Geography, Transport and the Composition of Trade

Russell Hillberry* Manuel I. Jimenez* Bilgehan Karabay†

June 20, 2025

Abstract

We develop a general equilibrium model of international trade and find support for it empirically. In the model a global transport sector allocates a fixed stock of shipping capital across bilateral routes. Sectors vary in the intensity in which they use transport capital. With intuition matching that of von Thünen's model of regional agricultural specialization around an exogenous city, our model predicts that geographically central/large countries specialize in transport-intensive sectors, while remote regions specialize in goods that are less transport-intensive. The model improves on von Thünen's because it is a complete general equilibrium. Our empirical results demonstrate that larger and more central regions specialize in products that rely on air shipment and shipment on containerized vessels. Remote regions export products that rely on less costly forms of shipping capital (e.g. bulk vessels). The results suggest another reason for primary commodity specialization in the periphery, and an important potential constraint on remote regions' ability to grow by producing increasingly complex goods.

Keywords: von Thünen model, Revealed Comparative Advantage, Transportation, Geography of Trade
JEL Codes: F12, F14, O18, R10, R40

*Department of Agricultural Economics, Purdue University; e-mail: rhillber@purdue.edu, jimene19@purdue.edu.
†Department of Economics and Management and School of International Studies, University of Trento, e-mail: bilgehan.karabay@gmail.com.

1 Introduction

New Zealand's top export products are all lightly processed primary commodities.[1] Given New Zealand's strong institutions and its high levels of human capital and economic development, it seems obvious that the country's export product mix would be quite different were its geographic location that of, for example, Ireland.[2] Besides New Zealand, many other countries in the global periphery – including Australia and much of Latin America and Africa – are also specialized in primary commodities or lightly processed versions thereof. This paper investigates the joint role of transport and geography in determining revealed comparative advantage. We develop a general equilibrium model of international export specialization that follows the intuition of von Thünen (1826). Transport-intensive activities locate in countries with better market access, while remote countries produce products that are less transport-intensive. We then show that geographically remote countries have revealed comparative advantage (RCA) in products that move long distances by non-containerized maritime vessels. These findings are robust to the inclusion of a wide variety of controls related to other determinants of RCA, and to the replacement of transport mode shares with bins of products based on their weight-to-value ratios.

Some reasons for primary commodity specialization in the periphery have been addressed in the literature, and our theoretical model can be extended to incorporate two leading explanations: a Heckscher-Ohlin channel based on natural endowments, and a channel based on products' vertical position in the supply chain (as in Antràs and Chor (2013)). Our theoretical contribution is to develop a model that predicts a third force for remote specialization in primary commodities: commodities that can be transported long distances at relatively low cost are produced in the global periphery. The model's predictions for export specialization in central vs. peripheral countries is similar to those that a literature on home market effects (HME) predicts for large vs. small

[1] According to CEPII (2024), New Zealand's top five HS-6-level exports in 2022 were 1) milk and cream in powders, granules or other forms, 2) frozen cuts of beef, 3) untreated and unpainted coniferous wood; 4) frozen sheep meat cuts, and 5) kiwifruit. Together these five commodities represent about 29 percent of New Zealand's 2022 export value. Most of the other products that are important in New Zealand's exports are also primary commodities or lightly processed versions of the same.

[2] Ireland's top five export products include various pharmaceuticals and related inputs, integrated circuits, and chemical products according to CEPII (2024) data. Ireland's export bundle, of course, is affected by its role as a tax haven within the EU. We conjecture that New Zealand might have played a similar role for Europe if it were located in Ireland's position. Our claim here is more modest; simply that New Zealand's export mix would be quite different were it located in Ireland's geographic position. A similar comparison of the exports of Argentina or Chile with exports from Mexico would also make our point.

countries.

Our theory contains two important features that distinguish it from earlier von Thünen-type theories of specialization and trade. Our use of a monopolistic competition framework allows us to dispense with an exogenous central location that is the sole source of demand in the standard framework. As in the standard framework, market access matters for industry location, but we require only differences in market access across location, not an exogenous source of demand. Second, Krugman-style monopolistic competition provides us with an industry-level entry condition. The entry condition plays a very important role in the model, and is the reason the model can produce the sharp specialization patterns that arise in the standard von Thünen model. A third difference with much of the existing New Economic Geography literature is that we model transport costs as payments to transport capital, rather than as iceberg trade costs. This complicates the analytical exercises, but offers a stronger link to the von Thünen (1826) theory, and a more realistic representation of the transport sector's role in determining comparative advantage. Remote regions are less well-served by a transport sector that employs a fixed stock of transport capital. Efficient allocation of constrained transport capital leads to sectors that rely on transport capital to locate nearer the geographic center, in order to serve global demand.

Our empirical exercises are based on the idea that products have physical and/or market characteristics that determine the transport mode that is most suitable for intercontinental transport. Bulk and other non-containerized maritime vessels provide low-cost transport of high weight-to-value products over long distances. Container vessels are most suitable for less bulky products and provide other aspects of transport quality (security, refrigeration, etc.), but not speed. Air shipment delivers products that are high value, low-weight and embody some need for timely delivery. Because the focus of the paper is industry location, we abstract from inferring product characteristics from mode choices over land routes; such choices are highly dependent on distance shipped and on physical infrastructure. Amongst air, container and other maritime vessels, it is likely that the costs of air and container vessels' time are higher than those of non-containerized vessels'. The relevant implication from our model is that air and containerized maritime vessels would be made relatively less available to shippers in remote regions than vessels that carry primary commodities.[3] We hypothesize that a relative lack of supply of container vessels and air transport services in remote regions helps to determine revealed comparative advantage; goods that require high-cost transport services are produced in the geographic core, while the periphery specializes in goods that can be transported long distances inexpensively by non-containerized maritime vessels.[4] Since the mapping of products to the preferred mode of transport in long-distance shipment is conceptually novel, we also show that remote regions have comparative advantage in the products with the highest weight-to-value ratios, while more central commodities specialize in products with the lowest weight-to-value ratios.[5]

A small but influential literature has used the interaction of product- and country-characteristics to test hypotheses about the determinants of comparative advantage.[6] We use this framework to investigate the interaction of a) product characteristics related to the transport mode in which the products are shipped over long distances and b) geographic country characteristics that our model associates with production location decisions (market access, population size, and per capita income). We use data from U.S. imports - net of imports from Canada or Mexico - to calculate product-level statistics on transport mode choice, and apply these to HS6 product-level exports from countries other than the U.S. We test the hypothesis that products' mode of shipment in U.S. import data predict the spatial distribution of RCA across the globe, and find evidence supporting the hypothesis. Large countries are more likely to specialize in products that are typically air-shipped, while geographically central countries specialize in products that move over long distances by air

[3]This channel is consistent with our model, though the model makes predictions about transport intensity rather than type of transport vessel employed. In motivating regressions, we show that in U.S. import data, the share of transport in delivered cost is higher for products transported by container vessels or air shipment than for products that usually move as non-containerized freight. The transport share in delivered value also rises faster with distance in shipment by container vessels or air transport than it does in other kinds of maritime transport.

