

# Internal Geography, International Trade, and Regional Specialization\*

A. Kerem Coşar  
University of Chicago  
Booth School of Business

Pablo D. Fajgelbaum  
University of California  
Los Angeles and NBER

December 2014

## Abstract

We introduce an internal geography to the canonical model of international trade driven by comparative advantages to study the regional effects of external economic integration. The model features a dual-economy structure, in which locations near international gates specialize in export-oriented sectors while more distant locations do not trade with the rest of the world. The theory rationalizes patterns of specialization, employment, and relative incomes observed in developing countries that opened up to trade. We find regional specialization patterns consistent with the model in industry-level data from Chinese prefectures.

*JEL Codes:* F11, R12

*Keywords:* trade, internal trade costs, market access, regional specialization

---

\*We thank Andrew Atkeson, Ariel Burstein, Banu Demir Pakel, Chang-Tai Hsieh, Esteban Rossi-Hansberg, Gordon Hanson, Vernon Henderson, Michael Zheng Song, Ralph Ossa, Henry Overman, Stephen Redding, Daniel Yi Xu and participants at various seminars for useful comments. Coşar thanks Chicago Booth for summer financial support.

# 1 Introduction

The experience of many developing countries reveals uneven regional effects of international economic integration. For example, China's recent globalization process took place jointly with large movements of workers and export-oriented industries toward fast-growing coastal regions (World Bank, 2009). The standard view in international trade theory interprets countries as points in space, missing this type of phenomenon. In this paper we introduce an internal geography to the canonical model of international trade driven by comparative advantages to study the regional effects of external economic integration. The theory rationalizes patterns of specialization, employment, and relative incomes observed in developing countries that opened up to trade. A key feature of the model is that comparative-advantage industries locate closer to international gates. We find evidence consistent with this pattern in industry-level data from Chinese prefectures, and document that alternative explanations based on economies of scale or location fundamentals do not fully account for it.

We model a 2-sector economy where locations are arbitrarily arranged on a map and differ in trade costs to international gates such as seaports, airports, or land crossings. Within the country, trade is costly and international shipments must cross through international gates to reach foreign markets. To produce, firms in each location use both a mobile and an immobile factor. The geographical advantage of international gates acts as an agglomeration force, but decreasing returns to scale due to immobile factors create incentives to spread economic activity across locations. This structure allows us to represent all equilibrium outcomes in closed form and the model remains tractable under various extensions.

The key equilibrium feature is a dual-economy structure. Whenever the economy is not fully specialized in what it exports, two distinct regions necessarily emerge: locations near international gates specialize in export-oriented sectors, whereas more distant locations are incompletely specialized and do not trade with other regions or with the rest of the world. The first set of locations constitutes a commercially integrated *coastal region* with high employment density, while the second set forms an autarkic *interior region* where immobile factors are poorer. We examine how the interaction between internal and international trade costs determines each of these regions' weight in national income and economic activity. International integration causes migration of mobile factors to the coastal region and an increase in the measure of locations specialized in export-oriented sectors.

This specialization pattern leads to uneven regional effects of international trade. Higher openness, triggered by a reduction in either international or internal trade costs, has opposing welfare effects on immobile factors located in different points of the country. The interior region loses due to the reduced availability of mobile factors who emigrate to the coastal region, while some locations in the coastal region necessarily gain. When the country is open to trade and internal trade costs are reduced from a sufficiently high level, regional inequality follows an inverted-U shape. The theory also offers a rationale for why the gains from trade may be low in the presence of poor domestic infrastructure. At the aggregate level, the gains from international trade are a

weighted average of the gains accrued to each location, and can be broken down into a familiar term that captures the gains without internal geography and a term that captures domestic trade frictions. The latter depends on the size of the coastal region, so that, by reducing the measure of trading locations, domestic trade frictions hamper the aggregate gains from trade.

These qualitative results rationalize patterns of specialization, employment, and relative incomes observed in developing countries that opened up to trade. For example, economic integration led to larger employment growth in comparative-advantage industries located near borders in Mexico (Hanson 1996) and near ports in Vietnam (McCaig and Pavnick 2012). The dual-economy structure generated by the model is consistent with the coexistence of globally integrated regions and a hinterland with low international participation in large developing countries with a poor transportation infrastructure (Asian Development Bank Institute 2009). The theory rationalizes these facts by introducing internal trade costs to the neoclassical trade model. Limao and Venables (2001) and Atkin and Donaldson (2014) provide evidence that such costs are sizeable in developing countries.

In the empirical analysis, we assess the key prediction of the model about regional specialization, from which its welfare and distributional implications stem: due to the interaction between national comparative advantages and internal trade costs, export-oriented industries (i.e., industries in which the economy has comparative advantages relative to the rest of the world) locate closer to international gates, leading to uneven regional effects of trade. We assess this prediction using industry-level data from Chinese prefectures, proxying for industries' export orientation with national export-revenue ratios. To sidestep the concern that these ratios are endogenous to activity within China, we also rely on the argument that China's comparative advantages lie in labor-intensive industries and use industry-specific labor intensities from U.S. data as a second proxy for industry export orientation.

We find that economic activity at the industry-prefecture level is strongly correlated with the interaction between prefecture proximity to coastlines and industry export orientation. The direct effect of distance on economic activity is sizeable: moving inland by 275 miles (the median distance from the coast across prefectures) decreases industry employment by 17 percent for an industry with average export-revenue ratio, and by 13 percent for an industry with average labor intensity. But this negative distance gradient is stronger for export-oriented industries: employment shrinks by 32 percent for an industry that has an export-revenue ratio one standard deviation higher than average, and by 21 percent for an industry that has labor intensity one standard deviation higher than average.

Finally, we examine alternative explanations that can also generate this pattern. Our approach to explaining regional specialization complements explanations within the New Economic Geography tradition of Krugman (1991). In that view, it could be that industries with higher returns to scale locate closer to international markets and are more likely to export. Following Hanson and Xiang (2004), we use product differentiation as an industry-specific measure of returns to scale to control for this home-market effect. We also control for industry-specific trade costs.

While the data suggest that economies of scale may also be important in determining industry location, our main finding is not affected by this control. Our approach also complements explanations based on location fundamentals such as Courant and Deardorff (1992), and on external economies of scale as in Rossi-Hansberg (2005). In those views, it could be that, closer to international gates, either endowments or externalities drive low relative unit costs in specific industries which become export oriented. We argue that explanations of this type do not fully account for the regional specialization pattern that we find, for they imply a positive relative-productivity gradient in favor of comparative-advantage industries toward coastlines that is not strongly present in the Chinese data.

**Relation to the literature:** Few studies consider an explicit interaction between international and domestic trade costs. In a New Economic Geography context, Krugman and Livas-Elizondo (1996) and Behrens et al. (2006) present models where two regions within a country trade with the rest of the world. Henderson (1982) and Rauch (1991) model systems of cities in an open economy framework. Rossi-Hansberg (2005) studies the location of industries on a continuous space with spatial externalities. These papers are based on economies of scale, whereas we focus on the interaction between heterogeneous market access within countries and comparative advantages between countries.

Matsuyama (1999) studies a multi-region extension of Helpman and Krugman (1985). He focuses on home-market effects under different spatial configurations, but does not include factor mobility. In neoclassical environments, Bond (1993) and Courant and Deardorff (1993) present models with regional specialization where relative factor endowments may vary across discrete regions within a country. These papers do not include heterogeneity in access to world markets. Venables and Limao (2002) study international specialization in a setup with central and peripheral locations and no factor mobility.

More recently, Ramondo et al. (2014) quantitatively study the gains from trade and technology diffusion allowing for multiple regions within countries in the Eaton and Kortum (2002) model. Redding (2014) develops a multi-region version of the trade and labor mobility model in Helpman (1998), while Allen and Arkolakis (2014) study an Armington model extended with labor mobility and external economies. These papers do not study the effect of international trade on the pattern of industrial specialization across regions within countries.

Finally, a large empirical literature studies determinants of industry location. Holmes and Stevens (2004) offer a summary assessment of the forces that determine industry location in the United States, while Hanson (1998) reviews the literature on the effect of North American trade integration on the location of economic activity. Amiti and Javorcik (2008) and Lu and Tao (2009) document the determinants of firm and industry locations in China. Our empirical analysis complements these studies, suggesting that international trade is an important driver of industry location within China and that it may also be so in other developing countries that are open to international trade but face high internal trade costs.

## 2 Model

**Geography and Trade Costs** The country consists of a set of locations arbitrarily arranged on a map. We index locations by  $\ell$ , and assume that only some locations can trade directly with the rest of the world. Goods must cross through a port to be shipped internationally. As will be clear below, the nature of our model implies that only the distance separating each location  $\ell$  from its nearest port matters for the equilibrium. Therefore, we assume without loss of generality that  $\ell$  represents the distance separating each location from its nearest port, and we denote all ports by  $\ell = 0$ . We let  $\bar{\ell}$  be the maximum distance between a location inside the country and its nearest port.

There are two industries,  $i \in \{X, M\}$ . There is an industry-specific international iceberg trade cost such that for each unit of  $i$  shipped from  $\ell = 0$ , only  $e^{-\tau_0^i}$  units arrive in the rest of the world (RoW). Within the country, iceberg trade costs are constant per unit of distance and equal  $\tau_1^i$  in industry  $i$ . Therefore, for each unit of  $i$  that location  $\ell$  ships to the RoW, only  $e^{-\tau_0^i - \tau_1^i \ell}$  units arrive.

Given this geography, we can interpret each location  $\ell = 0$  as a seaport, airport, or international land crossing. What is key is that not all locations have the same technology for trading with the RoW. This will drive concentration near points with good access. Internal geography vanishes when  $\tau_1^i = 0$  for  $i = X, M$  or  $\bar{\ell} = 0$ .

**Endowments** There are two factors of production, a mobile factor and an immobile one. We refer to the mobile factor as workers, and to the immobile factor as land.<sup>1</sup> We let  $\lambda(\ell)$  be the total amount of land available in locations at distance  $\ell$  from their nearest port and normalize units so that the national land endowment also equals one.<sup>2</sup> Land is owned by immobile landlords who do not work. We let  $N$  be the labor endowment at the national level and  $n(\ell)$  denote employment density in  $\ell$ , which is to be determined in equilibrium since workers choose their location optimally.

**Preferences** Both workers and landlords spend their income locally. The indirect utility of a worker living in  $\ell$  is

$$u(\ell) = \frac{w(\ell)}{E(\ell)}, \tag{1}$$

where  $w(\ell)$  is the wage at  $\ell$ , and  $E(\ell)$  is the cost-of-living index. We let  $p(\ell) \equiv P_X(\ell)/P_M(\ell)$  be the relative price of  $X$  in  $\ell$ . Since preferences are homothetic, there exists an increasing and concave function  $e(p(\ell))$  that depends on the relative price of  $X$  such that  $E(\ell) = P_M(\ell)e(p(\ell))$ .

For landlords, income equals rents  $r(\ell)$  per unit of land and utility is therefore increasing in  $r(\ell)/E(\ell)$ . Accounting for both wages and rents, income generated by each unit of land at  $\ell$  is  $y(\ell) = w(\ell)n(\ell) + r(\ell)$ .

---

<sup>1</sup>The appendix presents an extension where both labor and capital are mobile factors used in production and sectors can differ in labor intensity. In the empirical analysis we use data on both employment and capital to evaluate the model's prediction about the location of the mobile factor in different industries.

<sup>2</sup>If the distribution of land is uniform,  $\lambda(\ell)$  represents the measure of locations at distance  $\ell$  from their nearest port.

**Technology** Production in each sector requires one unit of land to operate a technology with decreasing returns to scale in labor. We let  $n_i(\ell)$  be employment per unit of land in industry  $i$  at location  $\ell$ . Profits per unit of land in sector  $i$  at  $\ell$  are

$$\pi_i(\ell) = \max_{n_i(\ell)} \{P_i(\ell)q_i[n_i(\ell)] - w(\ell)n_i(\ell) - r(\ell)\}. \quad (2)$$

The production technology is

$$q_i[n_i(\ell)] = \kappa_i \frac{n_i(\ell)^{1-\alpha_i}}{a_i(\ell)}, \quad (3)$$

where  $a_i(\ell)$  is the unit cost of production in industry  $i$  in sector  $\ell$  and  $\kappa_i \equiv \alpha_i^{-\alpha_i} (1 - \alpha_i)^{-(1-\alpha_i)}$  is a normalization constant that saves notation later on. Decreasing returns to scale  $1 - \alpha_i$  measure the labor intensity in sector  $i$ , acting as a congestion force. From (3) it follows that the aggregate production function in sector  $i$  at  $\ell$  is Cobb-Douglas with land intensity  $\alpha_i$ . To simplify the exposition, we assume that land intensity is the same across both sectors,  $\alpha_X = \alpha_M \equiv \alpha$ , and discuss the implications of differences in land intensity across sectors in Section 2.4.

Industry-specific production costs  $\{a_M(\ell), a_X(\ell)\}$  may vary across locations, subject to the restriction that the relative cost of production is constant across the country:

$$\frac{a_X(\ell)}{a_M(\ell)} = a \text{ for all } \ell \in [0, \bar{\ell}]. \quad (4)$$

Therefore, while some locations might be more productive than others in every industry, comparative advantages are defined at the national level. In turn,  $a$  differs across countries, creating incentives for international trade. In this way we retain the basic structure of a neoclassical model of trade, where countries are differentiated by their comparative advantages.

The solution to the firm's problem yields labor demand per unit of land used by each sector  $i$  in location  $\ell$ ,

$$n_i(\ell) = \frac{1 - \alpha}{\alpha} \left( \frac{P_i(\ell)}{a_i(\ell) w(\ell)} \right)^{1/\alpha} \text{ for } i = X, M. \quad (5)$$

Finally, we let  $\lambda_i(\ell)$  be the total amount of land used by sector  $i = X, M$  at  $\ell$ .

## 2.1 Local Equilibrium

We first define and characterize a local equilibrium at each location  $\ell$  that takes prices  $\{P_X(\ell), P_M(\ell)\}$  and the real wage  $u^*$  as given.

**Definition 1** *A local equilibrium at  $\ell$  consists of population density  $n(\ell)$ , labor demands  $\{n_i(\ell)\}_{i=X,M}$ , patterns of land use  $\{\lambda_i(\ell)\}_{i=X,M}$ , and factor prices  $\{w(\ell), r(\ell)\}$  such that*

1. workers maximize utility,

$$u(\ell) \leq u^*, = \text{ if } n(\ell) > 0; \quad (6)$$

2. profits are maximized,

$$\pi_i(\ell) \leq 0, = \text{ if } \lambda_i(\ell) > 0, \text{ for } i = X, M, \quad (7)$$

where  $\pi_i(\ell)$  is given by (2);

3. land and labor markets clear,

$$\sum_{i=X,M} \lambda_i(\ell) = \lambda(\ell), \quad (8)$$

$$\sum_{i=X,M} \frac{\lambda_i(\ell)}{\lambda(\ell)} n_i(\ell) = n(\ell); \text{ and} \quad (9)$$

4. trade is balanced.

