Internal Geography, International Trade, and Regional Specialization

A. Kerem Coşar          Pablo Fajgelbaum
Chicago Booth             UCLA

October 6, 2013

Abstract

We introduce an internal geography to the canonical model of international trade driven by comparative advantages to study regional effects of external economic integration. The model features a dual economic 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 became more open. We find support for the theory’s predictions about regional specialization in industry-level data from Chinese prefectures. The data suggest that economies of scale or differences in relative productivity across prefectures do not fully account for the specialization patterns that we find.

JEL Codes: F11, R12
Keywords: trade, internal trade costs, market access, regional specialization

*For their discussions of this paper, we thank Gordon Hanson, Vernon Henderson, and Henry Overman. For their insightful comments, we thank Andrew Atkeson, Ariel Burstein, Andres Rodriguez-Clare, Banu Demir Pakel, Chang-Tai Hsieh, Esteban Rossi-Hansberg, Michael Zheng Song, Ralph Ossa, Stephen Redding and workshop participants at Bilkent University, Central Bank of Turkey, Chicago Booth, Columbia University, DePaul University, Maryland University, Minnesota Fed, Princeton University, Sabanci University, UCLA, UCSD, University of San Andres, University of Pennsylvania, Yale University, 2012 ASSA meeting, 2013 China Economic Summer Institute, 2012 Conference on Urban and Regional Economics, 2012 European Economic Association meeting, 2012 Midwest Trade Conference, 2013 Society of Economic Dynamics meeting. Coşar thanks Chicago Booth for summer financial support.

Correspondence: kerem.cosar@chicagobooth.edu, pfajgelbaum@econ.ucla.edu.
1 Introduction

The experience of many developing countries reveals uneven regional effects of international economic integration. The standard view in international trade theory, which interprets countries as points in space, misses 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 regional effects of external economic integration. We find support for the theory’s predictions about regional specialization in industry-level data from Chinese prefectures. The data suggest that explanations based on external economies of scale or location fundamentals do not fully account for the specialization patterns that we find.

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 the 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 agglomeration force, but decreasing returns to scale due to the immobile factors create incentives to spread economic activity across locations. This simple structure allows us to fully represent all equilibrium outcomes in closed form and the model remains tractable under various extensions.

The key feature of the model is a dual economic 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 the rest of the world. The first set of locations constitute a commercially integrated coastal region with high employment density, while the second set conforms 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 leads to migration of mobile factors to the coastal region and to a larger set of locations specialized in export-oriented sectors.

We demonstrate that 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. Immobile factors in the interior region lose due to reduced availability of mobile factors who emigrate to the coastal region, but some immobile factors 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 of location, and can be decomposed into a familiar term that captures gains without internal geography and a term that captures the effect of domestic trade frictions. Among other factors, 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 became more open. For example, economic integration led to larger employment growth in comparative-advantage industries located near borders in Mexico (Hanson 1996) and Vietnam (McCaig and Pavnick 2012). The dual-economy structure featured by the model is consistent with low international trade participation of distant villages in large developing countries with poor transportation infrastructure such as India (Asian Development Bank Institute 2009). The theory explains these facts just by introducing internal trade costs to the neoclassical trade model. Limao and Venables (2003), and Atkin and Donaldson (2012) provide evidence that such costs are sizeable in several developing countries.

The key prediction 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, leading to uneven regional effects of trade. We test this prediction for regional specialization using industry-level data from Chinese prefectures and we control for alternative explanations. We find that the interaction between prefecture proximity to coastlines and industry comparative advantages is a positive driver of specialization, as our model implies. Moving inland by 1 percent from the coast decreases regional employment by 7 percent for an industry with export-revenue ratio that is one standard deviation higher than the average. To mitigate the concern that export-revenue ratios are endogenous we use industry-specific labor intensities from the U.S. as instruments for export orientation of Chinese industries.

Finally, we discuss whether alternative explanations can reproduce 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 larger 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 effect. 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 (2003). In those views, it could be that, closer to international gates, either endowments or externalities drive high relative productivity in specific industries which become export oriented. We argue that this type of explanations cannot fully explain the regional specialization pattern that we find, for they would imply a positive relative-productivity gradient in favor of comparative-advantage industries toward coastlines. However, this pattern seems to be absent in the Chinese data.

The paper is structured as follows. Section 2 lays out and characterizes the model. Section 3 uses spatially disaggregated data on Chinese industries to test the model prediction on regional specialization. Section 4 concludes. Proofs and description of the data are in the appendix.

**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) embed system of cities models in open
economy frameworks. Rossi-Hansberg (2004) 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 geographic specialization across regions that trade with a central location but do not allow for factor mobility.

More recently, Ramondo et al. (2011) quantitatively study the gains from trade and technology diffusion allowing for multiple regions within countries in the Eaton and Kortum (2002) model. Their framework does not capture the interaction between domestic frictions and the gains from trade because it assumes away labor mobility within countries and differences in foreign market access across locations.

Redding (2012) extends the framework in Eaton and Kortum (2002) with labor mobility across regions and develops a multi-region version of the trade and labor mobility model in Helpman (1998). Allen and Arkolakis (2012) extend the Armington model with labor mobility, external economies of scale and congestion forces, but do not make a distinction between internal or international labor mobility. These analyses do not separately allow for export-oriented and import-competing industries, whose presence is essential to our model and empirical analysis.

Finally, a large empirical literature studies determinants of industry location. Holmes and Stevens (2004) offer a summary assessment of the forces determining industry location in the U.S., while Hanson (1998) reviews the literature investigating the effect of North American trade integration on the location of economic activity. Our model of regional specialization generates predictions that are consistent with his findings and complements common explanations such as natural advantages or agglomeration effects. Our empirical section features a novel approach to identifying the effect of market access by relying on the prediction that proximity to demand should be a stronger driver of location in comparative-advantage industries.