[4]We use three categories of transport in our initial specification because they are available in the U.S. import data. Recognizing that there are multiple kinds of non-containerized freight, we employ the tariff schedule to also separate products according to the type of non-containerized freight vessel in which they are most likely to move. Specifically, we separate out products moving by non-containerized freight that are very likely to move by either a) roll-on, roll-off vessels, or b) liquid vessels and tankers. Including these categories strengthens the evidence that the periphery specializes in products that travel by low-cost bulk carriers.

[5]These results highlight the importance of the transport mode channel. If there were only a single mode of transport available, and freight charges were rising in the weight-to-value ratio, one would likely expect the heaviest goods to be produced nearest the global center. In a study of road freight in the United States, Duranton et al. (2014) find a stronger negative correlation between that market access and weight than between market access and value. We find regions with better market access have RCA in products with the lowest weight-to-value ratios, which overlaps with the set of products that are typically transported by air in intercontinental transport.

[6]See in particular Romalis (2004), Nunn (2007), Levchenko (2007) and Debaere (2014).

freight or container vessels. These results are robust to the inclusion of controls representing other possible determinants of RCA. The controls include relative factor endowments and intensities, the interaction of institutions with relevant product characteristics, the tendency of products that are upstream in production chains to be exported by countries in the geographic periphery, as well as WTO member status, which we use as a general indicator of trade policy orientation.

Our insight that products' transportability affects where they are produced is not novel, but we believe the mechanism plays an underappreciated role in the international trade literature. von Thünen (1826) was the first to posit that this particular channel affected production location decisions. In his model, an isolated city purchases a variety of agricultural products from its nearby hinterlands. The model predicts that the interaction of product's transport dependence and distance to the city determines the production mix of a given location. Transport-intensive agricultural products are grown near the central city, while more distant locations produce goods that are more easily transported. Assuming constant returns to scale, Venables and Limao (2002) generalize the von Thünen implications to an international trade setting, showing theoretically that product-level variation in transport costs can reverse Heckscher-Ohlin specialization in the periphery Krugman and Venables (1995) develop a model of global trade in which the geographic core hosts industries with increasing returns to scale in production and intermediate input trade. In their model the periphery endogenously exports products produced with constant returns to scale (and no trade costs). Our model is most similar to Venables and Limao (2002), but our use of a monopolistic competition framework allows us to model a complete general equilibrium, rather than assuming that demand is exogenously located in the global center.

A key innovation of our model is the use of a formal (but simple) model of the transport sector.[7] Monopolistically competitive firms produce output that they leave at the factory gate. A constant-returns-to-scale transportation sector rents a fixed supply of shipping capital, and adds value by delivering the good to consumers in destination markets. Consumers value the composite of the

[7] Many papers in the geography-of-trade literature employ iceberg trade costs, which assume that the factor bundle used in production is the same as that used in transport. Because the transport sector in this model uses different factors than those used in production, the model produces novel insights regarding the way in which trade costs affect the geography of export specialization. A key assumption of the model is that the productivity of the transport sector falls with distance, an assumption that makes the model behave in a manner similar to a model with iceberg transport costs.

physical good and the transport service.[8] Products differ in the degree to which the transport sector adds value. Long distance transport requires the shipping sector to rent more shipping capital services to accomplish its task. Since this raises the cost of transport for longer distance shipments, the production of transport-intensive products is less viable in remote regions. As in von Thünen (1826), cross-product variation in transport intensities interacts with geography to generate a high degree of specialization across countries, even absent the other forces that push remote regions to specialize in primary commodities.[9] The von-Thünen channel is alone sufficient to generate a high degree of product specialization across countries.

Our model is also related to the literature on home market effects (HMEs). This literature posits that large countries tend to specialize in sectors with stronger increasing returns to scale because large countries are relatively larger sources of demand. Hanson and Xiang (2004) is the most similar paper in the HME literature to our work; they use a monopolistic competition model to show that sectors with low demand elasticities and high transport costs are most likely to exhibit HME effects. Our model produces the same prediction for large vs. small countries, but we focus on the related predictions for central versus remote countries. The remaining differences with Hanson and Xiang (2004) are that we model the transport sector directly, rather than relying on iceberg trade costs, and that we estimate conditional variation in RCA using Poisson Pseudo Maximum Likelihood (PPML) estimators rather than among bilateral trade flows with log-linear OLS.[10]

Our empirical work shows that goods that are moved long distances by air transport or container vessels tend to locate in locations with good market access (that is, nearer the geographic center of world trade). Evans and Harrigan (2005) develop a partial equilibrium model with similar implications, and show its relevance for U.S. apparel imports in the 1990s. In the model, production

[8] Francois and Wooton (2001) employ this approach, but leave the transport sector outside the general equilibrium trade framework they use to evaluate welfare. Their focus is on the effects of imperfect competition in the transport sector on economic welfare. We abstract from this important topic in order to generate a tractable model in which RCA depends on transport intensity.

[9] A version of the model with product-level variation in iceberg trade costs generates specialization of the same kind, though the degree of specialization is much reduced, relative to our model, which generates the full specialization results in von Thünen (1826) when the parameters that define the transport share of delivered cost are not set at unreasonably high levels.

[10] A recent literature on HMEs identifies the importance of external economies of scale (EES) for generating HMEs. Norris (2025) is the most recent. Elasticities of demand facing a good's producers are smaller when transport intensities are higher, leading to more entry in equilibrium, and hence larger implied industry scale. This mechanism operates in our model, though the transport channel plays a larger role in our framework.

of apparel products - especially of goods that require more frequent restocking is located nearer to the destination (U.S.) market than production of products for which restocking is less common. Frequently restocked goods, which would likely require air shipment were they to be produced in Asia, are more likely to be produced and exported from locations near the destination U.S. market. Our empirical work demonstrates a similar result - that products that move by air in intercontinental transport are produced in locations nearer the geographic center of global trade. Our model is different in important ways that are tailored to the particular research question we ask, but the empirical lessons are closely related to those of Evans and Harrigan.[11]

Several papers in the New Economic Geography (NEG) literature develop similar implications for agglomeration using the somewhat uncomfortable assumption that that trade in a CRS 'agricultural' good is costless.[12] In our model transport capital is fully mobile and serves as the numeraire, so we can dispense with the assumption of a costlessly traded numeraire good. Differences in transportability across goods producing sectors remain, but they are motivated by empirical evidence on the types of transport usually employed to carry the product types. Empirically our focus is on the composition of trade, rather than the implications of geography for per capita income, a frequent focus of the NEG literature.

