $\hat{S} > 0$, as illustrated by Figure 1. A key feature of this quantile approach is that it uses the entire productivity distributions, rather than fragmented information, to test the selection hypothesis.

5 Data and Variables

This paper primarily uses three datasets on the firm-level characteristics, the population size of cities, and GIS (geographic information system) data on locations and routes of controlled-access highways. We devoted substantial effort to minimizing potential errors in these data, as data noise could easily blur the estimated selection effect using the quantile approach (Ding and Niu, 2019).

5.1 Firm Productivity

The firm-level dataset derives from the Annual Survey of Industrial Firms (ASIF) during 1998-2007 conducted by China's National Bureau of Statistics (NBS). Included in the data are the so-called "above-scale" firms, which consist of all state-owned enterprises (SOEs) and those non-state-owned firms with sales exceeding 5 million RMB in the year. Above-scale firms account for the bulk of industrial activity. For example, in 2004, they employed about 78 percent of industrial workers and produced 90 percent of national industrial output. The dataset includes firm-level variables on output, capital, ownership, employment, intermediate input, value added, industry code, administrative division codes and detailed financial information. Online Appendix B details the procedures used to check and process the data, and it helps to reduce data noise. We drop firms in the sectors of mining and public utilities, and retain those in the manufacturing sectors for this research, which make up 90.6 percent of all above-scale firms in 1998 and 93.0 percent in 2007.

Using the above ASIF data, we measure firm productivity by estimating total factor productivity (TFP) at the firm-year level. Assuming that firms' production function takes a Cobb-

[8] We use the code provided at http://diegopuga.org/data/selectagg/ for the estimation.

Douglas form, we obtain the following estimating equation:

$$\log(V_{it}) = \beta_C + \beta_K \log(K_{it}) + \beta_L \log(L_{it}) + \mu_{it} \tag{6}$$

where L_{it} is employment of firm i in year t, V_{it} is value added, K_{it} is capital input. The ASIF data reports the employment at the end of the year and the average employment over the entire year, and we use the latter as L_{it}. Both value added and capital are measured after deflation, as described in Online Appendix B. To estimate coefficients in equation (6), we use the firm observations whose status is *normal operation* and exclude others that are temporarily closed, under preparation, or in other non-operational states. Using the approach developed by Ackerberg et al. (2015), we estimate function (6) separately by each of the 28 2-digit sectors, and thereby measure the TFP of firm i in year t by $\widehat{\mu}_{it} = \log(V_{it}) - \widehat{\beta}_C - \widehat{\beta}_K \log(K_{it}) - \widehat{\beta}_L \log(L_{it})$. Online Appendix C discusses the estimation of function (6) and displays the estimated results.

Finally, we take a two-year average of TFP for each firm; for instance, the TFP of firm i during 1998-1999 is computed as $\widehat{\mu}_{i,1998-1999} = (\widehat{\mu}_{i,1998} + \widehat{\mu}_{i,1999})/2$. Averaging across two years helps to average out year-by-year noise in TFP estimates and improves the precision of the firm selection estimates. As some firms own more than one establishment, we primarily use mono-establishment firms that account for 96.9 percent of all firm observations.

As is common in studies using firm-level data,[9] this paper excludes some small firms, as the ASIF data omit non-SOEs with annual sales below 5 million RMB. This sample restriction likely leads to an underestimation of the relative strength of firm selection, S in equation (5). Assuming a positive correlation between firm size and TFP, which we confirm in our data,[10] Ding and Niu (2019) demonstrate in their Figure 4 that omitting firms below a given size threshold reduces the observed difference in firm selection across two TFP distributions. In their simulations, Ding and Niu (2019) set $S = 0.1986$ and $log(size) = 10 + 0.7 \times log(TFP) + \epsilon$, $\epsilon \sim \mathcal{N}(0, 0.8)$, exclude the bottom 25 percent, 50 percent and 75 percent of firms by size, and obtain estimates of S (i.e., $\hat{S}$) equal to 0.0871, 0.0435 and 0.0164, respectively. In short, $\hat{S} < S$ when $S > 0$ and small firms are excluded. On the other hand, if $S = 0$, they find $\hat{S}$ very close to zero after excluding some small firms. Therefore, our estimated value for the relative strength of selection is likely a lower bound on the true value ($\hat{S} < S$), and a statistically significant $\hat{S} > 0$ provides evidence supporting the firm selection hypothesis.

[9] Typical challenges include low data quality among small firms, missing key variables, and lack of survey coverage.

[10] Using the ASIF data, regressing sales on TFP yields $log(sales) = 7.4818 + 0.6177 \times log(TFP)$ with $R^2 = 0.293$, where both coefficients are statistically significant at the 1 percent level.

5.2 Controlled-Access Highways

To construct the GIS data for controlled-access highways, we followed several steps. First, we obtained GIS data on China's controlled-access highways as of early 2008 from Baidu Map, along with three hardcopy editions of China's 2008 highway maps. These sources provide information on the routes and the full list of highways. Because the maps are not identical and contain varying errors, cross-checking among them helps to validate our highway data.

Second, because a highway (e.g., the Beijing–Shanghai Highway) is typically constructed simultaneously at multiple locations and opened to traffic segment by segment, we collected information on each segment's two endpoints, length, and date of opening to traffic. To obtain these data, we primarily relied on the following sources. The first is the Yearbook of China Transportation and Communications (YCTC, *Zhongguo Jiaotong Nianjian*) for various years, which provides detailed information on most highway segments completed before 1998 but covers fewer segments thereafter. The second source is the Newspaper and Journal Databases of the China National Knowledge Infrastructure (CNKI), which allow searches and downloads of articles from newspapers, magazines, and journals. Local newspapers and journals typically reported the opening of each highway segment, often including detailed segment-level information. For segments not covered by these sources, we searched online for additional reports and consulted the corresponding provincial gazetteers. Ultimately, we were able to identify the exact opening date for most highway segments, and for the remaining segments, at least the year and month——or, in some cases, the year alone. The length of a typical highway segment ranges from 50 to 200 km.

Third, we digitized the highway routes to create the GIS data. We began by downloading the GIS data on China's high-grade highways for 1999, 2005, and 2010 used in Baum-Snow et al. (2017, 2020) and employed it as the base layer. Because high-grade highways include not only controlled-access highways but also some first- and second-class highways, we removed the latter. Using the locations of the two endpoints and the length of each segment, we identified the corresponding segments in the Baum-Snow et al. data and verified their accuracy. Segments that were missing or deviated from their actual location by more than 3 km were manually redrawn in ArcGIS.

Finally, our highway data indicate that the total length of controlled-access highways at the end of 1999 and 2005 was 11,735 km and 40,386 km, respectively, closely matching the 11,605 km and 41,005 km reported by the NBS. Further details are provided in Online Appendix D.

5.3 Measuring City Size

In order to apply the quantile estimating equation (5), we need to construct a measure of city size in China. It involves two considerations: (i) how to define the spatial scope of a city, and (ii) which variable (e.g., population or employment) to use as a measure of city size.

For the first issue, although a city is usually considered as a unified labor market (Duranton, 2015), China has not delineated any boundaries of local labor markets such as metropolitan statistical areas in the US and employment areas in France identified based on commuting patterns. Fortunately, the urban area of each administratively designated *shi* (also known as administratively designated city) or administratively designated *xian* (also known as administratively designated county and county equivalent) roughly approximates an integrated labor market (Chan, 2007; Chen et al., 2024; Li and Mykhnenko, 2018). Therefore, we use the urban area in each *shi* or *xian* as the spatial scope of each city, and we are able to identify the *shi* or *xian* a firm is located in according to the county-level administrative code from the ASIF data.[11]

For the second issue, we prefer to use the urban part of the permanent resident population (PRP) in the years of national population censuses to measure city size. During a population census, the PRP is recorded through door-to-door household visits and defined as individuals who have resided in the location for at least six months during the year. In non-census years, PRP figures are estimated using sources such as the one-thousandth population survey. By contrast, *hukou* (registered) population data suffer from large discrepancies between actual and registered residence, and complete employment data are not available.[12]

Therefore, we measure the size of each city using the urban component of the PRP in 2000 from the fifth national population census in each *shi* or *xian* (referred to as a city in the remainder of this paper). This time point is in the middle of our primary study periods for the quantile approach: 1998-1999, 2000-2001 and 2002-2003. Using a minimum population threshold of 10,000, we identify 2208 cities in China, with an average population of 207,526. See more discussion on the measurement of city size in Online Appendix E.

[11] In 2000, there are 659 *shi*, including 259 prefecture-level cities and 400 county-level cities, and 1622 *xian*, including 1503 counties, 116 autonomous counties and 3 banners (data source: https://www.xzqh.org/html/show/cn/2000.html). Each prefecture-level city consists of at least one city district. Each city district, county-level city or *xian* corresponds to a unique county-level code, which is the first six digits of the administrative division code.

[12] Since 1998, China Urban Statistical Yearbook reports *danwei* employment (*danwei congye renyuan*) instead of total employment by *shi*. The *danwei* employment primarily involves employment in governments, as well as institutions, organizations, and enterprises sponsored by governments. The *danwei* employment in all prefecture regions that cover almost entire China accounted for only 12.1 percent of their total population in 2000, when the employment-to-population ratio in China was 74 percent according to the World Bank.

5.4 Grouping Cities

To test Proposition 1 using the quantile approach, we need to identify two groups of cities: small-population cities without highway connections (henceforth, *small unconnected cities*, corresponding to city d in Proposition 1 and equation (5)), and large-population cities with highway connections (henceforth, *large connected cities*, corresponding to city b in Proposition 1 and equation (5)). In the benchmark results, we use a 0.5 million population cutoff to distinguish small from large cities, and we will check the robustness of our results to the choice of the cutoff.

We define a city as having highway connections if a highway lies within its administrative boundaries one day before the beginning of the study period. Conversely, a city is considered without highway connections if it is located more than 50 km from the nearest highway to its boundaries by the last day of the study period. Because some isolated highway segments are short, particularly during the early period, we apply the following criteria: a highway crossing a city's boundaries must exceed 100 km for the city to be classified as having highway connections, and isolated highway segments shorter than 50 km are ignored when identifying cities without highway connections. To avoid misclassifying small cities that are effectively part of larger urban areas, we exclude small unconnected cities that share a boundary with a large city.

6 Main Results

Our empirical analyses use mono-establishment firms, drop trading companies,[13] and use the grouping of cities described in Section 5.4, unless otherwise stated. As the firm selection mechanism requires competitive market forces to operate effectively, we restrict our analyses to 3-digit manufacturing industries in which the employment share of state-owned enterprises was below 50 percent in 1998 in each industry. Firm observations in small unconnected cities decline rapidly with the extensive expansion of highways, so we primarily focus on the periods 1998-1999, 2000-2001 and 2002-2003, which provide sufficient observations in both city groups to implement the quantile approach.[14] We exclude four western provinces—Gansu, Qinghai, Xizang (Tibet), and Xinjiang—from our analyses due to their distinct geographic and economic characteristics.[15]

[13]Trading companies either purchase goods from domestic firms for export or import goods for sale to domestic firms. Following Ahn et al. (2011) and Yu (2015), we identify a firm as a trading company if its registered name contains any Chinese characters associated with trading, importing, or exporting—–such as, in pinyin (Romanized Chinese), "mao4yi4", "jin4chu1kou3", "jing1mao4", "gong1mao4", "wai4jing1", or "wai4mao4"—and does not include the Chinese character "chang3", meaning factory. The identified trade companies account for 1.01 percent of all firm observations.

[14]Gobillon and Roux (2010) recommend at least a few thousands observations in each group.

[15]These provinces are characterized by plateau environments, extremely low population density, and historical underdevelopment. While occupying 42.2 percent of China's land area, they accounted for only 6.9 percent of the

Since the provinces of Jiangxi, Hunan, and Hubei were heavily flooded in 1998, we also exclude them for the study period 1998-1999. Following CDGPR (2012), we trim one percent of the firm observations from each tail of the TFP distribution to remove outliers and normalize the average log TFP in small unconnected cities to zero by choosing units of value added. This normalization allows our estimates of A to be interpreted as the average productivity gain of firms in large connected cities relative to those in small unconnected cities.

6.1 TFP Distributions

We begin by presenting smoothed TFP distributions for large connected cities and small unconnected cities across three periods. According to the firm selection hypothesis, the TFP distribution of large connected cities should exhibit a stronger left truncation than that of small unconnected cities, as illustrated in Figure 1. We see some evidence of this in Figure 2: at low productivity levels, the distributions for large connected cities (solid line) have visibly thinner left tails than those for small unconnected cities (dashed line). At high productivity levels, by contrast, the distributions of large connected cities largely resemble right-shifted versions of those for small unconnected cities, consistent with the presence of agglomeration economies that we later control for.

Because we pool cities with different truncation points and measure TFP likely with noise (e.g., due to measurement error), Figure 2 does not display the sharp left truncation observed in Panel A of Figure 1. Instead, the observed distributions resemble those in Panel B of Figure 1, which incorporates the noise. The presence of such noise likely leads to an underestimation of the relative strength of selection, S in equation (5), but has much less influence on A, as suggested by Ding and Niu (2019).

[Figure 2 here]

6.2 Estimated Results of the Quantile Model

Table 1 reports estimates of the parameters in the quantile estimating equation (5) for the periods 1998-1999, 2000-2001 and 2002-2003. We begin with preliminary specifications which have only the selection effect (columns 1-2) and both selection and agglomeration effects (columns 3-5) to explain the differences in TFP distributions between small unconnected and large connected cities; finally, in columns 6-9, we estimate the full specification which allows for all three effects:

population and 2.8 percent of GDP in 2000, implying that they may violate the assumption of common underlying productivity distributions that underlies the quantile approach.

selection, agglomeration and dilation. The model fits the data well, as the pseudo-R^2 in column 9 remains 0.95 in different periods.

