Domestic Transportation Infrastructure and Export Performance of Multiproduct Firms: The Role of Domestic Intermediate Inputs

Gao, Longfei and Tang, Yao

Peking University, Peking University

1 May 2024

Domestic Transportation Infrastructure and Export Performance of Multiproduct Firms: The Role of Domestic Intermediate Inputs

Long-fei Gao[†], Yao Tang[‡]

May 1, 2024

Abstract

In a multi-product Melitz model, we demonstrate that after a drop in domestic trade costs, the cost savings on the shipping of domestic intermediate inputs dominate the pressure from increased competition, thus aiding surviving domestic firms in increasing the number of export varieties. The response of export revenue at the product level is heterogeneous; revenue from a firm's low-markup varieties will increase, while that from high-markup varieties will decrease. Total export revenue of a firm increases if its export varieties or its exports of low-markup products expand significantly. Using 2SLS regressions, we test this theory with data on Chinese manufacturing firms from 2000 to 2007 and find supportive evidence. As access to domestic intermediate inputs improves with the expansion of the railway network, the number of export varieties of domestic firms increases. Meanwhile, revenue per product drops. Because the positive effect on varieties dominates the negative effect on revenue per product, the total export revenue of firms increases on average. Furthermore, the entry of new firms also increases, lending additional support to our theoretical model.

Keywords: export performance, domestic transportation infrastructure, access to intermediate inputs, railway network, Chinese manufacturing firms

JEL: F10, F15, R4

[†] Department of Applied Economics, Guanghua School of Management, Peking University, 5 Yihuayuan Road, Haidian District, Beijing 100871, China. E-mail: jimlfgao@pku.edu.cn.

[‡] Corresponding author. Department of Applied Economics, Guanghua School of Management, and Institute of Carbon Neutrality, Peking University, 5 Yihuayuan Road, Haidian District, Beijing 100871, China. E-mail: ytang@gsm.pku.edu.cn.

1 Introduction

Geographical distance is a major barrier for firms engaging in trade activities. Transportation infrastructure helps overcome these barriers by facilitating the flow of goods, factors, and information across regions. This enhances market integration and reduces trade costs for businesses. There is substantial evidence that domestic transportation costs represent a significant portion of the total costs in international trade (Rousslang and To, 1993; Limao and Venables, 2001). As developing countries have invested heavily in domestic infrastructure, the trade literature has begun to explore how such investments in domestic transportation infrastructure can boost firms' export activities (Martincus and Blyde, 2013; Duranton et al., 2014; Coşar and Demir, 2016; Xu, 2016; Martincus et al., 2017; Donaldson, 2018; Fan et al., 2023). However, while existing literature focuses on transportation infrastructure and the trade of final goods, it has not yet explored the potential effects of infrastructure-related savings on domestic intermediate inputs in trade.

In the current study, we investigate how domestic infrastructure shapes the extensive and intensive margins of export by introducing the domestic trade of intermediate inputs into the multi-product Melitz model, following Mayer et al. (2014). In the setup, each firm has a core variety for which it possesses a high level of productivity. The addition of each new variety entails a loss in production efficiency. In equilibrium, the production of a product by a firm is determined by comparing the firm-specific production cost to the cutoff cost level required for survival in the market.

Introducing two domestic locations (inland and seaboard) and a foreign location into the setup, we demonstrate that reductions in domestic trade costs have three effects on firm exports: cost savings on the shipping of intermediate goods, increased competition due to better access to intermediate goods in both domestic locations, and cost savings on the shipping of final goods. By showing that the balance between the first two effects favors the former, we obtain three predictions following a drop in domestic trade costs. First, the number of a firm's export varieties increases. Second, the export revenue from a firm’s low-markup varieties will increase, while that from high-markup varieties will decrease. Third, the total export revenue of a firm will increase if there is a significant expansion in the number of varieties or a substantial increase in the exports of low-markup products.

We test the predictions of our model using data on the evolution of China’s railway network and firm-level export data from Chinese manufacturing firms between 2000 and 2007. To measure the improvement in access to domestic intermediate goods associated with the expansion of China's railway network, we construct an Upstream Market Access (UMA) index by incorporating inter-industry input-output relations into the Market Access (MA) index of Donaldson and Hornbeck (2016). We address the endogeneity of railway network formation by following the methodologies of Faber (2014), Banerjee, Duflo, and Qian (2020), and Fan et al. (2023), using historical and geographical information to construct two artificial transportation networks. The corresponding counterfactual UMA indices associated with these artificial networks serve as instrumental variables (IVs) for the actual UMA index in our Two-Stage Least Squares (2SLS) estimation.

The main empirical findings are as follows. First, a one-percent increase in UMA is associated with a 2.76% growth in firms' export values. This growth is driven by the expansion of the extensive margins for both seaboard and inland firms, as evidenced by increases in the number of export varieties and destinations for these firms. Second, at the intensive margins, the impact of UMA is overall negative, as product-level export regressions suggest that, on average, UMA reduces the export value per product. Consistent with our theory, this negative effect is more pronounced for firms with higher markup rates. Third, when examining the cost-saving and competition effects associated with access to intermediate goods, the results indicate that UMA reduces firms' costs for intermediate inputs but also intensifies local competition by promoting a higher net entry rate of firms at the city-industry level. Overall, the empirical results support our theory on how better access to intermediate inputs affects the export activities of multi-product firms.

The theory presented in this study enhances the understanding of trade costs within the multi-product Melitz model. Besides shaping competition in local and export markets through reduced transportation costs for final goods, as discussed by Mayer et al. (2014), we highlight the role of lower domestic trade costs in improving access to domestic intermediate goods. This improved access helps firms reduce input costs and enhance market competitiveness. We demonstrate that this channel may promote the export value of domestic firms through the expansion of export varieties, although responses at the product level are heterogeneous. Similar to Mayer et al. (2014), our model posits that the intrinsic driving force behind firms' adjustments in product scope is the efficiency (cost) ladder between their core and peripheral products. This differs from Bernard et al. (2011), who focus on heterogeneity in firm profitability at the product and country levels, and from Baldwin and Gu (2006), who assume product homogeneity but recognize product-level scale effects. Moreover, by allowing inland firms to engage in exports, our model presents a more flexible spatial economic structure compared to Coşar and Fajgelbaum (2016), who describe an economic configuration where coastal regions specialize in the export-oriented sector, and inland areas are excluded from international trade. Additionally, our model considers the extensive margin of traded varieties, contrasting with Coşar and Fajgelbaum's two-product setup, which abstracts from the extensive margin.

The empirical findings of this study highlight the effects of domestic intermediate inputs on the extensive and intensive margins of trade, thus enriching the empirical literature on transportation infrastructure and trade. Previous research has established that transportation infrastructure increases international trade volume at both the regional (Coşar and Demir, 2016; Duranton et al., 2014; Donaldson, 2018) and firm levels (Martincus and Blyde, 2013; Martincus et al., 2017; Liu et al., 2023; Tian and Yu, 2023). Specifically focusing on domestic transportation infrastructure, Donaldson (2018) finds that railroad construction facilitates inter-regional and international trade in colonial-era India. In contrast, Duranton et al. (2014) observe that highways mainly increase the weight of city exports with little impact on export value. The results of our study align more closely with those of Donaldson (2018), and we introduce domestic intermediate goods as a new channel influencing trade.

This paper also contributes to the literature on trade liberalization and firm performance by demonstrating that access to domestic intermediate inputs can improve productivity. Trade liberalization, measured by reductions in tariffs on intermediate goods, has been found to enhance firms' productivity (Amiti and Konings, 2007; Topalova and Khandelwal, 2011; Bas, 2012; Yu, 2015) and expand their product scopes (Goldberg et al., 2010). Our research is closely related to that of Tian and Yu (2013) and Feng et al. (2016), who find that the reduction of tariffs on intermediate goods increases firms' total exports or export intensity by providing access to lower-priced imported intermediate goods. In contrast to studies that examine the substitution of domestic intermediate goods with imported intermediates, we emphasize the cost-saving channel of improved access to domestic intermediate goods.

The remainder of the paper is structured as follows: Section 2 develops the theoretical model and decomposes the effects of domestic trade costs on firm exports. Section 3 presents the empirical strategies and data descriptions, while Section 4 reports the empirical results. Section 5 concludes the paper.

2 Model

In this section, we extend the model of multi-product firm of Mayer et al. (2014) by introducing two inputs, labor and intermediate goods, into the production function. we first develop the model of a closed economy.

2.1 Closed Economy

2.1.1 Model Setup

Preference. Consider an economy with L consumers. Preferences are defined over a continuum of differentiated varieties and a homogeneous good serving as the numéraire. All consumers have identical preferences described by the following utility function:

$$U = q_0^c + \alpha \int_{i\in\Omega} q_i^c di - \frac{1}{2}\gamma \int_{i\in\Omega} (q_i^c)^2 di - \frac{1}{2}\eta \left(\int_{i\in\Omega} q_i^c di\right)^2, \quad (1)$$

where q_0^c and q_i^c represent the individual consumption of the numéraire good and each variety $i \in \Omega$. The demand parameters α, γ, and η are all positive. The parameters α and η measure the degree to which consumers prefer the differentiated varieties relative to the numéraire. The parameter γ indexes the degree of product differentiation between the varieties. The substitutability between the varieties declines as γ increases, and consumers give increasing weight to smoothing consumption levels across varieties.

Each consumer supplies one unit of labor and makes consumption decisions subject to a budget constraint. The inverse demand for variety i is given by:

$$p_i = \alpha - \gamma q_i^c - \eta Q^c, \quad (2)$$

where p_i is the price of variety i and Q^c is the total demand of the differentiated product. we express the linear market demand function for i by:

$$q_i \equiv L q_i^c = \frac{\alpha L}{\eta M+\gamma} - \frac{L}{\gamma} p_i + \frac{\eta M L}{(\eta M+\gamma)\gamma} \bar{p}, \ \forall i \in \Omega^*, \quad (3)$$

where M is the measure of varieties in the subset $\Omega^* \subset \Omega$ that are consumed (with positive q_i^c),[1] and $\bar{p} = \frac{1}{N}\int_{i\in\Omega^*} p_i di$ is the average price over Ω^*. The demand elasticity here depends on the intensity of market competition, which is associated with the degree of product differentiation (γ), the measure of competing varieties (N) and their average price ($\bar{p}$).[2]

Production. The production of the differentiated varieties requires two inputs: labor (L) and intermediate goods (I), both of which are assumed to be inelastically supplied in a competitive market. The production technology exhibits constant returns to scale and is homogeneous across different varieties. Consider a firm producing a variety with the following production function:

---

[1] The upper price bound $p^{max} \equiv \frac{\alpha\gamma+\eta M\bar{p}}{\eta M+\gamma}$ is defined by letting $q_i = 0$ in Equation (3). Therefore, all varieties $i \in \Omega^*$ with positive demand must satisfy $p_i < p^{max}$.

[2] In contrast, the demand elasticity under the CES preference is constant given the degree of product differentiation.

$$q = \varphi L^{\beta} I^{1-\beta}, \quad (4)$$

where φ is firm's productivity on the variety. From the firm's cost minimization problem, the unit production cost of a variety is:[3]

$$c = \frac{1}{\varphi} w^{\beta} r^{1-\beta}, \quad (5)$$

where w and r are the factor cost of labor and intermediate goods, respectively. To enter the market, a firm must incur a fixed cost denoted as F before knowing its productivity φ of a variety. Assume that productivity of firms is drawn from a common distribution $G_{\varphi}(\varphi)$, with its support defined over the interval $[\varphi_M, \infty]$.

As in Mayer et. (2014), each firm has a core variety considered as its "core competency". A firm possesses the highest productivity on its core variety. The firm can expand its production line to other varieties, but each additional variety entails additional efficiency losses.

After entering the market, a firm can immediately know the productivity of its core product, φ. Other varieties are indexed by m ($m = 0$ for the core variety) according to their distance from the core competency. The productivity of a firm with core productivity φ on its m-th variety is assumed to be $\phi(m, \varphi) = \omega^m \varphi$ with $\omega \in (0,1)$. The parameter ω governs the efficiency gap between products. As ω approaches 1, firms are more likely to expand from their core variety, as production of varieties with larger m incurs minimal efficiency losses. Conversely, as ω goes to zero, firms should produce only their core product if they produce at all.

Equivalently, we can denote core competency in terms of production cost. Given Equation (5) and the productivity distribution, the unit cost c follows distribution $G(c)$ with support $(0, c_M]$, where $c_M = \frac{1}{\varphi_M} w^{\beta} r^{1-\beta}$. Denote $v(m, \varphi)$ as the marginal cost of the m-th variety of a firm with core productivity φ, it follows that $v(m, \varphi) = \frac{1}{\omega^m \varphi} w^{\beta} r^{1-\beta} = \omega^{-m} c \equiv v(m, c)$. As m moves away from core competency, the marginal cost $v(m, c)$ increases.

Firm Behavior. After investing in the (sunk) entry cost, a firm then decides the production of each variety. Product m will be produced if the firm can cover its marginal cost $v(m, c)$. Note that, a firm survives only if it earns non-negative profit on its core variety.

[3] The exact expression from cost minimization is $c = \frac{1}{\beta^{\beta}(1-\beta)^{1-\beta}} \frac{1}{\varphi} w^{\beta} r^{1-\beta}$, which equals Equation (5) multiplied by a constant term $\frac{1}{\beta^{\beta}(1-\beta)^{1-\beta}}$. With proper normalization of φ, one can obtain the simple expression in Equation (5).

Otherwise, the firm will immediately exit the market. Specifically, given a core marginal cost c, a surviving firm will maximize the following operating profit:

$$\max_{p(c),q(c)} \pi(c) = (p(c) - c)q(c), \tag{6}$$

where $q(c) = \frac{\alpha L}{\eta M+\gamma} - \frac{L}{\gamma}p(c) + \frac{\eta ML}{(\eta M+\gamma)\gamma}\bar{p}$. Denote $c_D = \frac{\alpha\gamma+\eta M\bar{p}}{\eta M+\gamma}$ as the cost cutoff of survival such that $\pi(c_D) = 0$. Only those firms with $c \leq c_D$ could survive and produce its core variety. It is similar for the survival of other products: a firm could profitably produce their m-th variety if $v(m,c) \leq c_D$, that is, $c \leq \omega^m c_D$. Therefore, the production scope of a firm with cost c, defined by the number of varieties it produces, is given by:

$$M(c) = \begin{cases} \max(m|c \leq \omega^m c_D) + 1 & if \quad c \leq c_D \\ 0 & if \quad c > c_D \end{cases}. \tag{7}$$

It is clear that $M'(c) \leq 0$ and $\partial M(c)/\partial c_D \geq 0$ such that firms with lower core marginal costs (from larger productivity φ) could serve more varieties in markets with less intense competition.