A key goal of our paper is to re-center geography as an important determinant of countries' export specialization. A macro-development literature has emphasized that countries' movement from less to more complex goods is an important channel of economic development (Rodrik, 2006; Hausmann et al., 2007; Hidalgo and Hausmann, 2009; Reed, 2024). We view our results as having moderately pessimistic implications for growth of this kind in the periphery; there are several mutually reinforcing reasons that the periphery specializes in primary commodities. The von-Thünen channel we propose is one of them. An international trade literature has emphasized institutions as an important determinant of comparative advantage (Levchenko, 2007; Nunn, 2007). We use similar methods to show that transport characteristics associated with different products also play a role in determining export specialization.

[11] Our focus is on implications for the periphery, where we find that revealed comparative advantage is in goods that are neither shipped by air transport nor container vessel. We also study the entire set of traded goods, rather than those in the apparel sector alone, as Evans and Harrigan do.

[12] See e.g., Krugman and Venables (1995), Krugman and Venables (1996), Redding and Schott (2003), and Redding and Venables (2004), among others.

The paper proceeds as follows. Section two develops the theoretical model and illustrates the mechanisms through which each of the three proposed channels generates primary commodity specialization in the periphery. We show theoretically how transport intensity can lead the periphery to produce products that mode by low-cost transport. A numerical simulation model is used to demonstrate the generalization of the model to other sources the specialization we see. Section three develops an empirical model for testing the hypothesis that remote regions export products that are less reliant on high-quality shipping. Section four describes the data. Section five reports the results. Section six concludes.

2 Theoretical Framework

2.1 Factor markets and income

Let households in country i be endowed with two factors. A factor of production L_i (labor) is mobile across production sectors, but immobile internationally. Households are also endowed with a portion of global transport capital, T_i. Transport capital can only be used to supply transport services, but it is fully mobile internationally.[13]

Let ℓ_i^k be the allocation of region i labor to sector k. A labor market clearance condition determines the wage, w_i.

$$L_i = \sum_k \ell_i^k \tag{1}$$

Moving product k along bilateral route ij requires the transport sector to purchase the services of transport capital. We formalize this relationship later, but in order to illustrate clearance in the transport capital services market we denote the transport capital required to carry sector k goods sold from i to j as t_{ij}^k. The associated factor market clearance condition for transport capital covers all routes and commodities and determines the rental rate r, which is common across the globe:

$$\sum_i T_i = \sum_i \sum_j \sum_k t_{ij}^k. \tag{2}$$

The income of the representative household in region i, Y_i, is the sum of factor incomes from labor

[13] Since our model is static, we translate it to mean that there is a given stock of transport capital at a moment in time. Transport providers hire transport capital to serve bilateral trade routes. Transport capital that is purchased to serve one route is not available for concurrent use on other routes.

and shipping capital:

$$Y_i = w_i L_i + rT_i \tag{3}$$

2.2 Consumption

Let consumers have Cobb-Douglas preferences over products k. Each product has a common expenditure share $\alpha_k = \alpha$. Within each product category k, consumers have Dixit-Stiglitz preferences over individual varieties produced by n_{ik} distinct firms located in every region i. Delivered quantities in region j from i are a composite of the produced variety from i and a transport service. As in Francois and Wooton (2001), consumers value composite varieties that bundle the produced good and the transport service; it is not possible to consume the good without also consuming the bundled transport service. The quantity of each composite variety of k sold on the bilateral route ij is q_{ij}^{Ck}. Sector k varieties originating in i are identical, so the quantity q_{ij}^{Ck} is common to all n_i^k varieties on the route.

It is useful to illustrate preferences through the indirect trade utility function:

$$V(p_i^k, n_i^k, Y_j) = \frac{Y_j}{\prod_k \left(\sum_i n_i^k (p_{ij}^{Ck})^{1-\sigma}\right)^{\frac{\alpha}{1-\sigma}}} \tag{4}$$

where p_{ij}^{Ck} is the price of a composite variety of k that includes the commodity produced in i and the transport service along the ij route and σ is the elasticity of substitution parameter across composite commodities. For purposes of analytical clarity, we take σ to be common across commodity groups.

The good-transport composite is constructed via a Cobb-Douglas technology, which we represent here as the zero-profit condition of a fictional freight forwarder:

$$p_{ij}^{Ck} = (p_i^k)^{1-\gamma^k} (p_{ij}^{Tk})^{\gamma^k}, \tag{5}$$

where p_i^k is the factory gate price of a k variety produced in i, p_{ij}^{Tk} the price of associated transport services from i to j, and γ^k a measure of the transport intensity of good k.

2.3 Demand functions

Application of Roy's identity to (4) recovers j's demand for n_i^k varieties of the composite good. Demand for a single variety of the composite good follows

$$q_{ij}^{Ck} = \frac{(p_{ij}^{Ck})^{-\sigma}\alpha Y_j}{\widetilde{P}_j^k} \tag{6}$$

where $\widetilde{P}_j^k = \sum_i n_i^k (p_{ij}^{Ck})^{1-\sigma}$.

Applying Shepard's lemma to (5) returns variety-specific demands for the transport service, and for each variety of the physical good produced in i. Demand for the quality of transport services is represented as

$$q_{ij}^{Tk} = \gamma^k \frac{p_{ij}^{Ck}}{p_{ij}^{Tk}} q_{ij}^{Ck}, \tag{7}$$

while bilateral demands for a variety of the physical good follow:

$$q_{ij}^{k} = (1-\gamma^k)\frac{p_{ij}^{Ck}}{p_i^k} q_{ij}^{Ck}. \tag{8}$$

It is worth noting that the price of the manufactured good p_i^k does not vary over destinations, as it would in a model with iceberg trade costs. We introduce a transport cost wedge by modeling the transport sector explicitly. Producing firms are indifferent to the destination of their sales when they set their prices.[14] The notional freight forwarder purchases the good at the factory gate and combines it with the transport services in order to deliver the composite.

2.4 The transportation sector

As indicated above, there is cross-commodity variation in the intensity with which products rely on transport services. We have also distinguished between the price of transport services p_{ij}^{Tk} and the quantity of services provided q_{ij}^{Tk}. This distinction is important for the operation of the model. The direct effect of distance on these variables is to increase the price of transport. Quantities of transport services q_{ij}^{Tk} are not rising with distance; in equilibrium q_{ij}^{Tk} falls as p_{ij}^{Tk} rises.[15]

[14]It is conventional when modeling large group monopolistic competition to assume that the producer makes production decisions without any ability to price discriminate over destinations. In iceberg trade cost models, this stems from a) each firm represents a small share of overall demand, and b) each firm faces a constant elasticity of demand across routes. We will also assert (a), and show that (b) also applies in our setting. Our assumption that the transport sector is competitive is useful in allowing us to avoid outcomes in which producers have an incentive to bargain over the price of freight.