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 international iceberg cost in industry \( i \) between \( \ell = 0 \) and the rest of the world (RoW) equal to \( e^{\tau_i} \). Within the country, iceberg trade costs are constant per unit of distance. The cost of shipping a good for distance \( d \) in industry \( i \) is \( e^{\tau_i d} \). This implies a cost of international trade equal to \( e^{\tau_i} + \ell \) from location \( \ell \).

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 goods access. Internal geography vanishes when \( \tau_i = 0 \) for \( i = X, M \) or \( \bar{\ell} = 0 \).

**Endowments** There are two factors of production, a mobile factor and a fixed factor. We refer to the mobile factor as workers, and to the fixed factor as land. 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 1.\(^1\) Land is owned by immobile landlords who do not work and who spend their rental income locally. We let \( N \) be the labor to land ratio at the national level and \( n(\ell) \) denote employment density in \( \ell \), which is to be determined in equilibrium.

**Preferences** Workers and landlords consume in the same location as they live. Indirect utility of a worker who lives in \( \ell \) is

\[
    u(\ell) = \frac{w(\ell)}{E(\ell)},
\]

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 owners of fixed factors, income equals rents \( r(\ell) \) per unit of land and utility is therefore increasing in \( r(\ell)/E(\ell) \). Landowners are immobile, but workers decide where to live. Income generated by each unit of land at \( \ell \) is \( y(\ell) = w(\ell)n(\ell) + r(\ell) \).

**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 industries \( X, M \) at location \( \ell \). Profits per unit of land in sector \( i = X, M \) at \( \ell \) are

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

The production technology is

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

\(^1\)If the distribution of land is uniform, \( \lambda(\ell) \) represents just the measure of locations at distance \( \ell \) from their nearest port.
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 just a normalization constant that helps to save notation. Decreasing returns to scale $1 - \alpha_i$ measure the labor intensity in sector $i$, acting as 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, $\alpha_X = \alpha_M \equiv \alpha$. We discuss implications of differences in land intensity across sectors below.

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,

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

(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 Ricardian 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 $\ell$,

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

(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,

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

(6)

2. profits are maximized,

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

(7)

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

---

\(^2\)The introduction of location-specific amenities as a multiplicative shifter of utility in equation (1) would be isomorphic to a common shifter in the productivity of all industries.
3. land and labor markets clear,

\[
\sum_{i=X,M,N} \lambda_i(\ell) = \lambda(\ell),
\]

(8)

\[
\sum_{i=X,M,N} \frac{\lambda_i(\ell)}{\lambda(\ell)} n_i(\ell) = n(\ell);
\]

(9)

4. trade is balanced.

Conditions (2) to (4) constitute a small Ricardian economy. In addition, in each local economy \( \ell \) the employment density \( n(\ell) \) is determined by (6).

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. As in a standard Ricardian model, 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 \( p(\ell) < p_A \). Since each location takes relative prices as given, a location that trades with either RoW or with other locations is (generically) fully specialized. Only if \( p(\ell) \) happens to coincide with \( p_A \) a trading location may be incompletely specialized. This logic also implies that an incompletely specialized location is (generically) in autarky.

To solve for the 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,\(^3\)

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

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

Trade affects density through the 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. Also, regardless of whether a location trades or stays in autarky, keeping relative prices constant an increase in the national real wage \( u^* \) causes workers to emigrate from \( \ell \).

\(^3\)We relegate the derivation of these expressions to the Appendix.
2.2 General Equilibrium

We move on to study how market access matters for the employment density and the specialization pattern in general equilibrium. We study a small economy that takes international prices \( \{ P^*_X, P^*_M \} \) as given, and we let

\[
p^* = \frac{P^*_X}{P^*_M}
\]

be the relative price at RoW. We assume that every port \( \ell = 0 \) faces the same international price \( p^* \). We also define the average international and domestic iceberg cost across sectors, \( \tau_j \equiv \frac{1}{2} \sum_{i=X,M} \tau^i_j \) 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

\[
P_i(\ell')/P_i(\ell) \leq e^{\tau_i \delta} \text{ for } i = X, M. \tag{11}
\]

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 and face the same world relative prices, (11) implies

\[
e^{-2\tau_0} \leq p(0)/p^* \leq e^{2\tau_0}. \tag{12}
\]

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

\[
e^{-2\tau_1 \ell} \leq p(\ell)/p(0) \leq e^{2\tau_1 \ell}, \tag{13}
\]

where the first inequality binds if \( \ell \) exports \( X \) to RoW, and 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 \( \ell \in [0, \ell]\);  
2. relative prices \( p(\ell) \) satisfy the no-arbitrage conditions (12) and (13) for all \( \ell, \ell' \in [0, \ell] \); and 
3. the real wage \( u^* \) adjusts such that the national labor market clears,

\[
\int_0^{\ell} n(\ell) \lambda(\ell) d\ell = N. \tag{14}
\]

Since Definition 1 of a local equilibrium includes trade balance for each location, trade must also balance 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.\(^4\) This implies that the country is in international autarky if and only if all locations are in autarky and incompletely specialized.\(^5\)

With this in mind, we can characterize the general equilibrium. We can partition the country into the set of locations that trades with RoW and those that stay in autarky. It follows that if the country is not in international autarky there must be some boundary \(b \in [0, \ell]\) such that all locations \(\ell < b\) are fully specialized in the export industry. In turn, all locations \(\ell > b\) do not trade with the RoW and stay in autarky. All locations at distance \(b\) from the nearest gate are indifferent between trading or not with RoW.\(^6\) 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 \(\ell \in (b, \ell]\).

Thus, a key feature of the model is that the distance separating each location from the nearest international gate \(\ell = 0\) 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, every bilateral trade flow in the country either originates from ports, or is directed toward them.\(^7\) These features allow us to represent a two-dimensional geography on the line, leading to closed form characterizations for aggregate equilibrium outcomes and making the model tractable for counterfactuals.