Now for each surviving variety m of firm with core marginal cost c, the equilibrium outcomes can be entirely expressed in terms of c_D and c:

$$q(m,c) = \frac{L}{2\gamma}(c_D - v(m,c)) = \frac{L}{2\gamma}(c_D - \omega^{-m}c), \tag{8}$$

$$p(m,c) = \frac{1}{2}(c_D + v(m,c)) = \frac{1}{2}(c_D + \omega^{-m}c), \tag{9}$$

$$\lambda(m,c) = \frac{1}{2}(c_D - v(m,c)) = \frac{1}{2}(c_D - \omega^{-m}c), \tag{10}$$

$$x(m,c) = \frac{L}{4\gamma}(c_D^2 - v(m,c)^2) = \frac{L}{4\gamma}(c_D^2 - (\omega^{-m}c)^2), \tag{11}$$

$$\pi(m,c) = \frac{L}{4\gamma}\big(c_D - v(m,c)\big)^2 = \frac{L}{4\gamma}(c_D - \omega^{-m}c)^2, \tag{12}$$

where $\lambda(m,c) = p(m,c) - v(m,c)$ and $x(m,c) = p(m,c)q(m,c)$ are the markup and product sales (revenue) of variety m. From the above results, lower firms and varieties have lower prices, higher market demand and higher markup rates, thus yielding higher profits. On the other hand, as an indicator of tougher market competition, a decrease in c_D certainly reduces the prices, outputs, and profits for all current varieties in production.

2.1.2 Free Entry Equilibrium

After entering the market, the aggregated profit of a firm with core marginal cost c can be derived from Equations (7) and (12):

$$\Pi(c) = \sum_{m=0}^{M(c)-1} \pi(m,c) = \sum_{m=0}^{M(c)-1} \frac{L}{4\gamma}(c_D - \omega^{-m}c)^2. \tag{13}$$

Before entering the market, the expected profit for a firm is $\int_0^{c_D} \Pi(c) dG(c) - F$, and hence the free entry equilibrium condition is:

$$\int_0^{c_D} \Pi(c) dG(c) = F \tag{14}$$

For simplicity in subsequent analysis, we parameterize the distribution of firm's core productivity φ. Specifically, φ is assumed to follow a Pareto distribution with lower bound φ_M and shape parameter $k \geq 1$, i.e., $G_\varphi(\varphi) = \left(\frac{\varphi}{\varphi_M}\right)^{-k}$. This implies a distribution of marginal cost c:

$$G(c) = \left(\frac{c}{c_M}\right)^k, c \in [0, c_M]. \tag{15}$$

The parameter k measures the degree of cost dispersion. With a larger k, the cost distribution is more densely concentrated at higher costs.

Consider a mass of prospective entrants in the local market. Under the assumption of Pareto distribution of costs, we can derive from the free entry condition the analytical solution for the cost cutoff c_D and the mass of surviving firms N in equilibrium:[4]

$$c_D = \left(\frac{\gamma\theta}{L\Omega}\right)^{\frac{1}{k+2}}, \tag{16}$$

$$N = \frac{2(k+1)\gamma}{\eta\Omega}\left(\frac{\alpha}{c_D} - 1\right), \tag{17}$$

where $\theta = 2(k+1)(k+2)(c_M)^k F$ and $\Omega = \frac{1}{(1-\omega^k)} > 1$. The equilibrium c_D is negatively correlated with the local market size L while positively correlated with entry costs F and the upper bound of marginal cost c_M. Note that, as $c_M = \frac{1}{\varphi_M} w^\beta r^{1-\beta}$, c_D decreases with the reduction of factor costs w and r. And from Equation (17) we know that the number of firms increases with a lower c_D.

2.2 Open Economy

In this section, we will derive the equilibrium in an open economy by introducing trade costs. In the open-economy model, trade costs affect both the transportation of final goods and the cost of intermediate inputs. we will discuss the implications of the reduction in trade costs in a simplified trade structure in Section 2.2.2.

2.2.1 Free Entry Equilibrium in an Open Economy

[4] See Appendix B for details of the proof.

We first consider a general open economy consisting of some asymmetric regions, denoted by $l = 1,2,\dots,J$. Exporting products from region l to region f incurs an iceberg trade cost $e^{\tau_{lf}} > 1$, for all $l \neq f$.[5] Trade costs are symmetric such that $e^{\tau_{lf}} = e^{\tau_{fl}}$.

In addition to final goods, trade costs also affect the accessibility of intermediate inputs. For simplicity, we assume that changes in trade costs only affect the prices of intermediate inputs without altering the trade conditions. Specifically, we assume that the intermediate good price, $r = r(e^{\tau_{lf}} | l \neq f\})$, is (weakly) increasing in each element of the vector of trade costs, i.e., and $\frac{\partial r}{\partial e^{\tau_{lf}}} \geq 0$. The assumption implies that reduction in the trade cost between any two locations will improve the overall intermediates network, thereby reducing the prices in the perfectly competitive market for intermediate goods.[6]

Similar to the setup in a closed economy, the core marginal costs follow the distribution $G(c) = \left(\frac{c}{c_M}\right)^k$ with support $(0, c_M]$, where $c_M = \frac{1}{\varphi_M} w^\beta r^{1-\beta}$.

For any region l, we first determine the cost cutoff for firms to survive in the local market, denoted by $c_{ll} = \frac{\alpha\gamma + \eta M_l \bar{p}_l}{\eta M_l + \gamma}$, where M_l is the measure of varieties sold in region l and $\bar{p}_l$ is the average price. Due to the existence of trade costs, for a for a firm in region l with a core marginal production cost c, the total marginal cost of selling in region f is $e^{\tau_{lf}} c$. Therefore, the cost cutoff for exporting from l to f is $c_{lf} = \frac{c_{ff}}{e^{\tau_{lf}}} = \frac{1}{e^{\tau_{lf}}} \frac{\alpha\gamma + \eta M_f \bar{p}_f}{\eta M_f + \gamma}$.

Therefore, the number of varieties a firm with cost c exports from l to f is given by:

$$M_{lf}(c) = \begin{cases} \max\left(m \middle| c \leq \omega^m c_{lf}\right) + 1 & if \quad c \leq c_{lf} \\ 0 & if \quad c > c_{lf} \end{cases}. \tag{18}$$

Note that when $l = f$, this equation represents the production scope for the local market.

Now we write the equilibrium outcomes after a firm at l with cost c enters the exporting market of f.

$$q_{lf}(m,c) = \frac{L_l}{2\gamma} e^{\tau_{lf}} \left(c_{lf} - \omega^{-m} c\right) = \frac{L_l}{2\gamma} \left(c_{ff} - e^{\tau_{lf}} \omega^{-m} c\right), \tag{19}$$

$$p_{lf}(m,c) = \frac{1}{2} e^{\tau_{lf}} \left(c_{lf} + \omega^{-m} c\right) = \frac{1}{2} \left(c_{ff} + e^{\tau_{lf}} \omega^{-m} c\right), \tag{20}$$

$$\lambda_{lf}(m,c) = \frac{1}{2} e^{\tau_{lf}} \left(c_{lf} - \omega^{-m} c\right) = \frac{1}{2} \left(c_{ff} - e^{\tau_{lf}} \omega^{-m} c\right), \tag{21}$$

$$x_{lf}(m,c) = \frac{L_l}{4\gamma} (e^{\tau_{lf}})^2 \left(c_{lf}^2 - (\omega^{-m} c)^2\right) = \frac{L_l}{4\gamma} \left(c_{ff}^2 - (e^{\tau_{lf}} \omega^{-m} c)^2\right), \tag{22}$$

---

[5] For small values of τ, it holds that $e^\tau \approx 1 + \tau$.

[6] Alternatively, Consider the intermediate price as an increasing function of trade cost index, that is, $r = r(t)$ and $r'(t) \geq 0$, where $t = \left(\sum_{f \neq l} (e^{\tau_{lf}})^{1-\sigma}\right)^{1/1-\sigma}$.

$$\pi_{lf}(m,c) = \frac{L_l}{4\gamma}(e^{\tau_{lf}})^2\left(c_{lf} - \omega^{-m}c\right)^2 = \frac{L_l}{4\gamma}\left(c_{ff} - e^{\tau_{lf}}\omega^{-m}c\right)^2. \tag{23}$$

And its total profits at the local and exporting markets are:

$$\Pi_{ll}(c) = \sum_{m=0}^{M_{ll}(c)-1} \pi_{ll}(m,c) = \sum_{m=0}^{M_{ll}(c)-1} \frac{L_l}{4\gamma}(c_{ll} - \omega^{-m}c)^2, \tag{24}$$

$$\Pi_{lf}(c) = \sum_{m=0}^{M_{lf}(c)-1} \pi_{lf}(m,c) = \sum_{m=0}^{M_{ll}(c)-1} \frac{L_l}{4\gamma}\left(c_{ff} - e^{\tau_{lf}}\omega^{-m}c\right)^2. \tag{25}$$

Prior to entry, the expected profit of a potential entrant at region l is given by $\int_0^{c_{ll}} \Pi_{ll}(c)dG(c) + \sum_{f\neq l}\int_0^{c_{lf}} \Pi_{lf}(c)dG(c) - F$. And the free entry condition at region l is:

$$\int_0^{c_{ll}} \Pi_{ll}(c)dG(c) + \sum_{f\neq l}\int_0^{c_{lf}} \Pi_{lf}(c)dG(c) = F. \tag{26}$$

For region $l = 1,2,\ldots,J$, the J free entry conditions form a system of equations. Applying the Cramer's rule, the cost cutoff of serving the local market for each region can be solved as:[7]

$$c_{ll} = \left(\frac{\gamma\theta}{L_l\Omega}\frac{|\mathcal{P}_l|}{|\mathcal{P}|}\right)^{\frac{1}{k+2}}, \tag{27}$$

where L_l is the market size of l, and $|\mathcal{P}|$ is the determinant of the following matrix:

$$\mathcal{P} \equiv \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1J} \\ \rho_{21} & 1 & \cdots & \rho_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{J1} & \rho_{J2} & \cdots & 1 \end{pmatrix},$$

where $\rho_{lf} \equiv \frac{1}{\left(e^{\tau_{lf}}\right)^k} < 1$ is negatively correlated with the trade cost $e^{\tau_{lf}}$ and hence measures the market integration between l and f.[8] And $|\mathcal{P}_l|$ is the determinant of the matrix which replaces the l-th column of $\mathcal{P}$ with a column vector with all elements equal to one.

The number of surviving firms in l is determined as:

$$N_l = \frac{2(k+1)\gamma}{\eta\Omega}\left(\frac{\alpha}{c_{ll}} - 1\right), \tag{28}$$

which is negatively correlated with the surviving cutoff c_{ll}.

2.2.2 Trade Cost in an Inland-Seaboard Trade Structure

Now we simplify the general setup of asymmetric open economy from the previous section to characterize the domestic trade pattern and firm exporting behavior focused on in this paper. Consider a country consists of two cities, an inland city (indexed by l) and a seaboard one (indexed by s). As in Cosar and Fajgelbaum (2016), we assume that only the

[7] See Appendix B for details on solution.

[8] Mayer et al. (2014) called it "freeness" of trade.

seaboard city can trade directly with the rest of the world (indexed by f),[9] because exports must cross through a port to be shipped internationally. As described in Figure 1, the domestic trade cost between l and s is denoted by e^{τ_0}, and the international trade cost between s and f is e^{τ_1}. It means exporting products from l to f incurs a trade cost $e^{\tau_0+\tau_1}$. As in the previous section, the intermediate goods price is an increasing function of trade costs such that $r = r(e^{\tau_0}, e^{\tau_1})$ with $\hat{r}_0 \equiv \frac{\partial r}{\partial e^{\tau_0}} \geq 0$ and $\hat{r}_1 \equiv \frac{\partial r}{\partial e^{\tau_1}} \geq 0$.

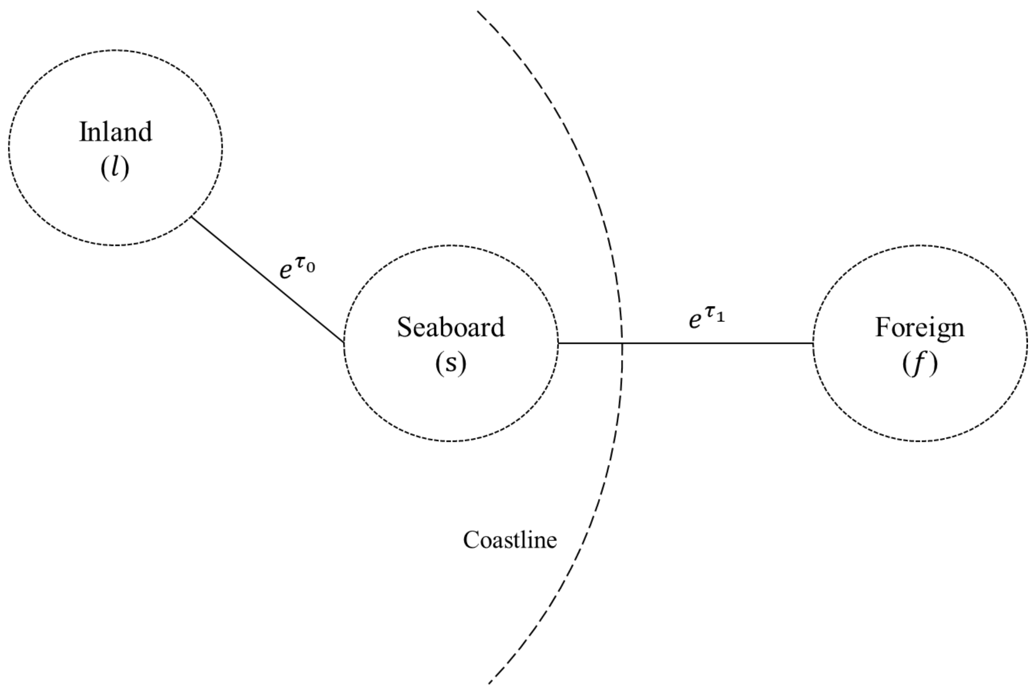


Figure 1 An Inland-Seaboard Trade Structure

Under this trade geography, we can derive the cutoffs to survive in the three markets:

$$c_{ss} = \left(\frac{\gamma\theta}{L_s\Omega}\frac{1-\rho_0\rho_1}{(1+\rho_0)(1+\rho_1)}\right)^{\frac{1}{k+2}}, \tag{29}$$

$$c_{ll} = \left(\frac{\gamma\theta}{L_l\Omega}\frac{1}{1+\rho_0}\right)^{\frac{1}{k+2}}, \tag{30}$$

$$c_{ff} = \left(\frac{\gamma\theta}{L_f\Omega}\frac{1}{1+\rho_1}\right)^{\frac{1}{k+2}}, \tag{31}$$

where $\rho_0 = \frac{1}{(e^{\tau_0})^k} < 1$ and $\rho_1 = \frac{1}{(e^{\tau_1})^k} < 1$ are the degrees of domestic and international market integration, respectively. Note that $\frac{c_{ll}}{c_{ss}} = \left(\frac{L_o}{L_l}\frac{1+\rho_1}{1-\rho_0\rho_1}\right)^{\frac{1}{k+2}} > 1$ if $L_s \geq L_l$, which means

[9] The rest of the world is treated as a foreign country.

that when the market size of seaboard city is at least not smaller than that of the inland city, firms require lower marginal costs to survive in the seaboard city.