Table 1 presents strong evidence for firm selection. Across different periods, the estimated values of S remain positive and statistically significant (column 6); $\hat{S}$ ranges between 2.14 percent and 3.15 percent, suggesting that small unconnected cities need to left-truncate 2.14-3.15 percent of their incumbent firms to achieve the same strength of selection as large connected cities. If S is underestimated, the true relative strength of firm selection should be much stronger. This formalizes the earlier visual evidence from the graphs in Figure 2.

[Table 1 here]

Regarding the other parameters, the estimated coefficients of A (in column 7 of Table 1) are all positive and statistically significant, suggesting that large connected cities have right-shifted their TFP distributions relative to small unconnected cities, consistent with the hypothesis of agglomeration economies in the form of input sharing, labor pooling and knowledge spillovers.[16] During 2000-2001, for instance, the right-shift effect alone increases the average TFP by $e^{0.2264} - 1$, or 25.4 percent. $\hat{D}$ is close to one and not significantly different from one, and it suggests large connected cities right-shift different quantiles of the TFP distribution to a similar extent. Other factors that can be captured by right-shift and dilation include natural advantages that simultaneously enlarge city size and increase TFP of local firms, as well as special economic zones which are disproportionately located in large cities and which boost TFP (Lu et al., 2019).

An identification condition of the quantile approach is that firms entering different cities draw from the same underlying distribution of productivity. This condition might be violated if firms relocate from large cities to small ones after finding their ideal production process (Duranton and Puga, 2001), firms locate their headquarters in large cities and production facilities in small cities (Duranton and Puga, 2005), or larger cities or markets attract more productive firms or entrepreneurs (Behrens, et al., 2014; Gaubert, 2018). Such effects may not be strong given relatively low rates of firm relocation: 1.9 percent of establishments relocated across American counties during 2009-2013 (Rupasingha and Marre, 2020); 4.7 percent of establishments relocated across French employment areas during 1993-1996, and only 2-4 percent in most manufacturing sectors (Duranton and Puga, 2001). If these effects encourage more productive firms to choose larger cities, it should be largely captured by the rightward shift and dilation. Therefore, we do not further investigate which factors are driving the rightward shift and dilation, but instead treat A and D primarily as control variables to improve the validity of our estimates of S.

[16] See Ellison et al. (2010).

6.3 Controlling for Endogeneity in the Location of Highways

The estimated results in Table 1 may be plagued by an endogeneity problem as the location of highways is rarely random. Cities with more competitive market environments—characterized by stronger selection effects and fewer low-productivity firms——might be favored by provincial governments, which then prioritize highway construction through these cities to further stimulate growth. If so, the value of S in equation (5) could be overestimated. Conversely, under China's Coordinated Regional Development Strategy and the objective of common prosperity, the central government might direct highways toward low-income cities, which typically have a disproportionately large share of low-productivity firms, in order to support their development. Consequently, the value of S in equation (5) could be underestimated.

To accommodate this potential endogeneity, for the large connected cities, we will only keep those that had highway access due to largely exogenous reasons. In 1992, China's State Council approved the construction of a national highway network with seven horizontal and five vertical axes aiming to connect all provincial capitals to administrative cities (*"shi"*) with non-agricultural population exceeding 500,000. Highways connecting these major target nodes must pass through intermediate locations; thus, cities located along these efficient point-to-point routes are more likely to gain highway access for geographic and engineering reasons, rather than because of their pre-existing economic conditions or firm-level productivity profiles (Faber, 2014).

Given these considerations, we identify large cities with plausibly exogenous highway access using the following criteria: among large cities with highways within their boundaries, we retain only those (*i*) that were not target nodes in China's 1992 highway plan, and (*ii*) whose boundaries lay close to (within 25 km of) a hypothetically efficient highway network connecting the target nodes, proxied by the Euclidean Spanning Tree (EST) suggested by Faber (2014). In 1997, 56.6 percent of large connected cities are target nodes, 9.4 percent are not target nodes but located more than 25 km from the nearest EST line, and the remaining 34.0 percent are non-target nodes within 25 km from the nearest EST line, which are retained in the group of large connected cities for estimation. For 2000-2001, the corresponding shares are 46.8 percent, 20.8 percent, and 32.5 percent. For 2002–2003, they are 41.0 percent, 27.0 percent, and 32.0 percent, respectively.

[Table 2 here]

Table 2 reports the estimated results based on the redefined group of large connected cities described above. Compared to Table 1, the number of observations for large connected cities (column 6) becomes smaller. The group of small cities remains the same, as none of them had

a highway within 50 km during the study period, and obviously the number of small unconnected cities (column 5) remains unchanged. Consistent with Table 1, we continue to find strong evidence for firm selection, as $\hat{S}$ is positive and statistically significant in all three periods. Importantly, the magnitudes of $\hat{S}$ in Table 2 are larger than in Table 1—for example, increasing from 0.0315 to 0.0544 in 1998–1999, from 0.0457 to 0.0563 in 2000–2001, and from 0.0214 to 0.0234 in 2002–2003—indicating that endogenous highway placement leads to a systematic underestimation of S in equation (5). Overall, the evidence for selection is not only robust but in fact stronger once we address potential endogeneity in the placement of highways.

We illustrate below how to interpret the estimates in Table 2. During 1998–1999, to achieve the same level of firm selection as in large connected cities, small unconnected cities would need to left-truncate $\hat{S} = 5.44\%$ of their incumbent firms (the least productive ones). This truncation would raise TFP in small unconnected cities by 5.60 percent at the mean, 21.97 percent at the bottom quartile, and 4.56 percent at the top quartile. Although these effects can be viewed as a lower bound of the selection effect, they explain a non-trivial share of the observed TFP differences between the two groups of cities: the TFP of large connected cities exceeds that of small unconnected cities by 29.30 percent at the mean, 62.31 percent at the bottom quartile, and 21.47 percent at the top quartile.

During 2000-2001, similarly, left-truncating $\hat{S} = 5.63\%$ of incumbent firms in small unconnected cities would raise the average TFP by 5.76 percent. During 2002-2003, left-truncating $\hat{S} = 2.34\%$ of incumbent firms in small unconnected cities would raise the average TFP by 2.37 percent.

6.4 Alternative Specifications

We examine several alternative specifications to assess the robustness of the above findings. To save space, we focus on the period 2000–2001, the middle of the three study periods, which provides relatively abundant observations for both city groups. Unless otherwise noted, all specifications are identical to those in Table 2.

Because the baseline results use mono-establishment firms, we examine to what extent including all firms alters the estimates. The first row of Table 3 shows that $\hat{S}$ equals 0.0562, only marginally lower than the estimate of 0.0563 in Table 2. The standard error increases slightly, but the estimate remains highly significant.

[Table 3 here]

In the second row of Table 3, we re-estimate the quantile model without excluding any outlier provinces. This increases the number of observations for small unconnected cities but leaves that for large connected cities unchanged, as highways and large cities in the outlier provinces are scarce. The estimated value of S decreases slightly to 0.0527, with larger standard errors, but remains statistically significant at the one percent level.

In Tables 1 and 2, we use a population threshold of 0.5 million to distinguish between large and small cities. This choice is straightforward and ensures sufficient observations in both city groups, but we also examine alternative thresholds. The third row in Table 3 reports estimates using 0.75 million as the cutoff, and the fourth row reports estimates using 0.25 million. The value of $\hat{S}$ fluctuates slightly around 0.05, and its statistical significance remains high.

We measure city size using urban population in each *shi* or *xian*, but in several cases, the major urban areas in a few adjacent *shi* or *xian* have expanded across their administrative boundaries and became contiguous, suggesting that their labor markets might be highly integrated. In the fifth row of Table 3, we treat such integrated *shi* or *xian* as a single city (see Online Appendix E for details), use their combined urban population as the city's population, and regroup cities accordingly. The estimated results indicate that $\hat{S}$ is slightly lower than that in Table 2, and is still statistically significant.

In Table 2, large connected cities are required to be located within 25 km of an EST line. In the sixth row, we instead require that the EST line cross the boundaries of large connected cities. This restriction reduces the number of firm observations in large connected cities from 7,450 to 4,499, and the estimated $\hat{S}$ decreases to below 0.05 but remains statistically significant. In the last row, we relax the threshold and require large connected cities to be located within 50 km of an EST line, which increases the number of observations. The evidence supporting the selection hypothesis continues to hold.

6.5 The Effects of Accessibility

The stronger selection effect found in large connected cities relative to small unconnected cities results from the combined effects of larger population size and better accessibility. In this section, we seek to isolate, to some extent, the role of accessibility.

Building on the theory in Section 3, we can derive a proposition that, for any two cities with identical population, the one with better accessibility should have stronger firm selection than the other.[17] While it is not feasible to find two cities with identical populations, we can compare

[17]Consider any two cities b and d such that $L_b = L_d$. Let city b have better accessibility ($b \in \Psi$, $d \in \Phi$, so that

two groups of cities within a narrow population range so that their differences in selection can be largely attributed to accessibility. Specifically, we focus on cities with populations below 0.5 million and divide them into two groups. The first group consists of the cities without any highway within 50 km of their boundaries by the end of the study period, which are identical to the small unconnected cities in Tables 2 and 3. The second group consists of the cities that have highways within their boundaries by the beginning of the study period and are located within 25 km of the nearest EST line.

Table 4 reports the estimation results of equation (5) using this grouping. $\hat{S}$ remains positive and statistically significant during 1998-1999 and 2000-2001, suggesting that accessibility alone likely strengthens firm selection. Compared with Table 2, however, $\hat{S}$ declines from 0.0544 to 0.0205 during 1998-1999, from 0.0563 to 0.0487 during 2000-2001, and from 0.0234 to 0.0107 during 2002-2003, losing statistical significance in the last period. Thus, once differences in city size are minimized, the selection gap between the two groups narrows significantly. These findings suggest that neither accessibility nor city size is trivial in shaping firm selection.

[Table 4 here]

6.6 Using Market Access

In the theory, we assess a city's accessibility to other cities largely through its access to highways, which enables us to derive a testable proposition. Other factors, such as access to non-highway roads and intercity distance, may also influence accessibility, market size, and consequently, the strength of firm selection. To incorporate more of these factors, we measure city i's market access to other cities (MA_i) following Donaldson and Hornbeck (2016):

$$MA_i = \sum_{i \neq j} L_j H_{ij}^{-\theta} \tag{7}$$

where L_j denotes the population of city j in 2000, as discussed above. Since we cannot estimate iceberg trade costs like Donaldson and Hornbeck (2016), we follow Jedwab and Storeygard (2022) and use driving time between city i and j as H_{ij}. The parameter θ represents the elasticity of the intercity trade with respect to driving time.

To estimate driving time between cities, we generated the centroid of each city's urban area. We then constructed road networks at the end of 1997, 1999, 2001, 2003, 2005 and 2007. Each

$\tau_{bk} < \tau_{dk}$, $\forall k$). City b has a lower marginal cost cutoff ($\overline{p}_b < \overline{p}_d$) and therefore exhibits stronger firm selection ($S_b > S_d$).

network includes three types of roads. The first consists of controlled-access highways that were opened to traffic by the end of the respective year. The second comprises regular roads based on the digitized maps used in Baum-Snow et al. (2017, 2020), which we verified and refined. Since the digitized maps are available only for January 1999 and March 2005, we use the 1999 regular road data or the road networks in 1997, 1999 and 2001, and use the 2005 regular road data for the road networks in 2003, 2005 and 2007. The third type consists of straight lines connecting each city's centroid to the centroids of neighboring cities that share an administrative boundary.

Similar to Baum-Snow et al. (2020), we assume speeds of 90 kph on controlled-access highways, 30 kph on regular roads, and 15 kph on straight lines. Using ESRI's Network Analyst with the Djikstra algorithm, we computed the minimum driving time between city centroids along the road network, denoted H_{ij} in function (7). From the end of 1997 to the end of 2007, the average value of driving time between cities (H_{ij}) declines from 55.1 to 31.6 hours, and the driving time between Beijing and Shanghai falls from 33.7 to 15.5 hours. Further details are provided in Online Appendix F.

We set the value of θ in equation (7) following the literature. Duranton et al. (2014) estimate that the elasticity of intercity trade with respect to highway distance in the United States is between –1.28 and –1.41, while Fan et al. (2023) estimate that the elasticity of export shipments from cities to ports with respect to road distance in China is between –1.93 and –2.25. Using these four values of θ, we compute four corresponding measures of the market access. Although not directly derived from the theoretical model of firm selection, the market access is widely used in the literature and provides complementary evidence for the selection hypothesis.

Using a strategy consistent with the above analyses, in this section, we treat a city with both a larger local market and higher market access than another city as having a larger overall market size. Accordingly, we group cities as follows. First, cities in large markets have greater market access than those in small markets: all four measures of market access for large-market cities are above the 75th percentiles, whereas those for small-market cities are below.[18] Second, cities in large markets are located within a one-day drive of the nearest major port, whereas cities in small markets are located beyond a one-day drive.[19] This criterion helps capture accessibility to international markets. Third, consistent with previous analyses, large-market cities have larger

[18]For example, during 2000–2001, all four measures of market access for large-market cities must be above the 75th percentiles at the ends of 1999 and 2001, while those for small-market cities must be below the 75th percentiles at both time points.