Since all locations \(\ell \in (b, \ell]\) 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) give the price distribution depending on the position of \(b\). Henceforth, we assume that the economy is net exporter of \(X\), and we provide below the conditions such that this is the case. Using (15) and (17), the regional price distribution is given by

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

\(^4\)To see why, suppose that there is bilateral trade between locations \(\ell_X, \ell_M\) at distance \(\delta > 0\). 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.

\(^5\)This result 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.

\(^6\)To determine which locations belong in each set, we note that if the country is 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).

\(^7\)In the model, international shipments depart from coastal locations near international gates, while interior locations only ship locally. This is broadly consistent with the view in Hillberry and Hummels (2008) that shipments in the U.S. are highly localized. Bilateral trade flows unrelated to foreign trade would arise if we allowed for product differentiation within industries or for differences in comparative advantages across locations.
Using (10) in the aggregate labor-market clearing condition (14) we can solve for the real wage,

\[ u^* \left[ \frac{1 - \alpha}{\alpha} \frac{1}{N} \int_0^T \lambda(\ell) \left( \frac{1}{a_X(\ell) e_T(p(\ell))} \right)^{1/\alpha} d\ell \right]^{\alpha}. \]  

(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:

\[ p(b) \geq p_A, \quad \text{if } b < \bar{\ell}. \]  

(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 fundamentals \( a_X(\ell) \) is constant across locations, so that international trade is the only force that shapes the distribution of population density. Since 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 autarky.
So far we have assumed a given trade pattern at the national level. It is simple to determine the national trade pattern by noting that 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 $e^{-2(\tau_0 + \tau_1)} < p_A/p^*$, i.e. when trade costs $\{\tau_1, \tau_0\}$ or the extension of land $\ell$ 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 interior region $[b, \ell]$ 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 $[c, \ell]$ 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. 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. 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 industries and the boundary moves inland.

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

Figure 2 illustrates the effects. The solid black line reproduces the initial equilibrium from Figure 1. The solid red line represents a new equilibrium with lower international trade costs.
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 \([c, b']\) loose population. The dashed 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 distance become more attractive when domestic trade costs decline. In both cases, population density is higher in \([0, c]\) in the new equilibrium.

These results reproduce the cases that we highlighted in the introduction of the paper: as trade costs decline, employment migrates to coastal areas that host comparative-advantage industries. The empirical literature also highlights the importance of proximity to ports for the level of economic activity. In our model, statements about the distribution of employment density apply as well to the distribution of real income per unit of land across locations, since both are proportional.

We move 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 the impact of trade costs on the average real returns to fixed factors is the same as the impact on mobile factors.

We can consider two extreme cases. As \( \tau_1 \to \infty \), domestic trade becomes prohibitive so that \( b \to 0 \) 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 \( b = \ell \) and all locations face the relative price \( p(\ell) = p(0) \). In that case, the real wage is

\[
\bar{u} = \left( \frac{1 - \alpha}{\alpha} \frac{Z}{N} \right)^\alpha \frac{p(0)}{e_T(p(0))}.
\]
where
\[
Z = \int_0^\ell \frac{\lambda(\ell)}{aX(\ell)^{1/\alpha}} d\ell
\]
measures the distribution of land and total productivity across locations. As in a standard Ricardian model, the real wage is increasing in the terms of trade. Here congestion 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 decomposed as follows,
\[
\frac{u^*}{u^a} = \Omega(b; \tau_1) \ast \frac{\overline{u}}{u^a}, \tag{18}
\]
where we define the potential gains of moving from autarky to international trade as
\[
\frac{\overline{u}}{u^a} = \frac{p(0)/e_T(p(0))}{p_A/e_T(p_A)},
\]
and the effects of domestic trade frictions is captured by
\[
\Omega(b; \tau_1) = \left[ \int_0^{\overline{\ell}} \frac{\lambda(\ell)/aX(\ell)^{1/\alpha}}{Z} \left( \frac{p(\ell)/e_T(p(\ell))}{p(0)/e_T(p(0))} \right)^{1/\alpha} d\ell \right]^\alpha. \tag{19}
\]

The actual gains from trade, \(u^*/u^a\), equal the potential gains from trade without domestic trade costs, \(\overline{u}/u^a\), 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. The weights across locations correspond to the importance of their fundamentals and land endowments. In turn, the location-specific losses from domestic trade costs are captured by the reduction in the terms of trade. The overall 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? Intuitively, 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, we should expect the gains from trade to decrease with domestic trade costs. 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, \(\varepsilon(p) = \frac{\varepsilon'(p)}{\varepsilon(p)} p\), and we also let \(s(\ell)\) be the share of location \(\ell\) in total employment. Total differentiation of the clearing condition (14) and the condition (17) for the determination of \(b\), given a shock to \(p^*, \tau_0\) or \(\tau_1\), yield the change in the real wage:
\[
\widehat{u}^* = \int_0^b [1 - \varepsilon(p(\ell))] s(\ell) \widehat{p}(\ell) d\ell, \tag{20}
\]
where \(\widehat{x}\) represent the proportional change in variable \(x\). This expression describes the aggregate gains from a reduction in trade costs, either domestic or international, as 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.

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:

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

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

$$\dot{v}(\ell) = \frac{1}{\alpha} [1 - \varepsilon(p(\ell))] \dot{p}(\ell) - \frac{1 - \alpha}{\alpha} \dot{u}.$$  \hspace{1cm} (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 emigration of mobile factors.

Consider first the interior locations $\ell \in (b, \bar{\ell}]$. In these places, $\dot{p}(\ell) = 0$ 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 loose 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 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 $\tau_1$.
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 \(\tau_1 \to \infty\), 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 new integrated locations. But as \(\tau_1 \to 0\), internal geography vanishes and outcomes are again symmetric in all locations within the country, eliminating income differences across locations.