As the focus of this paper is on the export behavior of domestic firms, the interest is on the cost cutoffs for firms located in domestic cities l and s to export to the foreign market f:

$$c_{sf} = \frac{c_{ff}}{e^{\tau_1}} = \left(\frac{\gamma\theta}{L_f\Omega} \frac{(\rho_1)^{\frac{k+2}{k}}}{1+\rho_1} \right)^{\frac{1}{k+2}}, \tag{32}$$

$$c_{lf} = \frac{c_{ff}}{e^{\tau_0+\tau_1}} = \left(\frac{\gamma\theta}{L_f\Omega} \frac{(\rho_0\rho_1)^{\frac{k+2}{k}}}{1+\rho_1} \right)^{\frac{1}{k+2}}. \tag{33}$$

Clearly, it holds that $c_{lf}/c_{sf} = \rho^{1/k} < 1$, indicating that firms in the inland city face tougher export competition relative to the seaboard city. With this inequality, we can compare the differences in export performance of firms between two locations with the same core cost. Recall that $\partial M(c)/\partial c_D \geq 0$ for any cutoff c_D, and thus we have $M_{sf}(c) \geq M_{lf}(c)$ given the marginal cost c.

The aggregated firm-level export revenue at city s and l are given by:

$$X_{sf}(c) = \sum_{m=0}^{M_{sf}(c)} x_{sf}(m,c) = \sum_{m=0}^{M_{sf}(c)} \frac{L_s}{4\gamma} \left(c_{ff}^2 - (e^{\tau_1}\omega^{-m}c)^2\right), \tag{34}$$

$$X_{lf}(c) = \sum_{m=0}^{M_{lf}(c)} x_{lf}(m,c) = \sum_{m=0}^{M_{lf}(c)} \frac{L_l}{4\gamma} \left(c_{ff}^2 - (e^{\tau_0+\tau_1}\omega^{-m}c)^2\right). \tag{35}$$

Clearly, it must hold that $X_{sf}(c) > X_{lf}(c)$ as $M_{sf}(c) \geq M_{lf}(c)$ and $x_{sf}(m,c) > x_{lf}(m,c)$ for $\forall m$.

We obtain the following lemma based on the above discussion:

Lemma 1 *In an open economy of inland-seaboard trade structure as described in this section, compared to being in the inland city, a firm in the seaboard city has: (i) a lower cost cutoff for survival, (ii) a larger number of export varieties, (iii) and higher export revenue.*

Now we conduct the comparative statics regarding the effects of the domestic trade cost on firm behavior. In this model, a reduction in e^{τ_0} affects a firm's local production as well as export to other markets through three channels. Firstly, it directly reduces the transportation costs of final goods in domestic transportation, to which we refer as the *cost-saving effect of final goods*. Note that this enhances the price advantage of a domestic firm's exporting products in the destination market. In the local market, however, it also faces stronger competition from final goods of other regions. Secondly, better access to domestic intermediate goods reduces the price of intermediate input, thereby lowering the firm's marginal production cost for a given level of productivity, which is reflected by $\frac{\partial c}{\partial e^{\tau_0}} = (1-\beta)\frac{\hat{r}_0}{r}c > 0$. We call it the *cost-saving effect of intermediate inputs*. In addition, as $\frac{\partial \theta}{\partial e^{\tau_0}} = k(1-\beta)\frac{\hat{r}_0}{r}\theta > 0$,[10] a firm would also face stronger competitors with lower intermediate costs on average, which is a *competition effect of intermediate inputs*. Different from the direct effect on transportation cost of final goods, the last two effects are both associated with improvement in the access to intermediate goods. We will separate them in the subsequent theoretical and empirical analysis.

We first look at the changes in local market competition. Differentiate the cutoff for firms to survive in the local markets o and l with respect to e^{τ_0}, we have:

$$\frac{\partial c_{ss}}{\partial e^{\tau_0}} = \frac{k}{k+2}c_{ss}[(1-\beta)\frac{\hat{r}_0}{r} + \frac{(1+\rho_1)(\rho_0)^{(k+1)/k}}{(1+\rho_0)(1-\rho_0\rho_1)}] \geq 0, \quad (36)$$

$$\frac{\partial c_{ll}}{\partial e^{\tau_0}} = \frac{k}{k+2}c_{ll}[(1-\beta)\frac{\hat{r}_0}{r} + \frac{(\rho_0)^{(k+1)/k}}{1+\rho_0}] \geq 0. \quad (37)$$

In each of the two equations above, the first term in the square bracket represents the competition effect of intermediate inputs, while the second term represents tougher local competition due to lower trade cost of imported final goods. These two forces in the same direction make the surviving cutoffs drops, leading to more intense local competition in domestic cities s and l. And according to Equation (28), we can also deduce that $\frac{\partial N_s}{\partial e^{\tau_0}} < 0$ and $\frac{\partial N_l}{\partial e^{\tau_0}} < 0$, implying an increase of the measure of surviving firms. We have the following lemma:

Lemma 2 *The reduction in the domestic trade cost will lower the cost cutoffs for firms to survive and increase the measure of active firms through both the competition effect of intermediate goods and the cost-saving effect of final goods.*

Of particular interest is how the export behavior of a surviving firm changes with domestic trade costs. First is the extensive margin of export which is measured by the number of export varieties. From Equation (7), we know that $M(c)$ is the largest m that satisfies

---

[10] Recall that $\theta = 2(k+1)(k+2)(c_M)^k F$ and $c_M = \frac{1}{\varphi_M}w^\beta r^{1-\beta}$.

$c \leq \omega^m c_D$, given a cost cutoff c_D. This condition can be rewritten as:

$$m \leq \frac{\ln c_D - \ln c}{\ln \omega^{-1}}, \tag{38}$$

where the right side $h(c, c_D) \equiv \frac{\ln c_D - \ln c}{\ln \omega^{-1}}$, with $\frac{\partial h}{\partial c} < 0$ and $\frac{\partial h}{\partial c_D} > 0$, determines the maximum value of m. It implies that $\partial M(c)/\partial h \geq 0$.[11] There are confounding effects of better access to intermediate inputs through lower domestic trade cost. On the one hand, low-cost domestic intermediate goods help the firms to surpass the export threshold for more varieties, thereby expanding the export product scope. On the other side, tougher export competition, from export competitors who also benefit from intermediate cost savings, may block the export of those marginal products. It can be proven that the former force is dominant such that $\partial M_{sf}(c)/\partial e^{\tau_0} \geq 0$ and $\partial M_{lf}(c)/\partial e^{\tau_0} \geq 0$. Note that, in addition to the channel of intermediates, the reduction in domestic trade costs also increases firms' export varieties through cost-saving effect of final goods.

The result is summarized in the following proposition:[12]

Proposition 1 *The increase access to domestic intermediate goods associated with the reduction in domestic trade costs will lead both seaboard and inland firms to expand their scope of export varieties.*

To understand the change in firms' total exports, we first analyze the intensive margin of export, that is, the revenue of a given export variety. By holding $M(c)$ constant, we take the derivatives of the export revenues of the m-th variety for a firm with cost c at both city s and l, $\partial x_{sf}(m,c)$ and $\partial x_{lf}(m,c)$, with respect to e^{τ_0}, respectively. It can be proved that:

$$\frac{\partial x_{sf}(m,c)}{\partial e^{\tau_0}} \lesseqgtr 0 \quad if \quad \frac{v(m,c)}{c_{sf}} \gtreqless \left(\frac{k}{k+2}\right)^{1/2}, \tag{39}$$

$$\frac{\partial x_{lf}(m,c)}{\partial e^{\tau_0}} \lesseqgtr 0 \quad if \quad \frac{v(m,c)}{c_{lf}} \gtreqless \left(\frac{k}{k+2}\frac{1}{1+\mu}\right)^{1/2}, \tag{40}$$

where $\mu = \frac{1}{e^{\tau_0}(1-\beta)\frac{\hat{r}_0}{r}}$. This implies that only the export revenues of higher-cost products will increase with the decline in domestic trade costs. Recall that the markup of the m-th variety of a firm with cost c decreases monotonically with $v(m,c)$. By Substituting $v(m,c)$ in Equations (39) and (40) with the expression of markup in Equation (21), we have:

[11] Here I treat $M(c)$ as a continuous variable for simplicity in mathematical proof.

[12] See Appendix B for the proof.

$$\frac{\partial x_{sf}(m,c)}{\partial e^{\tau_0}} \lesseqgtr 0 \quad if \quad \lambda_{sf}(m,c) \lesseqgtr \frac{1}{2}\left(1-\left(\frac{k}{k+2}\right)^{1/2}\right) e^{\tau_1} c_{sf}, \tag{41}$$

$$\frac{\partial x_{lf}(m,c)}{\partial e^{\tau_0}} \lesseqgtr 0 \quad if \quad \lambda_{lf}(m,c) \lesseqgtr \frac{1}{2}\left(1-\left(\frac{k}{k+2}\frac{1}{1+\mu}\right)^{1/2}\right) e^{\tau_0+\tau_1} c_{lf}, \tag{42}$$

As shown in the Appendix B, the net effect from cost saving of intermediate inputs increases with $v(m,c)$ and hence decrease with $\lambda(m,c)$, so the exports value of low-markup varieties of low-markup firms could benefit more from the improved access to intermediate goods.

We summarize the following proposition:

Proposition 2 *After the reduction of domestic trade cost, the cost-saving effect of intermediate goods is bigger (smaller) than the competition effect of intermediates for varieties with lower (higher) markups, leads to an increase (decrease) in their product-level export revenues.*

Finally, although it is difficult to obtain a comprehensive description about the relationship between a firm's total export revenue and domestic trade costs, the following conditional relationship can be inferred from the two propositions:

Corollary 1 *Following a reduction in domestic trade cost, better access to intermediate goods could raise the total export revenues of firms at seaboard or inland cities if at least one of the following two conditions holds: (i) the number of export varieties increases sufficiently, (2) given the export scope, the total export growth of low-markup varieties is greater than the export reduction of high-markup varieties.*

3 Empirical Design and Data

Our objective is to examine how the domestic trade integration brought about by the reduction in domestic trade costs affects firms' export performance. Specifically, we hope to distinguish the effect better access to domestic intermediate goods from the direct cost-saving effect during the transportation of final export products.

In this section, we first give the regression specification based on our model's predictions. We then introduce the method to construct an upstream market access index and address the identification questions. Finally, we describe the firm-level data used in the empirical analysis.

3.1 Regression Model

The regression model to be estimated is as follows:

$$\ln(export_{ijct}) = \beta_0 + \beta_1 \ln(UMA_{jct}) + \beta_2 \ln(dist\_port_{ct}) + \beta_3 X_{ct} + \beta_4 Z_{ijct} \\ +\varphi_i + \omega_t + \mu_{ct} , \quad (43)$$

where $export_{ijct}$ is the exports of firm i in industry j and city c at year t. Our dependent variable in the benchmark regression is the log of firm-level exports value. In further empirical analysis, we also use firms' export varieties, destination countries, variety-destination combinations, and firm-product-level export values as the dependent variables.

The variable UMA_{jct} represents the upstream market access (UMA) of industry j in city c which we construct by extending the method of Donaldson and Hornbeck (2016) by combining geographical information of railway network with inter-industry input-output relations. This index measures the spatial integration of firms with their upstream industries in the domestic transportation network, thereby depicting their access to domestic intermediate goods. We show the details on the UMA index in Section 3.2. The variable $dist\_port_{ct}$ is the is the distance from prefecture c to its nearest port city[13], which captures the transportation costs of the inland segment of exports through coastal ports. We are particularly interested in coefficients β_1 and β_2 because they correspond, respectively, to the effects of domestic trade costs on firm exports through the intermediate goods channel and the final goods channel.

We also include a series of control variables to capture the effects of both city-level and firm-level factors. X_{ct} is a cluster of prefecture-level control variables that may affect firms' export behaviors, including log of GDP ($lngdp$), log of GDP per capita ($lngdppc$), GDP ratio of the secondary to tertiary industries ($indserv$), and highway density measured as kilometers of highway per square kilometer ($highway$).[14] Z_{ijct} the vector of firm characteristics, including firm size (log of total assets, denoted $size$), capital intensity (fixed assets per worker, denoted $klratio$) and an indicator variable for state-owned enterprises (SOE). The variables φ_i, ω_t, and μ_{ct} represent the firm fixed effect, year fixed effect and the error term, respectively.

3.2 Construction of Upstream Market Access

Following the market access approach proposed by Donaldson and Hornbeck (2016), we

---

[13] It refers to the major ports engaged in international trade in mainland China, see Appendix C for list of major ports issued by the Ministry of Transport of China in 2004.

[14] Due to the lack of historical road network data at the city level, I use province-level road density as a proxy.

construct an upstream market access index based on connectivity to freight railway in China. We extend the original market access at region level to the region-industry level by incorporating inter-industry input-output information. Our index also differs in that we also use the annual average freight rate to capture the changes in transportation prices. Specifically, the upstream market access of industry j in city c at year t is given by:

$$UMA_{jct} = \sum_{c' \neq c} \tau_{cc',t}^{-\theta} \sum_{j' \neq j} a_{jj'} \sum_{i' \in I(j',c')} VA_{i'j'c',t}, \tag{44}$$

where $VA_{i'j'c',t}$ is the value added of firm i' of industry j' in domestic city c' at year t with $I(j',c')$ representing the set of firms of industry j' in city c'. Therefore, the part $\sum_{i' \in I(j',c')} VA_{i'j'c',t}$ represents the aggregated value added at the city-industry level. Our firm-level output data comes from the Annual Survey of Industrial Firms (ASIF) in China, which will be detailed in Section 3.4. The factor $a_{jj'}$ is the input-output coefficient, measured by the proportion of total output of industry j that comes from the intermediate inputs of industry j'. Therefore, j' is an upstream industry to j if $a_{jj'}$ is positive. The input-output coefficient data is sourced from the 2002 Chinese Input-output Tables. $\tau_{cc',t}^{-\theta}$ is an index of railway transportation cost per ton of goods shipped between city c and c' at year t. θ is known as the price elasticity of trade that measures the substitution of outputs between different regions. Following Donaldson (2018), we choose 3.8 as the benchmark value.

Following Baum-Snow et al. (2016), the expression for $\tau_{cc',t}$ is given by:

$$\tau_{cc',t} = 1 + p_t \left(d_{cc',t}\right)^{\sigma}, \tag{45}$$

where p_t is the annual average freight rate. We obtain the average freight rate by dividing the total revenue of railway transportation deflated with production price index by the total freight volume in tons. The variable $d_{cc',t}$ is the shortest railway distance between two cities calculated based on Dijkstra et al. (1959). σ is a parameter that captures the concave relationship between shipping cost and railway distance, which is set to 0.8 (Baum-Snow et al., 2016).

To calculate the shortest railway distance between any two cities, we construct the vector maps of the Chinese national railway network. The detailed steps for constructing the railway dataset are described in Appendix A.

Our index captures differences among firms in access to domestic upstream intermediate goods in three dimensions. Firstly, in terms of spatial dimension, geographic remoteness hinders connectivity between local firms and those in other regions. The second aspect is the heterogeneity in industrial distribution. Among enterprises within the same city, if neighboring regions are densely populated with upstream firms related to a particular industry, then local firms in that industry will have better accessibility to intermediate goods. The third aspect is time variation. Cities with rapid railway development will help their firms in obtaining more intermediate goods from other regions at lower costs.