[19]The nine Chinese ports handling the largest volumes of international trade during 1998–2007 are Shanghai, Ningbo, Guangzhou, Tianjin, Qingdao, Qinhuangdao, Dalian, Shenzhen, and Lianyungang, following Baum-Snow et al. (2020). Given that a truck driver typically drives 6–9 hours per day, we classify cities within five driving hours of the nearest port as within a one-day drive and those more than ten hours away as beyond a one-day drive.

populations than small-market cities, using 0.5 million as the cutoff.

[Table 5 here]

Table 5 presents the estimated results of equation (5) using the above grouping of cities. We find strong evidence for firm selection, as $\hat{S}$ is positive and statistically significant in all 5 periods. The first and last periods report relatively high values of $\hat{S}$, while the middle period (2002-2003) reports the lowest value. One of the potential reasons is that we use 1999 regular road data to proxy for 2001 and 2005 regular road data to proxy for 2003, which likely introduces additional noise that leads to an underestimation of the selection effect.

7 Highway Access, Firm Exit, and TFP Growth: Panel Evidence

While using the quantile approach ensures comparability of our results with the existing literature, it primarily exploits cross-sectional variation.[20] In this section, we exploit the panel dimension of our data to further investigate the selection mechanism. Similar to above analyses, we use 3-digit manufacturing industries whose employment shares of state-owned enterprises (SOEs) were below 50 percent in 1998.

7.1 Highway Access and Firm Exit

Since our identification of the relative strength of firm selection relies on the assumption that the underlying distribution of firm productivity, $\tilde{F}(\phi)$ in equation (4), is identical across cities, one may be concerned whether our estimated result, $\hat{S} > 0$, reflects differences in these underlying distributions rather than firm selection. To address this concern, we investigate an implication of the selection theory: if a city gains highway access that lowers trade costs with other cities, its TFP cutoff should rise, forcing some least productive firms to exit. We estimate the following model:

$$Exit_{i,c,t} = \beta_0 + \beta_1 Highway_{c,t-1} + \beta_2 Highway_{c,t-1} \times TFP_{i,t-1} + \beta_3 TFP_{i,t-1} + X'_{i,c,t}\gamma + \epsilon_{i,c,t} \quad (8)$$

where the dummy variable $Exit_{i,c,t}$ equals 1 if firm i in city c appears in year t in the ASIF data and disappears thereafter.[21] Highway access is measured by a dummy variable $Highway_{c,t-1}$,

[20] Using temporal variation to test the stronger left-truncation of the TFP distributions, however, may be more complex. For example, Melitz and Ottaviano (2008) show that in a three-city model, reducing trade costs between city 1 and city 2 can lower rather than raise the TFP cutoff in city 3.

[21] Because we cannot identify firms that exit in 2007, the last year covered by our ASIF data, we restrict the sample to 1998–2006 when estimating equation (8).

which equals one if a highway is located within city c in year $t-1$. $TFP_{i,t-1}$ denotes firm i's productivity in year $t-1$. The vector $X_{i,c,t}$ includes firm, city, and two-digit industry-year fixed effects, as well as firm characteristics such as an SOE dummy, a mono-establishment dummy, a trading firm dummy, firm age and its square, and the shares of paid-in capital from foreign sources, from Hong Kong, Macau, or Taiwan, from collectively owned entities, from individuals, and from legal persons, respectively.

If highway access strengthens firm selection, less productive firms should be more likely to exit once a city gains highway access. Therefore, we expect $\beta_1 > 0$ and $\beta_2 < 0$.

[Table 6 here]

The estimated results in Table 6 are consistent with the selection hypothesis. In columns (1)–(3), which include different sets of fixed effects and firm-level controls, the coefficients on the highway dummy are positive and statistically significant, while the coefficients on its interaction with $Log(TFP)$ are negative and significant. These results suggest that, following highway access, less productive firms are more likely to exit.

As discussed previously, highway placement may be endogenous. For instance, policymakers may have anticipated future trends in firm exit and prioritized highway construction in cities expected to experience higher exit rates. To address this concern, when estimating equation (8), we restrict the firm sample to cities that (1) obtained highway access between 1997 and 2006, (2) were not designated as target nodes in China's 1992 national highway plan, and (3) were located close to the EST lines (within 25 km of city boundaries). As discussed in Section 6.3, these cities likely gained highway access for largely exogenous reasons. By doing so, our estimation primarily relies on within-firm variation.

The results in columns (4) show that the coefficient on the interaction term remains negative and statistically significant, indicating that less productive firms are indeed more likely to exit following highway access. Moreover, the estimated coefficient is larger in magnitude than that in column (3), consistent with our earlier findings in Table 2 that the estimated selection effect becomes stronger once we address the endogeneity of highway placement.

7.2 Highway Access and TFP Growth

We investigate an alternative explanation to the main results $\hat{S} > 0$. If highway access disproportionately improves the productivity of firms below certain TFP threshold, it could also generate a pattern similar to left-truncation in the TFP distribution of large connected cities. To address

this concern, we estimate the following model:

$$\Delta TFP_{i,c,t} = \beta_0 + \beta_1 Highway_{c,t-1} + \beta_2 Highway_{c,t-1} \times TFP_{i,t-1} + \beta_3 TFP_{i,t-1} + X'_{i,c,t}\gamma + \epsilon_{i,c,t} \quad (9)$$

where $\Delta TFP_{i,c,t}$ is the growth rate of TFP, that is, $\Delta TFP_{i,c,t} = (TFP_{i,c,t+1} - TFP_{i,c,t})/TFP_{i,c,t}$. The other variables use the same definitions as equation (8).

We find no evidence that highway access increases firm productivity. In columns (1)-(3) of Table 7, we sequentially add some fixed effects and firm-level controls, yet the coefficients on the highway dummy and its interaction with $Log(TFP)$ remain small and statistically insignificant. This suggests that highway access does not significantly affect TFP. The negative coefficients on $Log(TFP)$ indicate that, on average, less productive firms experience faster TFP growth. In columns (4) and (5), we split the sample into small cities (population below 0.5 million) and large cities (population above 0.5 million), and re-estimate equation (9). The results again show no effects of highway access on TFP. We further test whether highway access disproportionately benefits low-productivity firms——those below the 10th percentile of the city's TFP distribution in a given year——and find no meaningful effects (column (6)). Finally, to mitigate concerns about endogenous highway placement, we restrict the sample to cities that gained highway access for exogenous reasons (the same set of observations as in columns (4) of Table 6) and re-estimate the models corresponding to columns (3)–(6). The results in columns (7)–(10) remain virtually unchanged. These findings are consistent with Ghani et al. (2016), who find that the effects of highway access on TFP are weak.

[Table 7 here]

8 Conclusions

In this paper we revisit the firm selection hypothesis and analyze the distributions of firm productivity between large cities with highway access and small cities without it. We find that large connected cities eliminate approximately 5 percent more least productive firms than small unconnected cities, thus supporting the firm selection hypothesis. The stronger selection raises the average TFP in large connected cities by about 5.5 percent, and these magnitudes should be interpreted as the lower bound of the true selection effect. Our further investigations suggest that the evidence for selection should not result from the endogeneity of highway placement or any particular model specification.

Complementing our cross-sectional quantile analyses, panel evidence shows that less productive firms are more likely to exit after their cities gain highway access, while the productivity of survivals remains largely unchanged. This pattern reinforces the interpretation that highways intensify competitive selection rather than shifting productivity through agglomeration.

This study is notable for several reasons. First, we provide strong evidence that market size generates the selection effect, which is quite debatable over the past decades. Second, our results imply that the lack of firm selection found in previous studies may result from the assumption of symmetric trade costs. We demonstrate how to incorporate asymmetric trade costs and generate a testable prediction in the structural model, and this framework can be applied to other empirical settings on firm selection or related to market size. Third, our results imply that transportation costs and transportation infrastructure are an important consideration for the proper measurement of market size and identification of the firm selection effect.

While this paper has pointed out that transportation networks, rather than city boundaries, may define a market for the purpose of assessing the firm selection hypothesis, there are other types of effects, such as agglomeration economies and sectoral specialization, which may be more prominent at the city-level. We are examining these issues in follow-up work.

References

Accetturo, A., Di Giacinto, V., Micucci, G., & Pagnini, M. (2018). Geography, productivity, and trade: Does selection explain why some locations are more productive than others?. *Journal of Regional Science*, 58(5), 949-979.

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

Ahn, J., Khandelwal, A. K., & Wei, S. J. (2011). The role of intermediaries in facilitating trade. *Journal of International Economics*, 84(1), 73-85.

Arimoto, Y., Nakajima, K., & Okazaki, T. (2014). Sources of productivity improvement in industrial clusters: The case of the prewar Japanese silk-reeling industry. *Regional Science and Urban Economics*, 46, 27-41.

Backus, M. (2020). Why is productivity correlated with competition? *Econometrica*, 88(6), 2415-2444.

Banerjee, A., Duflo, E., & Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in China. *Journal of Development Economics*, 145, 102442.

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.

Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2020). Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*, 115, 103-124.

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

Behrens, K., & Robert-Nicoud, F. (2014). Survival of the fittest in cities: Urbanisation and inequality. *Economic Journal*, 124(581), 1371-1400.

Behrens, K., & Robert-Nicoud, F. (2015). Agglomeration theory with heterogeneous agents. *Handbook of Regional and Urban Economics*, 5, 171-245.

Chan, K. W. (2007). Misconceptions and complexities in the study of China's cities: Definitions, statistics, and implications. *Eurasian Geography and Economics*, 48(4), 383-412.

Chen, T., Gu, Y., & Zou, B. (2024). China's commuting-based metropolitan areas. *Journal of Urban Economics*, 144, 103715.

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.

Ding, C., & Niu, Y. (2019). Market size, competition, and firm productivity for manufacturing in China. *Regional Science and Urban Economics*, 74, 81-98.

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

Duranton, G. (2015). Delineating metropolitan areas: measuring spatial labour market networks through commuting patterns. In T. Watanabe, I. Uesugi and A. Ono (Ed.), The economics of interfirm networks (pp. 107-133). Tokyo: Springer Japan.

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. (2001). Nursery cities: urban diversity, process innovation, and the life cycle of products. *American Economic Review*, 91(5), 1454-1477.

Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100(3), 1195-1213.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from China's

National Trunk Highway System. *Review of Economic Studies*, 81(3), 1046-1070.

Fan, J., Lu, Y., & Luo, W. (2023). Valuing domestic transport infrastructure: A view from the route choice of exporters. *Review of Economics and Statistics*, 105(6), 1562-1579.

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review*, 108(11), 3117-53.

Gobillon, L., & Roux, S. (2010). Quantile-based inference of parametric transformations between two distributions. Processed, CREST-INSEE.

Jia, H., Juan, Z., Zhang, X., & Ni, A. (2004). Determination and comparison of fuel consumption for expressway post-assessment. *Journal of Jilin University*, 34(2), 298-301.

Jedwab, R., & Storeygard, A. (2022). The average and heterogeneous effects of transportation investments: Evidence from Sub-Saharan Africa 1960–2010. *Journal of the European Economic Association*, 20(1), 1-38.

Kondo, K. (2017). Testing for agglomeration economies and firm selection in spatial productivity differences: The case of Japan. Research Institute of Economy, Trade and Industry (RIETI).

Li, H., & Mykhnenko, V. (2018). Urban shrinkage with Chinese characteristics. *Geographical Journal*, 184(4), 398-412.

Lu, Y., Wang, J., & Zhu, L. (2019). Place-based policies, creation, and agglomeration economies: Evidence from China's economic zone program. *American Economic Journal: Economic Policy*, 11(3), 325-360.

Melitz, M. J., & Ottaviano, G. I. (2008). Market size, trade, and productivity. *The Review of Economic Studies*, 75(1), 295-316.

Melitz, M. J., & Redding, S. J. (2014). Heterogeneous firms and trade. *Handbook of International Economics*, 4, 1-54.

Proost, S., & Thisse, J. F. (2019). What can be learned from spatial economics?. *Journal of Economic Literature*, 57(3), 575-643.

Rupasingha, A., & Marré, A. W. (2020). Moving to the hinterlands: Agglomeration, search costs and urban to rural business migration. *Journal of Economic Geography*, 20(1), 123-153.

Syverson, C. (2004). Market structure and productivity: A concrete example. *Journal of Political Economy*, 112(6), 1181-1222.

Wu, S. (2005). Analysis on Decrease Factors of Bus Transportation Cost—Taking Beijing-Shijiazhuang Expressway and No. 107 State Highway as Examples. *Transportation Standardization (Jiaotong Yunshu Yanjiu)*, (7), 47.

Yu, M. (2015). Processing trade, tariff reductions and firm productivity: Evidence from Chinese firms. *The Economic Journal*, 125(585), 943-988.

