Internal Geography, International Trade, and Regional Outcomes (PRELIMINARY AND INCOMPLETE)*

A. Kerem Cosar
Chicago Booth

Pablo Fajgelbaum
UCLA

July 10, 2013

Abstract

When trade is costly within countries, international trade leads to concentration of economic activity in locations with good access to foreign markets. Costly trade within countries also makes it hard for remote locations to gain from international trade. We investigate the role of these forces in shaping industry location, employment concentration and the gains from international trade. We develop a model that features comparative advantages between countries, coupled with differences in proximity to international markets across locations within a country. 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. We present motivating evidence from China which shows that export-oriented industries are more likely to locate closer to international ports. The model can be used to measure the importance of international trade in concentrating economic activity, and of domestic trade costs in hampering the gains from international trade.

JEL Codes: TBD
Keywords: TBD

*TBD

Correspondence: `kerem.cosar@chicagobooth.edu`, `pablofajgelbaum@ucla.edu`.

1 Introduction

TBD

2 Model

Geography and Trade Costs The country consists of a set of locations arbitrarily arranged on a map. We index locations by ℓ, and we 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 it will be clear below, given the nature of our model only the distance separating each location ℓ from its nearest port matters for the equilibrium. Therefore, we assume without loss of generality that ℓ represents the distance separating each location from its nearest port, and we denote all ports by $\ell = 0$. We let $\overline{\ell}$ be the maximum distance between a location inside the country and its nearest port.

There are two industries, $i \in \{X, M\}$.[1] International and domestic trade costs are industry specific. The international iceberg cost in industry i between $\ell = 0$ and the rest of the world (RoW) is $e^{\tau_0^i}$. Within the country, iceberg trade costs are constant per unit of distance. Therefore, the cost of shipping a good for distance d in industry i equals $e^{\tau_1^i d}$. This implies a cost of international trade equal to $e^{\tau_0^i + \tau_1^i \ell}$ in industry i from location ℓ.

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 $\tau_1^i = 0$ for both industries.

Endowments There are two factors of production, a perfectly mobile factor and a fixed factor. We refer to the mobile factor as workers, and to the fixed factor as land. We choose units such that the national land endowment equals 1, and we let $\lambda(\ell)$ be the total amount of land available in locations at distance ℓ from their nearest port.[2]

Land is owned by immobile landlords who do not work and who spend their rental income locally. Locations at distance ℓ are also endowed with a level of public amenities $m(\ell)$.

We let N be the measure of mobile factors in the country. We refer to this factor as labor, so that N is the labor to land ratio at the national level. We let $n(\ell)$ denote employment density in ℓ, which is to be determined in equilibrium.

Preferences Workers and landlords consume in the same location as they live. Utility of an agent who lives in ℓ is proportional to

$$m(\ell)\, C_T(\ell)^{\beta_T}\, C_N(\ell)^{1-\beta_T},$$

where $C_T(\ell)$ is a consumption basket of tradables that includes X and M, $C_N(\ell)$ is consumption of non-tradables, and $m(\ell)$ is the level of amenities at ℓ. Indirect utility of a worker who lives in ℓ therefore

[1] The analysis can as well accomodate a continuum of industries as in Dornbusch, Fischer and Samuelson (1977). Our results would carry on in that case.

[2] If the distribution of land is uniform, $\lambda(\ell)$ represents just the measure of locations at distance ℓ from their nearest port.

is

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

where $w(\ell)$ is the wage at ℓ, and $E(\ell)$ is the cost of living index which includes the price index for tradables $E_T(\ell)$ as well as the price of non-tradables $P_N(\ell)$,

$$E(\ell) = E_T(\ell)^{\beta_T} P_N(\ell)^{1-\beta_T}. \tag{2}$$

The price index of tradables $E_T(\ell) \equiv E_T(P_X(\ell), P_M(\ell))$ is defined over the prices in sectors X and M. We let $p(\ell) \equiv P_X(\ell)/P_M(\ell)$ be the relative price of X in ℓ. Since preferences are homothetic, there exists an increasing and concave function $e(p(\ell))$ that depends on the relative price of X such that

$$E_T(\ell) = P_M(\ell) e(p(\ell)).$$

For owners of fixed factors, income equals rents $r(\ell)$ per unit of land and utility is therefore increasing in $m(\ell) r(\ell)/E(\ell)$. Landowners are immobile, but workers decide where to live.

To characterize the equilibrium we will need to use consumption of non-tradables in each location. Demand for non-tradables per unit of land in location ℓ is

$$c_N(\ell) = (1-\beta_T) \frac{y(\ell)}{P_N(\ell)}, \tag{3}$$

where $y(\ell) = w(\ell) n(\ell) + r(\ell)$ is income generated by each unit of land at ℓ.

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 or in the nontradable sector N at location ℓ. Profits per unit of land in sector $i = X, M, N$ at ℓ are

$$\pi_i(\ell) = \max_{n_i(\ell)} \{P_i(\ell) q_i(n_i(\ell)) - w(\ell) n_i(\ell) - r(\ell)\}. \tag{4}$$

The production technology is

$$q_i(n_i(\ell)) = \kappa_i \frac{n_i(\ell)^{1-\alpha_i}}{a_i(\ell)}, \tag{5}$$

where $a_i(\ell)$ is the unit cost of production in industry i in sector ℓ 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 (5) it follows that the aggregate production function in sector i at ℓ is Cobb-Douglas with land intensity α_i.[3]

To simplify the exposition, we assume that land intensity is the same across tradeable sectors,

$$\alpha_X = \alpha_M \equiv \alpha_T,$$

but may differ between the tradeable and non-tradeable sectors, $\alpha_N \neq \alpha_T$. We discuss implications of

[3]It would have been equivalent to include consumption of two non-tradable sectors, housing (which only uses land) and services (which only uses labor), with expenditure share α_N on land. Our current notation is more compact.

differences in land intensity across tradeable sectors below.[4] For future use, we define the consumption-weighted average of sectorial land shares,

$$\overline{\alpha} = \beta_T \alpha_T + (1 - \beta_T)\, \alpha_N.$$

In the nontradable sector, unit costs $a_N(\ell)$ may vary across locations. In the tradeable sector, 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,

$$\frac{a_X(\ell)}{a_M(\ell)} = a \text{ for all } \ell \in \left[0, \overline{\ell}\right]. \tag{6}$$