[15]The consumer is indifferent to the distance over which the variety travels. She just responds to the higher price of transport over longer distances, and the freight forwarder's endogenous decision to buy fewer transport services.

The transportation sector has a Ricardian production technology. The delivery one unit of q_{ij}^{Tk} requires D_{ij}^{ρ} units of transport capital. D_{ij} is the distance from i to j, and ρ the elasticity of the unit factor requirement with respect to distance. We assume the transport market is competitive, so

$$p_{ij}^{Tk} = rD_{ij}^{\rho}. \tag{9}$$

Total demand for transport capital associated with moving sector k output along the ij route is the product of the number of varieties traded, the per unit factor demand and the associated quantity of transport services provided per variety:

$$t_{ij}^{k} = n_i^k D_{ij}^{\rho} q_{ij}^{Tk}. \tag{10}$$

2.5 Production

A firm produces a variety ω of sector k output in region i, with the quantities produced denoted $q_i^k(\omega)$. The production technology expresses the labor required to produce $q_i^k(\omega)$ as $\ell(\omega) = a + bq_i^k(\omega)$ where a and b are fixed and marginal labor requirements that are common across sectors and countries. $q_i^k(\omega)$'s will be identical across varieties, so region i sector k labor demand is $\ell_i^k = \sum_{\omega} a + bq_i^k(\omega)$. Since all varieties are identical we drop the ω representation hereafter and represent firm choices in terms of the representative firm and sectoral labor demands as the product of firm-level demands and the number of firms in the sector:

$$\ell_i^k = n_i^k \left(a + bq_i^k\right). \tag{11}$$

Global demand for the firm's variety is the summation of the bilateral demands, (8).

$$q_i^k = \sum_j q_{ij}^k = \frac{(1-\gamma^k)}{p_i^k} \sum_j \frac{\left((p_i^k)^{1-\gamma^k} (p_{ij}^{Tk})^{\gamma^k}\right)^{1-\sigma} \alpha Y_j}{\widetilde{P}_j^k} \tag{12}$$

In the conventional large group monopolistic competition framework, each firm takes the decisions of other firms as given, so the price index is unresponsive to decisions taken by the firm. The

q_{ij}^{Tk} might best be considered a bundle of aspects of transport quality. The consumer values having the good arrive intact and on-time, and the freight forwarder trades these preferences off against a costly decision to provide more transport services for a given route-product combination.

own-price elasticity of demand facing the firm is:

$$\epsilon^k = \sigma + \gamma^k - \sigma\gamma^k, \tag{13}$$

on each bilateral route. Since p_i^k is constant across destinations, ϵ^k is also the elasticity of total demand facing the manufacturing firm.

The firm maximizes profits by choosing q_i^k, free entry sets profits to zero. The conventional approach to solving this problem generates equilibrium values of the output price p_i^k and the quantity produced per variety q_i^k:

$$p_i^k = bw_i \frac{\epsilon^k}{\epsilon^k - 1} \tag{14}$$

$$q_i^k = \frac{a}{b}(\epsilon^k - 1). \tag{15}$$

A zero profit condition determines whether any region i firm chooses to produce a variety of k, and if so, how many. Following Balistreri and Rutherford (2013) we write the entry condition as a complementary slackness condition:

$$aw_i \geq \frac{p_i^k q_i^k}{\epsilon^k} \quad \perp \quad n_i^k \geq 0, \tag{16}$$

where aw_i represents the costs of hiring the fixed labor requirement, $\frac{p_i^k q_i^k}{\epsilon^k}$ is the revenue available to pay fixed costs, and $\perp$ is an operator that defines complementary slackness. Firms enter only under the condition that fixed costs are fully compensated. Equation (16) is quite important to the model, as it is the mechanism that determines commodity specialization, and whether or not specialization is complete. Adapting Krugman (1980) to our context, total labor demanded by sector k in region i is given by

$$\ell_i^k = n_i^k a \epsilon^k \tag{17}$$

2.6 Definition of the General Equilibrium

A general equilibrium of the model is defined as an allocation of labor across the sectors and transport capital across the routes and commodities transported that satisfies equations $(1) - (10)$, (14), (16) and (17). These equations solve, respectively, for the endogenous variables w_i, r, Y_j, V_j,

p_{ij}^{Ck}, q_{ij}^{Ck}, q_{ij}^{Tk}, q_i^k, p_{ij}^{Tk}, t_i^k, p_i^k, n_i^k and ℓ_i^k.[16]

2.6.1 Proof of existence of general equilibrium

To date, our proof of existence is numerical and computational. In the Mathiesen (1985) and Rutherford (1995) approach to computational modeling, existence can be proven through the exploitation of Walras' Law. The choice of a numeraire leaves the equilibrium system with one less variable than there are equations. If this constraint has no effect on the problem, Walras' Law is satisfied.

Numerical Proof: When we choose the price of shipping capital r as the numeraire, we find that the marginal on the associated constraint is zero. This means that Walras' Law is satisfied, a proof of existence. Moreover, a proportional change in the numerical value of the numeraire implies an equivalent proportional change in all nominal variables (prices and incomes), while leaving all real variables unaffected (input and output quantities and welfare levels). This verifies that the model is homogeneous degree one in prices and income, and homogeneous degree zero in quantities. We observe both outcomes with any change in the model numeraire.

2.7 Model implications for bilateral trade

Our model departs from the standard gravity-model-of-trade framework that relies on iceberg trade costs, but our closely related assumption that consumers purchase a composite of the manufactured good and the transport service produces a gravity-like trade pattern nonetheless. The assumptions that determine market conditions in the transport sector produce gravity-like predictions for the transport service. Physical goods and transport services are gross-complements in the demand system, so bilateral demand for the physical good decreases as the price of transport rises with distance, a mechanism that produces gravity in bilateral goods trade. Our functional form assumptions make this derivation of bilateral goods trade straightforward. Both the transport service and the physical good demands have constant elasticities of trade with respect to distance, though these elasticities differ. The demand elasticity facing the producing firm is different still.

[16] The computational model we use to explore deviations from the simpler analytical model represents the equilibrium as a mixed complementarity problem, using tools developed in Mathiesen (1985) and Rutherford (1995). Each model variable is represented as a Lagrange multiplier that is matched to a weak inequality representing an equilibrium condition defined by the theory (as we show in 16). Markusen (2021) offers an introduction to this approach. The MCP representation facilitates corner solutions, which are useful in generating full specialization outcomes like those that are important here. The Mathiesen-Rutherford framework uses expenditure functions rather than indirect utility, and contains an implicit market for produced utility, but these are not necessary for model explication.