3.3 Endogeneity Problems

In the benchmark regression of Equation (43), potential endogeneity arising from the upstream market access index comes from two sources. Firstly, there may exist some unobserved time trends at the city-industry level, such as industrial structural transformations and enterprise migrations, which simultaneously affect the value-added output and export performance of firms across regions. On the other hand, the selection of railway routes is evidently an endogenous decision, being related to some unobserved political, economic, and cultural factors at the city level, which in turn affect firms' production and export decisions.

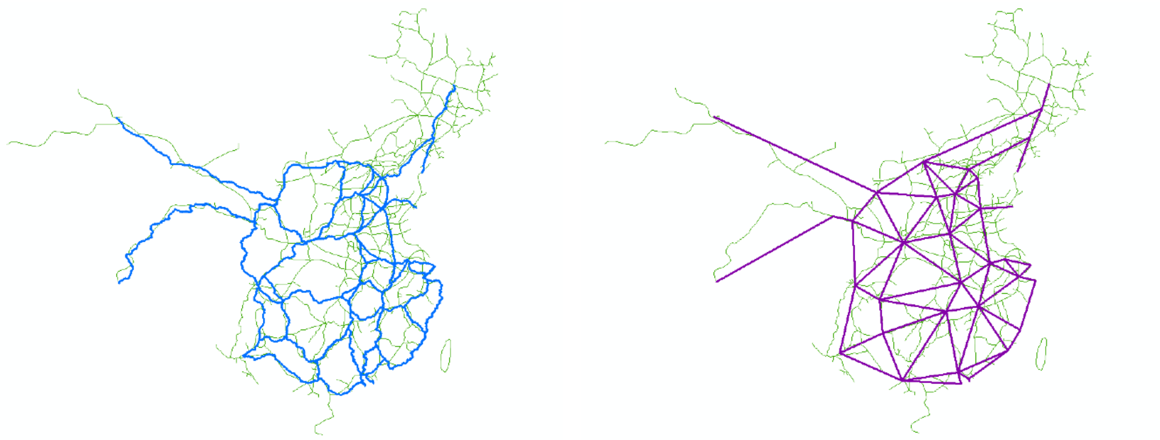
For the former concern, we use the value added of upstream firms in the first year of the sample (2000) as weights to construct the index. By doing so, the variation of out index over time is entirely driven by changes in the railway network.

To address potential endogeneity from the route choice of railways, we conduct an instrumental variable (IV) method to identify exogenous variations in railway construction. Specifically, we construct artificial transportation networks that aim to link 36 major cities in mainland China based on exogenous rules.[15] We construct two artificial transportation networks to link these major cities. Firstly, following the strategy of Faber (2014), we use the algorithm in the ArcGIS software to construct the least costly path between two cities. As in Fan, Lu, and Luo (2021), the development cost of each pixel in ArcGIS (a 1-meter-by-1-meter area) is posited to be proportional to the sum of the average gradient and 25 times the indicator function for presence of water body. The ArcGIS program computes the sum of development cost of all pixels associated with a path between two cities, and identify the least costly path which we take as the artificial transportation path. Secondly, following Banerjee, Duflo, and Qian (2020), we use the straight line between the cities as an artificial transportation network. Because the locations of the 36 major cities are determined by historical factors, whether any other city in the sample is likely to be better connected to the actual railway network is

[15] The major cities include the 27 provincial capitals in Mainland China, four municipalities directly under the Central Government (Beijing, Shanghai, Tianjin, and Chongqing) and five additional vice-provincial cities that are not provincial capitals (Dalian, Ningbo, Qingdao, Shenzhen, and Xiamen).

influenced by the exogenous geographic factors summarized by the straight-line-spanned artificial networks.

The artificial transportation networks are illustrated in Figure 2. The left and right panels correspond to the first and the second artificial network, respectively. In both panels, the denser network illustrated with thinner lines is the actual railway network in 2007. The artificial networks are clearly correlated with the actual network.



(a) Artificial network linked by least costly paths (network in blue)

(b) Artificial network linked by straight lines (network in purple)

Figure 2 Artificial Transportation Networks Predicted by Exogenous Factors and the Actual Railway Network in 2007

Based on the shortest railway distance on artificial networks and the upstream-downstream input-output relationship weighted by industrial value-added in the base year, we compute two counterfactual upstream market access measures. Note that these counterfactual indices do not exhibit temporal variations. We use the product of the two counterfactual upstream market access indices and the annual growth rate of nationwide fixed asset investment as the IVs for the actual upstream market access index. Clearly, faster investment growth contributes to infrastructure development, and the fluctuation in the nationwide investment growth rate is an exogenous shock for individual regions and firms.[16]

In addition to upstream market access, we also directly utilize bilateral distance information in the artificial transportation network to calculate the distance from each prefecture to its nearest port. This counterfactual distance, divided by the nationwide fixed

[16] Further, in order to avoid the endogeneity of macroeconomics, I filter out the trend of this series and keep the cyclical fluctuation.

investment growth rate, serves as the IV for the variable of actual distance to the nearest port.

3.4 Data

The sample data used in this study relies on two firm-level datasets, firm-level production data from the Annual Survey of Industrial Firms (ASIF) and trade transaction data from Chinese Customs Database (CCD) from 2000 to 2007.

Firm-level production data. The data are collected and maintained by China's National Bureau of Statistics (NBS) in an annual survey of manufacturing enterprises, covering all Chinses state-owned enterprises (SOEs) and non-SOEs with sales greater than five million yuan ($700,000 at current exchange rate). Due to the presence of noisy information caused by missing reports and misreports in the dataset, we follow the steps of Cai and Liu (2009), Brandt et al. (2012), Feenstra et al. (2014) and Yu (2015) to clean the sample. Specifically, we delete the observations with missing key financial variables (including total assets, fixed assets, number of employees, total sales, value of gross outputs and value added) and observation that do not comply with financial standards: (i) total fixed assets are greater than total assets; (ii) liquid assets are greater than total assets; (iii) the net value of fixed assets is greater than total assets; (iv) the depreciation for the current year exceeds the accumulated depreciation; and (v) any of the variables above is less than zero.

Trade transaction data. The Chinese Customs Database provides detailed information on China's trade transactions of both imports and exports. The dataset provides important variables for each transaction, such as trading price, quantity and value, product code at the HS eight-digit level, source or destination country. Most importantly, it records the identity information of the enterprises engaged in trade, allowing researchers to merge it with the dataset of firm-level production data.

Merged dataset. We first match the production data and trade transaction data following the steps of Yu (2015) by using firm's name, zip codes as well as phone numbers. Then they are merged by the datasets of upstream market access index we construct. To establish a one-to-one correspondence between the firms and their upstream industries, we convert the 4-digit Chinese industry classification (CIC) industry codes in the firm-level data to the industry classification in the 2002 Input-output Table with 128 industries.[17] Then we only retain the firms in the manufacturing sector as our interest is the relation of domestic intermediate inputs

[17] As the CIC experienced standard revisions in 2003, I first follow Brandt et al. (2012) to unify the pre- and post-revision industry classifications into an adjusted industry code.

and firms' export behaviors. We also merge the city-level characteristic variables, which are obtained from the China City Statistical Yearbook. Finally, we obtain an unbalanced panel of 271,975 firm-year observations, covering 94,201 export firms during 2000-2007. Summary of key variables can be found in Table 1.

Figure 3 depicts the trends of mean firm-level exports over time. The average export value of firms experienced continuous stable growth from 2000 to 2006, followed by a decline in 2007. Meanwhile, the average export scope, i.e., the number of export varieties, also increased by approximately 50% during this period. However, at the intensive margin, the average export value per variety remained almost unchanged or even experienced a slight decrease. This indicates that China's export growth during this period was primarily driven by the expansion of the extensive margin.

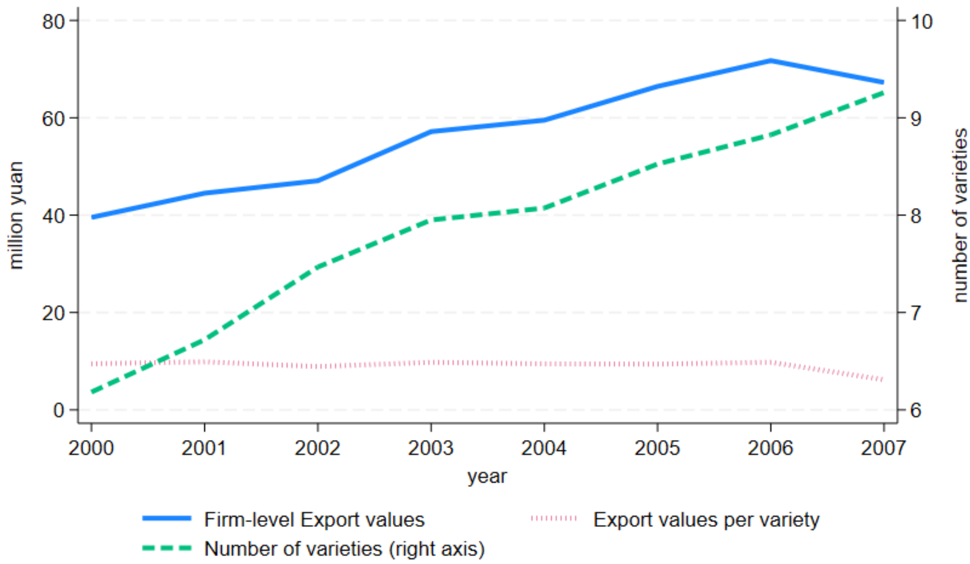


Figure 3 Trend of the Means of Firm-level Exports

Consistent with the model's setup, we divide the locations of sampled firms into seaboard and inland regions. Specifically, seaboard region includes eight coastal provinces (Liaoning, Shandong, Jiangsu, Zhejiang, Fujian, Guangdong, Guangxi, Hainan) and two province-level municipalities (Tianjin and Shanghai) and inland consists of the remaining provinces of the mainland China. Table 2 presents a detailed comparison of export performance between seaboard and inland firms across different dimensions: the number of exporters, average exports, export scope (number of varieties), export destinations (number of countries) and cost-to-sales ratios at the firm level. Consistent with Lemma 1 in the model, the seaboard firms outperform inland firms in exports. In average, seaboard firms have higher export revenues (61.38 compared to 59.82 million yuan) and export a broader scope of products (7.13 varieties vs. 7.01 varieties) and to more countries (8.54 countries vs. 7.34 countries). There are 74,096 seaboard exporters throughout the sample period, which is 3.7 times the number of inland (with 20,105 export firms). In addition, the average cost-to-sale ratio of seaboard export firms is higher than that of inland exporters (0.856 to 0.817), indicating a more stringent cost cutoff for entering the export market in the inland regions. All these facts are consistent with the model predictions. The t-values, however, indicate that disparities in export revenues and number of varieties are insignificant in the statistical sense.

4 Empirical Results

4.1 Benchmark Results

Table 3 reports the OLS estimation results for the benchmark specification. The simple unitary regression in Column (1) indicates a positive relation between upstream market access and firm exports. A one-percent increase in upstream market access correlates with a 0.915% increase in exports.

In Column (2) of Table 3, we included the distance from the firm to the nearest port city, of which the coefficient is positive but close to zero in both statistical and practical sense. The coefficient of upstream market access has hardly changed. In Column (3), as our preferred specification, we further add city-level and firm-level control variables. The estimate of upstream market access declines to 0.701 but still significantly positive. Now the distance to ports is negatively correlated with exports, though the coefficient remains insignificantly different from zero. In Column (4), we replace the year fixed effects with industry-year fixed effects to control for potential industry-specific time variations. The estimated coefficient for upstream market access is 0.680, which is similar to that in Column (3). Note that, firms' export levels are found to have a significant negative relation with the GDP of the city they are located in. This is consistent with the predictions of our model, that is, holding location and other conditions constant, a larger market size leads to more intense export competition.

To address the endogeneity problem of railway routes, we adopt an IV strategy to identify the exogenous change in upstream market access. As discussed in Section 3.2, we construct two counterfactual upstream market access indices for the real index. One instrument, $UMA\_geocost$, is constructed by the artificial railway network connected by the least costly paths. The other one, $UMA\_straight$, is based on the artificial railway network connected by straight lines. We also construct two instruments for the variable of distance to the nearest port using distances in these two artificial networks.[18] For ease of reference, we denote the instrumental variable combinations based on the networks of least costly paths and straight lines as $IV1$ and $IV2$, respectively.

Table 4 reports the 2SLS estimates for the benchmark specification. We use $IV1$ and $IV2$ as the instruments in Columns (1) and (2), and a one-percent increase in upstream market access is associated with a 2.757% and 2.302% increase in exports, respectively. The estimated coefficients from 2SLS are one to two times larger than that in Column (3) of Table 3, indicating that the OLS estimates underestimate the positive impact of upstream market access on firm exports. The coefficient for the distance to the nearest port has also increased in magnitude while remaining statistically insignificant. The K-P rk LM statistics and K-P rk Wald F statistics shows that we do not suffer from the weak IV problem.[19] In Column (3), both $IV1$ and $IV2$ are used for identification. The results are similar with those in the first two columns, and the Hansen J statistics do not reject the null that the IVs are valid. In subsequent empirical regressions, we use $IV1$ for identification as it incorporates more exogenous geographic information.

4.2 The Dual Margin of Exports

In this section, we test the predictions of Proposition 1 and 2, that is, how the upstream market access affects firm exports from the extensive as well as intensive margin.

In the literature, the extensive margin is usually defined as the scope of exports, while the meaning of export scope is divergent. For example, Hummels and Klenow (2005), Amiti and Freund (2008) and Shepherd (2010) define the extensive margin as the number of export varieties, Helpman et al. (2008) and Felbermayr and Kohler (2006) as the number of destination countries, and Evenett and Venables (2002) as the number of the combinations of countries and varieties.

According to our model, we first define export scope as the number of export varieties. The estimated results in Columns (1) and (2) of Table 5 show that a 1% increase in upstream market access leads to 2.135% and 1.786% increase in the number of export varieties at the

[18] Similar to the IV strategy for upstream market access, these two instrumental variables also induce temporal variation by dividing by the national fixed investment growth rate.

[19] See Appendix C for the estimated results at the first stage of 2SLS. For robustness checks, I also conduct the 2SLS estimation for the specification in Column (4) of Table 3 with industry-year fixed effects, which are also in Appendix C.

HS6 and HS4 product level, respectively. This result is consistent with the prediction of Proposition 1 that the net effect of better access to intermediate goods will enhance the export product scope. In the third column, we examine the impact of upstream market access on the number of destination countries for export, and it also reports a significantly positive estimate. The dependent variables in Columns (4) and (5) represents the number of combinations of export destinations and HS6-level and HS4-level export varieties. As expected, the estimated coefficients for upstream market access remains positive and the elasticity becomes larger than the first three columns. Further, in those columns, all the coefficients for port distance are significantly negative. As in our model, a decline in the trade cost to ports directly reduces transportation costs for exports, thereby enhancing the export scope through the cost-saving effects of final goods.

In Table 6, we represent the intensive margin as the average export value at each definition of extensive margin in Table 5. That is, we divide the total exports of firms by the number of export varieties and export destinations as well as their combinations. The results in the first three columns indicate that an increase in market access tends to increase the exports at intensive margin but the coefficients are insignificant. When we look at the last two columns, its coefficients even become negative (still being zero in the statistical sense). This implies that the effect of upstream market access on the intensive margin is weak, or there exists heterogeneity across firms.