2.4 Robustness

This benchmark model is purposely stylized to identify how internal and international trade costs interact to shape the regional pattern specialization, the distribution of economic activity, the aggregate gains from trade, and the distributional effects of trade across regions. In the appendix, we characterize the model allowing for location-specific amenities and a non-tradable sector. We have also assumed that returns to scale \(\alpha_i\) are the same across sectors. When \(\alpha_X \neq \alpha_M\), condition (4) is no longer sufficient to preserve the key feature of constant autarky prices across locations. Still in this case 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, independently from their labor endowment, if and only if \(a_X(\ell)^{\alpha_M}/a_M(\ell)^{\alpha_X}\) is constant for all \(\ell\).\(^{10}\) If this condition holds, we ensure that the economy retains the coastal-interior structure that we have described.\(^{11}\)

3 Comparative Advantage and Industry Locations in China

The key prediction 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, leading to uneven regional effects of trade. We now turn to an empirical examination of this prediction using data from China. Several factors make China a particularly suitable country for such an exercise. Its rising external trade has been largely driven by its comparative advantages based on factor endowments. Before market-oriented reforms, it was a closed agricultural economy with “an economic structure notably 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

\(^{10}\) 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 (26). Therefore, \(p_A(\ell)\) is independent from \(\ell\) if and only if \(a_X(\ell)^{\alpha_M}/a_M(\ell)^{\alpha_X}\) is constant.

\(^{11}\) When \(\alpha_M \neq \alpha_X\), conditions (2) to (4) of the local equilibrium conform a small Heckscher-Ohlin economy where the autarky price may in principle depend on factor endowments. However, (6) implies that the labor density is higher in places with more land abundance. This offsets factor proportions effects, turning each local economy into a small Ricardian economy.
regions. The systematic evidence presented in this section suggests that this mechanism may play an important role in shaping these developments.

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.\(^\text{12}\) Each prefecture-industry-year cell has information on employment, capital, revenue, value added, and exports by state-owned and private enterprises.\(^\text{13}\)

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. Table 1 in the Appendix provides further summary statistics. An alternative distance measure using as terminal points the prefectures hosting main international seaports generates similar results since these facilities are scattered somewhat uniformly across the Pacific seaboard (see map). Several characteristics of its geography and trade suggest that distance to the coast is a good measure of foreign market access in the Chinese setting. Maritime transportation is the primary mode of shipping in external trade. Exports over land to bordering countries account for a small share (6.7\%) 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\% (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\% of port capacity suitable for international trade.

3.2 Preliminary Analysis

As a preliminary check, we visualize the drop in regional openness as one moves toward the interior of the country. To uncover this cross-sectional pattern, we average industry-prefecture level observations over the data period and calculate export shares in prefectural revenues. Denoting industries by \(i\), prefectures by \(p\) and time by \(t\), average regional export share (in logarithm) is given by

\[
\text{ExportShare}_p = \ln \left( \frac{\sum_i (\sum_t X_{ipt})}{T} \right) - \ln \left( \frac{\sum_i (\sum_t R_{ipt})}{T} \right),
\]

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

\(^{12}\)Prefectures are the second level of the Chinese administrative structure, contained within 31 provinces in mainland China. Due to administrative re-classifications, number of prefectures and industries vary slightly over the years.

\(^{13}\)The underlying firm-level data has been recently used by Hsieh and Klenow (2009) and Roberts et al. (2011).
measure introduced above. As expected, regional openness declines with distance. The elasticity is -0.76 and highly significant.

Our model generates a negative distance gradient for regional export shares through intensive and extensive margins. The novel mechanism, however, is the latter, i.e. the change in industry composition as one moves toward the interior of the country. To assess whether declining export shares reflect 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

\[ \text{ExportShare}_i = \ln \left( \frac{\sum_t (\sum_p X_{ipt})}{T} \right) - \ln \left( \frac{\sum_t (\sum_p R_{ipt})}{T} \right). \]

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 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\(_i\), the farther away industry \(i\) is located from the coast. The figure shows that, on average, industries with higher export intensity at the national level are situated closer to the seaboard. The distance elasticity is -0.94 and variation in industry distance captures around 42% of the variation in industry export intensity.

While these cross-sectional results lend credibility to our model, they do not fully exploit the rich features of the data. To that purpose, the following subsection provides a more thorough statistical investigation with a richer set of controls.
3.3 Econometric Analysis

Our econometric analysis is based on panel data regressions of the form,

\[ Y_{ipt} = \beta Z_{ipt} + \theta \cdot Distance_p \times ExportShare_{it} + \epsilon_{ipt}. \]  

(23)

Depending on specification, the dependent variable equals the logarithm of employment \((L_{ipt})\) or capital stock \((K_{ipt})\) at industry-prefecture-year 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 non-linear specifications when the dependent variable is binary, taking the value 1 if \(L_{ipt} > 0\), and 0 otherwise.

On the right hand side, the variable of interest is the interaction between \(Distance_p\), logarithm of the time-invariant, prefecture-specific distance measure, and \(ExportShare_{it}\), time-varying aggregate industry export-revenue ratios that capture export orientation.\(^{14}\) Our model predicts a negative coefficient \(\theta < 0\), i.e. export-oriented industries should be more likely to locate closer to the coast. The set of other covariates \(Z_{ipt}\) includes a year-specific intercept and depending on specification, direct effects of interacted variables or various fixed effects.

Table 2 reports the baseline results when the dependent variable is the natural logarithm of employment (columns 1-4) or capital (columns 5-8), using non-zero observations. All specifications have year fixed effects. Standard errors are clustered at the prefecture level.

Columns 1 and 5 start by estimating the direct effects. Both regional employment and capital stock decline with distance, reflecting the gradient of population density in China. While industry-level export orientation is positively correlated with employment, it seems to have a negative effect on capital. Adding the interaction effect in Columns 2 and 6 doesn’t affect the distance coefficient but increases the export coefficients. The increase is large enough to turn the significant negative direct effect on capital stock presented in Column 5 to insignificant in Column 6. This effect is consistent with the result that the interaction effect itself is negative at a 1 percent level of significant. Columns 3 and 7 add prefecture fixed effects to control for forces such as amenities that drive workers in all industries into specific locations, and industry fixed effects to control for national labor demand by industry. The interaction coefficients remain highly significant and stable between -0.32 and -0.37.