Therefore, while some locations might be more productive than others in every industry or hold comparative advantages in non-tradables, comparative advantages in tradables are defined at the national level.[5] 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 ℓ,

$$n_i(\ell) = \frac{1 - \alpha_i}{\alpha_i} \left(\frac{P_i(\ell)}{a_i(\ell)\, w(\ell)} \right)^{1/\alpha_i} \quad \text{for } i = X, M, N. \tag{7}$$

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

2.1 Local Equilibrium

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

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

1. *workers maximize utility,*

$$u(\ell) \leq u^*, \quad = \text{ if } n(\ell) > 0, \tag{8}$$

with demand of non-tradables $c_N(\ell)$ given by (3);

2. *profits are maximized,*

$$\pi_i(\ell) \leq 0, \quad = \text{ if } \lambda_i(\ell) > 0, \text{ for } i = X, M, N, \tag{9}$$

where $\pi_i(\ell)$ is given by (4);

[4] See section (**??**). For our quantitative analysis, where we interpret the two tradeable sectors as export-oriented and import-competing industries in the U.S., setting $\alpha_X = \alpha_M$ is an appropriate assumption. See Section 4.

[5] When we introduce externalities, the levels of $[a_M(\ell), a_X(\ell)]$ will depend endogenously on population density in each location. See section (**??**).

3. land and labor markets clear,

$$\sum_{i=X,M,N} \lambda_i(\ell) = \lambda(\ell), \tag{10}$$

$$\sum_{i=X,M,N} \frac{\lambda_i(\ell)}{\lambda(\ell)} n_i(\ell) = n(\ell); \tag{11}$$

4. the non-tradeable market clears,

$$c_N(\ell) = q_N(n_N(\ell)); \text{ and} \tag{12}$$

5. trade is balanced.

Conditions 2 to 5 constitute a small Ricardian economy extended with a non-tradeable sector. In addition, in each local economy ℓ the employment density $n(\ell)$ is determined by (8).

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 the rest of the world, but when labor mobility is allowed across locations. We first note that, as in a standard Ricardian model, the autarky price p_A is the same and equal to a in all locations. Using (9), location ℓ 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 the rest of the world 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 (8) must be binding. Together with (7) and the clearing conditions, this gives the equilibrium population density in each location,[6]

$$n(\ell) = \begin{cases} \frac{1-\overline{\alpha}}{\overline{\alpha}} \left[\frac{z_X(\ell)}{u^*} \left(\frac{p(\ell)}{e_T(p(\ell))} \right)^{\beta_T} \right]^{1/\overline{\alpha}} & \text{if } p(\ell) \geq a, \\ \frac{1-\overline{\alpha}}{\overline{\alpha}} \left[\frac{z_M(\ell)}{u^*} \left(\frac{1}{e_T(p(\ell))} \right)^{\beta_T} \right]^{1/\overline{\alpha}} & \text{if } p(\ell) < a. \end{cases} \tag{13}$$

where $z_i(\ell)$ denotes the fundamentals of each location,

$$z_i(\ell) \equiv \frac{m(\ell)}{a_N(\ell)^{1-\beta_T} a_i(\ell)^{\beta_T}} \text{ for } i = X, M. \tag{14}$$

Expression (13) 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 preference are homothetic, agents employed in the industry-location pair (i, ℓ) necessarily enjoy a higher real income when the local relative price of industry X is higher in location ℓ. That is, the positive income effect from a higher relative price offsets cost-of-living effects. At the same time there are congestion forces, so that mobile workers avoid

[6]In Appendix A we present the solution for the local equilibrium for the more general case with differences in factor intensities across sectors. Equation (13) follows from that derivation.

places with high employment density. Congestion depend on the intensity of land use in non-tradables α_N and in tradables α_T. The larger the congestion, the smaller the population density. In the quantitative exercise, we will measure these parameters to assess the effects of trade on concentration of economic activity. Naturally, agents also prefer locations with better fundamentals $z_i(\ell)$.[7]

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 ℓ. Using these effects, we characterize the general equilibrium below. We summarize the properties of the local equilibrium as follows.

Proposition 2 (Local Equilibrium) *There is a unique local equilibrium where location ℓ is fully specialized in X when $p(\ell) > p_A$ and fully specialized in M when $p(\ell) < p_A$. Population density is given by (13), so that it increases with the relative price of the exported good, and it decreases with the real wage u^*.*

2.2 General Equilibrium

We have characterized the local equilibrium independently from a location's geographic position. 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,

$$\tau_j \equiv \frac{1}{2} \sum_{i=X,M} \tau_j^i \text{ for } j = 0, 1.$$

No arbitrage implies that for any pair of locations ℓ and ℓ' separated by distance $\delta \geq 0$, relative prices in industry i satisfy

$$P_i(\ell')/P_i(\ell) \leq e^{\tau_1^i \delta} \text{ for } i = X, M. \tag{15}$$

This condition binds if goods in industry i are shipped from ℓ to ℓ'. 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, (15) implies

$$e^{-2\tau_0} \leq p(0)/p^* \leq e^{2\tau_0}. \tag{16}$$

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

$$e^{-2\tau_1 \ell} \leq p(\ell)/p(0) \leq e^{2\tau_1 \ell}, \tag{17}$$

[7] In the last quantitative exercise of Section 4 we treat $z_i(\ell)$ as a residual to match the empirical distribution of employment.

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

We are ready to define the general equilibrium of the economy.[8]

Definition 3 (General Equilibrium) *An equilibrium in a small economy given international prices* $\{P_X^*, P_M^*\}$ *consists of a real wage* u^**, local outcomes* $n(\ell)$, $\{n_i(\ell)\}_{i=X,M,N}$, $\{\lambda_i(\ell)\}_{i=X,M,N}$, $c_N(\ell)$, $w(\ell)$, $r(\ell)$ *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* $\ell \in [0, \overline{\ell}]$*;*
2. *relative prices* $p(\ell)$ *satisfy the no-arbitrage conditions (16) and (17) for all* $\ell, \ell' \in [0, \overline{\ell}]$*; and*
3. *the real wage* u^* *adjusts such that the national labor market clears,*

$$\int_0^{\overline{\ell}} n(\ell)\,\lambda(\ell)\,d\ell = N. \tag{18}$$

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. Intuitively, 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 locations ℓ_X, ℓ_M at distance $\delta > 0$. If ℓ_X is the X-exporting location of the pair, part (i) from Proposition 2 implies that $p(\ell_M) \leq p_A \leq p(\ell_X)$. But, at the same time, the no-arbitrage condition (15) implies that the relative price of X is strictly higher in ℓ_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. This result is a spatial impossibility theorem, in the tradition of Starrett (1978).