Recall that in Proposition 2, after the reduction of domestic trade cost, the improvement of access to intermediate goods leads to an increase (decrease) in the product-level exports for those low-markup (high-markup) varieties. To test it, we conduct regressions at the HS6-product level in Table 7 to estimate the impact of upstream market access on single product's exports. To account for the effects of product-specific characteristics, we additionally control for product fixed effects at the HS6 level. The results in Column (1) indicate that upstream market access reduces exports at the product level on average.

It is challenging to measure the heterogeneity in firm-product unit markup as specified in the theoretical model. Due to the lack of detailed production-side data on exports, we are unable to capture the with-firm cost differences across export varieties. Hence, we opt to examine the effects of upstream market access on firms with different markup levels. Specifically, we use the cost rate of firms, measured by the total costs per unit of revenue, as a proxy variable for unit markup, as a larger cost rate is associated with a lower markup of firms. It is worth noting that this proxy has limitations so that we need to interpret the empirical results cautiously. On one hand, the cost variable is sourced from the ASIF, which includes the financial status of firms not only in export markets but also in all domestic markets. On the other hand, unlike the markup per output quantity in the model, the cost rate defined here actually measures the weighted average markup across all products within a firm. Nevertheless, this strategy can help identify the heterogeneous effects of upstream market access at the firm level. The Column (2) in Table 7 suggests a negative correlation between the cost rate of firms and product-level exports. Specifically, a one percentage point increase in the cost rate is associated with a 0.425% decline in product exports.[20] Furthermore, the significantly positive coefficient of the interaction term between upstream market access and cost rate indicates that, consistent with the predictions in our model, high-cost firms are more likely to achieve an increase in exports at the product level with improved access to domestic intermediate goods. In Column (3), we replace the continuous cost rate with an indicator variable which takes a value of one for firms with cost rates higher than the 50th percentile and zero otherwise. The estimation results show that, relative to the low-cost group, high-cost firms' product-level exports are subject to an additional positive effect of upstream market access with an elasticity of 0.087. In Column (4), we construct the cost indicator variable using the 75th percentile, and the estimate for the interaction term becomes larger as expected.[21] Besides, in all columns, although not significant, the coefficients of distance to ports range from -0.05 to -0.06, implying that lower transportation costs to ports tend to increase the exports of existing products.

Now we could cautiously conclude that upstream market access has heterogeneous effects on export intensive margins, where varieties from higher-cost firms are more likely to achieve export expansion. According to our model, higher-cost firms and varieties are more likely to benefit from the cost-saving effects of intermediate goods, while lower-cost ones are more responsive to the competition effects of intermediate goods. After all, on average across the sample, product-level exports decrease with the growth of upstream market access.

4.3 Heterogeneity of Seaboard and Inland

In this section we further test the heterogeneous effects of upstream market access in the seaboard and inland regions. In Panel A of Table 8, we conduct the baseline 2SLS regression

---

[20] Note that the coefficient of $ln(UMA)$ may not have practical significance because it represents the effect on firms with costs near zero.

[21] I also replicate these regressions at HS4-product level in Appendix C, and the results are similar.

in the first column for seaboard firms. The total exports of seaboard firms exhibit a significant and large elasticity (4.461) with respect to upstream market access. Further, we examine the effects on extensive margins of exports in Columns (2)-(4), where the dependent variables are number of HS6-level export varieties, HS4-level varieties, and export destinations. The estimated coefficients with range from 1.8 to 2.5 suggest that improved access to domestic intermediate goods significantly promotes the increase in all extensive margins. Note that, according to the estimates in all columns, the exports of seaboard firms also benefit from shorter distances to ports. This is because the seaboard firms are not all located in port cities, and therefore also experience variations in export transportation costs due to railway construction.

We replicate the regressions as above for inland firms in Panel B. Similar to seaboard firms and consistent with model predictions, the extensive margins of export for inland firms are found in Column (2)-(4) to increase with the improvement of upstream market access. The insignificant coefficient for their total exports in the first column, however, implies that the expanding export scopes are offset by contraction in the intensive margin.

To understand the changes in export intensive margins, in Table 9, we replicate the last three columns of Table 7 using subsamples of seaboard and inland firm. In general, the estimation results based on seaboard (in Columns (1)-(3)) and inland (in Columns (4)-(6)) firms show the same patterns as the full sample. That is, the increase in upstream market access tends to increase the product-level exports of higher-cost firms. For the seaboard firms, for example, relative to the low-cost group, the product-level exports of firms with costs higher than the 75th percentage are subject to an additional increase of 0.101% as upstream market access increase by one percent. Returning to the predictions of the model, high-cost products of high-cost firms are expected to benefit more from a positive net cost-saving effect of intermediate goods, thereby increasing exports. However, the empirical results suggest that, on average at the product level, the competition effects of intermediate goods dominate, thereby constraining the intensive margin expansion of both seaboard and inland enterprises. In practical terms, inland firms face more intense export competition. While the improved intermediate goods transportation network brings lower-cost intermediate goods, it may also lead to more competition from seaboard firms and their final products in the local market, thereby worsening the export conditions for existing inland firms. We will test the existence the two forces of access to domestic intermediated goods in the next section.

4.4 The Cost-saving and Competition Effects

In this section we investigate whether improved access to intermediate goods creates cost-saving effects for the exporters and whether it changes the competition environment.

Firstly, to test the changes in firm costs associated with intermediates access, the dependent variable in Column (1) of Table 10 is firms' input costs, measured by the log of total values of intermediate inputs. We find that an 1% increase in upstream market access is associated with a 0.851% drop in input costs. To control for the impact of firm size on total inputs, we use the intermediate input cost per unit of total revenue as the explanatory variable in Column (2). The interpretation of coefficients in this semi-elasticity model is that for every one percent increase in upstream market access, the cost per unit of revenues decreases by 0.0017 yuan. The dependent variable in Column (3) in replaced by the intermediate input cost per unit of total cost, and the estimate implies a decrease of 0.46 percentage points in the proportion of intermediate input costs in total costs. from the last column, the firms' total costs seem to be negatively correlated with upstream market access while the estimated coefficient is not significant.

Then we test the existence of competition effects. We use the full sample of manufacturing firms without matching the Customs data, and aggregate firms' status of survival and exports into the city-industry level. The dependent variable in Column (1) of Table 11 is entry rate, which is measured by the proportion of new entrant firms in the current year in the total number of firms at the city-industry level, and similarly, the variable in Column (2) is exit rate that the proportion of firms exiting the market in subsequent years. The estimated results for those two columns suggest that improved upstream market access promotes the entry of new firms while reducing the exit rate of incumbent firms.[22] The third column is a regression for the difference between Column (1) and Column (2), indicating that the net entry rate rises by 0.36 percentage points as upstream market access increases by a unit percentage. As in our model, the reduction in trade costs will lower the local survival cutoff, intensify local competition, and promote the net increase of firms. The reason for the negative coefficient of upstream market access on the exit rate is, while the cost cutoff for firm survival reduced, firms also become more competitive due to the decrease in intermediate costs. When the latter effect is greater, existing firms are more likely to survive.

[22] In reality, I need to interpret this result cautiously because both entry and exit are defined at the level of above-scale manufacturing enterprises. For example, the exact meaning of "new entrant firms" refers to firms whose production scale first meets the criteria for inclusion in the database.

Finally, in Column (4) of Table 11, we examine the impact on the export rate of firms measured by the proportion of exporting firms, which reflects the combined effect of cost-saving and competition. Consistent with previous discussions, better access to domestic intermediate goods and shorter transportation distances to ports will increase the probability of firm exporting.

4.5 Additional Robustness Checks

In Table 12, we examine the lagging effects of upstream market access. Column (1) replicates the benchmark regression in Table 4, while the last three columns regress firm exports on lagged upstream market access by one, two and three years, respectively. From the estimates, the positive coefficient of upstream market access on exports diminishes in magnitude with lagged periods and remains significant within two years.

We then show the results in the benchmark regression remain robust for several additional robustness checks in Table 13.

Firstly, to ensure our results are not driven by firms in the major cities, we exclude firms in the 36 major cities in Column (1) of Table 13, including the 4 municipalities directly under the central government, 27 provincial capitals and another 5 major cities as described in Section 3.3.

Secondly, to eliminate potential impact of compositional changes of exporters, we restrict the sample to a balanced panel of exporting firms. Specifically, we only retain the firms that consistently had positive export values from 2000 to 2007 in Column (2).

Thirdly, we exclude the effects of economic agglomeration on exporting firms. Following the method of Ruan and Zhang (2022), we construct a clustering index at the prefecture level that simultaneously incorporates information on market size and industry relatedness. To obtain the time-varying clustering index data, we use the annual aggregated asset scale of industrial enterprise in ASIF as weights. We add the log of this index as an additional control in Column (3).

Fourthly, as our upstream market access index in benchmark is based on the industrial value-added of firms as weights, we construct another index for upstream market access, $UMA^{total\ ouput}$, by using total output value instead of value added. We substitute the core independent variable with this new index in Column (4) as a robustness check. As we can see, in all columns of Table 13, the coefficients on upstream market access remain significant and the magnitude is similar to the benchmark.

And finally, we adjust the parameters in Equations (44) and reconstruct the upstream market access indices. We report the regressions with alternative upstream market access indices in Table 14, and the estimated coefficients in all columns remain significantly positive.

5 Conclusion

In conclusion, this study has explored how domestic trade integration, facilitated by improvements in transportation infrastructure, impacts the export performance of Chinese manufacturing firms, particularly through enhanced access to domestic intermediate goods. Building on the multi-product firm model by Mayer et al. (2014), we extended the analysis to incorporate the dynamics of intermediate goods access, revealing how reductions in domestic trade costs influence firm exports. Our findings demonstrate that the cost-saving effects, net of competition from intermediate goods, lead to an increase in the number of export varieties per firm. This expansion exhibits heterogeneity at the product level, where only low-markup varieties experience significant export growth. Consequently, total firm-level exports can rise significantly when either the number of exported product varieties increases, or individual products achieve substantial export growth through improved access to intermediate goods.

Utilizing firm-level export data from Chinese manufacturing firms from 2000 to 2007, along with the upstream market access (UMA) index that we developed, we tested the model's predictions. Our analysis addresses the endogeneity of railway expansions and provides robust evidence that improved upstream market access significantly boosts total export values for both seaboard and inland firms, primarily through the expansion of extensive export margins. Consistent with our model, UMA tends to reduce product-level exports for firms with higher markups. Furthermore, the study confirms the theoretical cost-saving and competition effects of accessing domestic intermediate goods, illustrating that despite increased competition, firms can enhance their export performance by leveraging lower-cost domestic inputs.

This research enriches the empirical literature by delineating the nuanced impacts of domestic trade integration on export dynamics, highlighting the pivotal role of infrastructure in enhancing economic competitiveness through strategic access to intermediate goods.

Table 1 Summary Statistics

Variables	Unit	Obs.	Mean	Min	Max	S.D.
$export$	million yuan	271975	61.05	0.00	176117.30	777.80
$ln(UMA)$	\	35596	20.14	17.84	20.62	0.30
$ln(UMA)\_geocost$	\	35596	19.86	17.51	20.46	0.36
$ln(UMA)\_straight$	\	35596	19.88	17.94	20.43	0.30
$dist\_port$	km	2015	658.00	0	4381.15	620.83
gdp	billion yuan	2015	6.03	0.24	121.89	8.98
$gdppc$	yuan per capita	2015	14670.04	1475.92	320254.80	19063.27
$indserv$	\	2015	1.40	0.11	10.55	0.80
$highway$	km per km2	237	6.84	0.35	32.05	6.00
$size$	million yuan	271975	147.35	0.00	154517.30	946.33
$klratio$	thousand yuan per capita	271975	108.23	0.00	209520.50	517.32
SOE	binary	271975	0.06	0	1	0.23

Note: $ln(UMA)$ is the log of UMA index, $ln(UMA)\_geocost$ is the log of counterfactual UMA that is constructed by the artificial transportation network connected by the least costly paths, and $ln(UMA)\_straight$ is the log of counterfactual UMA that is constructed by the artificial transportation network connected by straight lines. $dist\_port$ is the distance to the nearest port city. gdp and $gdppc$ are the GDP and GDP per capita at the prefecture level, and $indserv$ is the GDP ratio of the secondary and tertiary industries. $highway$ is province-level highway density measured by highway length per area. $size$ is the firm size measured by total assets, $klratio$ is the firm's capital intensity measured by the fixed assets per worker, and SOE is an indicator variable for state-owned enterprises.

Table 2 Comparison of Seaboard and Inland Firms' Exports

Statistics	Unit	Seaboard (1)	Inland (2)	Diff. (3)	t-value (4)
observations	count	213783	58192	155591	
number of exporters	count	79316	14885	64431	\
average exports	million yuan	65.93	27.49	38.44	20.35
export scope	number of varieties	7.32	5.59	1.73	17.83
export destination	number of countries	8.58	6.24	2.34	48.00

Note: Columns (1) and (2) represent the means of export performance of seaboard and inland firms. Column (3) calculates the differences between the first two columns with values of the t-statistic displayed in Column (4).

Table 3 Upstream Market Access and Firm Exports (OLS)

Dependent Variables:	$ln(export)$			
	(1)	(2)	(3)	(4)
$ln(UMA)$	0.915***	0.914***	0.701**	0.680***
	(0.346)	(0.346)	(0.315)	(0.239)
$ln(dist\_port)$		0.007	-0.008	-0.025
		(0.059)	(0.059)	(0.060)
$lngdp$			-0.308***	-0.337***
			(0.053)	(0.047)
$lngdppc$			0.035	0.034
			(0.038)	(0.037)
$indserv$			0.004	0.032
			(0.043)	(0.035)
$lnhighway$			0.051	0.055*
			(0.037)	(0.032)
$lnsize$			0.578***	0.555***
			(0.016)	(0.015)
$klratio$			-0.000	-0.000
			(0.000)	(0.000)
SOE			-0.035	-0.038
			(0.038)	(0.038)
Dependent Variable's Mean	13.704	13.704	13.704	13.704
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	No
Industry-year fixed effects	No	No	No	Yes
R^2	0.805	0.805	0.813	0.816
Observations	236,707	236,707	236,707	236,707

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable is the log of firm exports. $ln(UMA)$ is the log of upstream market access index. $ln(dist\_port)$ is the log of distance to the nearest ports. $lngdp$ and $lngdppc$ are the logs of GDP and GDP per capita. $indserv$ is the GDP ratio of the secondary and tertiary industries. $lnhighway$ is the log of province-level highway density measured by highway length per area. $lnsize$ is the log of firm size measured by total assets. $klratio$ is the firm's capital intensity measured by the fixed assets per worker. SOE is an indicator variable for state-owned enterprises.