Conditions 2-4 from this definition constitute a small Ricardian economy where the amount of labor is given. In addition, employment density  $n(\ell)$  at each location is determined by condition 1.

We let  $p_A$  be the autarky price in each location. By this, we mean the price prevailing in the absence of trade with any other location or with RoW, but when labor mobility is allowed across locations. The autarky price  $p_A$  is the same and equal to  $a$  in all locations. Using (7), location  $\ell$  must be fully specialized in  $X$  when  $p(\ell) > p_A$ , and fully specialized in  $M$  when  $p(\ell) < p_A$ . Because each location takes relative prices as given, this means that a trading location is fully specialized. Only if  $p(\ell)$  happens to coincide with  $p_A$  may a trading location be incompletely specialized. This logic also implies that an incompletely specialized location is in autarky.<sup>3</sup>

To solve for the local wage  $w(\ell)$  we note that whenever a location is populated, the local labor supply decision (6) must be binding. Together with (5) and the market clearing conditions, this gives the equilibrium population density in each location,<sup>4</sup>

$$n(\ell) = \begin{cases} \frac{1-\alpha}{\alpha} \left( \frac{1}{a_X(\ell)u^*} \frac{p(\ell)}{e(p(\ell))} \right)^{1/\alpha} & \text{if } p(\ell) \geq a, \\ \frac{1-\alpha}{\alpha} \left( \frac{1}{a_M(\ell)u^*} \frac{1}{e(p(\ell))} \right)^{1/\alpha} & \text{if } p(\ell) < a. \end{cases} \quad (10)$$

Expression (10) conveys the various forces that determine the location decision of workers. Agents care about the effect of prices on both income and cost of living. Since preferences are homothetic, agents employed in the industry-location pair  $(i, \ell)$  necessarily enjoy a higher real income when the local relative price of industry  $i$  is higher in location  $\ell$ . That is, the positive income effect from a higher relative price necessarily offsets cost-of-living effects. At the same time, there are congestion forces captured by the intensity of land use  $\alpha$ , so that workers avoid places with

<sup>3</sup>In knife-edge cases where  $p(\ell) = p_A$ , a location can be indifferent between autarky or trade. In that case, a location may trade and still be incompletely specialized.

<sup>4</sup>This is a special case of the expression for density derived under more general conditions in equation (35) in the appendix.

high employment density. The larger the congestion, the smaller the population density. Naturally, agents also prefer locations with better fundamental productivity, i.e., lower  $a_i(\ell)$ .

We highlight that trade affects density through its effect on the relative price  $p(\ell)$ . When  $p(\ell) \neq p_A$ , locations are fully specialized and necessarily export. In this circumstance  $n(\ell)$  increases with the relative price of the exported good and decreases with the national real wage  $u^*$ .

## 2.2 General Equilibrium

We move on to the effect of market access (distance to the nearest international gate) on employment density and the specialization pattern in general equilibrium. The country is a small economy taking international prices  $\{P_X^*, P_M^*\}$  as given.<sup>5</sup> We denote the relative price at RoW by  $p^* = P_X^*/P_M^*$ .

We also define the average international and domestic iceberg costs across sectors,  $\tau_j \equiv \frac{1}{2} \sum_{i=X,M} \tau_j^i$  for  $j = 0, 1$ . No arbitrage implies that for any pair of locations  $\ell$  and  $\ell'$  separated by distance  $\delta \geq 0$ , relative prices in industry  $i$  satisfy

$$P_i(\ell')/P_i(\ell) \leq e^{\tau_1^i \delta} \text{ for } i = X, M. \quad (11)$$

This condition binds if goods in industry  $i$  are shipped from  $\ell$  to  $\ell'$ . A similar condition holds with respect to RoW. Since all locations  $\ell = 0$  can trade directly with RoW at the same relative price, (11) implies

$$e^{-2\tau_0} \leq p(0)/p^* \leq e^{2\tau_0}. \quad (12)$$

The first inequality is binding if the country exports  $X$  to RoW, while the second is if it imports  $X$ . Therefore, for any location  $\ell$  we have

$$e^{-2\tau_1 \ell} \leq p(\ell)/p(0) \leq e^{2\tau_1 \ell}, \quad (13)$$

where the first inequality binds if  $\ell$  exports  $X$  to RoW, and the second does if  $\ell$  imports  $X$ .

We are ready to define the general equilibrium of the economy:

**Definition 2 (General Equilibrium)** *An equilibrium in a small economy given international prices  $\{P_X^*, P_M^*\}$  consists of a real wage  $u^*$ , local outcomes  $n(\ell)$ ,  $\{n_i(\ell)\}_{i=X,M}$ ,  $\{\lambda_i(\ell)\}_{i=X,M}$ ,  $w(\ell)$ ,  $r(\ell)$ , and goods prices  $\{P_i(\ell)\}_{i=X,M}$  such that*

1. given  $\{P_i(\ell)\}_{i=X,M}$  and  $u^*$ , local outcomes are a local equilibrium by Definition 1 for all  $\ell \in [0, \bar{\ell}]$ ;
2. relative prices  $p(\ell)$  satisfy the no-arbitrage conditions (12) and (13) for all  $\ell, \ell' \in [0, \bar{\ell}]$ ; and

---

<sup>5</sup>It is straightforward to extend the model to a two-country general equilibrium in which the world prices are endogenously determined.



3. the real wage  $u^*$  adjusts such that the national labor market clears,

$$\int_0^{\bar{\ell}} n(\ell) \lambda(\ell) d\ell = N. \quad (14)$$

Since Definition 1 of a local equilibrium includes trade balance for each location, trade is also balanced at the national level. To characterize the regional patterns of specialization, we first note that the no-arbitrage conditions rule out bilateral trade flows between any pair of locations within the country. Since all locations share the same relative unit costs, there are no gains from trade within the country. To see why, suppose that there is bilateral trade between two distinct locations  $\ell_X$  and  $\ell_M$ . If  $\ell_X$  is the  $X$ -exporting location of the pair, then  $p(\ell_M) \leq p_A \leq p(\ell_X)$ . But, at the same time, the no-arbitrage condition (11) implies that the relative price of  $X$  is strictly higher in  $\ell_M$ , which is a contradiction. This implies that the country is in international autarky if and only if all locations are in autarky and incompletely specialized.<sup>6</sup>

With this in mind, we can characterize the general equilibrium. The country can be partitioned into those locations that trade with RoW and those that stay in autarky. It follows that if the country is not in international autarky there must be some boundary  $b \in [0, \bar{\ell}]$  such that all locations  $\ell < b$  are fully specialized in the export-oriented industry. In turn, all locations  $\ell > b$  do not trade with the RoW and stay in autarky, while locations at distance  $b$  from their nearest international gate are indifferent between trading or not with RoW.<sup>7</sup> Therefore, an internal boundary  $b$  divides the country between a trading *coastal region* comprising all locations  $\ell \in [0, b)$  whose distance to the nearest international gate is less than  $b$  and an autarkic *interior region* comprising the remaining locations  $\ell \in (b, \bar{\ell}]$ .<sup>8</sup>

This reasoning implies that the distance separating each location from its nearest international gate is the only local fundamental that matters for specialization. This justifies our initial statement that locations may be arbitrarily arranged on a map, as well as our decision to index locations by their distance to the nearest port. In addition, any bilateral trade flow in the country either originates from ports, or is directed toward them. These features allow us to represent a two-dimensional geography on the line, leading to closed-form characterizations for aggregate

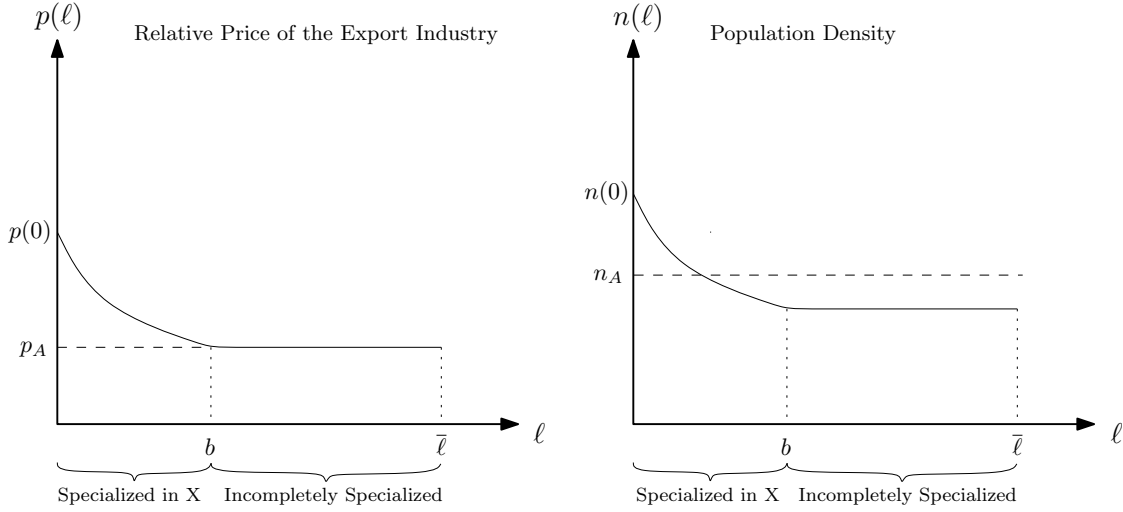
---

<sup>6</sup>This outcome is akin to the spatial impossibility results in the tradition of Starrett (1978), whereby homogeneous space and constant-returns-to-scale technologies lead to autarkic locations. In our setting, the only heterogeneity across space is the distance to the port. Thus, if trade occurs, it is with the RoW. Of course, this lack of bilateral internal trade is not a feature of the data. The model could be extended with product differentiation within sectors to generate internal trade without affecting the regional specialization pattern.

<sup>7</sup>To determine which locations belong in each set, we note that if the country is an exporter of  $X$  then all locations that trade with RoW must also export  $X$ . Therefore, all locations  $\ell$  such that  $e^{-2(\tau_0 + \tau_1 \ell)} p^* < p_A$  must stay in autarky, for if they specialized in  $X$  then the relative price of  $X$  would be so low that it would induce specialization in  $M$ . In the same way, all locations  $\ell$  such that  $e^{-2(\tau_0 + \tau_1 \ell)} p^* > p_A$  must specialize in  $X$ , for if they stayed in autarky then the relative price of  $M$  would be so high that it would induce domestic consumers to import from abroad, violating the no-arbitrage condition (13).

<sup>8</sup>The regional specialization pattern resembles a Von-Thünen type outcome. In the standard version of the Von-Thünen model, as featured for example in Chapter 3 of Fujita and Thisse (2013), the pattern of specialization across locations that trade with a central district is dictated by differences in trade costs or land intensity across goods. In our context, differences between countries lead to specialization rings within the country.

Figure 1: **Relative Prices and Population Density over Distance**



equilibrium outcomes and making the model tractable for counterfactuals.

Since all locations  $\ell \in (b, \bar{\ell}]$  are in autarky, they are incompletely specialized and their relative price is  $p_A = a$ . Given this price in the autarkic region and the regional pattern of production, the no-arbitrage conditions (12) and (13) generate the regional price distribution as function of the internal boundary  $b$ . Henceforth, we assume that the economy is a net exporter of  $X$ , and below we provide the conditions under which this is the case. Therefore, the distribution of local relative prices is given by

$$p(\ell) = p^* e^{-2(\tau_0 + \tau_1 \min[\ell, b])}. \quad (15)$$

Using (10) in the aggregate labor-market clearing condition (14), we can solve for the real wage,

$$u^* = \left[ \frac{1 - \alpha}{\alpha} \frac{1}{N} \int_0^{\bar{\ell}} \lambda(\ell) \left( \frac{1}{a_X(\ell)} \frac{p(\ell)}{e(p(\ell))} \right)^{1/\alpha} d\ell \right]^\alpha. \quad (16)$$

Since the relative price function in (15) depends on  $b$ , so does the real wage. To find the location of the boundary  $b$ , we use the continuity of the relative price function:

$$p(b) \geq p_A, = \text{ if } b < \bar{\ell}. \quad (17)$$

When  $p(\bar{\ell}) > p_A$ , then  $b = \bar{\ell}$ , so that the interior region does not exist. The general equilibrium is fully characterized by the pair  $\{u^*, b\}$  that solves (16) and (17). All other variables follow from these two outcomes.

Figure 1 illustrates the structure of the equilibrium when the economy exports good  $X$  but is not fully specialized. On the horizontal axis, locations are ordered by their distance to their nearest port. As we have already established, this is the only geographic aspect that determines the local equilibrium. In the left panel, the relative price of the exported good declines with distance to the

port until it hits the autarky relative price, and remains constant afterward. The economy is fully specialized in  $X$  in the coastal region, but incompletely specialized in the interior. Only the coastal locations  $\ell \in [0, b]$  are commercially integrated with RoW.

In the right panel, we plot population density assuming that the fundamental  $a_X(\ell)$  is constant across locations, so that international trade is the only force that shapes the distribution of population density. As the relative price of the export industry decreases away from international gates, so does population density in the coastal region until it reaches the interior region. Hence, international trade causes population to agglomerate near international gates relative to the autarkic density level  $n_A = N/\bar{\ell}$ .

So far we have assumed a given trade pattern at the national level. It is simple to determine the national trade pattern: it must be consistent with the trade pattern at ports. Therefore, as in a model that lacks internal geography, the country exports  $X$  if  $p_A/p^* < e^{-2\tau_0}$  and is in international autarky if  $e^{-2\tau_0} < p_A/p^* < e^{2\tau_0}$ . These results imply that domestic trade costs, while capable of affecting the gains and the volume of international trade, are unable to impact the pattern or the existence of it. In this context, the interior region exists if and only if  $e^{-2(\tau_0 + \tau_1 \bar{\ell})} < p_A/p^*$ , i.e., when trade costs  $\{\tau_1, \tau_0\}$  or the extension of land  $\bar{\ell}$  are sufficiently large, or when comparative advantages, captured by  $p_A/p^*$ , are sufficiently weak.

We summarize our findings so far as follows.