With this in mind, we are ready to 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 $b \in [0, \overline{\ell}]$ 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.[9] 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 $\ell \in (b, \overline{\ell}]$.

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

[8]Since Definition 2 of a local equilibrium includes trade balance for each location, trade must also balance at the national level.

[9]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 ℓ 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 ℓ 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 (17).

either originates from ports, or is directed toward them.[10] 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.[11]

Since all locations $\ell \in (b, \overline{\ell}]$ 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 (16) and (17) give the price distribution depending on the position of b. Using (19) and (21) we have

$$p(\ell) \;=\; \begin{cases} p^* e^{-2(\tau_0+\tau_1 \min[\ell,b])} & \text{if the country is net exporter of } X, \\ p^* e^{2(\tau_0+\tau_1 \min[\ell,b])} & \text{if the country is net exporter of } M. \end{cases} \tag{19}$$

From now on, we assume that the economy is net exporter of X and net importer of M. Below, we provide conditions such that this is the case. Using (13) in the aggregate labor-market clearing condition (18) we can solve for the real wage,

$$u^* = \left[\frac{1-\overline{\alpha}}{\overline{\alpha}} \frac{1}{N} \int_0^{\overline{\ell}} \lambda\left(\ell\right) z\left(\ell\right)^{1/\overline{\alpha}} \left(\frac{p(\ell)}{e_T\left(p(\ell)\right)} \right)^{\beta_T/\overline{\alpha}} d\ell \right]^{\overline{\alpha}}. \tag{20}$$

Since the relative price function in (19) 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:

$$p(b) \geq p_A, \;\; = \;\text{ if } b < \overline{\ell}. \tag{21}$$

When $p(\overline{\ell}) > p_A$ then $b = \overline{\ell}$, so that the interior region does not exist. The general equilibrium is fully characterized by the pair $\{u^*, b\}$ that solves (20) and (21). 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 shrinks away from 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.

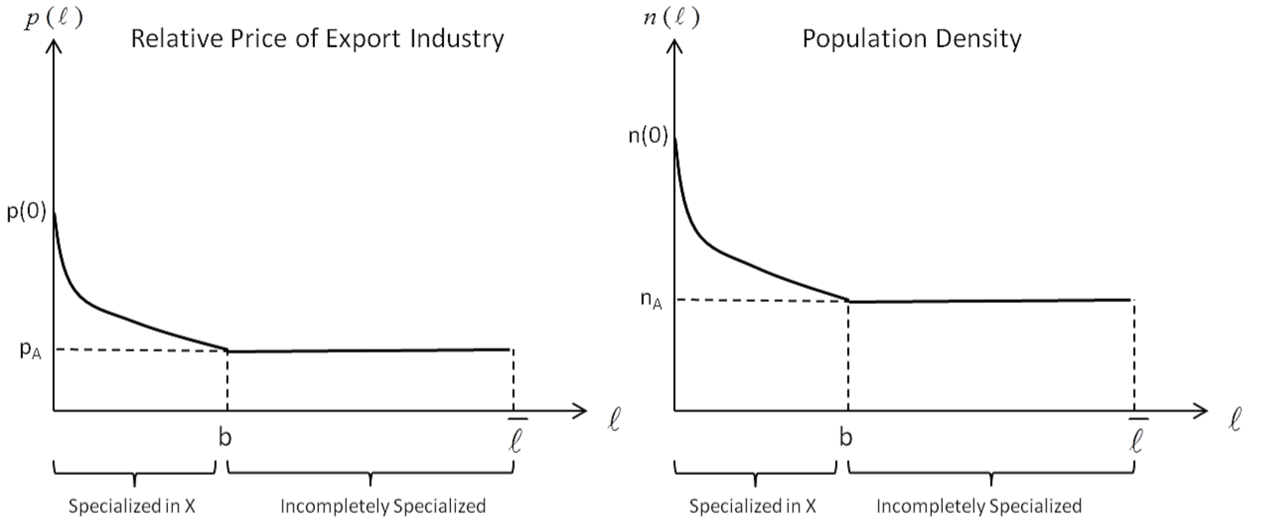
In the right panel, we plot population density assuming that the fundamentals $z\left(\ell\right)$ defined in (14) are constant across locations, so that international trade is the only force that shapes the distribution of population density. Since the relative price of the export industry decreases away from international gates, so does population density in the coastal region until it reaches the interior region.

We summarize our findings so far as follows.

[10] 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 Hilberry and Hummels (2008) that shipments in the U.S. are highly localized. In the calibration, we target the domestic shipping costs of export-bound shipments in the U.S. to measure τ_1.

[11] Naturally, 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. With these features, representing the two-dimensional geography on the line would be generically unfeasible.

Figure 1: Relative Prices and Population Density over Distance



Proposition 4 (Population and Industry Location in General Equilibrium) *There is a unique small-country equilibrium, where: (i) if the country trades internationally but is not fully specialized in what it exports, there exists an interior region* $\left[b, \overline{\ell}\right]$ *that is incompletely specialized and a coastal region* $[0, b)$ *that trades with RoW and specializes in the export-oriented industry; and (ii) if* $z(\ell)$ *is constant, the distribution of population is uniform under international autarky, but it increases toward international gates if the country trades.*

These results demonstrate that international trade drives concentration of economic activity and industry location. When the economy trades but is not fully specialized, two discrete regions emerge: a coastal region surrounding international gates that is densely populated, connected to international markets and specialized in the export-oriented industry; and an interior region that is sparsely populated, disconnected from the rest of the world and incompletely specialized.

In our reasoning so far we have assumed a given trade pattern at the national level. We establish the conditions on the parameters that determine the national trade pattern and existence of the interior region.