Table 4 Upstream Market Access and Firm Exports (2SLS)

Dependent Variables:		$ln(export)$	
	IV1	IV2	Double IVs
	(1)	(2)	(3)
$ln(UMA)$	2.757***	2.302***	2.827***
	(0.589)	(0.844)	(0.578)
$ln(dist\_port)$	-0.067	-0.063	-0.065
	(0.064)	(0.065)	(0.063)
Dependent Variable's Mean	13.704	13.704	13.704
Time-varying firm controls	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
K-P rk LM statistic	386.34	232.49	423.44
K-P rk Wald F statistic	510.20	250.04	279.59
Hansen J statistic	\	\	1.37
Observations	236,707	236,707	236,707

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable is the log of firm exports. The coefficients for Column (1) are estimated using IV1, the logs of counterfactual UMA and $dist\_port$ that are constructed by the artificial transportation network connected by the least costly paths. The coefficients for Column (2) are estimated using IV2, the logs of counterfactual UMA and $dist\_port$ that are constructed by the artificial transportation network connected by connected by straight lines. Column (3) uses all IVs from the first two columns and conducts an overidentification test.

Table 5 Upstream Market Access and Exporting Extensive Margin

Dependent Variables:	lnM^{hs6}	lnM^{hs4}	lnD	$ln(M^{hs6} \times D)$	$ln(M^{hs4} \times D)$
	(1)	(2)	(3)	(4)	(5)
$ln(UMA)$	2.135***	1.786***	2.250***	2.691***	2.656***
	(0.197)	(0.186)	(0.302)	(0.251)	(0.268)
$ln(dist\_port)$	-0.035**	-0.027**	-0.059**	-0.066***	-0.066***
	(0.016)	(0.013)	(0.023)	(0.024)	(0.024)
Dependent Variable's Mean	1.619	1.396	1.803	2.228	2.159
Time-varying firm controls	Yes	Yes	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
K-P rk LM statistic	386.34	386.34	386.34	386.34	386.34
K-P rk Wald F statistic	510.20	510.20	510.20	510.20	510.20
Observations	236,707	236,707	236,707	236,707	236,707

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variables in Columns (1) and (2) are the logs of number of export varieties at the HS6 and HS4 level, respectively. The dependent variable in Column (3) is the log of number of exporting destination countries. The dependent variables in Columns (4) and (5) are the logs of number of combinations of exporting destinations and export varieties at the HS6 and HS4 level, respectively.

Table 6 Upstream Market Access and Exporting Intensive Margin

Dependent Variables:	$ln\frac{export}{M^{hs6}}$	$ln\frac{export}{M^{hs4}}$	$ln\frac{export}{D}$	$ln\frac{export}{M^{hs6}\times D}$	$ln\frac{export}{M^{hs4}\times D}$
	(1)	(2)	(3)	(4)	(5)
$ln(UMA)$	0.118	0.453	0.155	-0.269	-0.227
	(0.529)	(0.530)	(0.719)	(0.536)	(0.576)
$ln(dist\_port)$	-0.028	-0.032	0.001	0.007	0.008
	(0.056)	(0.058)	(0.050)	(0.048)	(0.048)
Dependent Variable's Mean	12.410	12.692	12.188	11.684	11.763
Time-varying firm controls	Yes	Yes	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
K-P rk LM statistic	386.34	386.34	386.34	386.34	386.34
K-P rk Wald F statistic	510.20	510.20	510.20	510.20	510.20
Observations	236,707	236,707	236,707	236,707	236,707

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variables in Columns (1) and (2) are the logs of exports per export varieties at the HS6 and HS4 level, respectively. The dependent variable in Column (3) is the log exports per exporting destination country. The dependent variables in Columns (4) and (5) are the logs of exports per combination of exporting destinations and export varieties at the HS6 and HS4 level, respectively.

Table 7 Upstream Market Access and Product-level Exports (2SLS)

Dependent Variables:	$ln(export^{hs6})$			
	(1)	(2)	(3)	(4)
$ln(UMA)$	-2.160***	-4.042***	-2.211***	-2.180***
	(0.435)	(0.817)	(0.436)	(0.435)
$ln(dist\_port)$	-0.052	-0.054	-0.052	-0.053
	(0.041)	(0.041)	(0.041)	(0.041)
$costr$		-42.464***		
		(10.143)		
$ln(UMA) \times costr$		2.104***		
		(0.503)		
$\mathbf{1}(costr_{50th})$			-1.750**	
			(0.835)	
$ln(UMA) \times \mathbf{1}(costr_{50th})$			0.087**	
			(0.041)	
$\mathbf{1}(costr_{75th})$				-2.569***
				(0.840)
$ln(UMA) \times \mathbf{1}(costr_{75th})$				0.126***
				(0.042)
Dependent Variable's Mean	9.913	9.913	9.913	9.913
Time-varying firm controls	Yes	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Product fixed effects	Yes	Yes	Yes	Yes
K-P rk LM statistic	217.82	152.40	217.64	217.87
K-P rk Wald F statistic	404.31	90.45	391.91	402.02
Observations	1,738,897	1,738,624	1,738,897	1,738,897

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable is the log of exports at the HS6-product level. $costr$ is firm's cost rate measured by the total costs per unit of revenue. $\mathbf{1}(costr_{50th})$ is an indicator variable for firms with cost rates higher than the 50th percentile, and $\mathbf{1}(costr_{75th})$ is an indicator those with cost rates higher than the 75th percentile.

Table 8 Upstream Market Access and Firm Exports at Different Regions

	$ln(export)$	lnM^{hs6}	lnM^{hs4}	lnD
Dependent Variables:	(1)	(2)	(3)	(4)
	Panel A. Seaboard			
$ln(UMA)$	4.461***	2.192***	1.793***	2.463***
	(0.717)	(0.269)	(0.251)	(0.365)
$ln(dist\_port)$	-0.133**	-0.050***	-0.036***	-0.080***
	(0.067)	(0.017)	(0.014)	(0.025)
Dependent Variable's Mean	13.769	1.632	1.403	1.830
K-P rk LM statistic	262.01	262.01	262.01	262.01
K-P rk Wald F statistic	252.43	252.43	252.43	252.43
Observations	186,236	186,236	186,236	186,236
	Panel B. Inland			
$ln(UMA)$	-0.530	2.099***	1.823***	1.905***
	(0.944)	(0.236)	(0.214)	(0.448)
$ln(dist\_port)$	0.195	0.151	0.102	0.119***
	(0.191)	(0.106)	(0.099)	(0.035)
Dependent Variable's Mean	13.465	1.572	1.371	1.704
K-P rk LM statistic	3.88	3.88	3.88	3.88
K-P rk Wald F statistic	126.00	126.00	126.00	126.00
Observations	50,465	50,465	50,465	50,465
Time-varying firm controls	Yes	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable in Column (1) is the log of exports at the firm level. The dependent variables in Columns (2) and (3) are the logs of number of export varieties at the HS6 and HS4 level, respectively. The dependent variable in Column (4) is the log of number of exporting destination countries. c) Panel A presents the estimation results for firms located in seaboard cities and Panel B shows results for firms in inland cities.

Table 9 Upstream Market Access and Product-level Exports at Different Regions

Dependent Variables:	$ln(export^{hs6})$					
		Seaboard			Inland	
	(1)	(2)	(3)	(4)	(5)	(6)
$ln(UMA)$	-1.942**	-0.946**	-0.913*	-5.905***	-4.695***	-4.737***
	(0.969)	(0.470)	(0.470)	(0.730)	(0.629)	(0.606)
$ln(dist\_port)$	-0.074*	-0.073*	-0.075*	-0.259	-0.255	-0.255
	(0.041)	(0.042)	(0.042)	(0.219)	(0.219)	(0.218)
$costr$	-23.605*			-29.673***		
	(12.970)			(8.532)		
$ln(UMA) \times costr$	1.171*			1.465***		
	(0.643)			(0.423)		
$\mathbf{1}(costr_{50th})$		-1.021			-2.658*	
		(0.943)			(1.570)	
$ln(UMA) \times \mathbf{1}(costr_{50th})$		0.051			0.130*	
		(0.047)			(0.078)	
$\mathbf{1}(costr_{75th})$			-2.060**			-3.600*
			(0.870)			(2.046)
$ln(UMA) \times \mathbf{1}(costr_{75th})$			0.101**			0.177*
			(0.043)			(0.101)
Dependent Variable's Mean	9.966	9.966	9.966	9.714	9.714	9.714
Time-varying firm controls	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Product fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
K-P rk LM statistic	110.04	166.62	168.26	12.42	3.17	3.18
K-P rk Wald F statistic	63.11	226.53	231.46	646.22	205.90	206.13
Observations	1,376,982	1,376,982	1,376,982	361,424	361,424	361,424

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable is the log of exports at the HS6-product level. $costr$ is firm's cost rate measured by the total costs per unit of revenue. $\mathbf{1}(costr_{50th})$ is an indicator variable for firms with cost rates higher than the 50th percentile, and $\mathbf{1}(costr_{75th})$ is an indicator for firms with cost rates higher than the 75th percentile. c) Columns (1)-(3) presents the estimation results for firms located in seaboard cities and Columns (4)-(6) are for firms in inland cities.

Table 10 Upstream Market Access and Cost Reduction

Dependent Variables:	$ln(cost_{input})$	$\frac{cost_{input}}{cost}$	$\frac{cost_{input}}{sales}$	$ln(cost)$
	(1)	(2)	(3)	(4)
$ln(UMA)$	-0.851***	-0.455***	-0.172***	-0.317
	(0.322)	(0.077)	(0.053)	(0.310)
$ln(dist\_port)$	-0.002	0.017**	0.002	-0.026
	(0.023)	(0.008)	(0.005)	(0.023)
Dependent Variable's Mean	10.342	0.939	0.773	10.454
Time-varying firm controls	Yes	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
K-P rk LM statistic	386.34	386.34	386.34	386.34
K-P rk Wald F statistic	510.20	510.20	510.20	510.20
Observations	236,707	236,707	236,707	236,707

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable in Column (1) is the log of input costs. The dependent variables in Columns (2) and (3) are input costs per unit of total costs and total revenue, respectively. The dependent variable in Column (4) is the log of total costs.

Table 11 Upstream Market Access and Market Competition

Dependent Variables:	$\frac{N_{export}}{N_{city\times ind}}$ (1)	$\frac{N_{entry}}{N_{city\times ind}}$ (2)	$\frac{N_{exit}}{N_{city\times ind}}$ (3)	$\frac{N_{entry}-N_{exit}}{N_{city\times ind}}$ (4)
$ln(UMA)$	0.154***	0.107***	-0.251***	0.358***
	(0.039)	(0.034)	(0.040)	(0.057)
$ln(dist\_port)$	-1.928***	-0.047	-0.157	0.110
	(0.494)	(0.458)	(0.428)	(0.658)
Dependent Variable's Mean	0.194	0.159	0.122	0.036
Time-varying firm controls	Yes	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes	Yes
City-industry fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
K-P rk LM statistic	524.62	524.62	524.62	524.62
K-P rk Wald F statistic	408.91	408.91	408.91	408.91
Observations	94,808	94,808	94,808	94,808

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable in Column (1) is exporting rate measured by the proportion of exporting firms in the total number of firms at the city-industry level. The dependent variables in Columns (2) and (3) are entry rate and exit rate, measured by the proportion of new entrant firms in the current year and firms exiting the market in subsequent years, respectively. The dependent variable in Column (4) is the net entry rate of firms which is the difference of dependent variables between Columns (2) and (3).

Table 12 Lagging Effects of Upstream Market Access

Dependent Variables:	$ln\ (export)$			
	(1)	(2)	(3)	(4)
$ln(UMA)$	2.757***			
	(0.589)			
$ln(UMA)\_lag1$		2.049***		
		(0.545)		
$ln(UMA)\_lag2$			1.383**	
			(0.646)	
$ln(UMA)\_lag3$				1.336
				(0.854)
$ln(dist\_port)$	-0.067	-0.078	0.062	0.050
	(0.064)	(0.063)	(0.107)	(0.134)
Dependent Variable's Mean	13.704	13.721	13.733	13.742
Time-varying firm controls	Yes	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes	Yes
City-industry fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
K-P rk LM statistic	386.34	350.20	290.53	202.36
K-P rk Wald F statistic	510.20	1009.37	808.97	546.65
Observations	236,707	150,186	101,856	61,752

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable is the log of exports at the firm level. $ln(UMA)\_lag1$, $ln(UMA)\_lag2$ and $ln(UMA)\_lag3$ represent the lagged $ln(UMA)$ by one, two and three periods, respectively.

Table 13 Robustness Checks

Dependent Variables:	$ln(export)$			
	(1)	(2)	(3)	(4)
$ln(UMA)$	2.658***	2.945***	2.752***	
	(0.623)	(0.832)	(0.592)	
$ln(dist\_port)$	-0.141	-0.112	-0.065	-0.058
	(0.202)	(0.088)	(0.064)	(0.064)
$ln(cluster)$			0.014	
			(0.057)	
$ln(UMA^{total\ ouput})$				2.516***
				(0.540)
Dependent Variable's Mean	13.614	14.781	13.704	13.704
Time-varying firm controls	Yes	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
K-P rk LM statistic	261.11	218.32	378.09	369.52
K-P rk Wald F statistic	332.79	252.32	512.28	500.82
Observations	123,271	35,376	236,691	236,691

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable is the log of exports at the firm level. Column (1) excludes the firms in the 36 major Chinese cities as described in Section 3.3. Column (2) presents the estimation based on the balanced panel sample, which only includes firms that export consistently throughout the sample period. Column (3) controls $ln(cluster)$, the logs of cluster index at city level. Column (4) substitutes the baseline upstream market access with $UMA^{total\ ouput}$ which is constructed based on upstream firms' total output rather than value added.

Table 14 Robustness Checks: Alternative Parameters of UMA

Dependent Variables:	$ln(export)$					
	(1)	(2)	(3)	(4)	(5)	(6)
$ln(UMA)\_t2$	2.515***					
	(0.537)					
$ln(UMA)\_t3$		1.202***				
		(0.253)				
$ln(UMA)\_t4$			1.622***			
			(0.343)			
$ln(UMA)\_s2$				8.226***		
				(1.796)		
$ln(UMA)\_s3$					1.286***	
					(0.271)	
$ln(UMA)\_ppi$						5.606***
						(1.258)
$ln(dist\_port)$	-0.068	-0.084	-0.075	-0.080	-0.080	-0.095
	(0.064)	(0.066)	(0.064)	(0.063)	(0.065)	(0.066)
Dependent Variable's Mean	13.704	13.704	13.704	13.704	13.704	13.704
Time-varying firm controls	Yes	Yes	Yes	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
K-P rk LM statistic	392.04	391.96	409.15	310.84	395.53	239.77
K-P rk Wald F statistic	548.63	848.43	778.77	262.44	879.95	145.99
Observations	236,707	236,707	236,707	236,707	236,707	236,707

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable is the log of exports at the firm level. The dependent variable is the log of sulfur dioxide emission level. The core independent variables in Columns (1)-(5) are logs of upstream market access constructed with alternative parameters of the price elasticity of trade θ, and the concavity of shipping cost to railway distance σ. These variables include $ln(UMA)\_t2$ ($\sigma = 0.8$; $\theta = 4.1$), $ln(UMA)\_t3$ ($\sigma = 0.8$; $\theta = 8.3$), $ln(UMA)\_t4$ ($\sigma = 0.8$; $\theta = 6$), $ln(UMA)\_s2$ ($\sigma = 0.6$; $\theta = 3.8$) and $ln(UMA)\_s3$ ($\sigma = 1$; $\theta = 3.8$). $ln(UMA)\_ppi$ has the same parameter values as $ln(UMA)$ but replace nominal freight rate with freight price after PPI adjustment.