**Proposition 1 (Regional Specialization Pattern)** *There is a unique general equilibrium, in which: (i) if the country trades internationally but is not fully specialized, there exists an autarkic interior region  $[b, \bar{\ell}]$  that is incompletely specialized and a coastal region  $[0, b)$  that trades with RoW and specializes in the export-oriented industry; and (ii) the national trade pattern is determined independently from internal geography.*

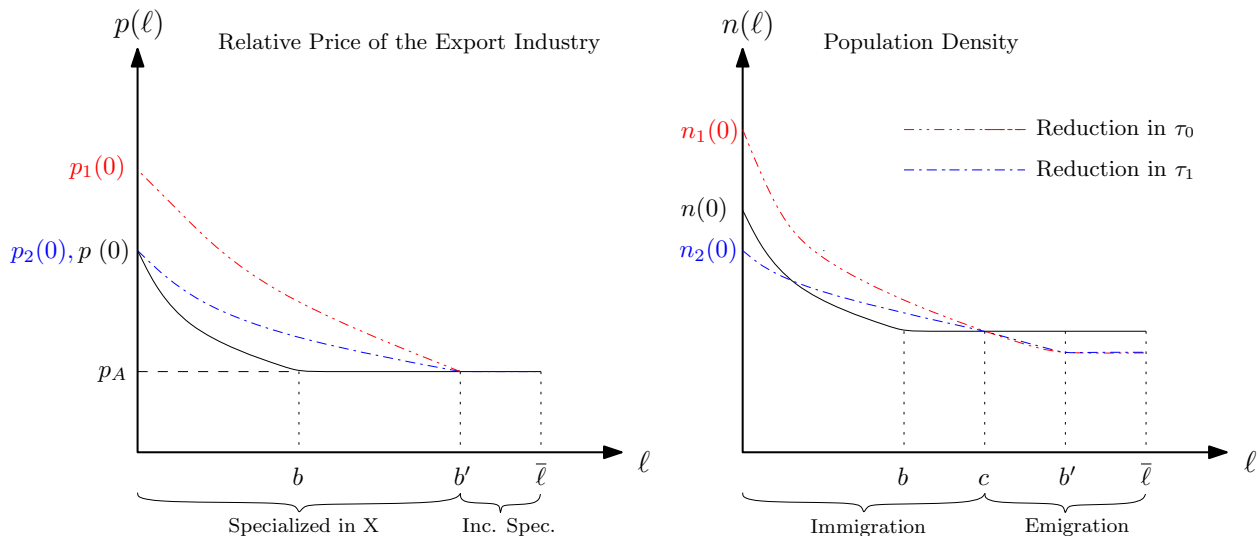
### 2.3 Impact of International and Domestic Trade Costs

The experience of many developing countries shows that international trade integration may be associated with shifts in economic concentration and industry location. In our model, population density varies across locations based on proximity to international gates, and population density in the coastal region relative to the interior region is endogenous. We summarize the impact of a discrete change in trade costs on these outcomes as follows.

**Proposition 2 (Internal Migration)** *A reduction in international or in domestic trade costs moves the boundary inland to  $b' > b$  and causes migration from region  $[c, \bar{\ell}]$  into region  $[0, c]$  for some  $c \in (b, b')$ . A lower  $\tau_0$  causes population at the port  $n(0)$  to increase, but a lower  $\tau_1$  causes  $n(0)$  to decrease.*

The direct impact of a reduction in trade costs is that the relative price of the exported industry increases in the coastal region,  $[0, b)$ . In the case of a reduction in  $\tau_0$ , the shift is uniform across locations, while a lower  $\tau_1$  results in a flattening of the slope of relative prices toward the interior.

Figure 2: **Changes in Domestic and International Trade Costs**



In both cases, the change in prices causes the relative price at  $b$  to be larger than the autarky price  $p_A$ , so that locations at the boundary now find it profitable to specialize in export-oriented industries and the boundary moves inland.

What are the internal migration patterns associated with these reductions in trade costs? A natural consequence of lower trade costs is an increase in the real wage,  $u^*$ . Since in the autarkic interior region relative prices remain constant, labor demand shrinks. As a result, workers migrate away from locations that remain autarkic toward the coast, and relative population density increases in the coastal region as result.

Figure 2 illustrates these effects. The solid black line reproduces the initial equilibrium from Figure 1. The dashed-double-dotted red line represents a new equilibrium with lower international trade costs. The price function shifts upward and the intercept increases from  $p(0)$  to  $p_1(0)$ , increasing population density at the port to  $n_1(0)$ . Locations in  $[b, b']$  start trading, but the newly specialized locations  $[c, b']$  lose population. The dashed-dotted blue line shows the effect of a reduction in domestic trade costs. Prices at the port stay constant, but the slope flattens. As a result, population density at the port shrinks to  $n_2(0)$ . In relative terms, locations at intermediate distances become more attractive when domestic trade costs decline. In both cases, population density is higher in  $[0, c]$  in the new equilibrium.

We move now to the impact of domestic trade costs  $\tau_1$  on the gains from international trade. We study the impact of trade costs on the real wage  $u^*$ , but we note that average real returns to land as well as national real income are proportional to the real wage  $u^*$ . Therefore trade costs have the same impact on the average real returns to both mobile and immobile factors.

We can consider two extreme cases. As  $\tau_1 \rightarrow \infty$ , domestic trade becomes prohibitive so that  $b \rightarrow 0$  and the country approaches international autarky. In that case, all locations face the same relative price  $p(\ell) = p_A$ . We let  $u_A$  be the real wage in that circumstance. In the other extreme,

when  $\tau_1 = 0$ , then  $b = \bar{\ell}$ , and all locations face the relative price  $p(\ell) = p(0)$ . In that case, the real wage is

$$\bar{u} = \left( \frac{1 - \alpha}{\alpha} \frac{Z}{N} \right)^\alpha \frac{p(0)}{e(p(0))},$$

where

$$Z = \int_0^{\bar{\ell}} \frac{\lambda(\ell)}{a_X(\ell)^{1/\alpha}} d\ell$$

measures the distribution of land and total productivity across locations. As in a standard neoclassical model, the real wage is increasing in the terms of trade, which are captured by  $p(0)$ . Diminishing returns to labor causes the real wage to decrease with the national labor endowment.

Using the solution for the real wage  $u^*$  from (16), the actual gains of moving from autarky to trade can be broken down as follows,

$$\frac{u^*}{u_A} = \Omega(b; \tau_1) \cdot \frac{\bar{u}}{u_A}, \quad (18)$$

where we define the potential gains of moving from autarky to international trade as

$$\frac{\bar{u}}{u_A} = \frac{p(0)/e(p(0))}{p_A/e(p_A)},$$

and the effects of domestic trade frictions are captured by

$$\Omega(b; \tau_1) = \left[ \int_0^{\bar{\ell}} \frac{\lambda(\ell)/a_X(\ell)^{1/\alpha}}{Z} \left( \frac{p(\ell)/e(p(\ell))}{p(0)/e(p(0))} \right)^{1/\alpha} d\ell \right]^\alpha. \quad (19)$$

The actual gains from trade,  $u^*/u_A$ , equal the potential gains from trade without domestic trade costs,  $\bar{u}/u_A$ , adjusted by  $\Omega(\tau_1, b)$ . This function captures the impact of internal geography on the gains from trade. It is a weighted average of the losses caused by domestic trade costs in each location, with weights that correspond to each location's absolute productivity level and land endowment relative to the country aggregate. In turn, the location-specific losses from domestic trade costs are captured by the reduction in the terms of trade caused by distance. The total friction,  $\Omega(\tau_1, b)$ , is strictly below 1 as long as  $\tau_1 > 0$ , and it equals 1 if  $\tau_1 = 0$ .

How do the gains from trade depend on domestic trade costs? Naturally, the larger the size of the export-oriented region, the more a country should benefit from trade. Since  $\tau_1$  causes the export-oriented region to shrink, the gains from trade decrease with domestic trade costs. Intuitively, a lower  $\tau_1$  makes exporting profitable for locations further away from the port, allowing economic activity to spread out and mitigate the congestion forces in dense coastal areas.

To formalize this, we define the elasticity of the consumer price index,  $\varepsilon(p) = \frac{e'(p)}{e(p)}p$ , and let  $s(\ell)$  be the share of location  $\ell$  in total employment. Given a shock to  $p^*$ ,  $\tau_0$ , or  $\tau_1$ , total differentiation

of the labor clearing condition (14) and the condition (17) for  $b$  yield the change in the real wage:

$$\widehat{u}^* = \int_0^b [1 - \varepsilon(p(\ell))] s(\ell) \widehat{p}(\ell) d\ell, \quad (20)$$

where  $\widehat{x}$  represents the proportional change in variable  $x$ . This expression describes the aggregate gains from a reduction in trade costs, either domestic or international, as a function of the relative price change faced by export-oriented locations weighted by their population shares  $s(\ell)$ . Reductions in domestic or international trade costs cause the relative price of the exported good to increase. This has a positive effect on revenues and a negative effect on the cost of living. The latter is captured by the price-index elasticity  $\varepsilon(p(\ell))$ , and mitigates the total gains. In this context, (20) implies that the gains from an improvement in the terms of trade, caused by either lower  $\tau_0$  or larger  $p^*$ , are bounded above by the share of employment in export-oriented regions.<sup>9</sup>

It follows from this reasoning that domestic and international trade costs are complementary, in that the gains from international trade are decreasing in domestic trade costs. Larger  $\tau_1$  causes relative export prices to decrease faster toward the interior, reducing the gains from trade:<sup>10</sup>

$$\frac{d(u^*/u_A)}{d\tau_1} < 0. \quad (21)$$

This aggregate effect of trade-cost reductions hides distributional effects between immobile factors located in different points of the country. The real returns to immobile factors at location  $\ell$  are  $v(\ell) \equiv r(\ell)/E(\ell)$ . The change in these returns at  $\ell$  when there are marginal changes in international or domestic trade costs is

$$\widehat{v}(\ell) = \frac{1}{\alpha} [1 - \varepsilon(p(\ell))] \widehat{p}(\ell) - \frac{1 - \alpha}{\alpha} \widehat{u}^*. \quad (22)$$

The first term measures the impact of relative price changes through both revenues and cost of living. The second part considers the economy-wide increase in real wages, which captures the emigration of mobile factors.

Consider first the interior locations  $\ell \in (b, \bar{\ell}]$ . In these places,  $\widehat{p}(\ell) = 0$  because distance precludes terms of trade improvements, but mobile factors emigrate with trade reform because of the real wage increase. As a result, immobile factors in the interior  $(b, \bar{\ell}]$  region lose from lower trade costs. However, some immobile factors located in the coastal areas  $\ell \in [0, b)$  necessarily gain. Hence, reductions in both international and domestic trade costs generate redistribution of resources away from the interior to the coastal region.

However, outcomes for immobile resources in the coastal region are not uniformly positive. While on average the coastal region necessarily gains from improvements in international or domestic trade conditions, reductions in domestic trade costs  $\tau_1$  necessarily hurt immobile factors located at the

---

<sup>9</sup>An increase in the terms of trade implies  $\widehat{p}(\ell) = \widehat{p}$  for all  $\ell \in [0, b]$ . Then, (20) gives  $\widehat{u}^*/\widehat{p} = \beta_T \int_0^b [1 - \varepsilon(p(\ell))] s(\ell) d\ell < \int_0^b s(\ell) d\ell$ .

<sup>10</sup>Formally,  $d\tau_1^i > 0$  implies  $\widehat{p}(\ell) = -\ell d\tau_1^i < 0$  for all  $\ell \in [0, b]$  so that  $\widehat{u}^*/d\tau_1 < 0$ , implying (21).

port. In other words, coastal areas are better off if places further inside the country have poorer access to world markets.

We summarize these results as follows.

**Proposition 3 (Distributive Effects of Trade Across Regions)** *Real income of immobile factors located in the interior region  $(b, \bar{\ell}]$  decreases with reductions in international or internal trade costs, while on average the coastal locations  $[0, b)$  gain. There is some  $\varepsilon < b$  such that real income of immobile factors in locations  $[0, \varepsilon)$  decreases with reductions in domestic trade costs.*

These distributional effects give rise to an inverted-U shape for inequality across immobile factors with internal integration. When the country is open to trade and  $\tau_1 \rightarrow \infty$ , all locations other than the port are autarkic. Reducing internal trade costs leads to an initial increase in regional income inequality by expanding the coastal region and thus benefiting newly integrated locations. But as  $\tau_1 \rightarrow 0$ , internal geography vanishes and outcomes are again symmetric in all locations within the country, eliminating income differences across locations.

## 2.4 Extensions

The benchmark model developed so far is purposely stylized to highlight the key forces of our analysis. In Appendix A we characterize the local equilibrium and specialization patterns in more general model that allows for a non-tradable sector, location-specific amenities, capital as an additional (mobile) factor of production, and differences in capital and land intensity across sectors.

When the benchmark model is extended to allow for  $\alpha_X \neq \alpha_M$ , so that land intensity may vary across sectors, the equilibrium features highlighted in Proposition 1 remain unchanged under a condition similar to (4). In that case, local autarky prices are the same in all locations, independent of their labor endowments, if and only if  $a_X(\ell)^{\alpha_M} / a_M(\ell)^{\alpha_X}$  is constant for all  $\ell$ .<sup>11</sup> If this condition holds, we ensure that the economy retains the dual-economy structure that we have described.<sup>12</sup>

In turn, when the benchmark model is extended to allow for capital as an additional factor of production, comparative advantages at the national level are also determined by factor endowments as in the Heckscher-Ohlin model. In Appendix A we show that, if sector  $X$  is labor intensive, then it is the export-oriented sector if the national labor-land ratio  $n$  is sufficiently large. In the empirical analysis, we use industry labor intensity as a proxy for industry export orientation.

---

<sup>11</sup>From Appendix A we have that the autarky price  $p_A(\ell)$  corresponds to the unique value of  $p(\ell)$  such that  $\omega_X(\ell) = \omega_Y(\ell)$ , where  $\omega_i(\ell)$  is defined in (34).

<sup>12</sup>When  $\alpha_M \neq \alpha_X$ , conditions (2)-(4) of the local equilibrium constitute a small Heckscher-Ohlin economy where the autarky price may in principle depend on factor endowments. However, labor mobility implies that the labor density is higher in places with more land abundance. This offsets the factor proportions effect, turning each local economy into a small Ricardian economy.

### 3 Comparative Advantage and Regional Specialization in China

The central mechanism of the model, from which its welfare and distributional implications stem, is the regional specialization pattern in export-oriented industries: due to the interaction between national comparative advantages and internal trade costs, export-oriented industries locate closer to international gates. This industry-location margin is not present in recent models of trade and internal geography such as Ramondo et al. (2014), Redding (2014), or Allen and Arkolakis (2014).

We now turn to an empirical examination of this mechanism using data from China. We examine whether industries in which China has comparative advantages relative the rest of the world are more likely to locate in places with better access to international markets. We proxy for comparative advantages with national export-output ratios, and also rely on the argument that China’s comparative advantages lie on labor-intensive industries and proxy for comparative advantages using industry labor intensities from the United States.