The negative interaction coefficient is consistent with our model in that it reveals increasing levels of regional economic activity for export oriented industries toward the coast. The economic importance of an elasticity around -0.35 (at the mid-point of the estimates in Columns 2-3 and 6-7) can be interpreted in the following way: consider an industry with an export-revenue ratio that is

\(^{14}\)We add one to distance in order to keep coastal prefectures in the sample. As an alternative measure of export orientation, we also used the Balassa (1965) revealed comparative advantage (RCA) index. The correlation between the two measures is around 0.7, and the results are both qualitatively and quantitatively very similar. RCA based results are available from the authors upon request.
one standard deviation higher than the average.\textsuperscript{15} Regional employment (or capital stock) in this industry increases by 7 percent in response to a 1 percent decrease in distance. This is a fairly sizeable impact on regional specialization.

Our next specifications in Columns 4 and 8 replace separate industry and prefecture fixed effects with industry-prefecture fixed effects in order to control for unobserved, time-invariant factors that affect economic outcomes at the prefecture-industry level. Identification in these specifications is thus coming from within prefecture-industry variation. 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 during the data period increased relatively more in prefectures close to China’s seaboard.

A natural concern with the baseline specification is the potential endogeneity of the export orientation measure. To address this concern, we draw on the argument that China’s comparative advantage lies in labor intensive industries and instrument export orientation of industries with their labor-capital ratio measured from the US data. Being exogenous to the Chinese economy, this variable is a good candidate for capturing industry characteristics that determine China’s comparative advantages given its endowments.

To implement this approach, we concord Chinese CSIS industries to 1997 NAICS industries in the NBER-CES Manufacturing Industry Database (Becker et al. 2013). The NBER database provides annual information on employment and capital stock in US industries for each year in our dataset. This helps us to calculate labor-capital ratio $L_{it}^{us}/K_{it}^{us}$ for US industries. This variable and its interaction with distance are used as instruments for $ExportShare_{it}$, and its interaction with distance, in a two-stage least squares specification. Table 3 reports the results. The first-stage statistics show that our instrument is strong. The interaction coefficient is still negative and highly significant, suggesting that our baseline results are not driven by the potential endogeneity of the export orientation measure of Chinese industries.

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 74% 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 1 if industry $i$ has positive employment in prefecture $p$ at time $t$, and 0 otherwise.\textsuperscript{16} Using the same explanatory variables, we estimate the probability of industry presence under various specifications and report the results in Table 4. The results from a linear probability model (LMP) in Columns 1-2 imply that the probability of observing an industry with an export-revenue ratio that is one standard deviation higher than the average increases by around 1 percent in response to a 1 percent decrease in distance. The third column replaces the separate fixed effects with industry-prefecture fixed effects. The interaction coefficient remains significantly negative while its

\textsuperscript{15} Mean export-revenue ratio across all years is 0.22 with a standard deviation of 0.2.

\textsuperscript{16} Out of 1,363,012 observations, there are only 4,361 cells with zero employment and a strictly positive capital stock, and 3,284 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.
value drops by an order magnitude. In order to avoid the incidental parameters problem, we do not include fixed effects in the probit model presented in Column 4, but re-introduce industry-prefecture fixed effects in the conditional logit model in Column 5. The marginal effect of the interaction term is negative and significant at 1 and 5 percent levels of significance, respectively.

We finish our main 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 toward estimating gravity equations without excluding bilateral country pairs with zero trade flows, but the methodology is applicable to a wide range of estimations with a nonnegative and continuous dependent variable. Table 5 reports the results. 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.

We conclude by noting that our baseline results proved to be robust to various specifications with sample restrictions, proxying export orientation by the Balassa (1965) measure of revealed comparative advantage, clustering at various levels (province, industry, and industry-prefecture pairs), and including industry- and prefecture-specific time trends. In particular, we replicate the estimation in Table 2 by excluding coastal prefectures or three large interior provinces (Inner Mongolia, Tibet Autonomous Region and Xinjiang Uyghur Autonomous Region) from our sample and find only minor changes in coefficients except one case. When coastal prefectures are excluded from the specifications with industry-prefecture fixed effects (Columns 4 and 8), the interaction coefficient loses its significance for both employment and capital stock. The implication is that export industries in coastal prefectures drew most of the migrant workers and investment over the data period under study.

3.4 Alternative Mechanisms

We finish by inspecting 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 (2003). Following these views, it could be that, closer to international gates, either endowments or externalities drive high relative productivity in specific industries which become export oriented. In the light of these mechanisms, we should observe that relative productivity in export-oriented industries decreases as we move away from international gates. To inspect this pattern we replace the dependent variable in the estimation equation (23) with measures of local industry productivity. In particular, define value added per worker and per unit of capital as

\[ VL_{ipt} = \frac{Value\, Added_{ipt}}{L_{ipt}}, \quad VK_{ipt} = \frac{Value\, Added_{ipt}}{K_{ipt}}. \]

\(^{17}\)These results are available from the authors upon request.
We regress the natural logarithms of these variables on the interaction variable of interest while controlling for fixed effects. The results are in Table 6. A negative interaction coefficient would be supportive of these alternative explanations.

While there is no gradient for labor productivity (first column), capital productivity increases toward coastal prefectures for industries with higher export orientation (second column). This finding, however, is not robust to excluding state-owned enterprises when constructing our regional data. Doing this turns the interaction coefficient on capital productivity to insignificant (fourth column), 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 capital-labor ratio and export-revenue ratio is -0.3. So, interior state-owned enterprises in heavy-industries create the impression that coastal prefectures have higher capital productivity in export-oriented industries. Therefore, the data suggest that the negative market-access gradient for the location of export-oriented industries is not associated with a negative gradient in the value added of these industries.

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 a positive relationship between industry export-revenue ratio at the national level (due to higher intra-industry trade) and returns to scale.