References

Amiti, M., & Konings, J. (2007). Trade liberalization, intermediate inputs, and productivity: Evidence from Indonesia. American economic review, 97(5), 1611-1638.

Amiti, M., & Freund, C. (2008). The Anatomy of China's Export Growth. The World Bank.

Bas, M. (2012). Input-trade liberalization and firm export decisions: Evidence from Argentina. Journal of Development Economics, 97(2), 481-493.

Banerjee. A., Duflo, E., & Qian, N. (2020). On the Road: Access to Transportation Infrastructure and Economic Growth in China. Journal of Development Economics, 145, 1-36.

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.

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

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

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

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., Morrow, P. M., & Turner, M. A. (2014). Roads and Trade: Evidence from the US. *Review of Economic Studies*, *81*(2), 681-724.

Evenett, S. J., & Venables, A. J. (2002). Export Growth in Developing Countries: Market Entry and Bilateral Trade Flows (pp. 1-42). University of Bern working paper, mimeo.

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.

Faber, B. (2014). Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. Review of Economic Studies, 81, 1046-1070.

Felbermayr, G. J., & Kohler, W. (2006). Exploring the Intensive and Extensive Margins of World Trade. Review of World Economics, 142(4), 642-674.

Feng, L., Li, Z., & Swenson, D. L. (2016). The connection between imported intermediate inputs and exports: Evidence from Chinese firms. Journal of International Economics, 101, 86-101.

Goldberg, P. K., Khandelwal, A. K., Pavcnik, N., & Topalova, P. (2010). Imported intermediate inputs and domestic product growth: Evidence from India. The Quarterly journal of economics, 125(4), 1727-1767.

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. The Quarterly Journal of Economics, 123(2), 441-487.

Hummels, D., & Klenow, P. J. (2005). The Variety and Quality of a Nation's Exports. American Economic Review, 95(3), 704-723.

Limao, N., & Venables, A. J. (2001). Infrastructure, geographical disadvantage, transport costs, and trade. The world bank economic review, 15(3), 451-479.

Liu, D., Sheng, L., & Yu, M. (2023). Highways and firms' exports: Evidence from China. Review of International Economics, 31(2), 413-443.

Martincus, C. V., & Blyde, J. (2013). Shaky roads and trembling exports: Assessing the trade effects of domestic infrastructure using a natural experiment. Journal of International Economics, 90(1), 148-161.

Martincus, C. V., Carballo, J., & Cusolito, A. (2017). Roads, exports and employment: Evidence from a developing country. Journal of Development Economics, 125, 21-39.

Mayer, T., Melitz, M. J., & Ottaviano, G. I. (2014). Market size, competition, and the product mix of exporters. American Economic Review, 104(2), 495-536.

Melitz, M. J., & Ottaviano, G. I. (2008). Market size, trade, and productivity. The review of economic studies, 75(1), 295-316.

Rousslang, D. J., & To, T. (1993). Domestic trade and transportation costs as barriers to international trade. Canadian Journal of Economics, 208-221.

Shepherd, B. (2010). Geographical Diversification of Developing Country Exports. World Development, 38(9), 1217-1228.

Tian, L., & Yu, Y. (2023). Geographic Integration and Firm Exports Evidence from China. Available at SSRN 4724154.

Topalova, P., & Khandelwal, A. (2011). Trade liberalization and firm productivity: The case of India. Review of economics and statistics, 93(3), 995-1009.

Xu, H. (2016). Domestic railroad infrastructure and exports: Evidence from the Silk Route. China Economic Review, 41, 129-147.

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

Appendix

A Construction of Chinese railway network dataset

As the foundation of empirical work in this paper, we build a dataset of Chinese railway network on detail geographical information and calculate the time-varying shortest railway distances between prefecture-level cities in China from 1998 to 2021.

Construction of the railway network database begins with vector maps of railroads. Firstly, I collect the original maps of printed version from the China Railway Yearbook issued by the Ministry of Railways of China,[23] which describes the operating railway lines and introduces the projects under construction in each year. Note that limited to scale, these maps ignore unimportant feeder lines, which are generally terminal lines leading to county and lower-level administrative districts. Since our objective is to connect prefecture level cities, these omissions have only trivial effect on regional accessibility. I digitize these printed maps by manually drawing the outlines of each railway line with the aid of geographic information software (ArcGIS). Specifically, I first select 2016 as the benchmark year, drawing the complete digitalized railway map of this year, and then update it according to the yearbooks in other years.

The second step adds train stations to the map. I obtained the 2016 national railway station list from CNRDS, which contains the longitude and latitude information of 7042 stations.[24] The textual information is converted into geographic location point that can be matched with the railway network. Then, with the help of Python, I crawl the information about the opening years of the stations from Baidu Baike,[25] covering more than 5000 stations. I manually complement the rest and those newly opened stations after 2016 by searching relevant yearbooks and newspapers. Through these efforts, we are able to identify the stations in operation each year.

Thirdly, I match the prefectures to the network of railroads. Different from Donaldson and Hornbeck (2016) who connect all city centers directly to the nearest routes, I combine the information on administrative divisions and stations. Specifically, if there exists only one station in the city's jurisdiction, a single match is adopted; for multiple stations, the city will be matched with all of them. This practice is out of consideration of a realistic situation, that is, for a newly opened station that is closer to the city center, it may only lead to a certain

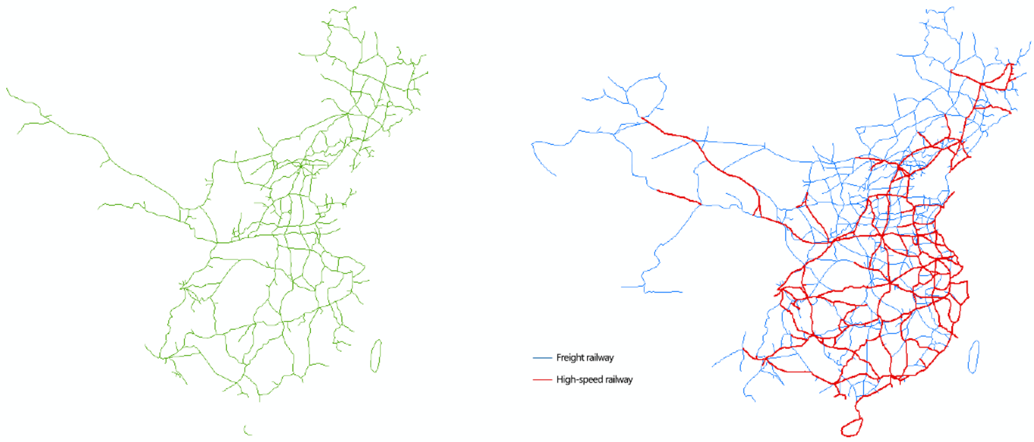
[23] These documents are printed publications. To access the resource, see https://cyfd.cnki.com.cn/N2020010195.htm.

[24] It is noted that the railway network I construct only includes train stations in mainland China without encompassing the Hong Kong and Macau Special Administrative Regions and Taiwan Province.

[25] It is a Chinese information collection platform covering knowledge and entries in various fields.

direction or connects only a small part of the railway network.[26]

In the fourth step, I distinguish the lines of conventional freight railway and high-speed railway (HSR). The reason is that only conventional railways have freight transportation functionality, which can be utilized to measure the transportation costs of goods and physical factors between regions. In contrast, high-speed railways primarily serve as passenger-dedicated lines, facilitating the cross-regional transfer of labor and information. For different research purposes, I construct separate network maps for conventional railways from 1998 to 2021 and high-speed railways from 2008 to 2021. Figure A1 shows the railway networks in 1998 and 2021, respectively. The spatial distribution of railways is uneven across regions. In 1998, it is obvious that the railway density was highest in the north plain and northeast regions, while the western region had only a few sparse railway lines. By 2021, the conventional railway network in the eastern and southern regions had seen significant development, and there has also been an increase in railway density in the western region. The routes of the high-speed railway lines were generally consistent with the existing railway network and were more developed in the eastern regions by 2021.



(a) Network in 1998 (b) Network in 2021

Figure A1 Chinese Major Railway Networks in 1998 and 2021

Source: author's calculation

In the last step, based on the railway vector maps, I use ArcGIS to calculate the optimal route (shortest path) between each pair of cities. The calculation method is from Dijkstra (1959), who proposed a way to find the shortest paths from a single node to all other nodes in the graph, producing a shortest-path tree.

[26] For example, the West and South Railway Station of Beijing are set to go in different directions. If the starting point is limited to the West Railway Station, it is inevitable to take a detour at least through central provinces to go to East China.

The logic to calculate the routes is as follows: (1) Define a starting point, called the initial node and denoted by s. (2) Create a set of visited nodes, V, as well as a set of unvisited nodes, U. At the beginning, V is an empty set. (3) Assign to every node a tentative distance value, i.e., set it to zero for our initial node and to infinity for all other nodes. Set initial node as current node. (4) Consider all of unvisited neighbors of current node and calculate their distances (based on the real distance exogenous given) to update the tentative distance values if newly calculated distance is smaller. (5) Select the node, called k for example, with smallest tentative distance values to s. Move k from U to V. (6) Recalculate the distance from s to all points through U, and compared with the route through V. For example, for two points $i, j \in U$, compare $dis(s,k) + dis(k,j)$ with $dis(s,i) + dis(i,j)$, and move i to V if the later distance is shorter. (7) Repeat the above steps until all points are visited. In the context of this paper, the initial node is the original prefectural-level city, and the distance between two neighbor nodes is the actual railway distance between the original and destination city. In this way, I calculate the shortest path along railways between any pair of cities.

There are some caveats on the limitations of this railway dataset. First, manual drawing and matching process obviously increase the potential for error. I alleviate this problem by using as much as possible auxiliary location points and ArcGIS topology tools. Second, the transport distance abstracted from the digitalized network could not perfectly measures the distance in reality. In fact, economists only need to focus on the relative differences in railway access across cities and variation over time. I make sure that all data are horizontally and vertically comparable by setting a benchmark year of digitalized maps. Third, since our focus is on the enhancement of interregional connectivity brought by railways, our matching strategy ignores the intra-city distance from the city center to the railway station, which is usually transported by city roads.

B Proofs of the Model

B1 Free Entry Equilibrium in a Closed Economy

Our steps to the solution of free entry equilibrium follow Melitz and Ottaviano (2008) and Mayer et al. (2014). In the closed economy, the free entry equilibrium condition in Equation (14) can be rewritten as:

$$
\begin{aligned}
\int_0^{c_D} \Pi(c)\, dG(c) &= \int_0^{c_D} \sum_{m=0}^{M(c)-1} \pi(m,c)\, dG(c) \\
&= \int_0^{c_D} \sum_{\{m|c\le\omega^m c_D\}} \pi(m,c)\, dG(c) \\
&= \sum_{m=0}^{\infty} \left[\int_0^{\omega^m c_D} \pi(m,c)\, dG(c)\right] \\
&= F
\end{aligned}
\quad, \qquad \text{(B.1)}
$$

where $\pi(m,c) = \frac{L}{4\gamma}(c_D - \omega^{-m}c)^2$. Given the parameterization of $G(c)$, we know that $dG(c) = \frac{kc^{k-1}}{c_M}dc$, and we can solve the closed-form solution of c_D:

$$
c_D = \left[\frac{2(k+1)(k+2)(c_M)^k(1-\omega^k)\gamma F}{L}\right]^{\frac{1}{k+2}}, \qquad \text{(B.2)}
$$

where we denote $\theta = 2(k+1)(k+2)(c_M)^k F$ and $\Omega = \frac{1}{1-\omega^k}$.

Recall that $c_D = \frac{\alpha\gamma+\eta M\bar{p}}{\eta M+\gamma}$ which indicates $M = \frac{\gamma}{\eta}\frac{\alpha-c_D}{c_D-\bar{p}}$. We can solve both the total number of varieties that are produced M and their average price $\bar{p}$ if we could write either of them as an expression for c_D. Note that $\bar{p} = \frac{1}{2}(c_D + \bar{v})$,[27] where $\bar{v}$ is the average marginal cost of all varieties produced.

Define $M_v(c)$ as the measure of varieties that could be produced by the Economy with production costs lower than c. Then the average cost is defined by $\bar{v} = \frac{1}{M}\int_0^{c_D} v\, dM_v(v)$. Given a mass of entrant firms N_E in the Economy, then $H_v(c) = M_v(c)/N_E$ is the measure of the number of varieties produced by a unit of entrants. In other words, for each entrant, $H_v(c)$ measures its probability of producing any potential variety with cost v that is lower than c, given its cost ladder in terms of core competency across varieties. Therefore, we have:

$$
H_v(c) = \sum_{m=0}^{\infty} G(\omega^m c) = \sum_{m=0}^{\infty} \omega^m G(c) = \Omega G(c), \qquad \text{(B.3)}
$$

which follows a sum of the same distribution as $G(c)$. On the other hand, the number of varieties in equilibrium M are produced by those N_E entrants with a real cutoff c_D, that is, $M = N_E H_v(c_D) = N_E \Omega G(c_D)$.

Now we obtain the solution of $\bar{v}$:

$$
\begin{aligned}
\bar{v} &= \frac{1}{M}\int_0^{c_D} v\, dM_v(v) = \frac{1}{N_E H_v(c_D)}\int_0^{c_D} v N_E\, dH_v(v) \\
&= \frac{1}{\Omega G(c_D)}\int_0^{c_D} v\Omega\, dG(v) = \frac{1}{G(c_D)}\int_0^{c_D} v\, dG(v) \\
&= \frac{k}{k+1}c_D
\end{aligned}
\quad, \qquad \text{(B.4)}
$$

From which we obtain the solution of $\bar{p}$, M and N_E:

[27] It holds as $p(v) = \frac{1}{2}(c_D + v)$ for any v in equilibrium.

$$\bar{p} = \frac{1}{2}(c_D + \bar{v}) = \frac{2k+1}{2(k+1)} c_D, \tag{B.5}$$

$$M = \frac{\gamma}{\eta} \frac{\alpha - c_D}{c_D - \bar{p}} = \frac{2(k+1)\gamma}{\eta} \frac{\alpha - c_D}{c_D}, \tag{B.6}$$

$$N_E = \frac{M}{H_v(c_D)} = \frac{2(k+1)\gamma}{\Omega\eta} \frac{\alpha - c_D}{c_D} \left(\frac{c_M}{c_D}\right)^k, \tag{B.7}$$

Finally, the number of surviving firms in equilibrium N equals to the number of entrants with core costs below c_D, that is:

$$N = N_E G(c_D) = \frac{2(k+1)\gamma}{\eta\Omega} \frac{\alpha - c_D}{c_D}. \tag{B.8}$$

Q.E.D.