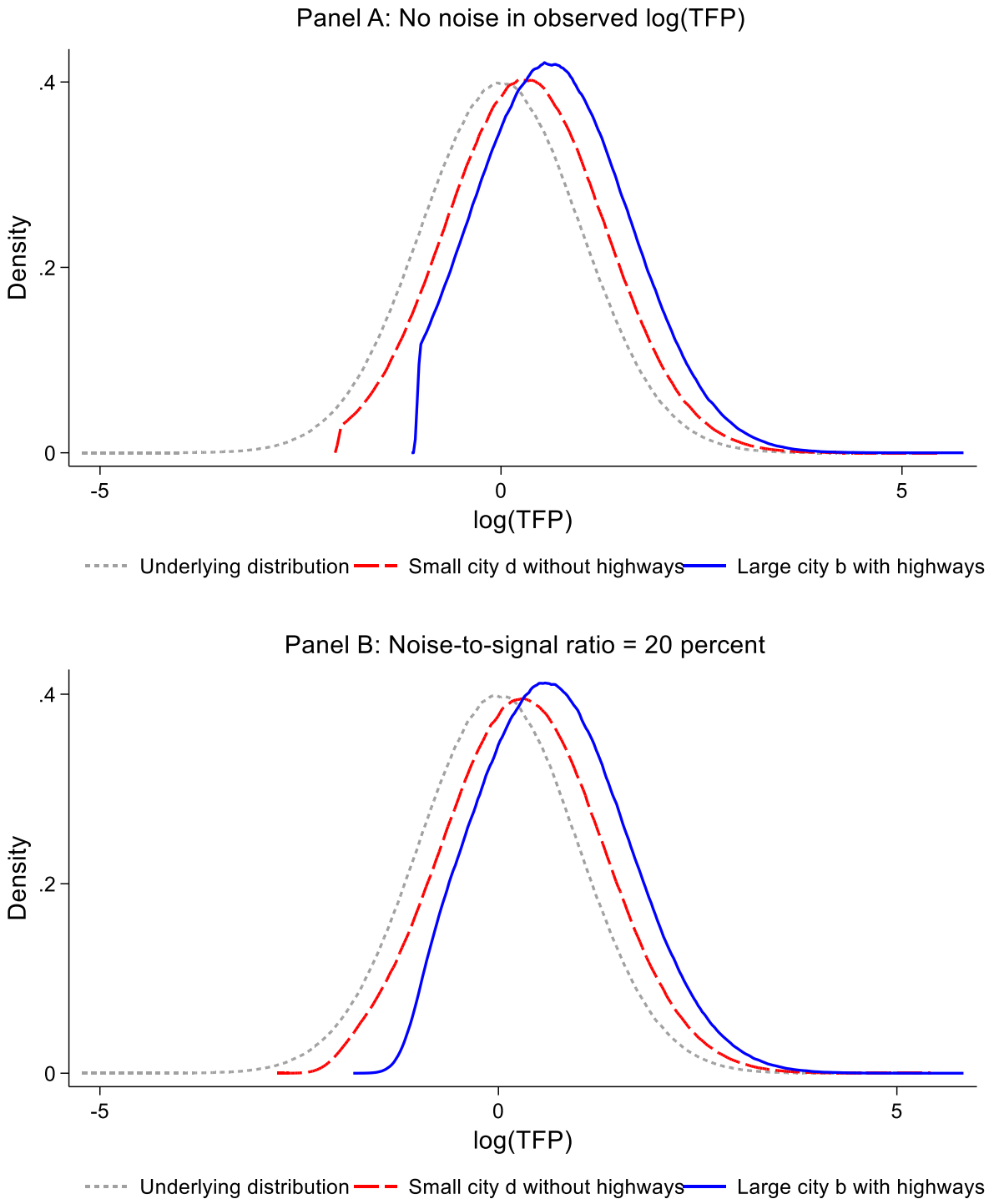

Figure 1: The Effects of Selection and Agglomeration on the Productivity Distribution

Notes: For illustrative purposes, we assume the underlying distribution of log(TFP) follows the standard normal distribution (dotted line). The dashed line represents the observed distribution of log(TFP) in the small city d without highway access, and the solid line represents that in the large city b with highway access. We set $S_b = 0.05$, $S_d = 0.01$, $A_b = 0.6$, $A_d = 0.3$, $D_b = D_d = 1$. In Panel A, no noise is added to the observed log(TFP). In Panel B, The observed log(TFP) is the sum of true log(TFP) and a noise randomly drawn from a normal distribution with zero mean, and the noise-to-signal ratio—the standard deviation of the noise relative to that of true log(TFP)—is set to 20 percent.

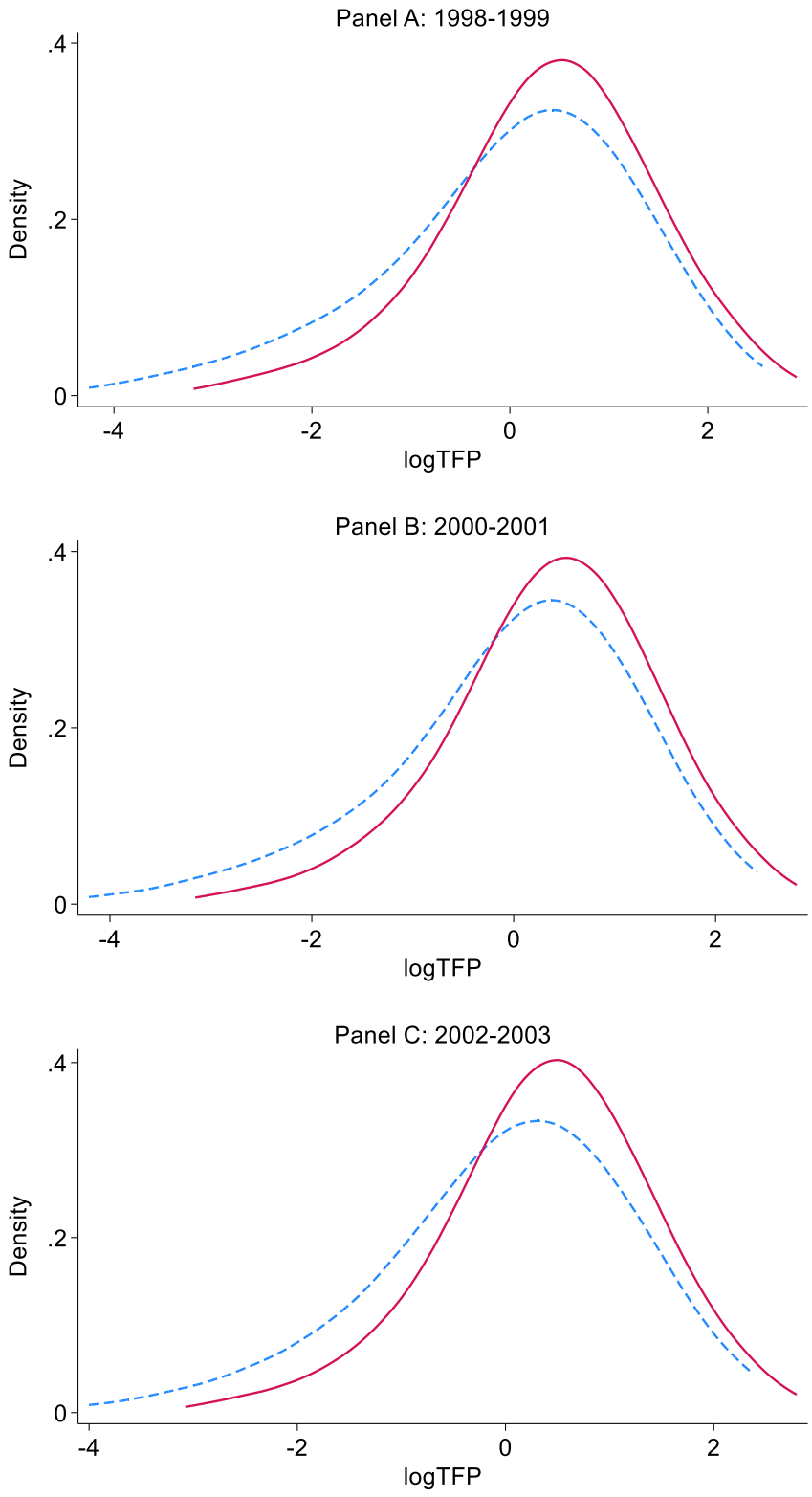

Figure 2: TFP Distributions of Large Connected Cities (Solid) and Small Unconnected Cities (Dashed)

Notes: (1) Dashed lines represent the TFP distributions for small unconnected cities, and solid lines represent those for large connected cities. (2) Small unconnected cities are those with populations below 0.5 million in 2000, not sharing an administrative boundary with any city exceeding 0.5 million, and without highways within 50 km of their boundaries by the end of the period. (3) Large connected cities are those with populations above 0.5 million in 2000, with highways within their boundaries as of one day prior to the beginning of the period. (4) We exclude four outlier provinces in Western China—Xizang, Gansu, Qinghai, and Xinjiang. In 1998-1999, we additionally exclude the provinces of Jiangxi, Hunan, and Hubei, which were heavily flooded in 1998. (5) The bandwidth is estimated using the plug-in method.

Table 1: Estimated Selection Effect, Large Connected Cities vs. Small Unconnected Cities

Period	$\hat{S}$	R^2	$\hat{S}$	$\hat{A}$	R^2	$\hat{S}$	$\hat{A}$	$\hat{D}$	R^2	Obs. in small cities	Obs. in large cities
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1998-1999	0.0836***	0.793	0.0413***	0.2403***	0.951	0.0315***	0.2739***	0.9624	0.953	5,098	23,659
	(0.0093)		(0.0068)	(0.0406)		(0.0100)	(0.0441)	(0.0342)			
2000-2001	0.0734***	0.721	0.0336***	0.2660***	0.959	0.0457***	0.2264***	1.0446	0.963	4,029	31,147
	(0.0096)		(0.0066)	(0.0412)		(0.0153)	(0.0576)	(0.0514)			
2002-2003	0.0902***	0.712	0.0312***	0.2802***	0.965	0.0214**	0.3112***	0.9595	0.969	2,834	38,463
	(0.0137)		(0.0075)	(0.0525)		(0.0098)	(0.0559)	(0.0377)			

Notes: (1) Small unconnected cities are those with populations below 0.5 million in 2000, not sharing an administrative boundary with any city exceeding 0.5 million, and without highways within 50 km of their boundaries by the end of the period. (2) Large connected cities are those with populations above 0.5 million in 2000, with highways within their boundaries as of one day prior to the beginning of the period. (3) We exclude four outlier provinces in Western China—Xizang, Gansu, Qinghai, and Xinjiang. In 1998-1999, we additionally exclude the provinces of Jiangxi, Hunan, and Hubei, which were heavily flooded in 1998. (4) In parentheses are standard errors computed from 100 bootstrapped replications. *, ** and *** indicate the statistical significance at the 10%, 5% and 1% levels, respectively. Tests are against the null hypotheses that $\hat{S} = 0$, $\hat{A} = 0$ or $\hat{D} = 1$.

Table 2: Controlling for the Potential Endogeneity of Highway Location, Large Connected Cities vs. Small Unconnected Cities

Period	$\hat{S}$	$\hat{A}$	$\hat{D}$	R^2	Obs. in small cities	Obs. in large cities
	(1)	(2)	(3)	(4)	(5)	(6)
1998-1999	0.0544**	0.2461***	0.9452	0.966	5,098	4,790
	(0.0276)	(0.0821)	(0.0683)			
2000-2001	0.0563***	0.2467***	1.0077	0.983	4,029	7,450
	(0.0109)	(0.0466)	(0.0360)			
2002-2003	0.0234**	0.3714***	0.9465	0.982	2,834	10,066
	(0.0117)	(0.0569)	(0.0375)			

Notes: (1) Small unconnected cities, as defined in Table 2, are those with population below 0.5 million in 2000, not sharing an administrative boundary with any city exceeding 0.5 million, and without highways within 50 km of their boundaries by the end of the period. (2) Large connected cities are those with population above 0.5 million in 2000, with highways within their boundaries as of one day prior to the beginning of the period, and with an EST line within 25 km of their boundaries. (3) We exclude four outlier provinces in Western China—Xizang, Gansu, Qinghai, and Xinjiang. In 1998-1999, we additionally exclude the provinces of Jiangxi, Hunan, and Hubei, which were heavily flooded in 1998. (4) In parentheses are standard errors computed from 100 bootstrapped replications. *, ** and *** indicate the statistical significance at the 10%, 5% and 1% levels, respectively. Tests are against the null hypotheses that $\hat{S} = 0$, $\hat{A} = 0$ or $\hat{D} = 1$.

Table 3: Robustness to Alternative Specifications

Specification	$\hat{S}$	$\hat{A}$	$\widehat{D}$	R^2	Obs. in small cities	Obs. in large cities
	(1)	(2)	(3)	(4)	(5)	(6)
Using all firms instead of mono-establishment firms	0.0562*** (0.0120)	0.2529*** (0.0440)	1.0094 (0.0400)	0.983	4,121	7,523
Not excluding any outlier provinces	0.0527*** (0.0194)	0.3767*** (0.0612)	0.9641 (0.0491)	0.983	4,592	7,450
The cutoff of city size = 0.75 million	0.0490*** (0.0088)	0.1868*** (0.0374)	1.0090 (0.0320)	0.983	5,657	6,208
The cutoff of city size = 0.25 million	0.0520*** (0.0149)	0.4028*** (0.0512)	0.9585 (0.0501)	0.990	2,018	14,804
Using the alternative measure of city size	0.0527*** (0.0130)	0.3784*** (0.0453)	0.9426 (0.0376)	0.988	3,874	13,478
Large cities crossed by EST lines, excluding target cities	0.0496*** (0.0122)	0.2167*** (0.0500)	0.9715 (0.0393)	0.982	4,029	4,499
Large cities located within 50 km of EST lines, excluding target cities	0.0536*** (0.0112)	0.2494*** (0.0433)	1.0018 (0.0384)	0.981	4,029	8,221

Notes: (1) The study period is 2000-2001. Unless otherwise specified, the city size cutoff is 0.5 million. (2) Small unconnected cities are those with population below the cutoff in 2000, not sharing an administrative boundary with any city exceeding the cutoff, and without highways within 50 km of their boundaries by the end of the study period. (2) Large connected cities are those with population above the cutoff in 2000, with highways within their boundaries as of one day prior to the study period, and with an EST line within 25 km of their boundaries (unless otherwise specified). (3) We exclude four outlier provinces in Western China—Xizang, Gansu, Qinghai, and Xinjiang, except in the second row. (4) In parentheses are standard errors computed from 100 bootstrapped replications. *, ** and *** indicate the statistical significance at the 10%, 5% and 1% levels, respectively. Tests are against the null hypotheses that $\hat{S} = 0$, $\hat{A} = 0$ or $\widehat{D} = 1$.