Several factors make China a particularly suitable country for this exercise. Its rising external trade has been largely driven by its comparative advantages based on factor endowments. Before market-oriented reforms, it has been described as a closed agricultural economy with an economic structure lacking in industrial specialization and agglomeration (Chan et al. 2008). After the reforms, increased industrial output and exports have been fueled by—among other factors—a sustained wave of migration of workers into coastal regions (World Bank 2009).<sup>13</sup>

#### 3.1 Data

Our regional data of Chinese industries is aggregated up from the firm-level *Annual Survey of Industrial Production* conducted by the Chinese government’s National Bureau of Statistics. The *Annual Survey* is a census of private firms with more than 5 million yuan (about \$600,000) in revenue and all state-owned firms. It covers 338 mainland prefectures and 425 manufacturing industries in 4-digit CSIS Chinese classification system between the years 1998-2007.<sup>14</sup> Each industry-prefecture-year cell has information on employment, capital, revenue, value added, and exports by state-owned and private enterprises.<sup>15</sup>

To this data, we add the *distance* of each prefecture to China’s coastline. More precisely, we calculate the Euclidian distance from the administrative center of a prefecture to that of the nearest coastal prefecture. There are 53 coastal prefectures with zero distance and the median is 275 miles. An alternative distance measure using as terminal points those prefectures which host main international seaports generates similar results since these facilities are scattered somewhat

---

<sup>13</sup>China’s Household Registration System, also known as the Hukou system, was relaxed during 1980s as part of market-oriented reforms. The current Hukou system limits the access of migrant workers to certain amenities, such as education and health, that are financed by local taxes. While the Hukou constitutes a cost to labor mobility, regional wage differentials have been sufficiently large to promote large migration flows to coastal provinces. See World Bank (2009).

<sup>14</sup>Prefectures are the second level of the Chinese administrative structure, contained within 31 provinces in mainland China. Due to administrative reclassifications, mostly before 2002, the number of prefectures and industries varies slightly over the years.

<sup>15</sup>The underlying firm-level data has been recently used by Hsieh and Klenow (2009) and Roberts et al. (2011).



uniformly across the Pacific seaboard (see the map in the online appendix). As reported below, an alternative measure based on travel times also generates similar results.

Several characteristics of China’s geography and trade suggest that distance to the coast is a good measure of foreign market access. Maritime transportation is the primary mode of shipping in external trade, while exports over land to bordering countries account for a small share (6.7 percent) of total exports (authors’ calculation using 2006 UN Comtrade data). The share of air shipments in exports to top-20 trade partners is just 17.4 percent (Harrigan and Deng 2010). While inland rivers play an important role in freight transportation in general, their export share is limited. Inland river ports, most of which are also in close proximity to the sea, constitute only 20 percent of port capacity (in terms of tonnage) suitable for international trade (authors’ calculation using data from Chinese Statistical Agency).

To carry out the empirical exercise, we need an industry-level measure of Chinese comparative advantages against the rest of the world. One possibility is to use proxies constructed from actual exports data such as the industry export-revenue ratio. While this measure correlates with the underlying variation in relative productivity in which we are interested, to avoid endogeneity concerns it is desirable to use an industry measure that both proxies for comparative advantages and is exogenous to the Chinese environment. In what follows, we use export-revenue ratios for descriptive purposes, but also incorporate a second industry-specific measure, labor-capital ratio in the US, to proxy for Chinese comparative advantages. In doing so, we draw on the notion that China has a comparative advantage in labor intensive industries.<sup>16</sup>

To construct industry-specific labor-capital ratios in the US, we use the publicly available NBER-CES Manufacturing Industry Database and divide production worker hours by the total real capital stock for each industry.<sup>17</sup> The NAICS 2002 classification—in which the US data is reported—can be mapped to the CSIS Chinese classification by using the official concordance between CSIS and ISIC Rev. 3 as a bridge. This reduces the number of industries to 120, still leaving us with enough variation to conduct our analysis. The US labor-capital ratio has a correlation coefficient of 0.6 with the Chinese labor-capital ratio, and a correlation coefficient of 0.4 with the Chinese export-revenue ratio. Table 1 in the appendix provides further summary statistics.

### 3.2 Preliminary Analysis

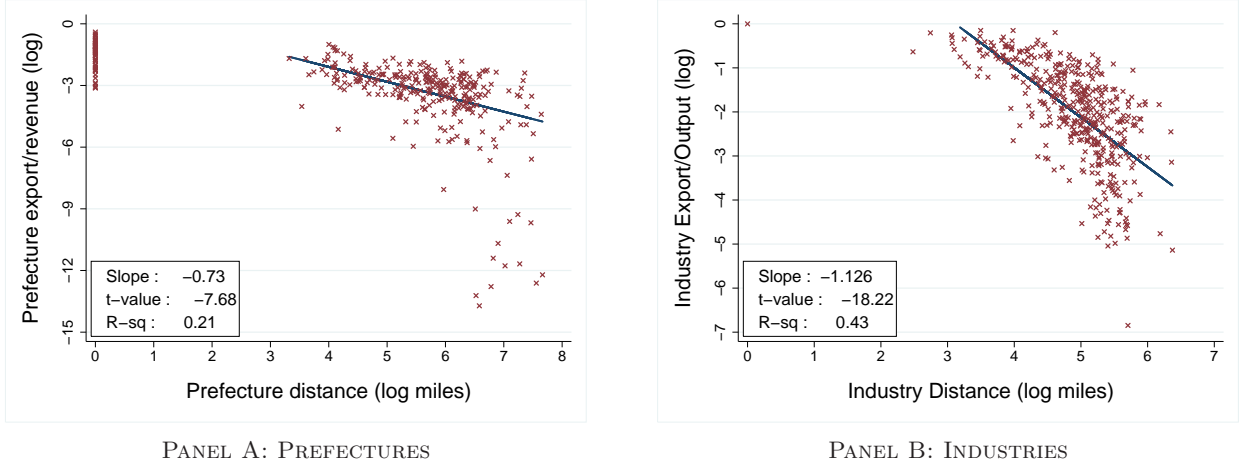
As a preliminary check, we visualize the drop in regional openness toward the interior of the country. We average industry-prefecture level observations over time and calculate export shares in prefectures’ revenues. Denoting industries by  $i$ , prefectures by  $p$ , and time by  $t$ , average regional

---

<sup>16</sup>In Appendix A we develop an extension to the model where capital is also used in production. At the national level, comparative advantages are determined as in the Heckscher-Ohlin model, so that  $X$  is the export-oriented sector if it is labor intensive and the country is sufficiently labor abundant.

<sup>17</sup>With Cobb-Douglas technologies and common factor prices across sectors, industry labor intensities can be measured by labor-capital ratios.

Figure 3: Distance and Export Intensity



*Notes:* In both panels, x-axis is  $\ln(\text{Distance}_p + 1)$ , i.e. we add one to distance before taking the logarithm in order to keep coastal prefectures in the figure. In panel A, each observation is a prefecture, in panel B, each observation is an industry. In panel B, distance refers to average industry distance defined in the text. While all prefectures and industries have positive aggregate revenue, some report zero exports. We thus also add 1 to aggregate exports to include them in figures, but fit the regression lines to observations with strictly positive distance and exports.

export share (in logarithm) at the prefecture level is

$$\text{ExportShare}_p = \ln \left( \frac{\sum_{t=1}^T (\sum_i X_{ipt})}{T} \right) - \ln \left( \frac{\sum_{t=1}^T (\sum_i R_{ipt})}{T} \right),$$

where  $(X_{ipt}, R_{ipt})$  stand for total exports and revenue at the industry-prefecture-year level. Panel A of Figure 3 plots this variable against the natural logarithm of the prefecture-specific distance measure introduced above. As expected, regional openness declines with distance. The elasticity is -0.73 and highly significant.

The model generates this negative distance gradient for regional export shares through intensive and extensive margins. The mechanism of interest, however, is the latter, i.e., the change in industry composition against comparative-advantage industries toward the interior of the country. To preliminary assess whether declining regional export shares reflect this type of industry composition, Panel B of Figure 3 plots industries' export-revenue ratio at the national level (i.e., the share of exports in total industry revenue) against “industry distance” in logarithmic scale. The former variable equals

$$\text{ExportShare}_i = \ln \left( \frac{\sum_{t=1}^T (\sum_p X_{ipt})}{T} \right) - \ln \left( \frac{\sum_{t=1}^T (\sum_p R_{ipt})}{T} \right).$$

Industry distance is defined as the employment-weighted average of prefecture distances:

$$\text{Distance}_i = \sum_p \left( \frac{L_{ip}}{\sum_p L_{ip}} \right) \cdot \text{Distance}_p,$$

where  $\text{Distance}_p$  is prefecture  $p$ 's distance and  $L_{ip}$  is industry  $i$ 's employment in prefecture  $p$

averaged over time. Therefore, the larger  $Distance_i$  is, the farther industry  $i$  is located from the coast. If the gradient from Panel A was exclusively generated by an intensive margin there would not be a definite relationship between the industry export share and the industry distance. However, Panel B shows that, on average, industries with higher export intensity at the national level are situated closer to the seaboard. The distance elasticity is -1.126 and variation in industry distance captures around 43 percent of the variation in industry export intensity.<sup>18</sup>

The following subsection provides a more thorough examination using a rich set of controls and an exogenous measure of industry comparative advantages.

### 3.3 Regression Analysis

Our econometric analysis is based on regressions of the form,

$$Y_{ip} = \beta Z_{ip} + \theta \cdot Distance_p \times ExportOrientation_i + \epsilon_{ip}. \quad (23)$$

Depending on the specification, the dependent variable is a function of employment ( $L_{ip}$ ), capital stock ( $K_{ip}$ ) or export orientation in industry-prefecture cells. The baseline linear specification uses observations with strictly positive employment or capital levels. We also estimate a Poisson pseudo-maximum likelihood regression using all observations, including those with zero activity. In a separate exercise, we investigate the extensive margin of industry presence using linear and nonlinear specifications when the dependent variable is binary, taking the value one if  $L_{ip} > 0$ , and zero otherwise.

On the right hand side, the variable of interest is the interaction between  $Distance_p$ , the prefecture-specific distance measure, and  $ExportOrientation_i$ , an industry-specific export orientation measure. As discussed above, we use two variables here: Chinese export-revenue ratios ( $CNXR_i$ ) and US labor-capital ratios ( $USLK_i$ ).<sup>19</sup>

Our model predicts a negative coefficient  $\theta < 0$ , i.e., export-oriented industries should be more likely to locate closer to the coast.<sup>20</sup> Depending on the specification, the set of other covariates  $Z_{ip}$  includes direct effects of the interacted variables or various fixed effects.

In all the regressions where  $USLK_i$  is used, we only exploit the cross-sectional variation from our data by taking time averages over 1998-2007. The data on export-revenue ratios ( $CNXR_{it}$ ) allows us to investigate our model prediction on regional specialization not only in cross-section but also in the panel dimension. The model predicts that, in a period of trade liberalization, industries that increase export-output ratios increase their presence closer to coastlines. In addition

---

<sup>18</sup>We also plotted the non-parametric fit of log employment over log distance separately for industries with above or below median labor intensity. As expected from the theory and consistently with other empirical results, labor intensive industries have a steeper gradient. The figure is presented in the online appendix.

<sup>19</sup>As an alternative measure of export orientation, we also use the Balassa (1965) revealed comparative advantage ( $CNRCA$ ) index. The correlation between  $CNRCA$  and  $CNXR$  is around 0.7, and the results are both qualitatively and quantitatively very similar. RCA based results are available from the authors upon request.

<sup>20</sup>In the model, the export-revenue ratio at the national level is positive for the comparative-advantage industry and zero for the import-competing industry. Adding product differentiation to the model would generate positive export-revenue ratios for both industries.

to the cross-sectional specifications described above, we estimate a separate panel regression with industry-specific fixed effects using the export-revenue ratios.

Table 2 reports the baseline cross-sectional results using time averages for all variables when the dependent variable is the natural logarithm of employment (Columns 1-2), capital (Columns 3-4) or the level of export-output ratio (Columns 5-6), using nonzero observations. Columns 2, 4 and 6 feature industry and prefecture fixed effects. Panel A uses export-revenue ratios and Panel B uses industry capital-labor ratios as proxy for industry comparative advantages. Standard errors are clustered at the prefecture level. As expected, regional employment, capital stock and export orientation decline with distance, reflecting the gradient of population density and economic activity in China. The coefficient of interest on the interaction term is negative and highly significant in all specifications.

To give a sense of the economic importance of the results, we use the interaction coefficients equal to -0.288 and -2.9 reported in the second column of panel A and B, respectively. Consider an industry with an average export-revenue ratio of 0.21. Moving inland by 275 miles, which is the median distance from the coast across prefectures, decreases regional employment by 17 percent for such an industry. The same decline is 32 percent for an industry with an export-revenue ratio one standard deviation (equal to 0.19) higher than the average. These magnitudes are 13 and 21 percent for the average and high labor-intensity industries, respectively. This is a fairly sizeable impact on regional specialization.<sup>21</sup>

Given the high level of spatial and industry disaggregation in our data, it is natural that not all industries will be observed in every prefecture. Approximately 58 percent of all prefecture-industry combinations in our data report zero employment. To exploit this extensive margin variation, we define the dependent variable in the estimation equation (23) as a binary variable that takes the value one if industry  $i$  has positive employment in prefecture  $p$  at time  $t$ , and zero otherwise.<sup>22</sup> Using the same explanatory variables, we estimate the probability of industry presence under various specifications and report the results in Table 3. The results from a probit model are reported in column 1. To avoid the incidental parameter problem, we introduce fixed effects through conditional logit and linear probability models in columns 2-4. Except for the conditional logit models using *USLK* as the export orientation measure, the interaction coefficient is negative and highly significant, indicating an increasing probability of regional presence for export oriented industries as one gets closer to the ocean. The linear probability model in the fourth column implies that the probability of observing an industry with an export-revenue ratio that is one standard deviation higher than the average increases by around one percent in response to a one percent decrease in distance.

Our next specification, reported in Table 4, identifies the effect within industry-prefectures over

---

<sup>21</sup>As Table 1 shows, mean export-revenue ratio across all years is 0.21 with a standard deviation of 0.19. Since *distance* is measured in 100 miles, we multiply  $-0.288 \cdot 0.40 \cdot 2.75 = -0.317$ .

<sup>22</sup>Out of 147,030 observations, there are only 509 cells with zero employment and a strictly positive capital stock, and 104 cases with the reverse pattern. Thus, using an alternative definition of industry presence as having positive levels of both employment and capital doesn't have a noticeable effect on results.

time using the Chinese export-revenue ratio as the trade measure. Instead of separate industry and prefecture fixed effects, now we use industry-prefecture fixed effects in order to control for unobserved, time-invariant factors that affect economic outcomes at the industry-prefecture level. In columns 2 and 4, we also include industry-year and prefecture-year fixed effects. The interaction coefficient is still significantly negative, implying that the previously estimated effects are not solely due to cross-sectional variation: employment (or capital stock) in industries that have become more export-oriented at the national level during the data period increased relatively more in prefectures close to China’s seaboard.