Proposition 5 (National Trade Pattern and Existence of Interior Region) *(i) The country exports X if* $p_A/p^* < e^{-2\tau_0}$*; in that case, the interior region exists if and only if* $e^{-2(\tau_0+\tau_1\overline{\ell})} < p_A/p^*$*; (ii) the country exports M if* $e^{2\tau_0} < p_A/p^*$*; in that case, the interior region exists if and only if* $p_A/p^* < e^{2(\tau_0+\tau_1\overline{\ell})}$*; and (iii) the country is in international autarky if* $e^{-2\tau_0} < p_A/p^* < e^{2\tau_0}$*.*

These results imply that domestic trade costs $\left\{\tau_1^X, \tau_1^M\right\}$, while capable of affecting the gains and the volume of international trade, are unable to impact the pattern or the existence of it. In other words, the conditions that determine when international trade exists as well as the direction of international trade flows are the same as in an environment without domestic geography. The second implication is that, when then country trades, the interior region exists when trade costs $\{\tau_1, \tau_0\}$ or the extension of land $\overline{\ell}$ are sufficiently large, or when comparative advantages, captured by p_A/p^*, are not sufficiently strong.

2.3 Impact of Changes in International and Domestic Trade Costs

We use the model to characterize the impact of international and domestic trade costs on the concentration of economic activity and the gains from trade. In the quantitative section we measure the importance of these effects.

2.3.1 Trade Costs and Internal Migrations

Our motivating examples from the introduction show that international trade integration is 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 6 (Internal Migration) *A reduction in international or in domestic trade costs moves the boundary inland to $b' > b$ and causes migration from region $\left[c, \overline{\ell}\right]$ into region $[0, c]$ for some $c \in (b, b')$. A lower τ_0 causes population at the port $n(0)$ to increase, but a lower τ_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 τ_0, the shift is uniform across locations, while a lower τ_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: Effects of Changes in Domestic and International Trade Costs

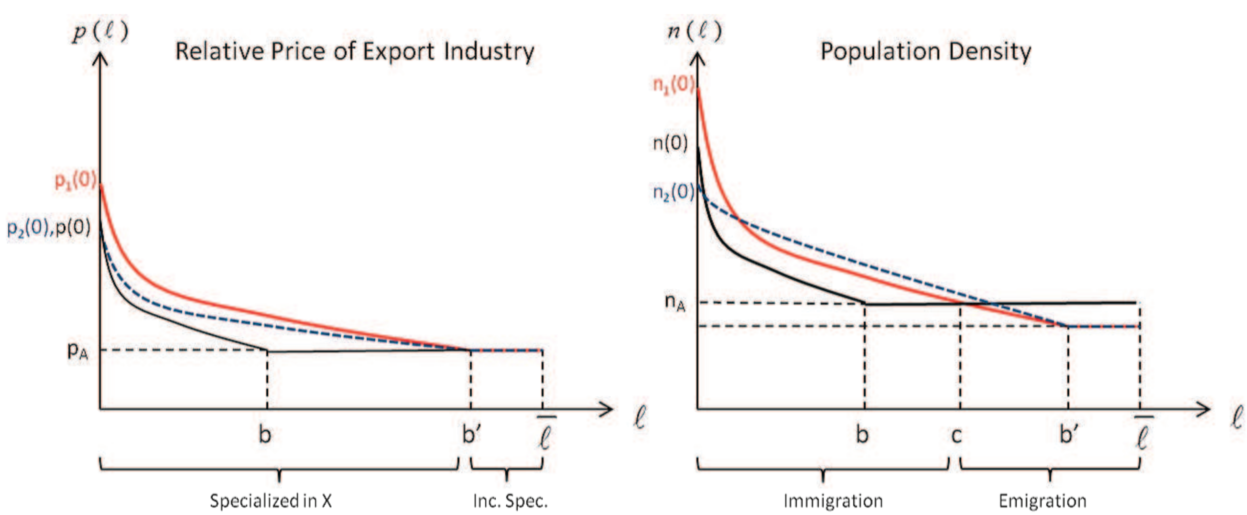


Figure 2 illustrates the effects. The solid black line reproduces the initial equilibrium from Figure 1. The solid red line represents a new equilibrium with lower international trade costs. The price function shifts upward and the intercept increases from $p(0)$ to $p_1(0)$, increasing population density at the port to $n_1(0)$. Locations in $[b, b']$ start trading, but the newly specialized locations $[c, b']$ 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 highlight in the introductory paragraphs 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. In the quantitative section, we use the model to measure the contribution of international trade to the excess concentration of employment in coastal relative to interior areas in the U.S.

2.3.2 Internal Geography and the Gains form Trade

We move to the impact of domestic trade costs τ_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^*.[12] 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 $b = \overline{\ell}$ and all locations face the relative price $p(\ell) = p(0)$. In that case, the real wage is

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

where

$$Z = \int_0^{\overline{\ell}} \lambda(\ell)\, z(\ell)^{1/\overline{\alpha}}\, d\ell$$

measures the distribution of land land and total productivity across locations. As in a standard Ricardian model, the real wage is increasing in the terms of trade. Here, it also depend on the share of tradables in total expenditures, while congestion causes the real wage to decrease with the national labor endowment.

Using the solution for the real wage u^* from (20), the actual gains of moving from autarky to trade

[12] Adding up real returns to labor and land across locations gives the aggregate production function,

$$Y \equiv \int_0^{\overline{\ell}} \frac{w(\ell)\, n(\ell) + r(\ell)}{E(\ell)} \lambda(\ell)\, d\ell = \left\{ \frac{1}{\overline{\alpha}} \int_0^{\overline{\ell}} \lambda(\ell)\, z(\ell)^{1/\overline{\alpha}} \left(\frac{p(\ell)}{e_T\left(p(\ell)\right)} \right)^{\beta_T/\overline{\alpha}} d\ell \right\}^{\overline{\alpha}} \left(\frac{N}{1-\overline{\alpha}} \right)^{1-\overline{\alpha}},$$

which in turn implies that real income per worker is proportional to the real wage, $Y/N = u^*/(1-\overline{\alpha})$. Following the same steps, we find that the real returns to land deflated at local prices equal $(\overline{\alpha}/1-\overline{\alpha})\, N u^*$ and are also proporitional to the real wage.

can be decomposed as follows,

$$\frac{u^*}{u^a} = \Omega\left(b;\tau_1\right) * \frac{\overline{u}}{u^a}, \tag{22}$$

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

$$\frac{\overline{u}}{u^a} = \left(\frac{p\left(0\right)/e_T\left(p\left(0\right)\right)}{p_A/e_T\left(p_A\right)}\right)^{\beta_T},$$

and where the effects of domestic frictions is

$$\Omega\left(b;\tau_1\right) = \left\{\int_0^{\overline{\ell}} \frac{\lambda\left(\ell\right) z\left(\ell\right)^{1/\overline{\alpha}}}{Z}\left[\frac{p\left(\ell\right)/e_T\left(p\left(\ell\right)\right)}{p\left(0\right)/e_T\left(p\left(0\right)\right)}\right]^{\beta_T/\overline{\alpha}} d\ell\right\}^{\overline{\alpha}}. \tag{23}$$

The actual gains from trade, u^*/u^a, equal the potential gains from trade without domestic trade costs, $\overline{u}/u^a$, adjusted by $\Omega\left(\tau_1, b\right)$. 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 $\Omega\left(\tau_1, b\right)$ is strictly below 1 as long as $\tau_1 > 0$, and it equals 1 if $\tau_1 = 0$. In the quantitative section we measure the magnitude of each component in (22) to assess the importance of domestic trade costs for the gains from trade.