B2 Free Entry Equilibrium in an Open Economy

In an open economy with J regions, the expected profit of a potential entrant at region l is a sum of expected profits in the local region and all export markets. And the free entry condition in Equation (26) can be rewritten as:

$$\begin{aligned}
&\int_0^{c_{ll}} \Pi_{ll}(c)\, dG(c) + \sum_{f \neq l} \int_0^{c_{lf}} \Pi_{lf}(c)\, dG(c) \\
&= \int_0^{c_{ll}} \sum_{\{m | c \leq \omega^m c_{ll}\}} \pi_{ll}(m, c)\, dG(c) + \sum_{f \neq l} \int_0^{c_{lf}} \sum_{\{m | c \leq \omega^m c_{lf}\}} \pi_{lf}(m, c)\, dG(c) \\
&= \sum_{m=0}^{\infty} \left[\int_0^{\omega^m c_{ll}} \pi_{ll}(m, c)\, dG(c)\right] + \sum_{f \neq l} \sum_{m=0}^{\infty} \left[\int_0^{\omega^m c_{lf}} \pi_{lf}(m, c)\, dG(c)\right] \\
&= F
\end{aligned}, \tag{B.9}$$

where $\pi_{ll}(m, c) = \frac{L}{4\gamma}(c_{ll} - \omega^{-m} c)^2$ and $\pi_{lf}(m, c) = \frac{L}{4\gamma}\left(c_{ff} - e^{\tau_{lf}} \omega^{-m} c\right)^2$. Given the distribution $G(c)$, we have the following free entry condition for region l:

$$\sum_{f=1}^{J} \rho_{lf} L_f c_{ff}^{k+2} = \frac{\gamma\theta}{\Omega}, \tag{B.10}$$

where $\rho_{lf} = \frac{1}{\left(e^{\tau_{lf}}\right)^k} < 1$ is the "freeness" of trade between l and f. Note that, this equation holds for each $l = 1, 2, \dots, J$, yielding a system of J equations. Define $\mathcal{P}$ as the matrix of freeness of trade between any two regions:

$$\mathcal{P} \equiv \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1J} \\ \rho_{21} & 1 & \cdots & \rho_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{J1} & \rho_{J2} & \cdots & 1 \end{pmatrix}, \tag{B.11}$$

With the Cramer's rule, we can solve the cutoffs for firms to survive in the local market:

$$c_{ll} = \left(\frac{\gamma\theta}{L_l \Omega} \frac{|\mathcal{P}_l|}{|\mathcal{P}|}\right)^{\frac{1}{k+2}}, \tag{B.12}$$

where $|\mathcal{P}|$ is the determinant of matrix $\mathcal{P}$, and $|\mathcal{P}_l|$ is the determinant of matrix which replaces the l-th column of $\mathcal{P}$ with a column vector with all elements equal to one.

The steps for solving other variables in equilibrium are similar to those in a closed economy. For region l we have:

$$\bar{v}_l = \frac{k}{k+1} c_{ll}, \quad \text{(B.13)}$$

$$\bar{p}_l = \frac{2k+1}{2(k+1)} c_{ll}, \quad \text{(B.14)}$$

$$M_l = \frac{2(k+1)\gamma}{\eta} \frac{\alpha - c_{ll}}{c_{ll}}. \quad \text{(B.15)}$$

$$N_l = \frac{2(k+1)\gamma}{\eta\Omega} \frac{\alpha - c_{ll}}{c_{ll}}. \quad \text{(B.16)}$$

Note that, for a market f, the measure of varieties M_h are produced by surviving entrants from all regions $l = 1,2, \dots, J$, that is:

$$M_h = \sum_{l=1}^{J} N_{E,l} H_v(c_{lh}) = \sum_{l=1}^{J} \rho_{lh} N_{E,l} \Omega G(c_{hh}), \ \forall h = 1,2, \dots, J, \quad \text{(B.17)}$$

from which we can solve the equilibrium entrant in each region based on the Cramer's rule:

$$N_{E,l} = \frac{2(k+1)\gamma}{\eta\Omega} \sum_{h=1}^{J} \left(\frac{c_M}{c_{hh}}\right)^k \frac{\alpha - c_{hh}}{c_{hh}} \frac{|\mathcal{P}_h|}{|\mathcal{P}|}. \quad \text{(B.18)}$$

Q.E.D.

B3 Proof of Proposition 1

Given the definition of $h(c, c_D)$ for any cutoff c_D, we take the derivative of the export cutoff for seaboard firms with respect to the domestic trade cost:

$$\begin{aligned} \frac{\partial h(c, c_{sf})}{\partial e^{\tau_0}} &= -\frac{1}{c} c(1-\beta)\frac{\hat{r}_0}{r} + \frac{1}{c_{sf}} c_{sf} \frac{k}{k+2}(1-\beta)\frac{\hat{r}_0}{r} \\ &= -\frac{2(1-\beta)}{k+2}\frac{\hat{r}_0}{r} \leq 0 \end{aligned}, \quad \text{(B.19)}$$

where the first and second term in the first equality represents the cost-saving and competition effect of final goods, respectively. The sign of second equality implies that the cost-saving effect dominates the other.

For the inland city, we have:

$$\begin{aligned} \frac{\partial h(c, c_{lf})}{\partial e^{\tau_0}} &= -\frac{1}{c} c(1-\beta)\frac{\hat{r}_0}{r} + \frac{1}{c_{lf}} c_{lf} \frac{k}{k+2}(1-\beta)\frac{\hat{r}_0}{r} - \frac{1}{c_{lf}} c_{lf} \frac{1}{e^{\tau_0}} \\ &= -\frac{2(1-\beta)}{k+2}\frac{\hat{r}_0}{r} - \frac{1}{e^{\tau_0}} \leq 0 \end{aligned}. \quad \text{(B.20)}$$

Similarly, domestic trade costs have a negative net effect on $h(c, c_{lf})$ through the channel of intermediate. In addition, the cost-saving effect of final goods $1/e^{\tau_0}$ ensures the negativity of Equation (B.20).

As $\partial M(c)/\partial h \geq 0$, we have proven that $\frac{\partial h(c, c_{of})}{\partial e^{\tau_0}} \leq 0$ and $\frac{\partial h(c, c_{lf})}{\partial e^{\tau_0}} \leq 0$.

Q.E.D.

B4 Proof of Proposition 2

Given the equilibrium expression of the export revenue of the m-th variety of a seaboard firm, $x_{of}(m,c)$, we have:

$$\begin{aligned}\frac{\partial x_{sf}(m,c)}{\partial e^{\tau_0}} &= \frac{L_s}{4\gamma}\left[2c_{ff}\frac{\partial c_{ff}}{\partial e^{\tau_0}} - (e^{\tau_1}\omega^{-m})^2 2c\frac{\partial c}{\partial e^{\tau_0}}\right] \\ &= \frac{(1-\beta)L_s}{2\gamma}\frac{\hat{r}_0}{r}\left[\frac{k}{k+2}c_{ff}^2 - (e^{\tau_1}\omega^{-m}c)^2\right] \quad , \\ &= \frac{(1-\beta)L_s}{2\gamma}\frac{\hat{r}_0}{r}(e^{\tau_1})^2\left[\frac{k}{k+2}c_{sf}^2 - v^2(m,c)\right]\end{aligned} \tag{B.21}$$

where the first term in the square bracket represents the competition effect of intermediate goods in the exporting market f, and the latter term represents the cost-saving effect of intermediate goods. As the last equality shows, given the level of market competition, the cost-saving effect is larger for varieties with higher production cost $v(m,c)$. Therefore, the cost-saving (competition) effect dominates for the high-cost (low-cost) firms and varieties, leading to an increase (decrease) of product-level export revenue. More specifically, we have:

$$\frac{\partial x_{sf}(m,c)}{\partial e^{\tau_0}} \lesseqgtr 0 \quad if \quad \frac{v(m,c)}{c_{sf}} \gtreqless \left(\frac{k}{k+2}\right)^{1/2}. \tag{B.22}$$

It is similar for firms in the inland city. There is：

$$\begin{aligned}\frac{\partial x_{lf}(m,c)}{\partial e^{\tau_0}} &= \frac{L_l}{4\gamma}\left[2c_{ff}\frac{\partial c_{ff}}{\partial e^{\tau_0}} - (e^{\tau_0+\tau_1}\omega^{-m})^2 2c\frac{\partial c}{\partial e^{\tau_0}} - 2e^{\tau_0+2\tau_1}(\omega^{-m}c)^2\right] \\ &= \frac{(1-\beta)L_l}{2\gamma}\frac{\hat{r}_0}{r}\left[\frac{k}{k+2}c_{ff}^2 - (e^{\tau_0+\tau_1}\omega^{-m}c)^2\right] - \frac{L_l}{2\gamma e^{\tau_0}}(e^{\tau_0+\tau_1}\omega^{-m}c)^2 \\ &= \frac{(1-\beta)L_l}{2\gamma}\frac{\hat{r}_0}{r}(e^{\tau_0+\tau_1})^2\left[\frac{k}{k+2}c_{sf}^2 - v^2(m,c)\right] - \frac{L_l}{2\gamma}e^{\tau_0+2\tau_1}v^2(m,c), \\ &= \frac{(1-\beta)L_l}{2\gamma}\frac{\hat{r}_0}{r}(e^{\tau_0+\tau_1})^2\left[\frac{k}{k+2}c_{sf}^2 - (1+\mu)v^2(m,c)\right]\end{aligned} \tag{B.23}$$

where $\mu = \frac{1}{e^{\tau_0}(1-\beta)\frac{\hat{r}_0}{r}} > 0$. In the third equality, the terms in the square bracket also measures the relative size of competition effect of intermediate goods relative to its cost-saving effect. Again, after the reduction of domestic trade cost, only high-cost products can obtain positive net benefits from intermediate goods channels. Besides that, parameter μ in the last equality measures the relative size of cost-saving effect of final goods compared to that of intermediate goods. Overall, the export revenue of a variety in city l will increase if one of the two cost-saving effects dominate the competition effect of intermediates, that is:

$$\frac{\partial x_{lf}(m,c)}{\partial e^{\tau_0}} \lesseqgtr 0 \quad if \quad \frac{v(m,c)}{c_{lf}} \gtreqless \left(\frac{k}{k+2}\frac{1}{1+\mu}\right)^{1/2}. \tag{B.24}$$

Q.E.D.

C Supplementary Figures and Tables

Table C1 List of Major Port Cities

Port	Province	Port	Province
DaLian	LiaoNing	NingBo	ZheJiang
YingKou	LiaoNing	ZhouShan	ZheJiang
QinHuangDao	HeBei	WenZhou	ZheJiang
TianJin	TianJin	FuZhou	FuJian
YanTai	ShanDong	ShaMen	FuJian
QingDao	ShanDong	ShanTou	GuangDong
RiZhao	ShanDong	ShenZhen	GuangDong
ShangHai	ShangHai	GuangZhou	GuangDong
LianYunGang	JiangSu	ZhuHai	GuangDong
NanTong	JiangSu	ZhanJiang	GuangDong
SuZhou	JiangSu	FangChengGang	GuangXi
ZhenJiang	JiangSu	HaiKou	Hainan
NanJing	JiangSu		

Table C2 Market Access and Firm Exports (2SLS, first stage)

	Panel A. First stage for UMA		
Dependent Variables:		$ln(UMA)$	
	(1)	(2)	(3)
$ln(UMA)\_geocost$	0.648***		0.741***
	(0.020)		(0.026)
$ln(UMA)\_straight$		0.469***	-0.110***
		(0.021)	(0.023)
Dependent Variable's Mean	20.180	20.180	20.180
Observations	236,707	236,707	236,707
	Panel B. First stage for distance to ports		
Dependent Variables:		$ln(dist\_port)$	
	(1)	(2)	(3)
$ln(dist\_port)\_geocost$	0.837***		1.014***
	(0.025)		(0.163)
$ln(dist\_port)\_straight$		0.850***	-0.184
		(0.032)	(0.182)
Dependent Variable's Mean	2.413	2.413	2.413
Observations	236,707	236,707	236,707
Time-varying firm controls	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable in Panel A is the log of upstream market access and that in Panel B is the log of distance to the nearest port. $ln(UMA)\_geocost$ and $ln(dist\_port)\_geocost$ are the logs of counterfactual UMA and $dist\_port$ that are constructed by the artificial transportation network connected by the least costly paths. $ln(UMA)\_straight$ and $ln(dist\_port)\_straight$ are the logs of counterfactual UMA and $dist\_port$ that are constructed by the artificial transportation network connected by straight lines.

Table C3 Upstream Market Access and Firm Exports with Industry-year Fixed Effects (2SLS)

Dependent Variables:	$ln(export)$		
	IV1	IV2	Double IVs
	(1)	(2)	(3)
$ln(UMA)$	2.222***	1.524***	2.320***
	(0.416)	(0.570)	(0.413)
$ln(dist\_port)$	-0.073	-0.065	-0.070
	(0.064)	(0.065)	(0.064)
Dependent Variable's Mean	13.703	13.703	13.703
Time-varying firm controls	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Industry-year fixed effects	Yes	Yes	Yes
K-P rk LM statistic	453.18	278.08	481.29
K-P rk Wald F statistic	587.00	270.04	315.31
Hansen J statistic	\	\	5.85
Observations	236,707	236,707	236,707

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable in Panel A is the log of upstream market access and that in Panel B is the log of distance to the nearest port. $ln(UMA)\_geocost$ and $ln(dist\_port)\_geocost$ are the logs of counterfactual UMA and $dist\_port$ that are constructed by the artificial transportation network connected by the least costly paths. $ln(UMA)\_straight$ and $ln(dist\_port)\_straight$ are the logs of counterfactual UMA and $dist\_port$ that are constructed by the artificial transportation network connected by straight lines.

Table C4 Upstream Market Access and HS4-product-level Exports (2SLS)

Dependent Variables:	$ln(export^{hs4})$			
	(1)	(2)	(3)	(4)
$ln(UMA)$	-2.155***	-3.891***	-2.202***	-2.166***
	(0.439)	(0.847)	(0.441)	(0.440)
$ln(dist\_port)$	-0.073	-0.075	-0.073	-0.074
	(0.052)	(0.052)	(0.052)	(0.052)
$costr$		-39.343***		
		(10.754)		
$ln(UMA) \times costr$		0.002***		
		(0.001)		
$\mathbf{1}(costr_{50th})$			-1.461	
			(0.924)	
$ln(UMA) \times \mathbf{1}(costr_{50th})$			0.073	
			(0.046)	
$\mathbf{1}(costr_{75th})$				-1.646*
				(0.907)
$ln(UMA) \times \mathbf{1}(costr_{75th})$				0.081*
				(0.045)
Dependent Variable's Mean	10.171	10.171	10.171	10.171
Time-varying firm controls	Yes	Yes	Yes	Yes
Time-varying city controls	Yes	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Product fixed effects	Yes	Yes	Yes	Yes
K-P rk LM statistic	260.04	172.49	258.62	259.64
K-P rk Wald F statistic	432.10	96.20	416.87	428.98
Observations	1,265,991	1,265,760	1,265,991	1,265,991

Note: a) *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are clustered at the city-industry level and shown in parenthesis. b) The dependent variable is the log of exports at the HS4-product level. $costr$ is firm's cost rate measured by the total costs per unit of revenue. $\mathbf{1}(costr_{50th})$ is an indicator variable for firms with cost rates higher than the 50th percentile, and $\mathbf{1}(costr_{75th})$ is an indicator for firms with cost rates higher than the 75th percentile.