Table 4: Comparing Small Cities With and Without Highway Access

Period	$\hat{S}$	$\hat{A}$	$\widehat{D}$	R^2	Obs. in cities without highways	Obs. in cities with highways
	(1)	(2)	(3)	(4)	(5)	(6)
1998-1999	0.0205**	0.5347***	0.7927***	0.980	5,098	7,792
	(0.0100)	(0.0331)	(0.0353)			
2000-2001	0.0487***	0.4126***	0.9166*	0.981	4,029	12,415
	(0.0172)	(0.0501)	(0.0494)			
2002-2003	0.0107	0.5058***	0.8322***	0.990	2,834	18,584
	(0.0097)	(0.0540)	(0.0329)			

Notes: (1) We only use firm observations located in cities smaller than 0.5 million population. (2) Cities without highways consist of the cities without any highway within 50 km of their boundaries by the end of the study period. Cities with highways consist of the cities that have highways within their boundaries by the beginning of the study period and are located within 25 km of the nearest EST line. (3) We exclude four outlier provinces in Western China—Xizang, Gansu, Qinghai, and Xinjiang. In 1998-1999, we additionally exclude the provinces of Jiangxi, Hunan, and Hubei, which were heavily flooded in 1998. (4) In parentheses are standard errors computed from 100 bootstrapped replications. *, ** and *** indicate the statistical significance at the 10%, 5% and 1% levels, respectively. Tests are against the null hypotheses that $\hat{S} = 0$, $\hat{A} = 0$ or $\widehat{D} = 1$.

Table 5: Estimated Selection Effect Using Market Access

Period	$\hat{S}$	$\hat{A}$	$\hat{D}$	R^2	Obs. in small-market cities	Obs. in large-market cities
	(1)	(2)	(3)	(4)	(5)	(6)
1998-1999	0.0516***	0.4888***	0.9769	0.979	4,901	19,587
	(0.0181)	(0.0691)	(0.0507)			
2000-2001	0.0312***	0.3023***	1.0126	0.973	6,916	22,210
	(0.0086)	(0.0506)	(0.0337)			
2002-2003	0.0134**	0.3741***	0.9138***	0.981	7,026	26,739
	(0.0059)	(0.0485)	(0.0255)			
2004-2005	0.0269**	0.2250***	0.9604	0.979	6,081	33,332
	(0.0135)	(0.0641)	(0.0264)			
2006-2007	0.0404**	0.1732**	0.9787	0.968	7,710	30,406
	(0.0173)	(0.0718)	(0.0325)			

Notes: (1) Large-market cities are those with all four measures of market access above the 75th percentiles, located within a one-day drive of the nearest major port, and with populations above 0.5 million. (2) Small-market cities are those with all four measures of market access below the 75th percentiles, located beyond a one-day drive of the nearest major port, and with populations below 0.5 million. (3) We exclude four outlier provinces in Western China—Xizang, Gansu, Qinghai, and Xinjiang. In 1998–1999, we additionally exclude Jiangxi, Hunan, and Hubei, which were heavily flooded in 1998. (4) In parentheses are standard errors computed from 100 bootstrapped replications. *, ** and *** indicate the statistical significance at the 10%, 5% and 1% levels, respectively. Tests are against the null hypotheses that $\hat{S} = 0$, $\hat{A} = 0$ or $\hat{D} = 1$.

Table 6: Linear Regressions on the Highway Access and Firm Exit

Independent Variables	(1)	(2)	(3)	(4)
Highway: Highway inside the city boundaries.	0.0514***	0.0460***	0.0457***	0.0582**
	(0.0144)	(0.0141)	(0.0141)	(0.0291)
Highway × Log(TFP)	-0.0155***	-0.0143***	-0.0140***	-0.0217***
	(0.0026)	(0.0025)	(0.0026)	(0.0061)
Log(TFP)	-0.0047**	-0.0053**	-0.0055**	0.0065
	(0.0024)	(0.0023)	(0.0023)	(0.0063)
Fixed effects: firm; year.	Yes	Yes	Yes	Yes
Fixed effects: 2-digit sector by year; city.	No	Yes	Yes	Yes
Firm characteristics	No	No	Yes	Yes
Number of observations	480,673	480,493	474,246	66,049

Notes: (1) The dependent variable is a dummy for firm exit. It equals 1 in year *t* if the firm appears in year *t* in the ASIF data and disappears thereafter. Otherwise, it equals 0. Firms exiting in 2007, the last year of the data, cannot be identified, so we drop that year and use observations from 1998–2006. (2) All independent variables are measured at $t-1$. (3) Firm characteristics include a dummy for SOEs, a dummy for mono-establishment firms, age, age squared, as well as the shares of paid-up capital from foreign countries, from Hong Kong, Macau or Taiwan, from the collective-owned entities, from individuals, and from the legal persons, respectively. (4) In parentheses are standard errors clustered at the city-year level. *, ** and *** indicate the statistical significance at the 10%, 5% and 1% levels, respectively.

Table 7: Linear Regressions on the Highway Access and TFP Growth

Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Highway: Highway inside the city boundaries.	-0.0057	-0.0219	-0.0159	-0.0324	-0.0018	0.0004	0.0738	-0.0345	0.1842	0.1335
	(0.0420)	(0.0428)	(0.0434)	(0.0489)	(0.0755)	(0.0187)	(0.1438)	(0.1839)	(0.1803)	(0.0866)
Highway × Log(TFP)	0.0007	0.0054	0.0039	0.0100	-0.0107		0.0139	0.0607	-0.0365	
	(0.0106)	(0.0109)	(0.0110)	(0.0122)	(0.0195)		(0.0368)	(0.0441)	(0.0431)	
Log(TFP)	-0.0897***	-0.0818***	-0.0809***	-0.0889***	-0.0619***		-0.0519	-0.0885**	-0.0077	
	(0.0092)	(0.0091)	(0.0092)	(0.0104)	(0.0178)		(0.0383)	(0.0404)	(0.0464)	
Highway × Low-productivity						-0.0277				-0.1813
						(0.0341)				(0.1333)
Low-productivity: Below the 10th TFP percentile in the city.						0.0523*				0.1484
						(0.0272)				(0.1297)
Fixed effects: firm; year.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Fixed effects: 2-digit sector by year; city.	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm characteristics	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	401,947	401,940	397,285	217,710	178,844	397,285	57,436	34,293	23,120	57,436

Notes: (1) The dependent variable is the growth rate of TFP, computed as $\left(TFP_{i,t+1} - TFP_{i,t}\right)/TFP_{i,t}$. To reduce the influences of outliers, we exclude observations whose growth rate of TFP is below the 1st percentile or above the 99th percentile. (2) Only firms in cities with populations below 0.5 million are used in column (6). (3) All independent variables are measured at $t-1$. (4) Firm characteristics include a dummy for SOEs, a dummy for mono-establishment firms, age, age squared, and the shares of paid-up capital from foreign countries, from HongKong, Macau or Taiwan, from the collective-owned entities, from the individuals, and from the legal persons, respectively. (5) In parentheses are standard errors clustered at the city-year level. *, ** and *** indicate the statistical significance at the 10%, 5% and 1% levels, respectively.

Online Appendix

A Proof for the Proposition

Consider any two cities b and d such that $L_b > L_d$, $b \in \Psi$ and $d \in \Phi$ so that $\tau_{bk} < \tau_{dk}$, $\forall k$. The equilibrium condition (3) for the two cities can be rewritten as

$$\begin{aligned}&\frac{L_b}{4\gamma}\int_0^{\overline{p}_b}(\overline{p}_b - c)^2 g(c)\,dc + \frac{L_d}{4\gamma}\int_0^{\frac{\overline{p}_d}{\tau_{bd}}}(\overline{p}_d - \tau_{bd}c)^2 g(c)\,dc \\ &+ \sum_{u\in\Psi, u\neq b}\frac{L_u}{4\gamma}\int_0^{\frac{\overline{p}_u}{\tau_{bu}}}(\overline{p}_u - \tau_{bu}c)^2 g(c)\,dc + \sum_{v\in\Phi, v\neq d}\frac{L_v}{4\gamma}\int_0^{\frac{\overline{p}_v}{\tau_{bv}}}(\overline{p}_v - \tau_{bv}c)^2 g(c)dc = f_E \end{aligned} \tag{A1}$$

$$\begin{aligned}&\frac{L_d}{4\gamma}\int_0^{\overline{p}_d}(\overline{p}_d - c)^2 g(c)\,dc + \frac{L_b}{4\gamma}\int_0^{\frac{\overline{p}_b}{\tau_{db}}}(\overline{p}_b - \tau_{db}c)^2 g(c)\,dc \\ &+ \sum_{u\in\Psi, u\neq b}\frac{L_u}{4\gamma}\int_0^{\frac{\overline{p}_u}{\tau_{du}}}(\overline{p}_u - \tau_{du}c)^2 g(c)\,dc + \sum_{v\in\Phi, v\neq d}\frac{L_v}{4\gamma}\int_0^{\frac{\overline{p}_v}{\tau_{dv}}}(\overline{p}_v - \tau_{dv}c)^2 g(c)dc = f_E \end{aligned} \tag{A2}$$

Subtract (A2) from (A1) and we obtain

$$\begin{aligned}&L_b e(\overline{p}_b, \tau_{db}) - L_d e(\overline{p}_d, \tau_{bd}) \\ &+ \sum_{u\in\Psi, u\neq b} L_u\left[f(\overline{p}_u, \tau_{bu}) - f(\overline{p}_u, \tau_{du})\right] + \sum_{v\in\Phi, v\neq d} L_v\left[f(\overline{p}_v, \tau_{bv}) - f(\overline{p}_v, \tau_{dv})\right] = 0 \end{aligned} \tag{A3}$$

where $e(x,\tau) = \int_0^x (x-c)^2 g(c)\,dc - \int_0^{\frac{x}{\tau}}(x-\tau c)^2 g(c)\,dc$, and $f(y,\tau) = \int_0^{\frac{y}{\tau}}(y-\tau c)^2 g(c)\,dc$.

As $\partial f(y,\tau)/\partial\tau < 0$ under $\tau > 1$, and $\tau_{bk} < \tau_{dk}$, $\forall k$, we have $f(\overline{p}_u, \tau_{bu}) - f(\overline{p}_u, \tau_{du}) > 0$, $\forall u \in \Psi$, $u \neq b$, and $f(\overline{p}_v, \tau_{bv}) - f(\overline{p}_v, \tau_{dv}) > 0$, $\forall v \in \Phi$, $v \neq d$. These results and equation (A3) suggest $L_b e(\overline{p}_b, \tau_{db}) - L_d e(\overline{p}_d, \tau_{bd}) < 0$. As $\partial e(x,\tau)/\partial x > 0$ under $\tau > 1$, $L_b > L_d$, and $\tau_{bd} = \tau_{db}$, we must have $\overline{p}_b < \overline{p}_d$. Given $S_b = 1 - G(\overline{p}_b)$ and $S_d = 1 - G(\overline{p}_d)$, we obtain $S_b > S_d$.

B Description and Checking of the Firm-Level Data

The firm-level data we use comes from China's Annual Survey of Industrial Firms (ASIF) during 1998-2007, collected by China's National Bureau of Statistics (NBS). Included in the data are the so-called above-scale firms, which consist of all state-owned firms enterprises (SOEs) and those non-state-owned firms with sales exceeding 5 million RMB in a year, in the sectors of mining,

manufacturing and public utilities. The data reported by the above-scale firms are rather reliable since they have independent accounting systems and are subject to a regular reporting system by the NBS (Holz, 2008). Along with China's rapidly growing economy, the industrial sectors, which make up a large component of the Chinese economy[1], had expanded substantially, with the number of above-scale firms increasing from 165,118 in 1998 to 336,768 in 2007; employment from 56.44 to 78.75 million; value added from 1.94 to 11.70 trillion RMB; and sales from 6.41 to 39.97 trillion RMB.

Comparing the ASIF data to the first National Economic Census in 2004 that surveyed all registered industrial firms, we found that, the above-scale firms employed approximately 78 percent of workers, produced 90 percent of the national industrial output, and made up 20 percent of all firms.

In manufacturing sectors that this paper investigates, above-scale firms also made up 20 percent of all firms in 2004, but this share should be higher in 1998, because the number of SOEs, all of which are included in the ASIF data, shrank by 47 percent during 1998-2004 under the market reform. Assuming the number of non-state-owned above-scale firms increased by the same proportion as the number of below-scale firms during 1998-2004, our estimated share of the above-scale firms in all manufacturing firms is approximately 27 percent in 1998.[2]

The noise in the firm-level data could generate notable attenuation bias of the estimated firm selection effect in the quantile approach (Ding and Niu, 2019), so we made the following efforts to reduce the noise.