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 τ_1 causes the export oriented region to shrink, we should expect the gains from trade to decrease with domestic trade costs. A lower τ_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,

$$\varepsilon(p) = \frac{de_T(p)/e_T\left(p\right)}{dp/p},$$

and we also define the share of location ℓ in total employment,

$$s\left(\ell\right) = \frac{n\left(\ell\right)\lambda\left(\ell\right)}{N}.$$

Using these definitions, we have the following result for the change in welfare when there is a change in the environment. We let $\widehat{x}$ represent the proportional change in variable x when there is a change in the environment.

Proposition 7 (Gains from International Trade) *Consider a shock to p^*, τ_0 or τ_1. Then, the change in the real wage is*

$$\widehat{u^*} = \beta_T \int_0^b \left[1 - \varepsilon(p(\ell))\right] s\left(\ell\right) \widehat{p(\ell)} d\ell. \tag{24}$$

Therefore, the gains from trade are decreasing with domestic trade costs τ_1*,*

$$\frac{d(u^*/u^a)}{d\tau_1} < 0. \tag{25}$$

Expression (24) 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, (24) implies that the gains from an improvement in the terms of trade, caused by either lower τ_0 or larger p^*, are bounded above by the share of employment in export-oriented regions. It also implies that domestic and international trade costs are complementary, in that the gains from international trade are decreasing in domestic trade costs. Larger τ_1 causes relative export prices to decrease faster toward the interior, reducing the gains from trade.

2.3.3 Distributive Effects of Trade between Immobile Factors

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

$$\widehat{v(\ell)} = \frac{\beta_T}{\overline{\alpha}}\left[1 - \varepsilon_T(p(\ell))\right]\widehat{p(\ell)} - \frac{1-\overline{\alpha}}{\overline{\alpha}}\widehat{u^*}. \tag{26}$$

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 $\ell \in (b, \overline{\ell}]$. 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 $(b, \overline{\ell}]$ 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 τ_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.

These results echo the well known distributional effects in specific factor models, where the factors that are specific to goods whose relative price increases gain from trade. Here, the new margin is that immobile factors experience differential relative price changes due to international and domestic trade reform depending on their geographic position. We summarize these results as follows.

Proposition 8 (Distributive Effects of Trade) *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.*

3 Comparative Advantage and Industry Locations in China

This section examines the empirical implications of the model using regional Chinese data. China is a particularly suitable country to look at through the lens of our framework since its rising external trade has been largely driven by its factor endowments. Before the market-oriented reforms, it was an agricultural closed economy with "an economic structure notably lacking in industrial specialization and agglomeration." (Chan, Henderson and Tsui) 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. POLISH: Most recently, interior regions and the federal government increased their spending on transportation infrastructure in order to relieve the congestion in coastal regions and allow export-oriented activity to move further inland. The systematic evidence presented in this section suggests that the mechanism we propose may play an important role in shaping these developments.

The analysis proceeds in several steps: the description of the data and key variables is followed by descriptive figures that verify some broad patterns that are consistent with the theory. A regression analysis then exploits both the cross-sectional and panel features of the data with a fuller set of controls. A set of exercises at the end intend to rule out alternative mechanisms that could generate the documented empirical patterns.

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.[13] Each prefecture-industry-year cell has information on employment, capital, output, value added, and exports by ownership. Due to size truncation, our dataset accounts for about 60% of total manufacturing employment in China. There is a fair amount of variation in the extensive margin. The median prefecture has positive employment in 96 out of 425 industries. Similarly, the median industry is present in 77 out of 338 prefectures. There is no activity, i.e. zero employment, in approximately 70% of all prefecture-industry combinations.[14] Across non-zero prefecture-industry-year cells, the median and mean employment are 332 and 1407, respectively. The largest prefecture-industry employment over the time period is observed for the *"Textile and garment manufacturing"* industry (with CSIS code 1810) located in Shanghai, which grows from 150,000 workers in 1998 to 210,000 in 2007.

To this data, we add the Euclidian *distance* of each prefecture to China's coastline and report summary statistics in Table 1. There are 53 coastal prefectures with zero distance. While the maximum distance is 2128 miles, only 23 prefectures are more than 1000 miles away from the seaboard. The median is 275 miles.

[13] Prefectures are the second level of the Chinese administrative structure, contained within 31 provinces in mainland China. Due to administrative reclassifications, number of prefectures and industries vary slightly over the years.

[14] This arises due to the size truncation in the underlying firm-level data as well as true absence of certain industries in some locations. In what follows, we treat empty cells as true zeros.

Table 1: **Summary Statistics of the Distance Measure (miles)**

Mean	373.29
St. Deviation	401.60
Median	274.95
Maximum	2127.86
Minimum	0

Is distance to coast a good measure of foreign market access in the Chinese setting? Several characteristics of China's geography and trade suggest so. 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. [15] The share of air shipments in exports to top 20 trade partners is 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 close to the sea, constitute only 20% of port capacity suitable for international trade.

FINISH

3.2 Preliminary Analysis

This subsection investigates the empirical validity of some basic implications of our framework. The first such implication is a negative gradient for regional openness as one moves toward the interior of the country. To document this cross-sectional pattern, we average industry-prefecture level observations over the data period. Denoting industries by i, prefectures by p and time by t, average regional openness is given by

$$\ln(xq_p) = \ln\left(\frac{\Sigma_t(\Sigma_i X_{ipt})}{T}\right) - \ln\left(\frac{\Sigma_t(\Sigma_i Q_{ipt})}{T}\right),$$

where (X_{ipt}, Q_{ipt}) stand for total exports and output at the industry-prefecture-year level. Panel A of Figure 3 plots this variable against the natural logarithm of $dist_p$, the prefecture-specific distance measure introduced above.[16] As predicted, regional openness declines with distance. The elasticity is -0.76 and highly significant.