To control for this channel, we follow Hanson and Xiang (2004) and use industry elasticities of substitution as measure of scale economies. 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 in play, one would expect \( EOS_i \) 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 7. While \( EOS_i \) has a negative and significant direct effect on employment, its interaction with distance doesn’t systematically vary with either of the dependent variables. Moreover, the interaction of export orientation with distance remains negative at 1 percent level of significance, and its magnitude is very close to its baseline estimate in Columns 2-3 and 6-7 in Table 2. Thus, the data suggests that home market effects within China are not the main driving force for the strong patterns of regional specialization that we have documented.

---

18 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).
19 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.
20 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 utilize Broda and Weinstein elasticities, we concord Chinese industry classification CSIS into 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.
We developed and tested a theory to characterize the interaction between international trade and the geographic distribution of economic activity within countries. The framework combines standard forces in international trade with an internal geography within countries. Locations within countries differ in access to international markets, and congestion forces deter economic activity from concentrating in a single point. The model reproduces salient facts on the interaction between international trade and regional outcomes, and it implies that the gains from international trade are larger when domestic trade costs are smaller.

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 fixed factors in the interior region. Qualitatively, these stylized 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 relatively isolated corners.

While previous work in economic geography and international trade highlights the importance of locational fundamentals and scale economies in explaining industry locations and the concentration of economic activity, we offer a novel mechanism through the interaction between country-wide comparative advantages and domestic trade costs. To test the prediction of the model on regional specialization patterns, we use Chinese regional data and find that export-oriented industries are more likely to locate close to the seaboard which is the main international gateway to foreign markets in China.
References


APPENDIX

A1: Derivation of Equilibrium Outcomes

We characterize the equilibrium outcomes using a model that allows for a nontradable sector and for location-specific amenities. Utility of an agent who lives in $\ell$ is proportional to $m(\ell) C_T(\ell) \ell^{\beta_T} C_N(\ell) \ell^{1-\beta_T}$, 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 model in the text is the special case when $\beta_T = m(\ell) = 1$.

Indirect utility of a worker who lives in $\ell$ therefore 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 now includes the price index for tradables $E_T(\ell)$ and the price of nontradables $P_N(\ell)$:

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

We let $p(\ell) \equiv P_X(\ell)/P_M(\ell)$ so that there exists an increasing and concave function $e(p(\ell))$ satisfying $E_T(\ell) = P_M(\ell) e(p(\ell))$. Demand for nontradables per unit of land in $\ell$ is $c_N(\ell) = (1-\beta_T) g(\ell) / P_N(\ell)$, where $g(\ell) = w(\ell) n(\ell) + r(\ell)$ is income generated by each unit of land at $\ell$. Production technologies are defined by (3) in the text, where the land intensity in sector $i = N, M, X$ is $\alpha_i$.

The local-equilibrium definition is the same as in (1) with the addition that non-tradeable market clears, $c_N(\ell) = q_N (n_N(\ell))$, and that the market-clearing conditions (8) and (9) include land and labor used in the non-traded sector. Defining $\omega(\ell) \equiv w(\ell) / r(\ell)$ as the wage to rental ratio at $\ell$, we note that the profit maximization condition (7) is equivalent to

$$P_i(\ell) \leq \frac{w(\ell)}{\omega(\ell)^{\alpha_i}} \lambda_i(\ell), \quad \text{if } \lambda_i(\ell) > 0$$  

for $i = X, M, N$. In every populated location, (24) binds for $i = N$ and $u(\ell) = u^*$ if $n(\ell) > 0$. Using the definition of $u(\ell)$ together with the price for nontradables $P_N(\ell)$ from (24) gives the wage at $\ell$ that leaves workers indifferent:

$$w(\ell) = E_T(\ell) \left( \frac{u^*}{m(\ell)} \right)^{\frac{1}{\beta_T}} \left[ \frac{a_N(\ell)}{\omega(\ell)^{\alpha_N}} \right]^{1-\beta_T}. \quad (25)$$

Using (24) and (25), we can solve for the wage-rental ratio for a location producing $i$,

$$\omega_i(\ell) = \left[ u^* a_i(\ell)^{\beta_T} a_N(\ell)^{1-\beta_T} \frac{P_M(\ell)}{P_i(\ell)} e(p(\ell)) \right]^{\beta_T} \left[ \frac{1}{\beta_T a_i(\ell) + (1-\beta_T) a_N(\ell)} \right]. \quad (26)$$

From (24), 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 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 $\omega_X(\ell) = \omega_M(\ell)$.

Since the local economy is incompletely specialized in autarky, $p_A(\ell)$ must be the autarky price. 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 $\ell$.

To solve for population density $n(\ell)$, we use (5) and the production function (3) to express the labor and
land market clearing, respectively, as functions of sectorial output and unit factor requirements:

\[
\begin{align*}
\sum_{i \in X,M,N} \alpha_i a_i(\ell) \lambda_i(\ell) \frac{q_i(\ell)}{\omega(\ell) \alpha_i} &= \lambda(\ell), \\
\sum_{i \in X,M,N} (1 - \alpha_i) a_i(\ell) \lambda_i(\ell) \frac{q_i(\ell)}{\omega(\ell) \alpha_i} &= \lambda(\ell) n(\ell).
\end{align*}
\]

Using market clearing for nontradables, when a local economy is fully specialized in \(i\) we can solve this system to obtain

\[
n(\ell) = \frac{1 - \overline{\alpha}_i}{\overline{\alpha}_i} \left[ \frac{z_i(\ell)}{u^*} \left( \frac{E_T(\ell)}{P_i(\ell)} \right)^{\delta_T} \right]^{1/\delta_T} \text{ for } i = X, M.
\]