### 3.4 Alternative Mechanisms

We now examine alternative mechanisms that could also generate this empirical pattern. First, our approach to explaining regional specialization complements explanations based on location fundamentals such as Courant and Deardorff (1992) and on external economies of scale as in Rossi-Hansberg (2005). Following these views, it could be that, closer to international gates, either endowments or externalities drive high relative productivity in specific industries which then become export oriented.

In light of these forces, we should observe that relative productivity in export-oriented industries decreases as we move away from international gates. To inspect if this pattern is present in China we replace the dependent variable in the estimation equation (23) with measures of local industry productivity. We define value added per worker and per unit of capital as

$$VL_{ip} = \frac{ValueAdded_{ip}}{L_{ip}}, \quad VK_{ip} = \frac{ValueAdded_{ip}}{K_{ip}},$$

and regress the natural logarithms of these variables on the interaction variable of interest and fixed effects. The results are in Table 5. We expect a negative interaction if these alternative explanations are driving the regional specialization patterns that we find.

While there is a negative gradient for labor productivity when *CNXR* is used to proxy export orientation (first column of panel A), capital productivity increases toward coastal prefectures for industries with higher export orientation using either measure (second column of panel A and B). However, this is not robust to excluding state-owned enterprises while constructing the regional data. Doing this renders the interaction coefficient on capital productivity significant at 10 percent only (fourth column, both panels), pointing toward the existence of unproductive, capital-intensive, state-owned enterprises located in interior prefectures, operating in industries with low export orientation. These capital-intensive industries are not major exporters: the correlation between the industry-wide capital-labor ratio (in China) and industry-wide export-revenue ratio is -0.3. Therefore, interior state-owned enterprises in capital-intensive industries create the impression that coastal prefectures have higher value added per unit of capital in export-oriented industries. A likely explanation for this pattern lies in Chinese regional planning strategy in the pre-reform era, which located heavy industries in interior provinces for national security purposes (see Gao

2004). Hence, while we find some association between the market-access gradient for the location of export-oriented industries and their productivity gradient, the relationship is weak once state-owned enterprises are excluded.

Second, our approach also complements explanations within the New Economic Geography tradition of Krugman (1991). In this view, industries with larger scale economies are more likely to locate in regions with high population density or close to gateways to large markets. Hanson and Xiang (2004) show that this mechanism is empirically relevant in international trade, i.e., more differentiated industries concentrate in large countries. Therefore, the specialization pattern that we find could result from industries with larger returns to scale situating closer to the seaboard, where many of the largest Chinese cities are historically located. Given that location, it is possible that these industries in turn become exporters to the rest of the world. Also, such industries are likely to have higher intra-industry trade, which would make the export-revenue measure less precise in capturing underlying comparative advantages.

To control for this channel, we follow Hanson and Xiang (2004) and use industry elasticities of substitution as a measure of scale economies.<sup>23</sup> We use the estimates of trade elasticities by Broda and Weinstein (2006) and let  $EOS_i$  denote the elasticity of substitution for sector  $i$ . If home-market effects are the only force that matters for the location of export-oriented industries at lower distances,  $EOS_i$  would be expected to pick up the distance gradient estimated before. We thus re-estimate our baseline relationship in equation (23) by including the interaction of  $EOS_i$  with distance and report the results in Table 6.<sup>24</sup> We also include controls for industry-specific measures of trade costs to control for the possibility that sectors with high trade-costs also locate closer large markets. In particular, we use two measures of trade costs: the weight/value ratio and the ratio of CIF charges to the value of imports calculated from the Census US Imports of Merchandise data.

The interaction between  $EOS_i$  and distance is positive and typically significant, implying that more differentiated industries (lower  $EOS_i$ ) locate relatively closer to seaboards. This confirms that home market effects based on product differentiation are an important determinant of industry location in China. However, the main interaction term between distance and industry export orientation remains stable with a negative sign and statistically significant.<sup>25</sup>

---

<sup>23</sup>In Krugman (1980), higher product differentiation, captured by a lower elasticity of substitution, leads to larger revenues per worker and accentuates the returns to scale. Clearly, there may be other industry-specific characteristics related to returns to scale, such as product differentiation on the inputs side.

<sup>24</sup>As expected, industry-level elasticities are negatively correlated with export orientation. A log-log regression of  $EOS_i$  on export-revenue ratio gives a statistically significant elasticity of  $-0.96$  with an  $R^2$  of 0.36. In order to use Broda and Weinstein elasticities, we concord Chinese industry classification CSIS to 3-digit SITC industries using as a bridge the official concordance between CSIS and ISIC Revision 3. This procedure reduces the number of observations since the mappings are not one-to-one.

<sup>25</sup>We also checked the robustness of results to using a Poisson pseudo-maximum likelihood estimator that allows us to include observations with zero employment in the estimation, to excluding autonomous interior provinces, coastal prefectures or state-owned enterprises, and to using as market access measures the travel time to the nearest international port. All results go through and additional insights into China's economic geography arise. These results are presented in the online appendix.

## 4 Conclusion

We developed a theory that combines a neoclassical international trade structure with an internal geography. In the model, international trade creates a partition between a commercially integrated coastal region with high population density and an interior region where immobile factors are poorer. Reductions in domestic or international trade costs generate migration to the coastal region and net welfare losses for immobile factors in the interior region. Qualitatively, these outcomes are consistent with the spatial structure of economic activity in large developing countries such as China or India where coastal regions integrated with the rest of the world coexist with remote and commercially isolated corners within the country.

A key feature of the model is that comparative-advantage industries at the national level locate closer to international gates. We find evidence consistent with this pattern in industry-level data from Chinese prefectures. Previous work in economic geography and international trade highlights the importance of relative productivities and scale economies for industry location, but these channels do not fully account for the patterns that we find. A full quantitative evaluation of the relative importance of the forces in our model relative to other mechanisms in shaping industry location is outside the scope of this paper and would be an interesting question for future research.

## References

- Allen, T. and Arkolakis C. (2014). Trade and the Topography of the Spatial Economy. *Quarterly Journal of Economics*. 129(3), 1085-1140.
- Amiti, M. and Javorcik, B. (2008). Trade Costs and Location of Foreign Firms in China. *Journal of Development Economics* 85.1.
- Asian Development Bank and Asian Development Bank Institute (2009). *Infrastructure for a Seamless Asia*. Tokyo: Asian Development Bank Institute.
- Atkin, D. and Donaldson D. (2014). Who's Getting Globalized? The Size and Nature of Intranational Trade Costs. Mimeo.
- Balassa, B. (1965). Trade Liberalisation and Revealed Comparative Advantage. *The Manchester School*, 33(2), 99-123.
- Becker, R., Gray W. and Marvakov J. (2013). NBER-CES Manufacturing Industry Database
- Behrens, K., Gaigne C., Ottaviano G., and Thisse JF. (2006). Is Remoteness a Locational Disadvantage? *Journal of Economic Geography*. 6, pp. 347-368.
- Bond, E. (1993). Trade, Factor Mobility, and Income Distribution in a Regional Model with Compensating Wage Differential. *Regional Science and Urban Economics*. 23, pp. 67-84.
- Broda, C., and Weinstein, D. E. (2006). Globalization and the Gains from Variety. *The Quarterly Journal of Economics*, 121(2), 541-585.

- Chan, K. W., Henderson, J. V., and Tsui, K. Y. (2008). Spatial Dimensions of Chinese Economic Development. In “China’s Great Economic Transformation.” Cambridge University Press.
- Courant, P. N., and Deardorff, A. V. (1992). International trade with lumpy countries. *Journal of Political Economy*, 198-210.
- Eaton, J. and Kortum S. (2003), Technology, Geography, and Trade. *Econometrica*. 70:5, pp. 1741-1775.
- Gao, T. (2004). Regional Industrial Growth: Evidence from Chinese Industries. *Regional Science and Urban Economics*, 34(1), 101-124.
- Fujita, M. and Thisse JF. (2013). *Economics of Agglomeration: Cities, Industrial Location, and Globalization*. Cambridge University Press.
- Hanson, G. (1996). Economic Integration, Intraindustry Trade, and Frontier Regions. *European Economic Review*. 40, pp. 941-949.
- Hanson, G. (1998). North American Economic Integration and Industry Location. *Oxford Review of Economic Policy*. 14:2, pp.30-44
- Hanson, G. H., and Xiang, C. (2004). The Home-Market Effect and Bilateral Trade Patterns. *The American Economic Review*, 94(4), 1108-1129.
- Harrigan, J., and Deng, H. (2010). China’s Local Comparative Advantage. In “China’s Growing Role in World Trade.” University of Chicago Press.
- Helpman, E. and Krugman P. (1985). *Market Structure and Foreign Trade*. Cambridge: MIT Press.
- Henderson, V. (1982). Systems of Cities in Closed and Open Economies. *Regional Science and Urban Economics*. 12, pp. 325-350.
- Holmes, T.J. and Stevens J. (2004). Spatial Distribution of Economic Activities in North America. *Handbook of Urban and Regional Economics*,. vol. 4, pp. 2797-2843.
- Hsieh, C. T., and Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, 124(4), 1403-1448.
- Krugman, P. (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *The American Economic Review*, 70(5), 950-959.
- Krugman, P. (1991). Increasing Returns and Economic Geography. *The Journal of Political Economy*, 99(3), 483-499.
- Krugman, P. and Livas-Elizondo R. (1996). Trade Policy and the Third World Metropolis. *Journal of Development Economics*. 49, pp. 137-150.
- Limao, N. and Venables A.J. (2001). Infrastructure, geographical disadvantage, transport costs, and trade. *The World Bank Economic Review* 15.3



- Lu, J. and Tao Z. (2009). Trends and Determinants of China's Industrial Agglomeration. *Journal of Urban Economics*. 65:2.
- Matsuyama, K. (1999). *Geography of the World Economy*, Northwestern University CMS-EMS Discussion Paper No. 1239.
- McCaig, B. and Pavnick N. (2012). *Export Markets and Labor Allocation in a Poor Country*. Mimeo.
- Ramondo, N., Rodriguez-Clare A. and Saborio-Rodriguez M. (2014). *Trade, Domestic Frictions, and Scale Effects?* Mimeo.
- Rauch, J. (1991). Comparative Advantage, Geographic Advantage and the Volume of Trade. *The Economic Journal*. 101, pp. 1230-1244.
- Redding, S. (2014). *Goods Trade, Factor Mobility and Welfare*, NBER WP 18008.
- Roberts, M. J., Xu, D. Y., Fan, X., and Zhang, S. (2012). *A Structural Model of Demand, Cost, and Export Market Selection for Chinese Footwear Producers*. NBER WP 17725.
- Rossi-Hansberg, E. (2005). A Spatial Theory of Trade. *American Economic Review*, 95(5), 1464-1491.
- Starrett, D. (1978). Market Allocations of Location Choice in a Model with Free Mobility. *Journal of Economic Theory*. 17, pp.21-37.
- Venables, A. and Limao N. (2002). Geographical Disadvantage: a Heckscher–Ohlin–von Thunen Model of International Specialization. *Journal of International Economics*. 58.
- World Bank (2009), *World Development Report: Reshaping Economic Geography*. Washington, DC: World Bank.

# APPENDIX

## A1: Derivation of Equilibrium Outcomes

We characterize local-equilibrium outcomes in a model that, in addition to the ingredients in the baseline model, also allows for a nontradable sector (with share  $1 - \beta_T$  in consumption), location-specific amenities  $m(\ell)$ , capital as an additional factor of production (with share  $1 - \gamma_i$  in production per unit of land in sector  $i$ ), and differences in land intensity  $\alpha_i$  across sectors. The model in the text corresponds to the case when  $\beta_T = \gamma_i = 1$ , for all  $i$ ,  $m(\ell) = 1$  for all  $\ell$ , and  $\alpha_X = \alpha_M$ .

Utility of an agent who lives in  $\ell$  is  $m(\ell) C_T(\ell)^{\beta_T} C_N(\ell)^{1-\beta_T}$ , where  $C_T(\ell)$  is a consumption basket of tradables that includes  $X$  and  $M$ ,  $C_N(\ell)$  is consumption of nontradables, and  $m(\ell)$  is the level of amenities at  $\ell$ . (The preferences in equation (1) in the text are the special case when  $\beta_T = m(\ell) = 1$ .)

The land endowment  $\lambda(\ell)$  in each location is owned by immobile landowners that consume where they live. There are two types of mobile agents: workers (who own only labor) and capital-owners (who own only capital).

Indirect utility of a worker who lives in  $\ell$  is  $u(\ell) = m(\ell) \frac{w(\ell)}{E(\ell)}$ , where  $w(\ell)$  is the wage at  $\ell$ , and  $E(\ell)$  is the cost-of-living index, which includes the (homothetic) price index for tradables  $E_T(\ell)$  and the price of nontradables  $P_N(\ell)$ :

$$E(\ell) = E_T(P_X(\ell), P_M(\ell))^{\beta_T} P_N(\ell)^{1-\beta_T}. \quad (24)$$

We let  $p(\ell) \equiv P_X(\ell)/P_M(\ell)$  be the relative price in sector  $X$ , so that there exists an increasing and concave function  $e(p(\ell))$  satisfying  $E_T(\ell) = P_M(\ell)e(p(\ell))$ . The (endogenous) real wage, common to all locations in equilibrium, is  $u^*$ .

Indirect utility of a capital-owner is  $m(\ell) \frac{r_k(\ell)}{E(\ell)}$ , where  $r_k(\ell)$  is the return to each unit of capital. Capital owners are mobile both across locations within the country and between the country and the rest of the world, where they can obtain an outside option  $r_k^*$ .

Demand for nontradables per unit of land in  $\ell$  is  $c_N(\ell) = (1 - \beta_T) y(\ell) / P_N(\ell)$ , where  $y(\ell)$  is income per unit of land at  $\ell$ . Income in each location is generated by land, capital, and labor, so that  $y(\ell) = [w(\ell) n(\ell) + r_k(\ell) k(\ell) + r(\ell)] \lambda(\ell)$ , where  $k(\ell)$  is the endogenous capital density.

The technology to produce output per unit of land is

$$q_i(\ell) = \frac{\kappa_i}{a_i(\ell)} \left( \left( \frac{n_i(\ell)}{\gamma_i} \right)^{\gamma_i} \left( \frac{k_i(\ell)}{1 - \gamma_i} \right)^{1-\gamma_i} \right)^{1-\alpha_i},$$

for  $i = N, M, X$ , where  $\kappa_i = (1 - \alpha_i)^{-(1-\alpha_i)} \alpha_i^{-\alpha_i}$ . Land intensity in production is  $\alpha_i$  and labor intensity per unit of land is  $\gamma_i$  (the production technologies defined in (3) in the text are the special case when  $\gamma_i \rightarrow 1$ ).