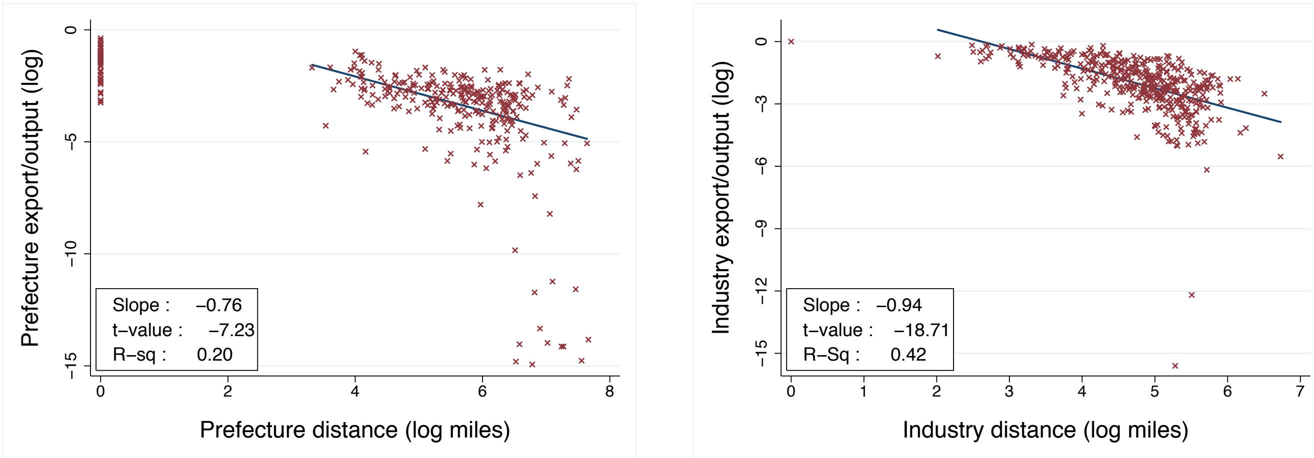
Note that our model generates a negative distance gradient for export intensity through two margins: in the coastal region, distance reduces the share of exports in output for the comparative advantage industry while the region itself remains specialized. This intensive margin gradient is not of primary interest in that it could be generated by a wide range of models featuring trade costs. The novel mechanism in our model is the extensive margin through which the industry composition changes as one moves toward the interior of the country. To assess whether the documented distance gradient of export intensity reflects industry composition, Panel B of Figure 3 plots industries' export/output ratio at the national level against "industry distance" in logarithmic scale. The former variable equals

$$\ln(xq_i) = \ln\left(\frac{\Sigma_t(\Sigma_p X_{ipt})}{T}\right) - \ln\left(\frac{\Sigma_t(\Sigma_p Q_{ipt})}{T}\right),$$

[15] Authors' calculation using 2006 UN Comtrade data.

[16] We add one to total exports and distance in order to preserve the zeros in the sample but fit the regression lines to observations with strictly positive exports and distance.

Figure 3: Distance and Export Intensity



Panel A: Export Intensity across Prefectures

Panel B: Export Intensity and Industry Distance

Notes: In both panels, x-axis is log(dist+1). In panel A, each observation is a prefecture with the y-axis being the log of the regional (export+1)/output ratio. In panel B, each observation is an industry with the y-axis being the log of its aggregate (export+1)/output ratio. Here, distance refers to average industry distance defined in the text. The regression lines are fitted to observations with strictly positive distance and exports.

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

$$dist_i = \sum_p \left(\frac{emp_{ip}}{\Sigma_p emp_{ip}} \right) \cdot dist_p,$$

where $dist_p$ is prefecture p's distance and emp_{ip} is industry i's employment in prefecture p averaged over time. Therefore, the larger is $dist_i$, the farther away industry i is located from the coast. The figure shows that, on average, industries with higher export/output ratios 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 some credibility on 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,

$$\ln(emp_{ipt}) \;=\; \gamma_t + \beta X_{ipt} + \theta \cdot \ln(dist_p) \times ExportOrientation_{it} + \epsilon_{ipt}, \tag{27}$$

where the dependent variable is employment at the industry-prefecture-year level and γ_t is a year-specific intercept. The variable of interest is the interaction between $dist_p$, the time-invariant, prefecture-specific distance measure, and $ExportOrientation_i$, a measure that is increasing in the export orientation of industry i. A negative coefficient $\theta < 0$ implies that export-oriented industries are more likely to locate closer to the coast. Depending on specification, the set of other covariates X_{ipt} includes direct effects of the interaction variables, or fixed effects. In what follows, we use two alternative measures of export orientation. The first one is aggregate industry export-output ratio. The second measure is based on the Balassa (1965) revealed comparative advantage (RCA) index, defined as the share of an industry's export in total Chinese exports relative to the export share of that industry in total world exports:

$$rca_{it} = \frac{X_{it}^{China}/\Sigma_i X_{it}^{China}}{X_{it}^{World}/\Sigma_i X_{it}^{World}}.$$

In order to construct this measure, we map Chinese CSIS industry classification to international Rev.3 ISIC classification via the official concordance. This allows us to use international trade data to calculate industry RCA.[17] The correlation between the two measures of export orientation is around 0.7.

Table 2 reports the baseline results. All variables enter the regression in natural logarithms. Standard errors are clustered by provinces and industry-prefecture grouping to deal with concerns about spatial and serial correlation. We add one to the distance measure in order to preserve coastal prefectures in the sample. The results are robust to excluding coastal prefectures with zero distance. We drop observations with zero employment and separately explore industry presence in the extensive margin below.[18] The first three columns use export-output ratio for $ExportOrientation_i$ while the last three use industry RCA. Column I and IV use direct effects of the interaction terms without fixed effects. Due to the inclusion of prefecture-year fixed effects (γ_p, γ_i) in Columns II and IV, the distance measure is dropped. Prefecture fixed effects control for forces such as amenities which drive workers in all industries into specific locations. We also include industry fixed effects to control for national labor demand by industry. Specifications in Columns III and VI use industry-wide labor-capital ratio as an instrument for export orientation.[19] The first-stage statistics show that the instruments are strong and we pass overidentification tests.