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

**A2: Proofs**

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

**Proposition 1** For (i), we have that if the country is 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 \(b < \ell\) such \(e^{-2(\tau_0 + \tau_1)b} p^* = p_A\). 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 \(b < \ell \leftrightarrow p_A/p^* < e^{2(\tau_0 + \tau_1)\ell}\). 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 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 \(\left| \frac{\partial n(\ell)'}{\partial \ell} \right| \leq \left| \frac{\partial n(\ell)'}{\partial \ell} \right| = \frac{1}{\overline{\alpha}(\ell)} \left[ \frac{z_i(\ell)}{u^*} \left( \frac{E_T(\ell)}{P_i(\ell)} \right)^{\delta_T} \right]^{1/\delta_T} \text{ for all } \ell \in [0, b)\). Therefore, \(n(b') > n(b)\) for otherwise \(\int_0^\ell n(\ell) ' \lambda(\ell) \, d\ell < N\). Also note that \(p(\ell) = p(\ell)' = a\) for all \(\ell \in (b, \ell]\) 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 \(n(\ell)' - n(\ell) > 0 \leftrightarrow \ell < c\), so that population density increases in \([0, c]\). When \(\tau_0' > n(0)\) for otherwise \(\int_0^\ell n(\ell) ' \lambda(\ell) \, d\ell < N\). 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 returns to immobile factors at \(\ell\) is \(v(\ell) = r(\ell)/E(\ell)\). Using (27) and \(u(\ell) = u^*\), we can write

\[
v(\ell) = \frac{\overline{\alpha}}{1 - \overline{\alpha}} n(\ell) u^*,
\]

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

\[
v = \frac{\overline{\alpha}}{1 - \overline{\alpha}} Nu^*.
\]
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

\[ \hat{n}(\ell) = \frac{\beta_T}{\alpha T} [1 - \varepsilon_T(p(\ell))] \hat{p}(\ell) - \frac{1}{\alpha} u^*. \] (29)

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 (28), there must be some \( v(\ell) > 0 \) for \( \ell < b \). In turn, when \( d\tau_0 = 0 > d\tau_1 \) then \( \hat{p}(0) = 0 \) but \( u^* > 0 \), implying \( \hat{v}(0) < 0 \). Since \( \hat{v}(\ell) > 0 \) for \( \ell < b \), there is some \( \varepsilon \) such that \( v(\ell) < 0 \) for all \( \ell \in [0, \varepsilon) \).
A3: Tables and Figures

Figure 4: Map of Chinese Prefectures and Ports

Notes: Prefecture boundaries and top seaports by cargo throughput in 2005 (Cullinane and Wang 2007).

Table 1: Summary Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employment, prefecture-level ($L_p$)</td>
<td>1,524,658</td>
<td>2,573,304</td>
</tr>
<tr>
<td>Employment, prefecture-industry level ($L_{ip}$)</td>
<td>1,434</td>
<td>5,098</td>
</tr>
<tr>
<td>Capital stock, prefecture-level ($L_p$), in 1000 USD</td>
<td>16,777,162</td>
<td>33,653,124</td>
</tr>
<tr>
<td>Capital stock, prefecture-industry level ($L_{ip}$), in 1000 USD</td>
<td>15,666</td>
<td>93,599</td>
</tr>
<tr>
<td>Export-revenue ratio, prefecture level ($ExportShare_p$)</td>
<td>0.094</td>
<td>0.11</td>
</tr>
<tr>
<td>Export-revenue ratio, prefecture-industry level ($ExportShare_{ip}$)</td>
<td>0.22</td>
<td>0.2</td>
</tr>
<tr>
<td>Distance (miles)</td>
<td>373</td>
<td>402</td>
</tr>
</tbody>
</table>
Table 2: Baseline OLS Estimates

<table>
<thead>
<tr>
<th></th>
<th>Employment, ln(L)</th>
<th>Capital, ln(K)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>ExportShare</td>
<td>0.323***</td>
<td>1.438***</td>
</tr>
<tr>
<td></td>
<td>(0.114)</td>
<td>(0.182)</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.0969***</td>
<td>-0.0999***</td>
</tr>
<tr>
<td></td>
<td>(0.0159)</td>
<td>(0.0161)</td>
</tr>
<tr>
<td>Distance × ExportShare</td>
<td>-0.338***</td>
<td>-0.369***</td>
</tr>
<tr>
<td></td>
<td>(0.0357)</td>
<td>(0.0379)</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Prefecture fixed effects</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Industry fixed effects</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ind-pref. fixed effects</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Observations</td>
<td>358289</td>
<td>358289</td>
</tr>
<tr>
<td>R²</td>
<td>0.028</td>
<td>0.038</td>
</tr>
</tbody>
</table>

Notes: All regressions are estimated with OLS. Prefecture-clustered standard errors in parentheses. Significance: * 10 percent, ** 5 percent, *** 1 percent. Dependent variable is ln(L) in columns 1-4, and ln(K) in columns 5-8. Distance is the logarithm of prefecture-specific distance measure defined in the text. ExportShare is industry-specific export-revenue ratio at the national level.

Table 3: Instrumental Variable Estimation

<table>
<thead>
<tr>
<th></th>
<th>Employment, ln(L)</th>
<th>Capital, ln(K)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>ExportShare</td>
<td>1.061***</td>
<td>-2.082***</td>
</tr>
<tr>
<td></td>
<td>(0.292)</td>
<td>(0.454)</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.0727***</td>
<td>-0.166***</td>
</tr>
<tr>
<td></td>
<td>(0.0272)</td>
<td>(0.0370)</td>
</tr>
<tr>
<td>Distance × ExportShare</td>
<td>-1.274***</td>
<td>-2.067***</td>
</tr>
<tr>
<td></td>
<td>(0.131)</td>
<td>(0.216)</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>183874</td>
<td>184983</td>
</tr>
<tr>
<td>First-stage F statistic</td>
<td>417.61</td>
<td>422.53</td>
</tr>
</tbody>
</table>

Notes: All regressions are estimated with 2SLS using labor/capital ratios of US industries as an instrument for export orientation of Chinese industries. Prefecture-clustered standard errors in parentheses. Significance: * 10 percent, ** 5 percent, *** 1 percent. Dependent variables are logarithms of employment L or capital K at industry-prefecture-year cells. Distance is the logarithm of prefecture-specific distance measure defined in the text. ExportShare is industry-specific export-revenue ratio at the national level instrumented by labor/capital ratio. First-stage F statistic is the Angrist and Pischke (2009) first-stage F statistics for weak identification of the endogenous variable.
Table 4: Extensive Margin