The local equilibrium definition is similar to Definition 1, with the following additions: the nontradable market clears, the market-clearing conditions include land and labor used in the

nontradable sector, there is an extra market clearing condition for capital, and real returns to capital are equalized throughout the country. Specifically, a local equilibrium at  $\ell$  consists of population density  $n(\ell)$ , capital density  $k(\ell)$ , labor and capital demands  $\{n_i(\ell), k_i(\ell)\}_{i=X,M}$ , patterns of land use  $\{\lambda_i(\ell)\}_{i=X,M}$ , and factor prices  $\{w(\ell), r(\ell), r_k(\ell)\}$  such that

1. workers maximize utility,

$$u(\ell) \leq u^*, \quad = \text{ if } n(\ell) > 0; \quad (25)$$

2. returns to capital are equalized

$$m(\ell) \frac{r_k(\ell)}{E(\ell)} \leq r_k^*, \quad = \text{ if } k(\ell) > 0; \quad (26)$$

3. profits are maximized,

$$\pi_i(\ell) \leq 0, \quad = \text{ if } \lambda_i(\ell) > 0, \text{ for } i = X, M, N \quad (27)$$

where  $\pi_i(\ell)$  is given by (2);

4. land, labor, and capital markets clear,

$$\sum_{i=X,M,N} \lambda_i(\ell) = \lambda(\ell), \quad (28)$$

$$\sum_{i=X,M,N} \frac{\lambda_i(\ell)}{\lambda(\ell)} n_i(\ell) = n(\ell), \quad (29)$$

$$\sum_{i=X,M,N} \frac{\lambda_i(\ell)}{\lambda(\ell)} k_i(\ell) = k(\ell), \quad (30)$$

5. non-traded sector clears, and

6. trade is balanced.

The profit maximization of landowners in  $\ell$  is

$$\pi_i(\ell) = \max_{n_i(\ell), k_i(\ell)} \{P_i(\ell) q_i [n_i(\ell)] - w(\ell) n_i(\ell) - r_k(\ell) k_i(\ell) - r(\ell)\}. \quad (31)$$

Defining  $\omega(\ell) \equiv w(\ell)/r(\ell)$  as the wage-to-land rental ratio at  $\ell$  and using the indifference conditions (25) and (26), the profit maximization condition (27) can be written:

$$P_i(\ell) \leq a_i(\ell) \frac{w(\ell)}{\omega(\ell)^{\alpha_i}} \left( \frac{r_k^*}{u^*} \right)^{(1-\gamma_i)(1-\alpha_i)}, \quad = \text{ if } \lambda_i(\ell) > 0 \quad (32)$$

for  $i = X, M, N$ . In every populated location, (32) binds for  $i = N$  if  $n(\ell) > 0$ . Using the definition of  $u(\ell)$  together with the price for nontradables  $P_N(\ell)$  from (32) gives the wage at  $\ell$  that leaves

workers indifferent:

$$w(\ell) = E_T(\ell) \left( \frac{u^*}{m(\ell)} \right)^{\frac{1}{\beta_T}} \left( \frac{a_N(\ell)}{\omega(\ell)^{\alpha_N}} \right)^{\frac{1-\beta_T}{\beta_T}} \left( \frac{r_k^*}{u^*} \right)^{(1-\alpha_N)(1-\alpha_N)\frac{1-\beta_T}{\beta_T}}. \quad (33)$$

Using (32), (33), and the definition of the price index (24) we solve for the wage-rental ratio in a location that produces in tradable sector  $i \in \{X, M\}$ ,

$$\omega_i(\ell) = \left[ a_i(\ell)^{\beta_T} a_N(\ell)^{1-\beta_T} \left( \frac{E_T(\ell)}{P_i(\ell)} \right)^{\beta_T} \frac{(u^*)^{1-\delta_i} (r_k^*)^{\delta_i}}{m(\ell)} \right]^{\frac{1}{\beta_T \alpha_i + (1-\beta_T) \alpha_N}}, \quad (34)$$

where  $\delta_i \equiv (1 - \alpha_N)(1 - \gamma_N)(1 - \beta_T) + (1 - \alpha_i)(1 - \gamma_i)\beta_T$ .

### Specialization Patterns

From (32), if  $\lambda_i(\ell) > 0$  then  $\omega_i(\ell) \leq \omega_j(\ell)$  for  $i, j = X, M$  and  $i \neq j$ . Also, note that  $\omega_X(\ell)$  is increasing in  $p(\ell)$  and  $\omega_M(\ell)$  is strictly decreasing in  $p(\ell)$ , so that the wage-rental ratio increases with the relative price of the good exported by the location. Therefore, the local economy is fully specialized in  $X$  when  $p(\ell) > p_A(\ell)$  and fully specialized in  $M$  when  $p(\ell) < p_A(\ell)$ , where  $p_A(\ell)$  is the unique value of  $p(\ell)$  such that  $\omega_X(\ell) = \omega_Y(\ell)$ . As mentioned in section (2.4), this implies that if  $\beta_T = m(\ell) = 1$  but  $\alpha_X \neq \alpha_M$ , then  $p_A(\ell)$  is independent from  $\ell$  if and only if  $a_X(\ell)^{\alpha_M} / a_M(\ell)^{\alpha_X}$  is independent from  $\ell$ .

Since the local economy is incompletely specialized in autarky,  $p_A(\ell)$  must be the autarky price. When  $\alpha_X = \alpha_M$  (so that both tradable sectors use capital with the same intensity) then:

$$p_A(\ell) = \left( \frac{u^*}{r_k^*} \right)^{\frac{\delta_M - \delta_X}{\beta_T}} \frac{a_X(\ell)}{a_M(\ell)}.$$

In this case, we have as in the main text that  $p_A(\ell)$  is the same across locations when  $\frac{a_X(\ell)}{a_M(\ell)} = a$ . In addition, comparative advantages at the national level are also determined by factor endowments as in the Heckscher-Ohlin model. Assuming that sector  $X$  is labor intensive ( $\gamma_X > \gamma_M$ , which implies  $\delta_M > \delta_X$ ) and the national labor-land ratio  $n$  is sufficiently large (so that  $u^*$ , and therefore the relative autarky price of  $X$ ,  $p_A$ , is small enough) implies that  $X$  is the export-oriented sector.

### Employment Density

To solve for population density  $n(\ell)$ , we combine the labor and capital demand derived from (31) with the market clearing conditions (28) to (30), the labor and capital mobility conditions (25) and (26), and the market clearing condition in the non-traded sector. When a local economy is fully specialized in the production of  $i \in \{X, M\}$  and  $N$ , we can solve this system to obtain:

$$n(\ell) = \frac{1 - \zeta_{1i} - \zeta_{2i}}{\zeta_{1i}} \frac{1}{\omega_i(\ell)}, \quad (35)$$

where  $\omega_i(\ell)$  is given in (34), and where  $\{\zeta_{1i}, \zeta_{2i}\}$  are function of the parameters.<sup>26</sup> This condition is the generalized version of equation (10) in the text. When capital is not used in production ( $\gamma_i = 1$  for  $i = X, M, N$ ) then labor density is:

$$n(\ell) = \frac{1 - \bar{\alpha}_i}{\bar{\alpha}_i} \left[ \frac{z_i(\ell)}{u^*} \left( \frac{E_T(\ell)}{P_i(\ell)} \right)^{\beta_T} \right]^{1/\bar{\alpha}_i} \quad \text{for } i = X, M \quad (36)$$

where  $\bar{\alpha}_i \equiv \beta_T \alpha_i + (1 - \beta_T) \alpha_N$  and  $z_i(\ell) \equiv m(\ell) / \left( a_i(\ell)^{\beta_T} a_N(\ell)^{1-\beta_T} \right)$ .

## A2: Proofs

The following results are shown under the assumption that  $\alpha_X = \alpha_M = \alpha$  so that  $\bar{\alpha}_i = \bar{\alpha}$  in the model from Appendix A1.

**Proposition 1** For (i), we have that if the country is a net exporter of  $X$ , all locations that trade with RoW must produce  $X$ . Condition (13) then implies that  $e^{-2(\tau_0 + \tau_1 \ell)} p^* \leq p(\ell)$ . Therefore, all locations such that  $e^{-2(\tau_0 + \tau_1 \ell)} p^* > p_A$  cannot be in autarky and must specialize in  $X$ . In turn, since the country is not fully specialized, there must be autarkic locations for which  $e^{-2(\tau_0 + \tau_1 \ell)} p^* < p_A$ . Therefore, there must exist  $b < \bar{\ell}$  such that  $e^{-2(\tau_0 + \tau_1 b)} p^* = p_A$ . For (ii), we have that, if  $p_A/p^* < e^{-2\tau_0}$  but the country is in international autarky or exports  $M$  then the no-arbitrage conditions (12) and (13) are violated. In that case, equilibrium condition (17) implies that  $b < \bar{\ell} \iff p_A/p^* < e^{2(\tau_0 + \tau_1 \bar{\ell})}$ . Similar reasoning applies when the country exports  $M$ . Finally, if  $e^{-2\tau_0} < p_A/p^* < e^{2\tau_0}$  but the country exports  $X$  or  $M$  then the no-arbitrage conditions (12) and (13) are violated.

**Proposition 2** From (17), it follows that  $b$  is decreasing with  $\tau_1$  or  $\tau_0$ . Therefore,  $b' > b$  when either  $\tau_1$  or  $\tau_0$  decreases. Consider the case when the economy is a net exporter of  $X$  and let  $p(\ell)'$  and  $n(\ell)'$  be the relative price and density in the new equilibrium with  $\tau_0' < \tau_0$  or  $\tau_1' < \tau_1$ . Note that  $\frac{\partial n(\ell)'}{\partial \ell} < 0$  and  $\left| \frac{\partial n(\ell)'}{\partial \ell} \frac{1}{n(\ell)'} \right| \leq \left| \frac{\partial n(\ell)}{\partial \ell} \frac{1}{n(\ell)} \right|$  for all  $\ell \in [0, b)$ . Therefore,  $n(b)' > n(b)$ , for otherwise  $\int_0^{\bar{\ell}} n(\ell)' \lambda(\ell) d\ell < N$ . Also note that  $p(\ell) = p(\ell)' = a$  for all  $\ell \in (b, \bar{\ell}]$  and that  $u^*$  is higher in the new equilibrium. Since  $n(b)' < n(b)$ , there must be  $c \in [b, b']$  such that  $n(c)' = n(b)$ . We conclude that  $n(\ell)' - n(\ell) > 0 \iff \ell < c$ , so that population density increases in  $[0, c]$ . When  $\tau_0$  shrinks,

<sup>26</sup>These coefficients are given by:

$$\zeta_{1i} = \alpha_N (1 - \beta_T) + \alpha_i \frac{1 - (1 - \beta_T) \left[ \alpha_N + (1 - \alpha_N) \left( \frac{\gamma_N}{1 - \gamma_N} \right)^{1 - \gamma_N} + (1 - \alpha_i) \left( \frac{\gamma_N}{1 - \gamma_N} \right)^{-\gamma_N} \right]}{\alpha_i + (1 - \alpha_i) \gamma_i^{-\gamma_i} (1 - \gamma_i)^{-(1 - \gamma_i)}},$$

$$\zeta_{2i} = (1 - \alpha_i) \left\{ (1 - \beta_T) \left( \frac{\gamma_N}{1 - \gamma_N} \right)^{-\gamma_N} + \left( \frac{\gamma_i}{1 - \gamma_i} \right)^{-\gamma_i} \frac{1 - (1 - \beta_T) \left[ \alpha_N + (1 - \alpha_N) \left( \frac{\gamma_N}{1 - \gamma_N} \right)^{1 - \gamma_N} + (1 - \alpha_i) \left( \frac{\gamma_N}{1 - \gamma_N} \right)^{-\gamma_N} \right]}{\alpha_i + (1 - \alpha_i) \gamma_i^{-\gamma_i} (1 - \gamma_i)^{-(1 - \gamma_i)}} \right\}.$$

$n(0)' > n(0)$ , for otherwise  $\int_0^{\bar{\ell}} n(\ell)' \lambda(\ell) d\ell < N$ . In contrast when  $\tau_1$  shrinks then  $n(0)' < n(0)$  since  $u^*$  is higher in the new equilibrium, but  $p(0) = p(0)'$ .

**Proposition 3** The real return to immobile factors at  $\ell$  is  $v(\ell) = r(\ell)/E(\ell)$ . Using (36) and  $u(\ell) = u^*$ , we can write

$$v(\ell) = \frac{\bar{\alpha}}{1 - \bar{\alpha}} n(\ell) u^*,$$

so that the real return to land at the national level,  $\int_0^{\bar{\ell}} v(\ell) \lambda(\ell) d\ell$ , equals

$$v = \frac{\bar{\alpha}}{1 - \bar{\alpha}} N u^*. \quad (37)$$

In turn, using (10), the percentage change in population density in each location in response to an arbitrary change in relative prices or in the real wage is

$$\hat{n}(\ell) = \frac{\beta_T}{\alpha} [1 - \varepsilon_T(p(\ell))] \hat{p}(\ell) - \frac{1}{\alpha} \hat{u}^*. \quad (38)$$

The percentage change in the real return to immobile factors at  $\ell$  is given by (22) in the text. In turn, the change in relative prices in response to a change in domestic or international trade costs is

$$\hat{p}(\ell) = \begin{cases} -2(d\tau_0 + \ell d\tau_1) & \text{if } \ell < b, \\ 0 & \text{if } \ell > b. \end{cases}$$

Therefore,  $\hat{v}(\ell) < 0$  if  $\ell > b$ . Since  $\hat{v} > 0$  from (37), there must be some  $\hat{v}(\ell) > 0$  for  $\ell < b$ . In turn, when  $d\tau_0 = 0 > d\tau_1$  then  $\hat{p}(0) = 0$  but  $\hat{u}^* > 0$ , implying  $\hat{v}(0) < 0$ . Since  $\hat{v}(\ell) > 0$  for  $\ell < b$ , there is some  $\varepsilon$  such that  $\hat{v}(\ell) < 0$  for all  $\ell \in [0, \varepsilon)$ .