Consider first bilateral demands for the transport service as in (7). Use $p_{ij}^{Tk} = rD_{ij}^{\rho}$ to generate bilateral demands for the transport service in terms of distance:

$$q_{ij}^{Tk} = \gamma^k \frac{\left((p_i^k)^{1-\gamma^k}(rD_{ij}^{\rho})^{\gamma^k}\right)^{1-\sigma}}{rD_{ij}^{\rho}} \frac{\alpha Y_j}{\widetilde{P}_j^k} \tag{18}$$

Demand for the transport service is rising in γ^k, but falling in distance. (18) produces a parametric negative elasticity of demand for the transport service with respect to distance

$$\frac{\partial q_{ij}^{Tk}}{\partial D_{ij}} \frac{D_{ij}}{q_{ij}^{Tk}} = \left(\gamma^k(1-\sigma) - 1\right)\rho \tag{19}$$

Commodities with larger γ^k's will have bilateral transport demands that are more sensitive to distance. One can interpret the elasticity of transport demand as the product of the demand for the transport service with respect to distance and the elasticity of factor demands with respect to distance.

One can do a similar exercise with (8) to produce bilateral demands for the good itself:

$$\frac{\partial q_{ij}^{k}}{\partial D_{ij}} \frac{D_{ij}}{q_{ij}^{k}} = \left(\gamma^k(1-\sigma)\right)\rho. \tag{20}$$

This elasticity is also negative, and becoming more so as γ^k rises. It can be understood as the product of the cross partial demand of the demand for goods with respect to the transport service price $\frac{\partial q_{ij}^k}{\partial p_{ij}^{Tk}} \frac{p_{ij}^{Tk}}{q_{ij}^k} = \gamma^k(1-\sigma)$ and the elasticity of transport demand with respect to distance $\frac{\partial t_{ij}^k}{\partial D_{ij}} \frac{D_{ij}}{t_{ij}^k} = \rho$. Note that we have generated gravity-like trade pattern in the produced good without the assumption of iceberg trade costs.

As noted above, the own price elasticity of demand with respect to price changes is

$$\frac{\partial q_{ij}^{k}}{\partial p_i^k} \frac{p_{ij}^{k}}{q_{ij}^{k}} = \epsilon^k = \gamma^k(1-\sigma) + \sigma.$$

This is the elasticity of demand facing the firm for its output. While related, it differs from the trade elasticity (20) in this framework.

The essence of the model intuition is that more central countries (as well as relatively larger countries) have shorter average distances to total demand than do peripheral and/or smaller countries. This has implications for the spatial distribution of the relative demands for two products with different transport intensities: as γ^k rises, demands in nearby regions become relatively more important for the output-producing firm. Firms in geographically central countries face relatively larger demands in high-γ^k products. Since high-γ^k products also have lower elasticities of demand, Dixit-Stiglitz preferences and monopolistic competition induce agglomeration economies that are stronger for high-γ^k products, causing production of those products to locate in countries where they face relatively larger local demands - geographically central and/or large countries. Peripheral/small countries specialize in the residual activities - those with low γ^k's. Products with lower γ^k generate lower markups and fewer, larger firms. In special cases where there are relatively few products with high average values of γ^k, the model solution can produce outcomes where all regions produce all products and have the same wages (RCA nonetheless operates according to the vT intuition). When the model is parameterized in a way that produces specialization, weaker agglomeration economies in the periphery produce lower wages there, though the effect is quite small.

2.8 The model as a four-equation system

Strategic substitutions allow us to rewrite the general equilibrium model as a much reduced set of non-linear equations that reproduce the results of the larger general equilibrium model exactly. Fixed relationships between distance, the price of shipping capital and the cost share of transport allow to focus the model on the allocation of labor across products. In the reduced system that solves for the labor allocation, the fixing of the price of shipping capital is sufficient to represent all model choices regarding transport.

The reduced system retains the labor market clearance condition (1) and the income definition equation (3) for each country i. These solve for w_i and Y_i respectively. Substituting (9), (14), (13) and (17) into the definition of $\widetilde{P}_j^k$ returns

$$\widetilde{P}_j^k = \sum_{i'} \frac{\ell_{i'}^k}{a(\sigma + \gamma^k - \sigma\gamma^k)} \left(\frac{(\sigma + \gamma^k - \sigma\gamma^k) w_{i'}}{(\sigma - 1)(1 - \gamma^k)} \right)^{(1-\sigma)(1-\gamma^k)} \left(r D_{i'j}^{\rho} \right)^{\gamma(1-\sigma)} . \quad (21)$$

This equation determines the price index P_j^k. The entry condition that determines n_i^k, (16), can be rewritten to determine ℓ_i^k:

$$w_i \geq \sum_j \frac{1 - \gamma^k}{a(\sigma + \gamma^k - \sigma\gamma^k)} \left(\frac{(\sigma + \gamma^k - \sigma\gamma^k) w_i}{(\sigma - 1)(1 - \gamma^k)} \right)^{(1-\sigma)(1-\gamma^k)} \left(r D_{ij}^{\rho} \right)^{\gamma(1-\sigma)} \frac{\alpha Y_j}{\widetilde{P}_j} \quad \perp \quad \ell_i^k \geq 0 \quad (22)$$

(22) can be interpreted as region i sector k's conditional inverse demand for labor.[17] With T_i's exogenous and r set to 1, we can solve this system of $2(I \times K) + 2I$ equations and $2(I \times K) + 2I$ unknowns. The objects of interest are the ℓ_i^k variables that determine the allocation of labor across sectors and the price indexes $\tilde{P}_j^k$. There are $(I \times K)$ of each of the two variables. The w_i and Y_i terms can be considered fixed in the analysis of any given sector. w_i is the Lagrange multiplier attached to the constraint defined by (1). Once this is determined, the Y_i is calculated with (3). Numerical solutions of this reduced problem are exactly those generated by the full general equilibrium system.

2.9 Model intuition

The key equilibrium relationship driving model outcomes is equation (22), which determines sector employment levels in each country-sector pair with a complementary slackness condition. Equilibrium wages (w_i) are always at least as large as the value marginal product of labor ($VMPL_i^k$). $VMPL_i^k$ jointly depend on regions' proximity to demand and wages, sector γ^k's, as well as the spatial distribution of demand and employment. To illustrate model mechanisms, it is helpful to do a representative numerical exercise with two countries, one large and one small. Because international trade costs are positive, firms in the larger region are closer to final demand. This advantage is relatively larger in sectors where γ^k is large.