The highly significant negative coefficient on the interaction term in all specifications is consistent with our model. To find out whether the interaction terms is economically significant, we answer the following question using the conservative coefficient of -0.0498 in Column I: by how much does employment in an industry at the 75th percentile export-output ratio increase relative an industry at the 25th percentile export-output ratio as it moves from a prefecture at median distance to a coastal one? The answer is about 43%: there is a large differential effect of distance on the level of activity in industries with varying levels of export orientation.

Our next specification replaces separate industry and prefecture fixed effects (γ_p, γ_i) with industry-prefecture fixed effects γ_{ip} in order to control for unobserved, time-invariant factors that affect economic outcomes at the prefecture-industry level. Identification is now coming from within prefecture-industry variation. The results are reported in Table 3.

FINISH THE DISCUSSION OF THE ABOVE TABLE.

[17] This procedure reduces the number of observations due to two factors. Firs, the concordance is valid from 2002 onward. Second, it provides a many-to-one mapping that collapses 425 CSIS industries to 112 ISIC industries with no overlaps. This allows simple aggregation of all the relevant observations to the ISIC level. Our source of trade data is http://wits.worldbank.org

[18] We also estimated a Poisson regression in which the dependent variable is the level (not the logarithm) of employment. The results are robust to this alternative specification which preserves observations with zero employment.

[19] Our dataset informs us about total fixed assets of firms at the industry-prefecture-year level. Industry-wide capital and employment are thus $K_{it} = \Sigma_p K_{ipt}$ and $L_{it} = \Sigma_p emp_{ipt}$, respectively. Although labor-capital ratio is an endogenous outcome itself, it can be considered pre-determined when compared to export orientation. Also, China's comparative advantage in labor intensive industries makes it a good proxy: its correlation with $\ln(xq)$ and $\ln(rca)$ is 0.66 and 0.56, respectively.

Table 2: Distance and Industry Employment Across Chinese Prefectures

	I	II	III	IV	V	VI
Dependent variable: ln(emp)						
ExportOrientation	0.206***	0.195***	1.824*	0.338***	0.328***	-0.369
	(0.0527)	(0.0388)	(1.045)	(0.0670)	(0.0399)	(0.721)
ln($dist$)	-0.211***		-0.254***	-0.196***		-0.244***
	(0.0199)		(0.0284)	(0.0214)		(0.017)
ln($dist$) × ExportOrientation	-0.0498***	-0.0524***	-0.0595***	-0.0551***	-0.0710***	-0.103***
	(0.00905)	(0.00852)	(0.0125)	(0.0108)	(0.0104)	(0.0182)
Regression	OLS	OLS	IV/2SLS	OLS	OLS	IV/2SLS
Fixed effects	No	Yes	No	No	Yes	No
ExportOrientation	ln(xq)	ln(xq)	ln(xq)	ln(rca)	ln(rca)	ln(rca)
Instrument	-	-	ln(L/K)	-	-	ln(L/K)
N	360,773	360,773	360,773	115,014	115,014	115,014
R^2	0.0369	0.311	-	0.0714	0.409	-
First-stage statistic	-	-	188.33	-	-	364.44
Over-identification statistic	-	-	367.03	-	-	729.07

Notes: Standard errors in parentheses, clustered by province and industry-prefecture grouping. Significance: * 10 percent, ** 5 percent, *** 1 percent. First-stage statistic is the Kleibergen-Paap rk Wald F statistic for test of weak identification of endogenous variables. Overidentification statistic is the Hansen J statistic

Table 3: Distance and Industry Employment Across Chinese Prefectures

	I	II
Dependent variable: ln(emp)		
ExportOrientation	0.0294	0.0755***
	(0.0244)	(0.0167)
ln($dist$) × ExportOrientation	-0.0131***	-0.0121***
	(0.0044)	(0.0042)
Regression	OLS	OLS
Fixed effects	Yes	Yes
ExportOrientation	ln(xq)	ln(rca)
N	360,773	115,014
R^2	0.058	0.071

Notes: Standard errors in parentheses, clustered by industry-prefecture grouping. Significance: * 10 percent, ** 5 percent, *** 1 percent.

Next, we estimate the probability of the presence of an industry in a given prefecture by exploiting the variation in the extensive margin:

$$\mathbf{1}(emp_{ipt} > 0) \;=\; \gamma_t + \beta X_{ipt} + \theta \cdot \ln(dist_p) \times ExportOrientation_{it} + \epsilon_{ipt}, \tag{28}$$

where the dependent variable takes the value one if industry i has positive employment in prefecture p at time t, and zero otherwise. Approximately 30% of all prefecture-industry combinations in our data have positive employment. Other variables are the same as in estimating equation 27. We estimate this equation using a linear probability model, as well as probit and instrumental variable probit regressions. As before, labor-capital ratio serves as an instrument for export orientation. Table 4 reports the marginal effects.

The significantly negative marginal effect of the interaction term in all specifications suggests that the probability of observing export oriented industries is indeed decreasing in distance. To make sense of magnitudes, consider the increase in the probability of observing an industry at 75th percentile of export-output ratio when we move from median distance to the coast, relative to an industry at 25 percentile.

The coefficient from the linear probability model in Column I yields a differential effect of 8.3 percentage points.

Table 4: Distance and Industry Presence Across Chinese Prefectures

	I	II	III	IV	V	VI
Dependent variable: $\mathcal{I}(emp > 0)$						
ExportOrientation	0.00119	0.005	-0.0038	0.0134	0.0174	-0.014
	(0.0168)	(0.013)	(0.0156)	(0.0098)	(0.0128)	(0.0312)
$\ln(dist)$	-0.0744***	-0.075***	-0.0807***	-0.0687***	-0.0749***	-0.076***
	(0.00913)	(0.008)	(0.0088)	(0.0096)	(0.0121)	(0.0116)
$\ln(dist) \times$ ExportOrientation	-0.00947***	-0.0049**	-0.0047	-0.0087***	-0.0076***	-0.0157**
	(0.00280)	(0.0023)	(0.0073)	(0.0016)	(0.002)	(0.0073)
Regression	LPM	Probit	IV-Probit	LPM	Probit	IV-Probit
ExportOrientation	$\ln(xq)$	$\ln(xq)$	$\ln(xq)$	$\ln(rca)$	$\ln(rca)$	$\ln(rca)$
Instrument	-	-	$\ln(L/K)$	-	-	$\ln(L/K)$
N	142619	142619	142619	37856	37856	37856
R^2(pseudo-R^2)	0.096	0.081	0.072	0.092	0.07	0.083
Wald stat. for exogeneity	-	-	60.65	-	-	97.59