## A3: Tables and Figures

Table 1: **Summary Statistics**

Variable	Mean	Median	St.Dev.
<b>Industry-prefecture level</b>			
Employment, ( $L_{ip}$ )	969	230	3,645
Capital stock, ( $K_{ip}$ ), in 1000 USD	10,591	1,379	65,003
Export-revenue ratio	0.1	0	0.22
<b>Prefecture level</b>			
Employment, ( $L_p$ )	173,783	91,987	261,038
Capital stock, ( $K_p$ ), in 1000 USD	1,899,058	851,308	3,435,081
Export-revenue ratio	0.093	0.058	0.11
Distance (miles)	373	275	402
<b>Industry level</b>			
Employment, ( $L_i$ )	135,032	70,170	234,274
Capital stock, ( $K_i$ ), in 1000 USD	1,475,590	575,106	3,381,592
Export-revenue ratio	0.21	0.15	0.19
US labor-capital ratio	0.016	0.015	0.010

*Notes:* U.S. labor-capital ratio is 1998-2007 average of production worker hours (in millions) divided by the total real capital stock (in million USD) as reported by the NBER-CES Manufacturing Industry Database. All other variables are calculated from the Chinese data. Nominal amounts for the capital stock are converted to USD using the 2005 Yuan-Dollar exchange rate.

Table 2: **Intensive Margin**

	Employment, $\ln(L)$		Capital, $\ln(K)$		Export/Output	
	(1)	(2)	(3)	(4)	(5)	(6)
<b>Panel A: Chinese Export-Revenue Ratio</b>						
<i>Distance</i>	-0.035*** (0.010)		-0.050*** (0.012)		-0.002** (0.001)	
<i>CNXR</i>	0.818*** (0.145)		-0.521*** (0.149)		0.683*** (0.028)	
<i>Distance</i> $\times$ <i>CNXR</i>	-0.293*** (0.037)	-0.288*** (0.037)	-0.283*** (0.041)	-0.216*** (0.036)	-0.086*** (0.009)	-0.073*** (0.007)
Observations	60,505	60,505	60,100	60,100	59,871	59,871
$R^2$	0.033	0.348	0.031	0.342	0.236	0.384
<b>Panel B: US Labor-Capital Ratio</b>						
<i>Distance</i>	-0.089*** (0.013)		-0.100*** (0.018)		-0.005*** (0.001)	
<i>USLK</i>	-10.66*** (1.598)		-37.59*** (1.696)		5.823*** (0.278)	
<i>Distance</i> $\times$ <i>USLK</i>	-2.997*** (0.418)	-2.900*** (0.448)	-2.809*** (0.527)	-2.550*** (0.569)	-0.645*** (0.064)	-0.546*** (0.060)
Observations	26,229	26,229	26,091	26,091	26,024	26,024
$R^2$	0.0796	0.486	0.0983	0.476	0.111	0.394
Industry fixed effects		Yes		Yes		Yes
Prefecture fixed effects		Yes		Yes		Yes

*Notes:* All regressions are estimated with OLS. Dependent variable is  $\ln(L)$  in Columns 1-2,  $\ln(K)$  in Columns 3-4, and export/sales ratio in Columns 5-6. Prefecture-clustered standard errors in parentheses. Significance: \* 10 percent, \*\* 5 percent, \*\*\* 1 percent.



Table 3: Extensive Margin

	(1)	(2)	(3)	(4)
<b>Panel A: Chinese Export-Revenue Ratio</b>				
<i>Distance</i>	-0.029*** (0.003)	-0.035*** (0.001)		
<i>CNXR</i>	-0.175*** (0.026)		-0.247*** (0.029)	
<i>Distance</i> × <i>CNXR</i>	-0.0661*** (0.007)	-0.081*** (0.007)	-0.082*** (0.008)	-0.01*** (0.004)
Observations	147,030	147,030	147,030	147,030
$R^2$ (pseudo- $R^2$ )	0.088	0.111	0.027	0.434
<b>Panel B: US Labor-Capital Ratio</b>				
<i>Distance</i>	-0.034*** (0.003)	-0.047*** (0.003)		
<i>USLK</i>	-3.061*** (0.347)		-5.005*** (0.466)	
<i>Distance</i> × <i>USLK</i>	-0.149** (0.076)	-0.125 (0.186)	-0.118 (0.1)	-0.137** (0.0689)
Observations	39,884	39,884	39,412	39,884
$R^2$ (pseudo- $R^2$ )	0.09	0.13	0.01	0.461
Regression	Probit	Clogit	Clogit	LPM
Industry fixed effects		Yes		Yes
Prefecture fixed effects			Yes	Yes

*Notes:* Dependent variable equals one if there is positive industry employment in a prefecture-year, zero otherwise. Specifications: Column 1 probit, columns 2-3 conditional logit, column 4 linear probability model. Pseudo- $R^2$  values of fit are reported for probit and conditional logit models. Marginal effects reported for probit and conditional logit. Prefecture-clustered standard errors in parentheses. Significance: \* 10 percent, \*\* 5 percent, \*\*\* 1 percent.

Table 4: **Time Variation**

	(1)	(2)	(3)	(4)	(5)
	$\ln(L)$	$\ln(L)$	$\ln(K)$	$\ln(K)$	Extensive Margin
<i>CNXR</i>	0.369*** (0.140)		0.155 (0.128)		-0.005 (0.008)
<i>Distance</i> $\times$ <i>CNXR</i>	-0.206*** (0.04)	-0.145*** (0.034)	-0.176*** (0.041)	-0.151*** (0.04)	-0.005*** (0.001)
Observations	361,580	361,580	362,652	362,652	1,363,315
$R^2$	0.818	0.841	0.800	0.817	0.732
Year fixed effects	Yes		Yes		Yes
Industry-prefecture fixed effects	Yes	Yes	Yes	Yes	Yes
Industry-year fixed effects		Yes		Yes	
Prefecture-year fixed effects		Yes		Yes	

*Notes:* All panel regressions are estimated with year and industry-prefecture fixed effects. Dependent variables are the logarithm of employment  $L$  or capital  $K$  at industry-prefecture-year cells with positive observations in the first four columns. Dependent variable in the fifth column is a binary variable that takes the value one if  $L > 0$  and zero otherwise. Prefecture-clustered standard errors in parentheses. Significance: \* 10 percent, \*\* 5 percent, \*\*\* 1 percent.

Table 5: **Alternative Mechanisms I: Local Productivity**

	(1)	(2)	(3)	(4)
	$\ln(AVL)$	$\ln(AVK)$	$\ln(AVL)$	$\ln(AVK)$
<b>Panel A: Chinese Export-Revenue Ratio</b>				
<i>Distance</i> $\times$ <i>CNXR</i>	-0.014 (0.014)	-0.078*** (0.017)	0.044*** (0.017)	-0.0302* (0.017)
Observations	58,999	58,932	54,171	54,119
$R^2$	0.279	0.183	0.251	0.156
<b>Panel B: US Labor-Capital Ratio</b>				
<i>Distance</i> $\times$ <i>USLK</i>	-0.718** (0.278)	-1.086*** (0.262)	0.0461 (0.238)	-0.541* (0.288)
Observations	25,728	25,705	23,969	23,946
$R^2$	0.339	0.222	0.316	0.194
Industry fixed effects	Yes	Yes	Yes	Yes
Prefecture fixed effects	Yes	Yes	Yes	Yes
Ownership	All	All	Private	Private

*Notes:* All regressions are estimated with OLS. *AVL* and *AVK* stand for *Average Value Added of Labor*, and of *Capital*, respectively. *Distance* is the prefecture-specific distance measure defined in the text. The first two columns use data for all ownership types whereas the last two columns exclude state-owned enterprises. Prefecture-clustered standard errors in parentheses. Significance: \* 10 percent, \*\* 5 percent, \*\*\* 1 percent.

Table 6: **Alternative Mechanisms II: Product Differentiation and Home Market Effects**

	Employment, $\ln(L)$		Capital, $\ln(K)$	
	(1)	(2)	(3)	(4)
<b>Panel A: Chinese Export-Revenue Ratio</b>				
<i>Distance</i> $\times$ <i>CNXR</i>	-0.426*** (0.0507)	-0.409*** (0.0496)	-0.361*** (0.0471)	-0.342*** (0.0465)
<i>Distance</i> $\times$ <i>EOS</i>		0.00131** (0.000595)		0.000972 (0.000700)
<i>Distance</i> $\times$ <i>Weight</i>		0.00380*** (0.00140)		0.00436** (0.00190)
<i>Distance</i> $\times$ <i>CIF</i>		-0.0421 (0.116)		-0.0286 (0.166)
Observations	25,928	25,928	25,792	25,792
$R^2$	0.489	0.489	0.486	0.486
<b>Panel B: US Labor-Capital Ratio</b>				
<i>Distance</i> $\times$ <i>USLK</i>	-3.425*** (0.527)	-3.331*** (0.532)	-2.911*** (0.600)	-2.782*** (0.608)
<i>Distance</i> $\times$ <i>EOS</i>		0.00274*** (0.000680)		0.00213*** (0.000750)
<i>Distance</i> $\times$ <i>Weight</i>		0.000171 (0.00136)		0.00132 (0.00181)
<i>Distance</i> $\times$ <i>CIF</i>		0.480*** (0.122)		0.405** (0.161)
Observations	25,928	25,928	25,792	25,792
$R^2$	0.483	0.484	0.482	0.483
Industry fixed effects	Yes	Yes	Yes	Yes
Prefecture fixed effects	Yes	Yes	Yes	Yes

*Notes:* All regressions are estimated with OLS. *EOS* is industry-specific elasticity of substitution calculated using Broda and Weinstein (2006) trade elasticities. *Weight* is industry-specific weight/value ratio of US imports. *CIF* is the industry-specific ratio of cost-insurance-freight charges to the value of US imports. Prefecture-clustered standard errors in parentheses. Significance: \* 10 percent, \*\* 5 percent, \*\*\* 1 percent.

# Online Appendix for “Internal Geography, International Trade, and Regional Specialization,” Coşar and Fajgelbaum (2014)

## Robustness

To check the robustness of the baseline results, we first replicate the intensive margin analysis with a specification that allows us to include observations with zero employment in the estimation. Santos Silva and Tenreyro (2006) have shown that the Poisson pseudo-maximum likelihood (PPML) estimator is well behaved in datasets with many zeros. Their implementation is motivated by the need to estimate gravity equations without excluding bilateral country pairs with zero trade flows, but the methodology is applicable to a wide range of estimations with nonnegative and continuous dependent variables. Table A1 reports the results. In these specifications, the dependent variable is the absolute level of employment and capital stock at the industry-prefecture level, including observations with a zero. While it is not possible to directly compare the coefficients from the non-linear PPML estimators with our baseline linear estimates in Table 2, we still find a highly significant negative effect of the distance-exports interaction on regional specialization patterns.

In Table A2, we replicate the specification presented in columns 2 and 4 of Table 2 by excluding from the sample 34 prefectures in three large, autonomous interior provinces of China (Inner Mongolia, Tibet and Xinjiang, marked with red triangles in Figure A1), or 53 coastal provinces with zero distance. The three interior regions that we exclude (columns 1 and 3) are among the least industrialized parts of China. Their autonomy may also subject them to a different policy environment. Since many large Chinese cities are historically on the eastern seaboard, excluding coastal prefectures (columns 2 and 4) is a robustness check on the effect of cities. Another concern addressed by this robustness check is that the first special export zones (SEZs) established in the early 1980s as part of China’s gradual reforms were all coastal.<sup>1</sup>

Excluding autonomous interior provinces significantly increases the absolute value of the interaction coefficient. For example, when employment is the dependent variable and US labor intensity is the trade measure, the distance gradient for export-oriented industries increases from 2.9 to 5.25. Similarly, when the dependent variable is capital, the gradient increases from 2.55 to 4.15. These changes suggest that while activity in export-oriented industries shrinks sharply from the coast to the immediate interior of the country, the decline is less pronounced as one reaches the remote interior.

Excluding coastal prefectures from the sample has the opposite effect: the interaction coefficient shrinks in absolute value, indicating the high levels of export orientation in coastal prefectures. In the case of capital, it drops from 0.216 to 0.057 and is only significant at the 10 percent level when Chinese industry-wide export-revenue ratio is the trade measure (4th columns in the top panel of Tables 2 and A2). Still, the coefficient is of the expected sign and highly significant in other cases. Thus, while coastal prefectures may display disproportionately high levels of activity in export-orientated industries, the baseline intensive margin results do not exclusively rely on their presence in the sample.

In the extensive margin, however, the message is different. Table A3 replicates the results of the fourth column in Table 3 (the linear probability model). The first two columns in Table A3 impose the same sample restrictions by excluding the remote interior and the coast. As in the intensive

---

<sup>1</sup>The four SEZs established in 1979 were coastal. An additional 14 port cities in eight coastal provinces were declared SEZs in 1984. The policy experiment, however, quickly spread to the rest of the country. By 1992, more than half of Chinese municipalities had a SEZ with the majority being in the interior (Wang 2013).

margin, the gradient of industry presence gets steeper when interior provinces are dropped, but the coefficient reverses sign and loses significance when coastal prefectures are excluded.

The last two columns of Table A3 check the robustness of the extensive margin results to excluding state-owned enterprises (column 3) and to clustering the standard errors at the more conservative level of provinces.<sup>2</sup> The loss of significance in the case of province-level clustering is consistent with the result obtained when coastal prefectures are excluded: most of the variation in the baseline extensive margin seems to be coming from the steep drop in the probability of observing an industry as one moves away from coastal prefectures.<sup>3</sup> Thus, our robustness checks suggest that immediate proximity to the seaboard and ports seems to be a big advantage in attracting export-oriented industries in China. Conditional on regional presence, however, there is still a significant decline in industry activity beyond the coast.

In Table A4, we assess the robustness of the panel data outcomes to the four alternative specifications tried above. The first two and bottom panels show that the results are robust to excluding state-owned enterprises and autonomous interior provinces from the sample, as well as clustering at the province-level. As in the extensive margin analysis, however, the panel data results are not robust to excluding coastal prefectures. All 3 columns in Panel C indicate the lack of a statistically significant relationship between export-orientation and industry location over time.

Finally, as an alternative measure of foreign market access, we use the natural logarithm of travel time (in hours) to the nearest international port obtained from Google Maps. Table A5 confirms that the interaction term remains negative and significant.

## References

- Cullinane, K., and Wang, T. F. (2006). Port Governance in China. *Research in Transportation Economics*, 17, 331-356.
- Silva, J. S., and Tenreyro, S. (2006). The Log of Gravity. *The Review of Economics and Statistics*, 88(4), 641-658.
- Wang, J. (2013). The Economic Impact of Special Economic Zones: Evidence from Chinese Municipalities. *Journal of Development Economics*. 101, pp. 133-147.