To illustrate the way in which this intuition translates into equilibrium labor allocation decisions, we construct a figure in a two-country, twenty-sector setting where the factor endowments of the large country are twice as large as those of the small country. In Figure 1 we graph the ratio of $VMPL_i^k$ to w_i across 20 sectors in the two countries. γ^k varies across sectors, rising from 0.05 to 0.2, and variation in γ^k is depicted along the figure's x-axis.

[17](22) can be understood as a first order Kuhn-Tucker condition from a problem that chooses ℓ_i^k to maximize global payments to labor, subject to the other three conditions in the reduced system of equations.

Figure 1: Ratio of $VMPL_i^k$ to w_i in equilibrium and with sectoral labor proportional to L_i

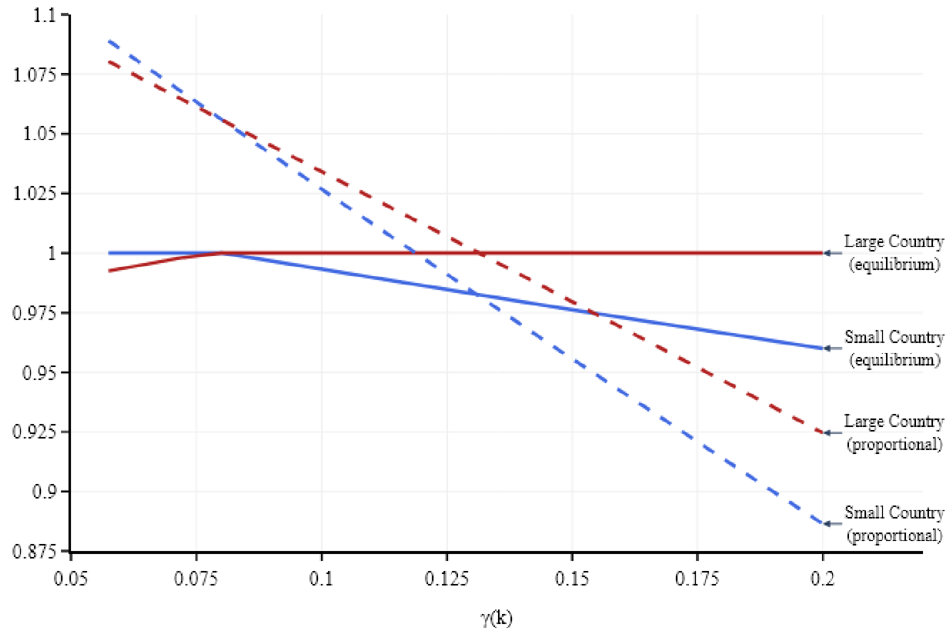

Figure notes: Numerical simulation with two countries and twenty sectors. The large country's factor endowment is five times that of the small country, and its endowment of shipping capital is twice as large. ($a = 2, b = 1, \sigma = 10$ and $\alpha = 1/20$). "Proportional" allocations of labor assign a fixed share of each countries' labor to each sector; $VMPL_i^k$ is calculated with equilibrium wages in both allocations.

The solid lines in the figure represent equilibrium ratios of $VMPL_i^k$ to w_i for both the large and small countries. The importance of (22) is evident in the figure. When wages equal sector VMPL's the ratio is one. In these sectors the country is active in production of sector k. In cases where w_i is less than $VMPL_i^k$, the ratio is less than one. The country would not participate in the production of sector k output in those industries. The figure shows that the equilibrium has the larger country fully specializing in the high-γ^k products, while the smaller country specializes in the low-γ^k products. Both countries participate in the production of a single 'cut-off' commodity in which the participation constraint is satisfied for both countries.

In order to better understand the sources of comparative advantage we also report ratios of $VMPL_i^k$

to w_i for an out-of-equilibrium allocation of labor. Holding fixed country wages, we reallocate labor across sectors so that each country's allocation of labor to each sector is proportional to its total labor supply. Since low-γ^k activities are relatively labor intensive, in equilibrium, the low-γ^k sectors employ relatively more labor. Moving from the equilibrium to the proportional allocations of labor reduces the labor available to low-γ^k sectors, raising both countries' $VMPL_i^k$'s in low-γ^k sectors, relative to the fixed wage. Similarly, allocating more labor to high-γ^k sectors reduces $VMPL_i^k$ in those sectors.

Figure 1 also illustrates a lesson about how revealed comparative advantage emerges in the model. The equilibrium wages used in the illustration are higher in the large country than in the small country. The right hand side of (22) sees both wages and D_{ij} enter with a negative exponent that includes γ^k. In low-γ^k sectors, the small country's lower wage gives it a larger $VMPL_i^k$. The large country has an advantage of closer proximity to final demand (as reflected by the $\sum_j (rD_{ij})^{\gamma^k(1-\sigma)} Y_j$ terms.) As γ^k rises, wages become less important to $VMPL_i^k$, while proximity to final demand becomes more important. We thus have a crossing in the ratio's of $VMPL_i^k$ to w_i under the proportional allocation of labor. For high values of γ^k the ratio is higher in the large country because proximity to final demand is more important than relatively lower wages. When labor is free to move towards the sectors with larger marginal products, there will be a general tendency in both countries to move toward low-γ^k sectors, but large country workers will move away from the lowest γ^k activities entirely, because the smaller country has higher $VMPL_i^k$ in those sectors. In equilibrium, the lowest γ^k sectors employ no workers in the large country. Similarly, workers in the high-γ^k sectors in the small country leave (entirely) because their $VMPL_i^k$ to w_i ratios are lower than in the large country. Only a single 'cut-off' sector with $VMPL_i^k$ to w_i ratios has employment in both countries. The precise allocation of labor across countries in the cut-off sector is that which equalizes the $VMPL_i^k/w_i$ ratios across the two countries.

2.10 Graphical representation of labor market equilibrium

The intuition behind the allocation of labor across sectors and countries is closely related to the HME, as expressed by Hanson and Xiang (2004), among others.[18] We now turn to showing how

[18] Hanson and Xiang (2004) show that products with low elasticities of substitution and high trade costs will tend to locate in the large country. Our γ^k parameter captures the effects of both transport intensity and lower elasticities of demand on location decisions.

the intuition follows to a setting more like von Thünen (1826)'s. We define a geography with 5 equally-endowed regions situated along a line. Distances between regions are calculated as $D_{ij} = 1 + 0.5 * |R_i - R_j|$, where R_i is the ordinal value of region i. Own-region distances are 1. We first consider a restricted version of the model, where labor is allocated across sectors in proportion to each region's total allocation of labor. In order to highlight implications for $VMPL_i^k$ we assume for now that the proportionally allocated labor is specific to its industry. With twenty sectors there are now 20 distinct labor markets in each country, with sector-specific wages w_i^k set equal to $VMPL_i^k$ in each country-sector pair.