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ExportShare</td>
<td>0.021</td>
<td>0.202***</td>
<td>-0.004</td>
<td>0.039</td>
<td>0.0503</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.031)</td>
<td>(0.013)</td>
<td>(0.027)</td>
<td>(0.193)</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.048***</td>
<td>-0.043***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance × ExportShare</td>
<td>-0.0479***</td>
<td>-0.0475***</td>
<td>-0.005**</td>
<td>-0.061***</td>
<td>-0.107***</td>
</tr>
<tr>
<td></td>
<td>(0.00626)</td>
<td>(0.00625)</td>
<td>(0.002)</td>
<td>(0.005)</td>
<td>(0.0430)</td>
</tr>
</tbody>
</table>

Regression LPM LPM LPM Probit Clogit
Year fixed effects Yes Yes Yes Yes Yes
Prefecture fixed effects - Yes - - -
Industry fixed effects - Yes - - -
Ind-pref. fixed effects - - Yes - Yes
Observations 1357066 1357066 1357066 1357066 390556
R²(pseudo-R²) 0.0764 0.366 0.733 0.066 0.021

Notes: Prefecture-clustered standard errors in parentheses, clustered by province. Significance: * 10 percent, ** 5 percent, *** 1 percent. Dependent variable equals 1 if there is positive industry employment in a prefecture-year, 0 otherwise. Specifications: Columns 1-3 linear probability model, Column 4 probit, Column 5 conditional logit. Pseudo-R² values of fit are reported for probit and conditional logit models. Marginal effects reported for probit and logit.

Table 5: Poisson Pseudo-Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ExportShare</td>
<td>1.890***</td>
<td>2.189***</td>
<td>-0.121</td>
<td>1.804***</td>
</tr>
<tr>
<td></td>
<td>(0.263)</td>
<td>(0.298)</td>
<td>(0.281)</td>
<td>(0.287)</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.264***</td>
<td>-0.316***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.027)</td>
<td>(0.036)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance × ExportShare</td>
<td>-0.685***</td>
<td>-0.765***</td>
<td>-0.803***</td>
<td>-0.730***</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.061)</td>
<td>(0.069)</td>
<td>(0.063)</td>
</tr>
</tbody>
</table>

Year fixed effects Yes Yes Yes Yes Yes
Industry fixed effects - Yes - - Yes
Prefecture fixed effects - Yes - - Yes
Observations 1362719 1362719 1362719 1362719
R² 0.021 0.366 0.007 0.314

Notes: All regressions are estimated with Poisson Pseudo-Maximum Likelihood. Prefecture-clustered standard errors in parentheses. Significance: * 10 percent, ** 5 percent, *** 1 percent. Dependent variables are employment L or capital K at industry-prefecture-year cells. Distance is the logarithm of prefecture-specific distance measure defined in the text. ExportShare is industry-specific export-revenue ratio at the national level. R² is computed as the square of the correlation between the dependent variable and its fitted values.
Table 6: Alternative Mechanisms I: Local Productivity

<table>
<thead>
<tr>
<th></th>
<th>ln((V_L))</th>
<th>ln((V_K))</th>
<th>ln((V_L))</th>
<th>ln((V_K))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance (\times) ExportShare</td>
<td>0.008</td>
<td>-0.033***</td>
<td>0.033***</td>
<td>-0.012</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.01)</td>
<td>(0.01)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Ownership</td>
<td>All</td>
<td>All</td>
<td>Private</td>
<td>Private</td>
</tr>
<tr>
<td>Observations</td>
<td>349058</td>
<td>349058</td>
<td>307575</td>
<td>307575</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.355</td>
<td>0.184</td>
<td>0.3</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Notes: All regressions are estimated with OLS. Prefecture-clustered standard errors in parentheses. Significance: * 10 percent, ** 5 percent, *** 1 percent. VL and VK stand for Value Added per Labor and Per Capital, respectively. Distance is the logarithm of prefecture-specific distance measure defined in the text. ExportShare is industry-specific export-revenue ratio at the national level. The first two columns use data for all ownership types whereas the last two columns exclude state-owned enterprises. All specifications have year, industry and prefecture fixed effects.

Table 7: Alternative Mechanisms II: Product Differentiation

<table>
<thead>
<tr>
<th></th>
<th>Employment, ln((L))</th>
<th>Capital, ln((K))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>ExportShare</td>
<td>1.609***</td>
<td>1.908***</td>
</tr>
<tr>
<td></td>
<td>(0.307)</td>
<td>(0.285)</td>
</tr>
<tr>
<td>Distance</td>
<td>-0.195***</td>
<td>-0.230***</td>
</tr>
<tr>
<td></td>
<td>(0.0237)</td>
<td>(0.0284)</td>
</tr>
<tr>
<td>EOS</td>
<td>-0.045***</td>
<td>-0.015</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>Distance (\times) ExportShare</td>
<td>-0.404***</td>
<td>-0.569***</td>
</tr>
<tr>
<td></td>
<td>(0.059)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>Distance (\times) EOS</td>
<td>0.003</td>
<td>0.003*</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>Year fixed effects</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Industry fixed effects</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td>Prefecture fixed effects</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>57439</td>
<td>57439</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.073</td>
<td>0.620</td>
</tr>
</tbody>
</table>

Notes: All regressions are estimated with OLS. Prefecture-clustered standard errors in parentheses. Significance: * 10 percent, ** 5 percent, *** 1 percent. Dependent variables are logarithms of employment L or capital K at industry-prefecture-year cells. EOS is demeaned within-industry elasticity of substitution calculated using Broda and Weinstein (2006) trade elasticities. Distance is the logarithm of prefecture-specific distance measure defined in the text. ExportShare is industry-specific export-revenue ratio at the national level.