B.1 Checking the Raw Data

While a growing number of studies use the ASIF data, their summary statistics are not always the same, perhaps because they obtain the data from different vendors. We obtained the ASIF data from two sources separately, hereinafter referred to as version 1 and version 2. Using the firm identifier (the unified code for legal person) assigned by China's NBS, as well as other information including firm name, legal representative, etc., we merged the two data versions

[1] In 2000, for example, the industrial sector accounted for 40.4 percent of China's GDP, while the construction, the primary and tertiary sectors contributed 5.6, 15.1 and 39.0 percent, respectively (data source: NBS, 2010, *China Compendium of Statistics 1949-2008*, China Statistical Press).

[2] In 2004, the number of all manufacturing firms is 1,258,586, consisting of 256,999 above-scale firms (27,071 SOEs + 229,928 non-state-owned firms) and 1,001,587 below-scale firms. In 1998, the number of above-scale firms in manufacturing sectors is 149,674 (57,139 SOEs + 92,535 non-state-owned firms), while the number of below-scale firms is unknown. Assuming the number of below-scale firms grow by the same proportion as the number of non-state-owned firms during 1998-2004 (229,928 / 92,535 = 2.4848), we estimate that in 1998 there are 403,091 below-scale firms (1,001,587 / 2.4848 = 403,091), and the share of above-scale firms in all manufacturing firms is approximately 27.1 percent (149,674 / (149,674 + 394,598) = 0.2708).

and checked whether the same firm reports the same values in the two versions. The main discrepancies pertain to a small number of firms' county-level codes (the first 6 digits of the administrative-division code, which represent a county, county-level city or city district), and the remaining variables have almost exactly the same values. The details are as follows.

The number of firms that have different county-level codes across the two versions is 593 out of 165,118 in 1998, 464 out of 162,033 in 1999, 461 out of 162,887 in 2000, 34 out of 181,557 in 2002, 46 out of 196,222 in 2003, 3 out of 301,961 in 2006, and 5 out of 336,768 in 2007. Checking the firm addresses, we found that for these years, the county-level codes in version 1 are always correct, so these codes are used.

The exception is the year 2001, in which the county-level codes in version 1 are all wrong, so we used the codes from version 2 instead. Because version 2 misses 2,225 firm observations in this year, we used the firms' zip codes to determine their county-level codes. Specifically, we first searched the zip codes among the remaining firm observations to identify the corresponding county-level codes. If this failed, we looked up the zip codes online to find the name of its administrative division and then matched it to the county-level code published by China's Ministry of Civil Affairs (MCA).[3]

For the year of 2004, version 1 has 2618 more observations than version 2. We found that some firms appear multiple times in version 1, and the number of firms in version 2 is exactly the same as the number reported in China Statistical Yearbook, so we used version 2 for 2004. Neither version reports value added in this year, and we computed it as: *value added* = *output* - *intermediate input* + *value added tax payable*.

For the year of 2005, version 1 has 1792 fewer observations than version 2. The number of firms in version 2 is exactly the same as the number in China Statistical Yearbook, so we use version 2 for 2005.

[Table B1 here]

Then we compared the statistics of our firm-level data with those published by the NBS as well as the firm-level data used in Table A1 of Brandt et al. (2014). Our Table B1 indicates that our data are equal or very close to the NBS data, and in a few cases such as 2001, our data are closer to the NBS data than Brandt et al. (2014).

[3]We often looked up the zip codes at https://www.youbianku.com/ for their administrative divisions. We used the county-level codes published by MCA in various years at https://www.mca.gov.cn/article/sj/xzqh/1980/. We found that the county-level code contains fewer errors than the zip code, possibly because firms reported their county-level codes more carefully than zip codes or because the government verified them more carefully after collecting the data. Therefore, we prefer the county-level code for identifying each firm's location whenever possible.

B.2 Harmonizing Industry Classification

We use the information on industry classification in the estimation of the production function and the investigation of the firm selection effect by sectors. The four-digit Chinese Industry Classification (CIC) was revised since 2003 in our study period, and a consistent classification over the years was constructed by Brandt et al. (2014), i.e., the original industry codes were adjusted and became consistent across the entire period.[4] However, we found that the harmonized classification overlooked five 4-digit industry codes began with "171" (processing of fibrous materials) during 1998-2002, which contain 5027 firm observations. After checking the definitions of the five industries and tracking how the firms reported their industry codes before and after the CIC adjustment in 2003, we added the five industries to the harmonized classification of Brandt et al. (2014) by adjusting the original industry code 1712 to 1721, adjusting 1713 and 1714 to 1730, adjusting 1719 to 1711, and leaving the original code 1711 unchanged. This updated industry classification is used in this research.

B.3 Checking the County-Level Code

We primarily rely on the county-level code (the first six digits of the administrative division code) to identify each firm's location.[5] Comparing to the county-level code published by China's Ministry of Civil Affairs (MCA),[6] we found many errors of this code in our data and made the corrections on most of them.

One type of the errors is that the last two digits of some county-level code were falsely reported as "00".[7] The number of firm observations having this problem is 5,021 (from 193 county-level divisions), including 906 firms (from 33 county-level divisions) in 1998, 135 (from 3 divisions) in 1999, 435 (from 12 divisions) in 2000, 1,382 (from 91 divisions) in 2001, 1,956 (from 36 divisions) in 2002, 28 (from 1 division) in 2003, 62 (from 5 divisions) in 2005, 26 (from 3 divisions) in 2006, and 91 (from 9 divisions) in 2007.

Besides the above firm observations, many other observations still contain invalid county-

[4] We downloaded all the codes used in Brandt et al. (2014 and 2019) at https://feb.kuleuven.be/public/N07057/CHINA/appendix/.

[5] While since 2004 the division code in our data contains 12 digits which correspond to a unique neighborhood committee (*juweihui*) or administrative village, during 1998-2003 the division code only reports six digits. The 1st and 2nd digits represent the province-level administrative division, the 3rd and 4th digits represent the prefecture-level division, and the 5th and 6th digits represent the county-level division.

[6] We used the county-level codes published by MCA in various years at https://www.mca.gov.cn/article/sj/xzqh/1980/.

[7] During our study period, only in the city proper of Dongguan shi and Zhongshan shi should the last two digits of the county-level code be "00", and reporting "00" in the last two digits for any other county-level divisions is false.

level codes that could not be found from the codes published by MCA.[8] The number of these firm observations is 17,872 (from 107 county-level divisions), including 10,683 firms (from 40 divisions) in 1998, 1,334 (from 27 divisions) in 1999, 316 (from 18 divisions) in 2000, 213 (from 8 divisions) in 2001, 51 (from 1 division) in 2002, 180 (from 7 divisions) in 2003, 37 (2 divisions) in 2006, and 5058 (4 divisions) in 2007.

Regarding the above types of errors, we corrected the county-level codes as much as possible. First, we used the first four digits of firms' zip codes, which typically correspond to a unique county-level division, to identify the correct county-level codes. To construct the correspondence between county-level codes and zip codes, we used firm observations that reported both variables correctly. For zip codes that could not be matched in this way, we searched them online (e.g., at www.youbianku.com) to identify the corresponding county-level divisions and then obtained their county-level codes from the MCA's official website. Second, for the remaining unmatched firms, we used the prefecture and county names reported by the firms to find the corresponding county-level codes on the MCA's website. Finally, since some firms reported outdated county-level codes, we compared the remaining unmatched firms to county-level codes in the 1995 version, identified the subsequent administrative changes, and updated them according to the information released by the MCA's website.

Because some cities changed their administrative codes over time, we harmonized the codes across different years to the 2000 version, corresponding to to the year for which city population is measured.

B.4 Matching Firms over Years

Estimating the production function and computing the firm's average TFP over years require matching firms over time and building a panel dataset in advance. To do so, we used and improved the approach of Brandt et al. (2014).

We matched firms in two consecutive years as follows. First, we matched firms using the firm identifier assigned by the NBS. For the unmatched firms, we used the firm name to find a match. These two steps are the same with Brandt et al. (2014).

In the third step, we used the combination of the county-level code, the name of the legal representative and the 2-digit industry code to link the unmatched firms in previous steps. This is different from Brandt et al. (2014), who use the combination of the name of legal representative and the prefecture-level code (the first four digits of the administrative division code) to link the

[8] One of the reasons for such differences is that some firms reported outdated county-level codes, which could change due to the adjustment of administrative divisions.

unmatched firms in previous steps, and their problem is that the same legal person sometimes own multiple firms in the same prefecture, leading to different firms being identified as the same one.

The third step of the matching could be influenced by the fact that, for firms that did not relocate, their county-level codes might change due to administrative division adjustments, including county changed to county-level city, county to city district, prefecture (*diqu*) to prefecture *shi*, etc. Therefore, based on the county-level codes published by MCA in various years, we created harmonized county-level codes that are consistent across the entire period, and used them in the third step of the matching. However, the changes of the county-level codes below the county-level still could not be addressed. For example, a town belonging to County A in the first year became part of County B in the second year. Therefore, in the fourth step, we used the combination of the first four digits of zip code, as well as the name of the legal representative and the 2-digit industry code, to link unmatched firms from the previous steps.

In the fifth step, we used the combination of the county-level code, the 2-digit industry code and phone number to link the unmatched firms in previous steps. This is different from Brandt et al. (2014), who use the combination of the prefecture-level code, the 3-digit industry code and phone number to link the unmatched firms, and their problem is that some firms changed their 3-digit industry code back and forth. We used the 2-digit industry code instead, which was rarely changed by firms.

Brandt et al. (2014) has one more step to link the unmatched firms in previous steps. The identifier is the combination of the founding year, county-level code, 4-digit industry code, name of town, and the first of the three main products. We found that this step sometimes recognizes different firms as the same one, so we did not use it.

As some firms might disappear from the sample and re-enter later, similar to Brandt et al. (2014), we subsequently matched remaining observations in data files two years apart (i.e., matching year t with year $t+2$). We added an additional step: for those firms that still appear only once in the data, we also used all other years to find matches (i.e., matching year t with $t+3$, $t+4$, ...).

B.5 Estimating the Real Values of Capital Stock, Output and Input

We need to use the firms' capital stock in the estimation of production function. As the firms only reported the nominal values of their capital stock, we followed Brandt et al. (2014)'s approach to estimate their real capital stock.

We also used the approach of Brandt et al. (2014) to deflate output and intermediate input. This approach uses the nominal and real values of output reported by each firm to construct an output deflator at the 4-digit industry level during 1998-2003. After 2003, firms did not report the real values of output anymore, and the 2-digit ex-factory price index from the China Statistical Yearbook is used as the output deflator during 2004-2007. With this output deflator and the input share calculated from the 2002 National Input-Output table, an input deflator was built following Brandt et al. (2019). Finally, we computed the real value added as: *real value added* = *real output - real intermediate input + real value added tax payable*.

B.6 Identifying the State-Owned Enterprises

We need to use an indicator for State-Owned Enterprises (SOEs) when estimating the production function and investigating the firm selection effect. Two main methods have been used to identify the SOEs in the ASIF data. The first method uses the information on the registered capital, as each firm reported its paid-up capital for six types of owners: state, collective, individual, legal person, Hong Kong-Macau-Taiwan, and foreign. For instance, the firms with positive capital from the state are identified as SOEs in Yu (2015), and the firms with more than 30 percent of capital from the state are identified as SOEs in Huang et al. (2017). The problem of this method, however, is the difficulty of tracing the sources of the capital from the legal person, which can capture a wide range of possibilities including state-controlled shareholding companies and private subsidiaries (Brandt et al., 2014). The share of legal person in all paid-up capital increased from 18 percent in 1998 to 33 percent in 2007.

The second method identifies SOEs according to the 23 types of the firm's legal registration. This method is used by, for example, Brandt et al. (2012). While the types of state-owned enterprise (registration code 110), state-owned jointly operated enterprise (141) and state-owned LLC (151) are usually recognized as SOEs, it is difficult to identify SOEs from some other types such as joint-stock cooperative enterprise (130), state and collectively owned jointly operated enterprise (143), stock limited company (160), etc. Moreover, many SOEs are legally registered as foreign firms, limited-liability firms, or publicly traded firms (Hsieh and Song, 2015). Dollar and Wei (2007) also suggest that some former SOEs do not change their registered ownership type even after ownership restructuring.

In our data, an indicator of the shareholding status suggests two types of state control, namely, absolute control (the state share exceeding 50 percent) and relative control (the state share less than 50 percent but being the largest shareholder). In 1998-2005, the indicator's value

of 1 denotes absolute control, and 2 denotes relative control; in 2006-2007, the value of 1 denotes both absolute control and relative control. Therefore, we identified the firms of the absolute control and the relative control as SOEs. This definition is probably the same with the NBS, as Table B2 shows that the number of SOEs from our data is almost the same with the NBS data, except in 1998 when our number is 5 percent higher than the NBS's number.

[Table B2 here]

According to our definition, 72.7 percent of SOEs observations are registered as state-owned enterprise (110), 10.2 percent are other LLC (159), 4.4 percent stock limited company (160), 3.0 percent state-owned LLC (151), 2.8 percent Chinese-foreign jointly owned enterprise (310), 2.8 percent Chinese-HMT jointly owned enterprise (210), and the rest types are all below 1 percent. This is consistent with Hsieh and Song (2015)'s finding that SOEs could be registered as foreign firms, limited-liability firms, or publicly traded firms. In the 389,905 (94.2 percent) SOE observations that reported positive values of paid-up capital, 76.1 percent reported that the state share was greater than 50 percent, and 76.7 percent reported that the state hold the largest share. While how much capital labeled as the legal person is actually from the state remains unknown, among SOE observations with positive paid-up capital, the capital share of the state and the legal person is larger than 50 percent in 93.4 percent observations, and the share of the state and the legal person is larger than any other type in 93.6 percent of observations.