In the set-up with proportional allocations of labor, firms in the more geographically central countries benefit from better market access, which raises their $VMPL_i^k$ relative $VMPL_i^k$ in the periphery. As an illustration, consider Figure 2, which shows $VMPL_i^k$ for each country in each of five identically-endowed regions for the two sectors with the highest and lowest values of γ^k. With proportionally allocated labor, $VMPL_i^k$ is everywhere higher in the labor-intensive sector ($\gamma^k = 0.0575$) than in the transport-intensive sector ($\gamma^k = 0.2$). Central regions also have higher values of $VMPL_i^k$ in both sectors than remote regions, since central regions' labor productivity benefits from proximity to final demand. Finally, the different slopes of the $VMPL_i^k$ lines show us that central regions' relative advantage in $VMPL_i^k$ is stronger in the high-γ^k sector than in the low-γ^k sector. Allowing labor to move across sectors will cause labor to move toward high-γ^k sectors in the geographic center, while labor in the periphery will tend to move to the low-γ^k sectors. This is how specialization operates in the model. As the central regions' allocations of labor to higher-γ^k activities rise, labor in those regions/sectors benefit disproportionally from increasing returns to scale, raising $VMPL_i^k$ in the geographically central, high-γ^k country sector pairs. Wages also change as labor moves, and the entry condition in (22) become the relevant condition that determines entry in every country-sector pair. Returns to scale are ultimately limited by the assumption that a fixed share of expenditure α goes to each product k.

Figure 2: $VMPL_i^k$ by region assuming proportional allocations of sector-specific labor

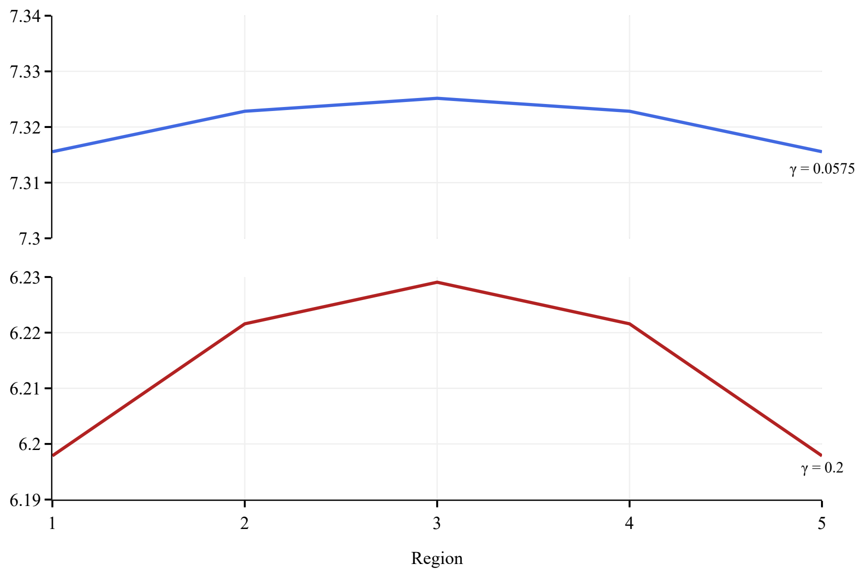

Figure notes: Numerical simulation with five regions and twenty sectors (only high- and low-γ^k sectors illustrated here). Endowments equal across countries and sectors. Sector-specific labor earns sector-specific wages. $D_{ij} = 1 + 0.5 \cdot |R_i - R_j|$, $\rho = 0.15, a = 2, b = 1, \sigma = 5$ and $\alpha = 1/20$.

In order to illustrate the von Thünen-like predictions of the model we provide a representative numerical example of the equilibrium. The numerical simulation is parameterized as follows. Each region is endowed as follows ($T_i = L_i = 10.$). Transport intensity is distributed across sectors according to $\gamma^k = 0.05 + 0.15(k/K)$. There are twenty sectors so $\alpha = 0.05$. Fixed and marginal labor requirements in production are $a = 2$ and $b = 2$, respectively. The elasticity of substitution between varieties is $\sigma = 5$, and $\epsilon^k = \gamma^k - (\gamma^k - 1)\sigma$. The price of transport capital is fixed at $r = 1$. All other prices and quantities are endogenous.

We solve the reduced model for the allocation of labor ℓ_i^k across countries and firms.[19] Figure 3 shows the results. Remote countries specialize in low-γ^k (high-ϵ^k) products. As in von Thünen (1826), countries that are more central produce products that are more transport-intensive. Another

[19] All model variables reach the same values if we solve the complete general equilibrium system.

similarity with von Thünen (1826) is the symmetry; countries that are equally distant from the world produce exactly the same output mix, in exactly the same proportions. $VMPL_i^k$'s are determined by the interaction of the product-specific ϵ^k's and region i's proximity to final demand. Countries that are equally distant to global demand will have equal wages and equal marginal products of labor at those wages, and so they will also have equal allocations of labor. Because we simulate an odd number of regions, the most central region (region 3) alone produces the sectors with the highest γ^k's. These sectors have the largest agglomeration economies (because of the low ϵ^k's firms are smaller so more firms enter). Equilibrium wages are higher in the geographic center, and lower in the periphery. Cross-country wage gaps are typically quite small in numerical terms; just large enough to induce specialization via (22).

Figure 3: Equilibrium allocation of labor across countries and sectors.

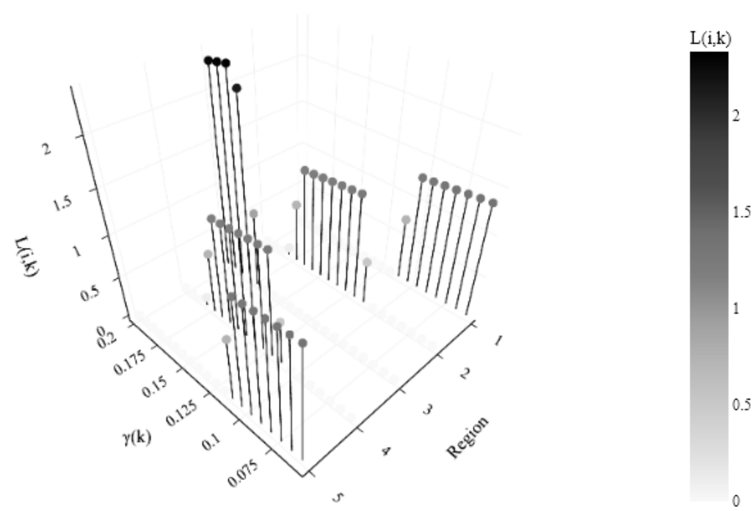

Figure notes: Numerical simulation with 5 countries and 20 sectors. Endowments equal across countries. $D_{ij} = 1 + 0.5 \cdot |R_i - R_j|$, $\rho = 0.15, a = 2, b = 1, \sigma = 5$ and $\alpha = 1/20$.