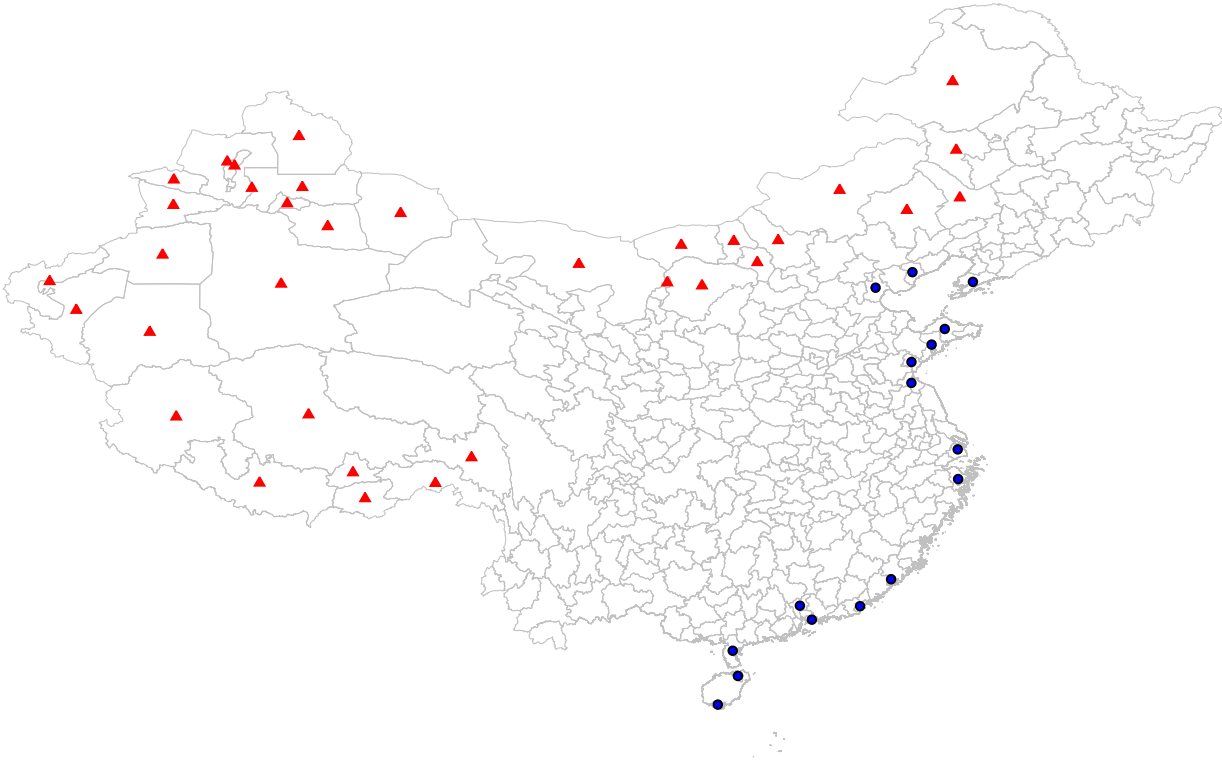
---

<sup>2</sup>The intensive margin results are robust to both of these changes. In order to keep the analysis focused, we do not report them in Table A2, but all additional results are available from the authors upon request.

<sup>3</sup>When estimated with the probit model, the extensive margin coefficient remains significantly negative when we cluster at the province-level. To be conservative, we only report the outcomes that are not robust to alternative specifications.

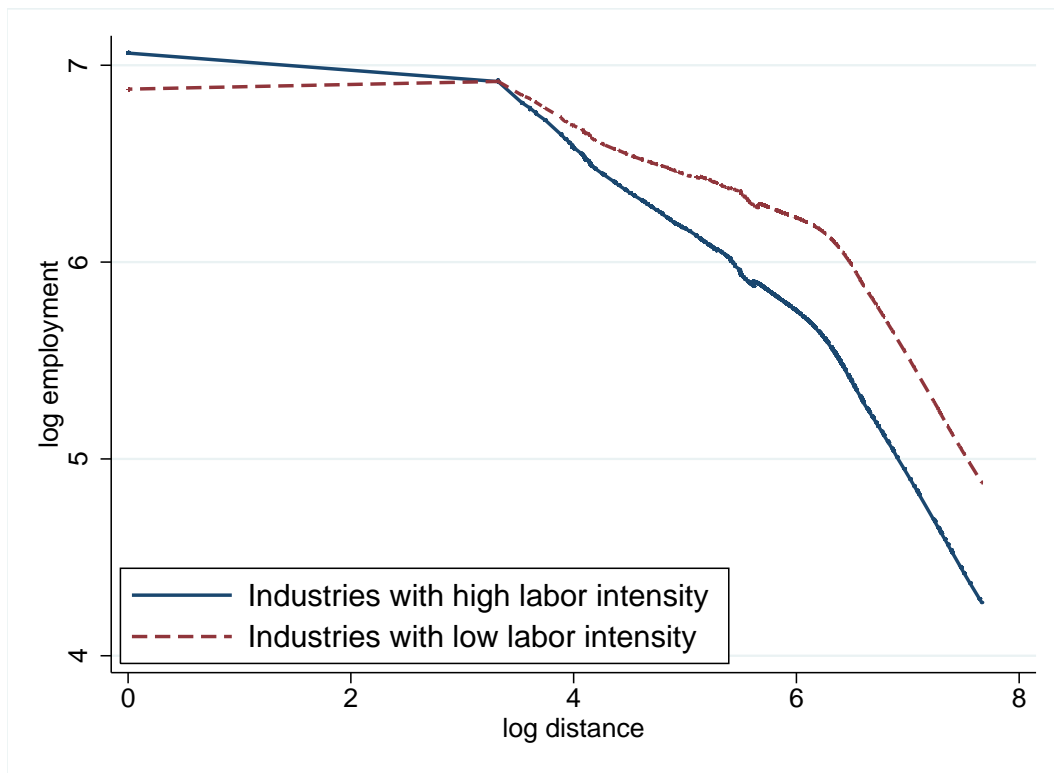
# Tables and Figures

Figure A1: Map of Chinese Prefectures and Ports



*Notes:* Prefecture boundaries. Blue dots are top seaports by cargo throughput in 2005 (Cullinane and Wang 2007). Red triangles are the prefectures of Inner Mongolia, Tibet, and Xinjiang provinces which are excluded from the sample in robustness checks.

Figure A2: Non-parametric Lowess Fit of Employment over Distance



Notes: Non-parametric fit of log employment over log distance for industries with above or below median labor intensity.

Table A1: Robustness: Poisson Pseudo-Maximum Likelihood

	Employment, $L$		Capital, $K$	
	(1)	(2)	(3)	(4)
<b>Panel A: Chinese Export-Revenue Ratio</b>				
<i>Distance</i>	-0.058*** (0.01)		-0.079*** (0.0264)	
<i>CNXR</i>	1.746*** (0.287)		-0.287 (0.361)	
<i>Distance</i> $\times$ <i>CNXR</i>	-1.421*** (0.169)	-1.111*** (0.117)	-1.615*** (0.237)	-0.925*** (0.128)
Observations	147,030	147,030	147,030	147,030
$R^2$	0.022	0.411	0.008	0.353
<b>Panel B: US Labor-Capital Ratio</b>				
<i>Distance</i>	-0.101*** (0.025)		-0.119*** (0.039)	
<i>USLK</i>	7.157** (2.849)		-46.17*** (4.566)	
<i>Distance</i> $\times$ <i>USLK</i>	-14.15*** (1.516)	-11.31*** (1.174)	-14.25*** (2.177)	-8.256*** (1.236)
Observations	39,884	39,884	39,884	39,884
$R^2$	0.037	0.442	0.024	0.450
Industry fixed effects		Yes		Yes
Prefecture fixed effects		Yes		Yes

*Notes:* All regressions are estimated with Poisson Pseudo-Maximum Likelihood. Dependent variables are employment  $L$  or capital  $K$  at industry-prefecture cells.  $R^2$  is computed as the square of the correlation between the dependent variable and its fitted values. Prefecture-clustered standard errors in parentheses. Significance: \* 10 percent, \*\* 5 percent, \*\*\* 1 percent.



Table A2: **Robustness to Sample Restrictions: Intensive Margin**

	Employment, $\ln(L)$		Capital, $\ln(K)$	
	(1)	(2)	(3)	(4)
<b>Panel A: Chinese Export-Revenue Ratio</b>				
<i>Distance</i> $\times$ <i>CNXR</i>	-0.410*** (0.045)	-0.123*** (0.026)	-0.310*** (0.045)	-0.057* (0.029)
Observations	58,157	45,762	57,780	45,387
$R^2$	0.347	0.328	0.343	0.319
<b>Panel B: US Labor-Capital Ratio</b>				
<i>Distance</i> $\times$ <i>USLK</i>	-5.253*** (0.531)	-1.703*** (0.410)	-4.145*** (0.609)	-1.708*** (0.607)
Observations	24,779	20,922	24,661	20,789
$R^2$	0.482	0.471	0.477	0.455
Robustness	No Remote Interior	Non- coastal	No Remote Interior	Non- coastal
Industry fixed effects	Yes	Yes	Yes	Yes
Prefecture fixed effects	Yes	Yes	Yes	Yes

*Notes:* All regressions are estimated with OLS. Dependent variable is  $\ln(L)$  in Columns 1-2, and  $\ln(K)$  in Columns 3-4. Columns 1-3 exclude the 34 prefectures in 3 large interior provinces (Inner Mongolia, Tibet, Xinjiang). Columns 2-4 exclude 53 coastal prefectures with zero distance. Significance: \* 10 percent, \*\* 5 percent, \*\*\* 1 percent.

Table A3: Robustness to Sample Restrictions: Extensive Margin

	(1)	(2)	(3)	(4)
<b>Panel A: Chinese Export-Revenue Ratio</b>				
<i>Distance</i> $\times$ <i>CNXR</i>	-0.033*** (0.007)	0.005 (0.003)	-0.01** (0.004)	-0.01 (0.01)
Observations	132,240	123,975	146,258	147,030
$R^2$	0.428	0.419	0.433	0.434
<b>Panel B: US Labor-Capital Ratio</b>				
<i>Distance</i> $\times$ <i>USLK</i>	-0.551*** (0.127)	0.0341 (0.069)	-0.273*** (0.074)	-0.137 (0.150)
Observations	35,872	33,630	39,766	39,884
$R^2$	0.445	0.465	0.480	0.461
Robustness	No Remote Interior	Non- coastal	Privately Owned	Province Clustered
Industry fixed effects	Yes	Yes	Yes	Yes
Prefecture fixed effects	Yes	Yes	Yes	Yes

*Notes:* All regressions are estimated with LPM. Dependent variable is  $\ln(L)$  in Columns 1-2, and  $\ln(K)$  in Columns 3-4. Column 1 excludes the 34 prefectures in three autonomous interior provinces (Inner Mongolia, Tibet, Xinjiang). Column 2 excludes 53 coastal prefectures with zero distance. Column 3 excludes state-owned enterprises in the construction of the regional data. Column 4 clusters standard errors at the province level. Significance: \* 10 percent, \*\* 5 percent, \*\*\* 1 percent.

Table A4: **Robustness to Sample Restrictions: Time Variation**

	(1)	(2)	(3)
	Employment, $\ln(L)$	Capital, $\ln(K)$	Extensive Margin
<b>Panel A: Private Enterprises Only</b>			
<i>CNXR</i>	0.233** (0.114)	-0.059 (0.119)	-0.026*** (0.007)
<i>Distance</i> $\times$ <i>CNXR</i>	-0.151*** (0.032)	-0.120*** (0.038)	-0.0033*** (0.001)
Observations	313,718	314,829	1,347,304
$R^2$	0.812	0.788	0.719
<b>Panel B: Exclude Autonomous Interior Provinces</b>			
<i>CNXR</i>	0.435*** (0.145)	0.215 (0.132)	0.008 (0.011)
<i>Distance</i> $\times$ <i>CNXR</i>	-0.259*** (0.0502)	-0.217*** (0.051)	-0.012*** (0.003)
Observations	350,977	351,814	1,225,857
$R^2$	0.817	0.800	0.728
<b>Panel C: Exclude Coastal Prefectures</b>			
<i>CNXR</i>	-0.306** (0.151)	-0.523*** (0.152)	-0.0249*** (0.008)
<i>Distance</i> $\times$ <i>CNXR</i>	-0.0473 (0.0360)	-0.0202 (0.0424)	-0.002 (0.001)
Observations	258,216	258,937	1,137,868
$R^2$	0.808	0.790	0.718
<b>Panel D: Clustering at Province Level</b>			
<i>CNXR</i>	0.369 (0.274)	0.155 (0.226)	-0.005 (0.012)
<i>Distance</i> $\times$ <i>CNXR</i>	-0.206** (0.078)	-0.176** (0.064)	-0.005* (0.003)
Observations	361,580	362,652	1,363,315
$R^2$	0.818	0.800	0.732
Year fixed effects	Yes	Yes	Yes
Industry-prefecture fixed effects	Yes	Yes	Yes

*Notes:* All panel regressions are estimated with year and industry-prefecture fixed effects. Dependent variables are the logarithm of employment  $L$  or capital  $K$  at industry-prefecture-year cells with positive observations in the first two columns. Dependent variable in the third column is a binary variable that takes the value one if  $L > 0$  and zero otherwise. Panel A excludes state-owned enterprises in the construction of the regional data. Panel B excludes the prefectures in three autonomous interior provinces (Inner Mongolia, Tibet, Xinjiang). Panel C excludes 53 coastal prefectures with zero distance. Panel D clusters standard errors at the province level. Significance: \* 10 percent, \*\* 5 percent, \*\*\* 1 percent.

Table A5: Robustness to Alternative Market Access Measure

	Employment, $\ln(L)$		Capital, $\ln(K)$	
	(1)	(2)	(3)	(4)
<b>Panel A: Chinese Export-Revenue Ratio</b>				
<i>TTime</i>	-0.013*** (0.003)		-0.017*** (0.004)	
<i>CNXR</i>	0.966*** (0.173)		-0.327* (0.175)	
<i>TTime</i> × <i>CNXR</i>	-0.099*** (0.015)	-0.090*** (0.014)	-0.103*** (0.016)	-0.072*** (0.013)
Observations	60,505	60,505	60,100	60,100
$R^2$	0.035	0.344	0.034	0.341
<b>Panel B: US Labor-Capital Ratio</b>				
<i>TTime</i>	-0.033*** (0.005)		-0.037*** (0.006)	
<i>USLK</i>	-10.96*** (2.047)		-37.65*** (2.245)	
<i>TTime</i> × <i>USLK</i>	-0.800*** (0.184)	-0.873*** (0.193)	-0.763*** (0.228)	-0.793*** (0.227)
Observations	26,229	26,229	26,091	26,091
$R^2$	0.080	0.476	0.100	0.471
Industry fixed effects		Yes		Yes
Prefecture fixed effects		Yes		Yes

*Notes:* All regressions are estimated with OLS. *TTime* is travel time (in hours) to the nearest international port. Dependent variable is  $\ln(L)$  in Columns 1-2, and  $\ln(K)$  in Columns 3-4. Prefecture-clustered standard errors in parentheses. Significance: \* 10 percent, \*\* 5 percent, \*\*\* 1 percent.