C Estimation of the Production Function

Ordinary least squares estimation of equation (6) may exhibit bias due to the endogeneity of inputs – for instance, a firm may adjust its input use after observing its productivity shocks. To handle this problem, Olley and Pakes (1996) model firm investment as a function of productivity and capital, and thereby use investment as a proxy for unobservable productivity shocks. However, this is problematic because investment is typically lumpy and it contains many "zero" observations. To solve this problem, Levinsohn and Petrin (2003) model intermediate inputs, instead of investment, as a function of productivity and capital, since intermediate inputs may be more responsive to productivity shocks than investment. However, these two estimators will fail to identify the labor coefficient if labor input and the proxy function for the unobserved productivity are perfectly collinear. Ackerberg et al. (2015) propose a correction based on inverting the intermediate demand functions conditional on labor inputs. We add an SOE dummy into the demand function for intermediate inputs, as SOEs might have different decision behaviors

than non-SOEs. Then we estimate function (6) separately by each of the 28 2-digit sectors. The estimation results of the production function for each sector are displayed in Table C1.

[Table C1 here]

D Description and Validation of Data on Controlled-Access Highways and Regular Roads

We first obtained the GIS (geographic information system) data on China's road network in 1999, 2005 and 2010, which were used by Baum-Snow et al. (2017) and Baum-Snow et al. (2020).[9] These data were created by digitizing published maps of China. Based on them, we constructed our data on controlled-access highways and other roads. Errors in the highway data could lead to incorrect grouping of cities used for the quantile approach, likely generating attenuation bias. Therefore, we carefully verified the data.

The map for January 1999 distinguishes three categories of roads: high-grade highways (*gao deng ji gong lu*), national roads (*guo jia ji gong lu*), and sub-national roads (*guo jia ji yi xia gong lu*). When digitizing the map to the GIS format, Baum-Snow et al. combined the latter two categories, so the resulting GIS data report only high-grade highways and regular roads. The map for March 2005, distinguishes two categories of roads: high-grade highways (*gao deng ji gong lu*), and regular roads (*gong lu*). The map for 2010 distinguishes three categories: high-grade highways (*gao deng ji gong lu*), national roads (*guo jia ji gong lu*), and sub-national roads (*yi ban gong lu*). The latter two categories were combined, so the resulting GIS data report only high-grade highways and regular roads.

We note the following issues with these maps. First, there is no clear definition of high-grade highways. Based on our inspection and comparison of the maps, this category appears to include all controlled-access highways, some first-class highways, and a small subset of second-class highways.[10] Which first- or second-class highways are depicted likely depends on their relative importance. For example, in provinces with several controlled-access highways, the maps often show no first- or second-class highways, whereas in provinces without any controlled-access highways, the maps usually display some first- and even second-class highways. Second, which national and sub-national roads are included on the maps likely depends on their relative

[9] We downloaded the data from https://matthewturner.org/research.htm.

[10] For instance, a national road G502 that connects Kedong *Xian* and Qiqihar *Shi* in Heilongjiang Province is a second-class highway (https://www.ggzy.gov.cn/information/html/b/230000/0101/202403/22/0023f7ebc25f555d413f9b37052524a6daf5.shtm), and it is indicated as high-grade highway in the 2005 map.

importance as well. Third, the highways under construction are mistakenly treated as completed highways. In the published maps of China, dashed lines denote highways under construction, whereas solid lines indicate those completed and open to traffic. However, both types of lines were digitized as completed highways in the downloaded GIS data. This problem is particularly severe in the 1999 dataset, which contains a large share of highways still under construction. We corrected this by deleting the highways under construction from the GIS data.

To construct the GIS data for controlled-access highways (hereafter highways), we undertook the following steps. For each highway segment, we identified its two endpoints, length, and date of opening to traffic. Note that a highway (e.g., the Beijing-Shanghai Highway) is typically constructed at multiple locations simultaneously and completed segment by segment at different times. We found that most highway segments are between 50 and 200 km in length.

Our primary data sources are as follows. The first is the Yearbook of China Transportation & Communications (YCTC, Zhongguo Jiaotong Nianjian) in various years, which reports the above information for most highway segments completed before 1998, but covers fewer segments thereafter. The second source is the Newspaper and Journal Databases of the China National Knowledge Infrastructure (CNKI). When a highway segment was completed and opened to traffic, local newspapers and journals typically reported the event, often including detailed information on the segment. For segments not covered by the above sources, we searched online for additional reports and articles,[11] and checked the corresponding provincial gazetteers. In the end, we identified the exact opening date for most highway segments, and at least the year and month——or the year alone——for the remaining ones.

After collecting the above information for each highway segment, we digitized the highway lines in ArcGIS. We used the GIS data of high-grade highways from Baum-Snow et al. as the base layer, as these are close to controlled-access highways and easy to modify. In addition, we purchased GIS data on China's controlled-access highways as of early 2008 from Baidu Map, as well as three hardcopy editions of China's highway maps published in 2008, and used them as references. Each of these references contains some errors of highways. Using the locations of the two endpoints and the length of each segment, we located the corresponding segment in the Baum-Snow et al. data and checked its accuracy. If a segment was missing or deviated from the actual location by more than 3 km, we redrew it manually in ArcGIS.[12]

Our highway data indicate that, for instance, the total length of controlled-access highways

[11] Some useful sources include: the website of the provincial government, Sina News: https://news.sina.com.cn/, and the Database of People's Daily: https://data.people.com.cn.

[12] We thank Tongwei Cui, Ruinan Du, Zhenhao Lei, Yang Liu, Yawen Liu, and Yumeng Song from the Capital University of Economics and Business for excellent research assistance on the above work.

at the end of 1999 and 2005 is 11,735 and 40,386 km, very close to 11,605 and 41,005 km reported by the NBS.

In addition, we constructed a dataset of regular roads, which mainly consist of first- and second-class highways, to be used for computing market access later. First, we extracted the relevant segments from our controlled-access highway data for January 31, 1999, and May 31, 2005, and compared them to the high-grade highways from Baum-Snow et al. for January 1999 and May 2005, respectively. Any high-grade highways that were not controlled-access highways were reclassified into the regular roads dataset from Baum-Snow et al. The total length of such roads is 3,743.6 km in 1999 and 1,018.7 km in 2005. Second, we compared the regular roads datasets from January 1999 and March 2005 provided by Baum-Snow et al. Any regular roads present in 1999 but missing in 2005 were deleted to improve comparability between the two periods. The total length of deleted roads is 27,808.4 km, mostly located in peripheral regions such as Inner Mongolia and Xinjiang.

The final dataset of regular roads indicates a total length of 165,991 km in January 1999 and 215,184 km in March 2005. For comparison, the NBS reports that at the end of 1998, first- and second-class highways measured 14,837 km and 125,245 km, respectively, for a total of 140,082 km; at the end of 2004, the corresponding figures were 33,522 km and 231,715 km, summing to 265,237 km. While our dataset of regular roads is likely less accurate than our controlled-access highway data, it still provides a reasonable approximation of the actual network of important roads beyond the controlled-access highways.

E Measuring City Size in China

There is no well-recognized ready-made dataset of city size in China. While the population data of the administratively designated *shi* (also known as the administratively designated city, including the prefecture-level city and the county-level city) published in China Urban Statistical Yearbook (CUSY) and China Urban Construction Statistical Yearbook (CUCSY) in various years are sometimes used to measure city size, they have several limitations.

In terms of coverage, the yearbooks omit cities located in administratively designated *xian* (also known as the administratively designated county and county equivalent), which contain 23.1 percent of China's urban population in 2000, for instance.

Moreover, the population indicators in the two yearbooks have several problems. First, most of the population indicators, including total population and non-agricultural population in CUSY and urban population in CUCSY, are constructed based on *hukou* (registered residence), which

has increasingly large difference from the actual residence. In 2000, for instance, 10.2 percent of the Chinese population live in townships that differ from their *hukou*; as a large part of such population move from rural areas and small cities to large cities, using *hukou* population would generate bias in the measure of city size in different directions and magnitudes for different cities. Second, *shi* contains much rural population, so using the total population in CUSY would overestimate city size to different extents for different cities. In 2000, the rural population accounts for 42.7 percent of the total population in all the 653 *shi*. Third, the population data in CUCSY contain much noise as they exhibit implausibly large fluctuations over years for many *shi*.

In this paper, we use the urban population in each *shi* or *xian* (referred to "city" in this paper) as the measure of population size of each city, as discussed in Section 5.3. Using the urban area of each *shi* or *xian* as the spatial scope of the city is supported by several empirical studies which use mobile phone data or Baidu Map's commuting information in the 2010s and report that most commuting in China did not cross the boundaries of *shi* and *xian* (Ding et al., 2015; Wang et al., 2018; Zhao, et al., 2019; Chen et al., 2024). In our study period, 1998-2007, the commuting distance should be shorter and the share of cross-boundary commuting trips should be lower.

We require that a city should have no less than 10,000 population, and thereby identify 2,208 cities in 2000. Their average size is 207.5 thousand, and the largest city (Shanghai) reaches 13.5 million.

In Section 6.4, we consider an alternative measure of city size because the major urban areas in several adjacent cities expanded across their administrative boundaries and became contiguous, and their labor markets are possibly integrated. We treat such integrated cities as a single city, use their total urban population as the city's population, and re-group cities.

To identify such cities, we use two data of land use and night light. The land-use data is from the Chinese Land Use Cover Change 100-Meters Grid Dataset, obtained from the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences (www.resdc.cn). It identifies 25 land-use types based on the Landsat TM/ETM satellite image, and we use both types of "urban land" (#51) and "other construction land" (#53) as urban area in this research.[13] The night-light data is from the website of the National Oceanic Atmospheric Administration, improved by Li et al. (2020). If both the land-use data and night-light data (with brightness greater than 30) indicate that the main urban areas of two adjacent cities are contiguous or in close proximity, we combine their populations and assign the total population

[13]Many development zones are classified as "other construction land."

to each city. We list such cities in Table E1. Most of these pairs of cities involve a very large city and a quite small city.

[Table E1 here]

F Measuring Travel Time Between Cities

We use the Network Analyst tool in ArcGIS to compute the shortest travel time between any pair of cities. This process requires converting city polygons into points and constructing a road network.

For each city, we created a centroid representing its urban area. To identify urban areas, we combined three datasets. The first is the 100-meter resolution urban land-use data from the Resource and Environment Science and Data Center of the Chinese Academy of Sciences. This dataset, derived from visual inspection of Landsat TM/ETM satellite images, is widely regarded as the most accurate source of land-use information for China (Xu et al., 2018). Each 100-meter pixel is classified into six primary and twenty-five secondary land-use categories. We define urban land as the secondary types "urban built-up land" (#51) and "other construction land" (#53).[14]

The second dataset is the 1-kilometer resolution night-time light data from Li et al. (2020). The third is the spatial boundary data of China's townships in 2000, produced by the China Data Center at the University of Michigan.[15]

In our definition, urban areas consist of townships that meet both of the following criteria in 2000: (1) the township's urban land area exceeds 1 square kilometer or 10 percent of its total area; and (2) the average brightness of the township's urban land is greater than 10. Because the urban land dataset includes some parcels located in rural areas——such as saltpans, quarries, and transportation infrastructure——relying solely on this dataset could misclassify certain rural areas as urban. Conversely, using only night-time light data could overlook urban areas in smaller cities or low-income regions with relatively low luminance, while also misidentifying some rural areas as urban due to spatial blurring (Abrahams et al., 2018). Combining both criteria improves the accuracy of identifying urban areas, a conclusion supported by our own data inspections.

[14]The six primary land-use types include cropland (#10), forest land (#20), grassland (#30), waters (#40), urban & rural, industrial & mining, and residential land (#50), and unutilized land (#60). The #50 category includes urban built-up land (#51), rural residential land (#52), and other construction land (#53). The #53, by definition, includes land for manufacturing, mining, oilfields, saltpans, quarries, roads, airports, and transportation stations, among others. When checking and inspecting the data, we found that many pixels in development zones are classified as #53, so we also included this type of land as part of urban land.

[15]We thank Jonathan Dingel for geneously sharing the GIS data of China's township boundaries.

If a city does not have any urban areas, we created a centroid of its administrative area.

Our road network consists of three components: controlled-access highways, regular roads, and straight lines connecting city centroids to their neighboring cities. Similar to Baum-Snow et al. (2020), we assigned travel speeds of 90 km/h, 30 km/h, and 15 km/h to these links, respectively. The data for controlled-access highways and regular roads are described in Online Appendix C. The straight-line connections are generated in ArcGIS using the city centroids defined above.

There are no bridges across the Qiongzhou Strait (between the Chinese mainland and Hainan Province) or the Zhoushan Strait (between the Chinese mainland and Zhoushan Prefecture in Zhejiang Province), and we set the travel time for each strait at seven hours according to news reports.[16] Seven additional cities are islands without bridges to the mainland; we exclude them from the analysis due to their small urban populations, which range from 16.8 to 73.9 thousand. The road network is divided into segments to allow turns at all intersections.

Finally, using the city centroids and the road network, we computed the shortest travel time between cities with ESRI's Network Analyst, applying Dijkstra's algorithm.

References

Abrahams, A., Oram, C., & Lozano-Gracia, N. (2018). Deblurring DMSP nighttime lights: A new method using Gaussian filters and frequencies of illumination. *Remote Sensing of Environment*, 210, 242-258.

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

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.

Baum-Snow, N., Henderson, J. V., Turner, M. A., Zhang, Q., & Brandt, L. (2020). Does investment in national highways help or hurt hinterland city growth? *Journal of Urban Economics*, 115, 103124.