Notes: Standard errors in parentheses, clustered by province. Significance: * 10 percent, ** 5 percent, *** 1 percent. Columns I and IV report results from a linear probability model. Columns III and VI instrument industry export orientation with labor-capital ratio. To avoid incidental parameters problem in fixed effect probit regressions, we estimate all regression separately for each year and report the results for 2005. Probit and IV-probit coefficients are marginal effects, interaction marginal effect is calculated using the methodology in Ai and Norton (2003). Pseudo R^2 values of fit are reported for probit models. Wald statistic for exogeneity is the χ^2 statistic for testing the null hypothesis of exogeneity for the instrumented variable.

3.4 Alternative Mechanisms

Could the empirical results obtained so far be driven by reverse causality? The concern is that certain industries locate in coastal regions for reasons other than those present in our model—such as location-specific fundamentals or agglomeration economies—and then become export-oriented by virtue of their proximity to foreign markets. Regardless of the underlying determinant of initial locations, such reverse causation is likely to manifest itself in coastal prefectures being relatively more productive in export-oriented industries. We formally explore this pattern by replacing the dependent variable in the estimation equation (27) with measures of local industry productivity. In particular, we define average revenue per labor

$$ARL_{ipt} = \frac{revenue_{ipt}}{emp_{ipt}},$$

and average value added per labor

$$AVL_{ipt} = \frac{value\ added_{ipt}}{emp_{ipt}}.$$

We regress the natural logarithms of these variables on the interaction term of interest while controlling for fixed effects. The results in Table 5 indicate the absence of a systematic relationship. The interaction coefficient has a positive sign in all specifications but it is not statistically significant at 10 percent level. While this analysis does not rule out reverse causality completely, it nevertheless rejects an important observable implication.

Our final empirical check addresses the possibility that industry export/output ratios, used as a measure of export orientation in our analysis, and industry locations are jointly determined by omitted factors. An important source of variation in export intensity across industries is the extent of intra-industry trade.

Table 5: Productivity and Industry Employment Across Chinese Prefectures

Dependent variable	$\ln(ARL)$	$\ln(AVL)$	$\ln(ARL)$	$\ln(AVL)$
$\ln(dist) \times$ ExportOrientation	0.002	0.0029	0.0035	0.0048
	(0.0024)	(0.0025)	(0.0044)	(0.004)
Regression	OLS	OLS	OLS	OLS
Fixed effects	Yes	Yes	Yes	yes
ExpOri	$\ln(xq)$	$\ln(xq)$	$\ln(rca)$	$\ln(rca)$
N	356,189	349,694	114,148	112,876
R^2	0.419	0.354	0.434	0.366

Notes: Standard errors in parentheses. Significance: * 10 percent, ** 5 percent, *** 1 percent. ARL and AVL stand for *Average Revenue per Labor* and *Average Value Added per Labor*, respectively. Observations with zero economic activity are dropped since ARL and AVL are not defined. Observations with negative value added are also dropped while taking natural logarithms.

A higher level of within-industry differentiation implies a higher export/output ratio for that industry (Krugman 1979). Turning to location choices, domestic "home market effects" working through scale economies could induce differentiated industries to agglomerate in regions with high population density, i.e. in large cities, as shown by Hanson and Xiang () for the case of industry concentration across countries. If such regions in China happen to be coastal for reasons outside of our model, the above mechanism could lead to an increasing presence of industries with high export/output ratio towards the seaboard. One way to check the validity of this mechanism is to include in our baseline specification estimates of within-industry substitution elasticities from the trade literature. In particular, we draw on the estimates in Broda and Weinstein () and let σ_i denote the substitution elasticity for sector i.[20] If the above mechanism holds, one would expect σ_i to pick up the distance gradient estimated before. To check this, we estimate our baseline estimation equation 27 including the interaction of σ_i with distance:[21]

$$\ln(emp_{ip}) \;=\; \beta X_{ip} + \theta \cdot \ln(dist_p) \times \ln(xq_i) + \psi \cdot \ln(dist_p) \times \ln(\sigma_i) + \epsilon_{ip}. \tag{29}$$

Column I in Table 29 reports the results when X_{ip} includes direct effects of the interaction variables. Column II features industry and prefecture fixed effects. A comparison of these results with Columns I-II in Table 2 shows that export-orientation still remains significant in accounting for industry locations, exerting its effect both directly as well as via distance. Industry elasticities of substitution fail to pick up any meaningful variation in regional employment patterns of industries.

[20] Unsurprisingly, industry-level substitution elasticities and export intensities are negatively correlated. A regression of $\ln(\sigma_i)$ on the logarithm of export/output ratio $\ln(xq_i)$ gives a statistically significant elasticity of -0.96 with an R^2 of 0.36.

[21] Since σ_i is time invariant, we collapse the panel to a cross section of averages. In order to utilize Broda and Weinstein's estimates, we map Chinese industries into 3-digit SITC industries. As before, this procedure reduces the number of observations.

Table 6: Industry Differentiation and Employment Across Chinese Prefectures

	I	II
Dependent variable: ln(*emp*)		
ln(*dist*)	-0.351***	
	(0.0485)	
ExportOrientation	0.173*	0.175***
	(0.0964)	(0.0663)
ln(σ)	-0.711***	
	(0.111)	
ln(*dist*) × ExportOrientation	-0.0565***	-0.0887***
	(0.0180)	(0.0126)
ln(*dist*) × ln(σ)	0.0388	0.0274
	(0.0259)	(0.0236)
Regression	OLS	OLS
Fixed effects	No	Yes
ExportOrientation	ln(*xq*)	ln(*xq*)
N	57,696	57,696
R^2	0.09	0.63

Notes: Standard errors in parentheses. Significance: * 10 percent, ** 5 percent, *** 1 percent. σ is industry-specific elasticity of substitution calculated using Broda and Weinstein (2006) trade elasticities. Both regressions are estimated with a year-specific intercept. Column II includes industry and prefecture fixed effects.

4 Conclusion

TBD