Brandt, L., Van Biesebroeck, J., Wang, L., & Zhang, Y. (2019). WTO accession and performance of Chinese manufacturing firms: Corrigendum. *American Economic Review*, 109(4), 1616-21.

[16] It usually took 4-5 hours for a truck to go through Qiongzhou Strait around 2020 (http://www.jjckb.cn/2018-12/27/c_137701093.htm; https://www.hinews.cn/news/system/2021/07/23/032586851.shtml), and it should take more hours around 2000 (http://www.hkwb.net/news/content/2014-03/28/content_2184610.htm).

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.

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

Chen, T., Gu, Y., & Zou, B. (2024). China's commuting-based metropolitan areas. *Journal of Urban Economics*, 144, 103715.

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.

Ding, C., & Niu, Y. (2019). Market size, competition, and firm productivity for manufacturing in China. *Regional Science and Urban Economics*, 74, 81-98.

Ding, L., Niu, X., & Song, X. (2015). Liyong shouji shuju shibie Shanghai zhongxincheng de tongqinqu [Identifying the commuting area of Shanghai central city using mobile phone data], *Chengshi guihua*, 9,100-106.

Dingel, J. I., Miscio, A., & Davis, D. R. (2021). Cities, lights, and skills in developing economies. *Journal of Urban Economics*, 125, 103174.

Dollar, D., & Wei, S. J. (2007). Das (Wasted) Kapital: Firm Ownership and Investment Efficiency in China. *IMF Working Paper No. 2007/009*.

Holz, C. (2008). How Can a Subset of Industry Produce More Output than All of Industry? Retrieved from http://carstenholz.people.ust.hk/CarstenHolz-industry-stats-07-web-27Nov08.pdf.

Hsieh, C. T., & Song, Z. M. (2015). Grasp the Large, Let Go of the Small: The Transformation of the State Sector in China. *Brookings Papers on Economic Activity*, Spring, 295-346.

Huang, Z., Li, L., Ma, G., & Xu, L. C. (2017). Hayek, local information, and commanding heights: Decentralizing state-owned enterprises in China. *American Economic Review*, 107(8), 2455-2478.

Li, X., Zhou, Y., Zhao, M., & Zhao, X. (2020). A harmonized global nighttime light dataset 1992–2018. *Scientific Data*, 7(1), 168.

Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studie*s, 70(2), 317-341.

Olley, G. S., & Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*,64(6), 1263-1297.

Wang, D., Gu, J., and Yan, L. (2018). Shanghai dushiqu bianjie huafen: jiyu shouji xinling shuju de tansuo [Delimiting the Shanghai metropolitan area: using mobile phone data]. *Dili xuebao*, 10, 1896-1909.

Yu, M. (2015). Processing trade, tariff reductions and firm productivity: Evidence from Chinese firms. *The Economic Journal*, 125(585), 943-988.

Zhao, P., Hu, H., Hai X., Huang, S. & Lyu, D. (2019). Jiyu shouji xinling shuju de chengshiqun diqu dushiquan kongjian fanwei duoweishibie: yi jingjinji weili [Identifying metropolitan edge in city clusters region using mobile phone data: a case study of Jing-Jin-Ji]. *Chengshi fazhan yanjiu*, 9, 69-79+2.

Table B1: Comparison of Sample Coverage with China Statistical Yearbook and Brandt et al. (2014)

Year	Source	Number of firms	Value added	Sales	Output	Employment	Export	Net value of fixed assets
1998	Our data	165,118	1.94	6.41	6.77	61.96	1.08	4.41
	NBS	165,080	1.94	6.41	6.77	61.96	1.08	4.41
	Brandt et al. (2014)	165,118	1.94	6.41	6.77	56.44	1.08	4.41
1999	Our data	162,033	2.16	6.99	7.27	58.05	1.15	4.73
	NBS	162,033	2.16	6.99	7.27	58.05	1.15	4.73
	Brandt et al. (2014)	162,033	2.16	6.99	7.27	58.05	1.16	4.73
2000	Our data	162,887	2.54	8.42	8.57	55.59	1.46	5.18
	NBS	162,885	2.54	8.42	8.57	55.59	1.46	5.18
	Brandt et al. (2014)	162,883	2.54	8.42	8.57	53.68	1.46	5.18
2001	Our data	171,256	2.83	9.37	9.54	54.41	1.62	5.54
	NBS	171,256	2.83	9.37	9.54	54.41	1.62	5.54
	Brandt et al. (2014)	169,030	2.79	9.24	9.41	52.97	1.61	5.45
2002	Our data	181,557	3.30	10.95	11.08	55.21	2.01	5.95
	NBS	181,557	3.30	10.95	11.08	55.21	2.01	5.95
	Brandt et al. (2014)	181,557	3.30	10.95	11.08	55.21	2.01	5.95
2003	Our data	196,222	4.20	14.32	14.23	57.49	2.69	6.61
	NBS	196,222	4.20	14.32	14.23	57.49	2.69	6.61
	Brandt et al. (2014)	196,222	4.20	14.32	14.23	57.49	2.69	6.61
2004	Our data	276,474	5.72	19.78	20.17	66.22	4.05	7.97
	NBS	276,474	5.48	18.78	20.17	66.22	4.05	7.38
	Brandt et al. (2014)	279,092	6.62	20.43	20.16	66.27	4.05	7.97
2005	Our data	271,835	7.22	24.69	25.16	68.96	4.77	8.95
	NBS	271,835	7.21	24.46	25.16	67.85	4.77	8.81
	Brandt et al. (2014)	271,835	7.22	24.69	25.16	68.96	4.77	8.95
2006	Our data	301,961	9.11	31.42	31.66	73.58	6.05	10.58
	NBS	301,961	9.11	31.36	31.66	73.58	5.96	10.58
	Brandt et al. (2014)	301,961	9.11	31.36	31.66	73.58	6.05	10.58
2007	Our data	336,768	11.70	40.06	40.51	78.75	7.34	12.34
	NBS	336,768	11.70	39.97	40.52	78.75	7.31	12.34
	Brandt et al. (2014)	336,768	11.70	39.97	40.52	78.75	7.34	12.34

Notes: our data is computed by summing all firms in the firm-level data; NBS's data are from China Statistical Yearbook, China Statistical Abstract, and China Industry Economy Statistical Yearbook; the unit of employment is million, and the unit of value added, sales, output, export, and net value of fixed assets are trillion yuan.

Table B2: Comparison of the Number of SOEs with China Statistical Yearbook

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
NBS	64,734	61,301	53,489	46,767	41,125	34,280	35,597	27,477	24,961	20,680
Our data	68,149	61,301	53,489	46,767	41,125	34,280	35,597	27,477	24,960	20,680

Notes: The NBS data are from Table 13-8 of China Statistical Yearbook 2009.

Table C1: Estimation Results of the Production Function

SIC	Industry	Emp.	Capital	Obs.	SIC	Industry	Emp.	Capital	Obs.
13	Food Processing	0.5158*** (0.0224)	0.2930*** (0.0241)	117,848	28	Chemical Fibers	0.3993*** (0.0103)	0.3285*** (0.0116)	9,752
14	Food Production	0.5800*** (0.0178)	0.3596*** (0.0151)	47,090	29	Rubber Products	0.4238*** (0.0087)	0.3306*** (0.0081)	23,111
15	Beverage Production	0.5189*** (0.0165)	0.4002*** (0.0131)	32,808	30	Plastic Products	0.4455*** (0.0075)	0.3253*** (0.0081)	90,197
17	Textile	0.4417*** (0.0055)	0.2663*** (0.0051)	167,092	31	Nonmetal Mineral Products	0.3098*** (0.0048)	0.3626*** (0.0087)	166,936
18	Garments & Other Fiber Products	0.5524*** (0.0052)	0.2318*** (0.0328)	93,653	32	Smelting & Pressing of Ferrous Metals	0.4466*** (0.0103)	0.3475*** (0.0094)	46,034
19	Leather, Furs, Down & Related Products	0.5139*** (0.0064)	0.2445*** (0.0065)	46,258	33	Smelting & Pressing of Nonferrous Metals	0.4842*** (0.0100)	0.2750*** (0.0092)	33,171
20	Timber Processing, Bamboo, Cane, Palm Fiber & Straw Products	0.4905*** (0.0136)	0.2436*** (0.0157)	42,103	34	Metal Products	0.4379*** (0.0085)	0.3172*** (0.0085)	105,177
21	Furniture Manufacturing	0.6377*** (0.0473)	0.2151** (0.0932)	22,559	35	Machinery & Equipment Manufacturing	0.3730*** (0.0129)	0.3168*** (0.0145)	146,900
22	Papermaking & Paper Products	0.4235*** (0.0070)	0.3382*** (0.0081)	57,856	36	Special Equipment Manufacturing	0.3847*** (0.0145)	0.2850*** (0.0131)	80,742
23	Printing & Record Pressing	0.4276*** (0.0090)	0.5467*** (0.0083)	40,972	37	Transportation Equipment Manufacturing	0.5153*** (0.0149)	0.3480*** (0.0146)	92,676
24	Stationery, Educational & Sports Goods	0.5070*** (0.0236)	0.2386*** (0.0280)	25,574	39	Electric Equipment & Machinery	0.4771*** (0.0071)	0.3387*** (0.0077)	114,897
25	Petroleum Processing, Coking Products & Gas Production	0.2618*** (0.0117)	0.4831*** (0.0129)	14,869	40	Electronic & Telecommunications	0.5348*** (0.0098)	0.3298*** (0.0119)	63,146
26	Raw Chemical Materials & Chemical Products	0.3379*** (0.0083)	0.3670*** (0.0079)	140,963	41	Instruments, Meters, Cultural & Official Machinery	0.4193*** (0.0079)	0.2753*** (0.0074)	27,147
27	Medical & Pharmaceutical Products	0.4319*** (0.0062)	0.4358*** (0.0064)	40,347	42	Handicrafts and Miscellaneous Manufacturing	0.4699*** (0.0077)	0.2243*** (0.0059)	38,044

Notes: Bootstrap standard errors from 100 repetitions are in the parentheses; *, ** and *** denote the statistical significance at the level of 10%, 5% and 1%, respectively.

Table D1: Cities Integrated with Each Other in 2000

City 1	City 2		City 1	City 2
Taiyuan Shi	Jinzhong Shi		Hefei Shi	Feidong Xian
Shijiazhuang Shi	Zhengding Xian		Hefei Shi	Feixi Xian
Shijiazhuang Shi	Luancheng Xian		Huaibei Shi	Suixi Xian
Shijiazhuang Shi	Gaocheng Shi		Tongling Shi	Tongling Xian
Shijiazhuang Shi	Luquan Shi		Fuzhou Shi	Changle Shi
Tangshan Shi	Fengrun Xian		Putian Shi	Putian Xian
Tangshan Shi	Fengnan Shi		Quanzhou Shi	Shishi Shi
Handan Shi	Handan Xian		Quanzhou Shi	Jinjiang Shi
Baoding Shi	Qingyuan Xian		Zhangzhou Shi	Longhai Shi
changzhi Shi	Lucheng Shi		Qingdao Shi	Jimo Shi
Linfen Shi	Xiangfen Xian		Dongying Shi	Kenli Xian
Liaoyang Shi	Liaoyang Xian		Anyang Shi	Anyang Xian
Nanjing Shi	Jiangning Xian		Xinxiang Shi	Xin Xiang Xian
Wuxi Shi	Xishan Shi		Puyang Shi	Puyang Xian
Xuzhou Shi	Tongshan Xian		Xuchang Shi	Xuchang Xian
Changzhou Shi	Wujin Shi		Luohe Shi	Yancheng Xian
Jiangyin Shi	Zhangjiagang Shi		Zhoukou Shi	Shangshui Xian
Suzhou Shi	Wujiang Shi		Guangzhou Shi	Zengcheng Shi
Suzhou Shi	Wuxian Shi		Shaoguan Shi	Qujiang Xian
Nantong Shi	Tongzhou Shi		Shantou Shi	Chenghai Shi
Huaiyin Shi	Huaiyin Xian		Foshan Shi	Gaoming Shi
Yancheng Shi	Yandu Xian		Jiangmen Shi	Xinhui Shi
Yangzhou Shi	Hanjiang Xian		Foshan Shi	Heshan Shi
Zhenjiang Shi	Dantu Xian		Maoming Shi	Dianbai Xian
Suqian Shi	Suyu Xian		Zhaoqing Shi	Gaoyao Shi
Hangzhou Shi	Xiaoshan Shi		Dongwan Shi	Boluo Xian
Hangzhou Shi	Yuhang Shi		Huizhou Shi	Huiyang Shi
Ningbo Shi	Yin Xian		Yangjiang Shi	Yangdong Xian
Cixi Shi	Yuyao Shi		Jieyang Shi	Jiedong Xian
Wenzhou Shi	Yongjia Xian		Nanning Shi	Yongning Xian
Cangnanxian	Pingyang Xian		Liuzhou Shi	Liujiang Xian
Wenzhou Shi	Ruian Shi		Guilin Shi	Lingchuan Xian
Wenzhou Shi	Leqing Shi		Wuzhou Shi	Cangwu Xian
Shaoxing Shi	Shaoxing Xian		Chengdu Shi	Shuangliu Xian
Shengzhou Shi	Xinchang Xian		Chengdu Shi	Pi Xian
Jinhua Shi	Jinhua Xian		Qujing Shi	Zhanyi Xian
Yiwu Shi	Dongyang Shi		Xi’an Shi	Changan Xian
Taizhou Shi	Wenling Shi			