While most products are produced in only one type of region or another there are some "cut-off" products produced in two bordering countries. Although it is not evident in the figure, the value marginal product of labor in these products would equal to the wage in the region in which it is produced. Finally note that while countries 1 and 2 (and 4 and 5) share only a single "cut-off" product region 3 shares two such products with countries 2 and 4. This occurs because the wage gap between bordering countries is small, so more than one product satisfies (22) with positive values of ℓ_i^k in neighboring countries.

2.11 Model predictions for RCA

In order to illustrate the implications of the model for revealed comparative advantage, we consider two geographies: a two-country geography in which one country is larger than the other, and a three-country geography with three equally-sized countries with a central country located halfway between the other two.

- Comparative advantage of large countries in transport-intensive sectors, with smaller countries specializing in the activities that are less transport-intensive.
- Comparative advantage of central countries in transport-intensive countries, with remote countries specializing in activities that are not transport-intensive.
- As in von Thünen (1826), the model predicts symmetric composition of trade for two countries that are equidistant to global demand.

2.12 Counterfactual experiments

Aside from questions about the existence of equilibrium and the model's implications for regional specialization, the question of how the model responds to shocks is an important and interesting one. It is particularly useful to know how developments in the shipping sector are likely to affect the spatial allocation of production.

In order to explore these questions we implement two shocks in the computational model. First, we consider a situation where there is growth in the stock of transport capital. The large subsidies that China has provided its shipbuilders would be represented in the model in this way.[20] We simulate a proportional increase in T_i across all countries. Since the price of shipping capital is the numeraire, the increase in T_i does not affect r. Wages, which are being compared to a fixed r, rise

[20] Kalouptsidi (2018) estimates the scale and effectiveness of these subsidies, and finds that they have substantially increased the supply of global bulk shipping capacity.

in direct proportion to the size of the transport capital shock. Incomes also rise in proportion to the shock. The allocation of productive activities is unaffected by the shock. Since shipping capital has become cheaper, there is a less-than-proportionate increase in the true-cost-of-living index, so easing the transport capital constraint increase utility across the globe. These latter effects are slightly stronger in the periphery, but the effects are not large. The lesson is that an increase in the global stock of transport capital is useful for all consumers, but slightly more so for those in the periphery, because they are net purchasers of transport-intensive products. the implication is that China's subsidized construction of bulk shipping capacity was more beneficial for countries in the periphery than for those in the center.

We also consider a shock to ρ, which is the distance elasticity of factor input requirements in the transport sector. A reduction in ρ represents distance-biased technical change. It might reflect new or larger maritime vessels or developments in air transport that allow longer flights between refueling.[21] The effect of reduced ρ on wage levels is small, but there is a slight compression of the small wage gap between core and periphery. The cross-country distribution of nominal income also compresses slightly. This is an effect, it seems, of the periphery becoming slightly more viable in the marginal good. The price of utility falls somewhat because the price of transport has fallen, and welfare rises slightly, and slightly more in the periphery than in the center.

In short, the effects of the two shocks appear to be rather similar. Both the increase in the capital stock and the distance-biased technical change improve welfare, with slightly larger benefits to the periphery. Neither shock materially changes the pattern of RCA. Specialization of the periphery in low γ^k activities is robust to changes in the stock of transport capital and to transport technology shocks.

2.13 Model extension

We can extend the model to incorporate other reasons for commodity specialization in the periphery. The most notable of these is the inclusion of a role for relative factor supplies and factor prices. If the periphery is better endowed with agricultural land and/or minerals, then the observation that primary products or lightly processed versions thereof are exported from the periphery might be

[21] One might also wish to consider it as representative of the consequences of reduced oil prices, though we do not explicitly model fuel as an input. Our experiment considers long-run changes that reduce the elasticity of transport costs to distance.

explained by other factors, a topic explored in Venables and Limao (2002).

In order to understand this question in our increasing returns to scale setting we include a role for factor proportions in production (transport continues to employ only the transport capital factor). We introduce a subscript f on the factor of production L_i, and vary the relative abundance of factors across geographies. We replace the the factor usage equation $\ell_i^k = n_i^k(a + bq_i^k)$ with

$$Z_i^k = n_i^k(a + bq_i^k) \tag{23}$$

where Z_i^k is a factor input composite assembled with Cobb-Douglas technology. The variable associated with (23) is PZ_i^k, the price of input bundle Z_i^k. The zero profit condition for assembling Z_i^k is

$$\prod_f (w_i^f)^{\mu^k} \geq PZ_i^k, \quad \sum_k \mu^k = 1 \tag{24}$$

where the left hand side is the cost function for assembling the factor bundle and μ^k the cost share of factor f in the input bundle.

Suppose there are two factors of production with $f = 1$ being Labor and $f = 2$ being Land. And suppose that land intensive products have lower transport costs than labor intensive manufacturing. In this case, relatively larger endowments of land in the global periphery will produce the specialization pattern we observe, even without the cost differences. Von Thünen forces will magnify the Heckscher-Ohlin forces, and vice versa. When we ask the question that Venables and Limao (2002) ask - can cross-product variation in transport costs overturn comparative advantage? - we do not find evidence that it can in this model. Transport cost differences can blunt comparative advantage, but we have yet to find a scenario where they overturn it. This outcome appears to arise because the transport intensity parameter is tied to the elasticity of demand facing the firm, and thus the strength of sector-level agglomeration economies.

3 Revealed comparative advantage and the PPML regression

Our primary empirical exercise is the estimation of the cross-country determinants of RCA. Following Balassa (1965), country i's RCA in product k is the share of product k in country i's exports divided by the share of product k in world exports. Our notation is:

$$RCA_i^k = \frac{X_i^k}{\overline{X}_i} \Big/ \frac{X_w^k}{\overline{X}_w} \tag{25}$$

where X_i^k is the value of exports of product k from region i, $\overline{X}_i$ is the total value of exports of region i, and X_w^k and $\overline{X}_w$ are counterparts for the world. The literature on the determinants of comparative advantage typically seeks to understand variation in (25) by taking the log of the formulation, and estimating with a linear combination of industry- and country-fixed effects, together with the interaction of industry and country characteristics. Our approach is conceptually equivalent, except for two main differences.

First, we replace industry characteristics with characteristics of HS6-digit products as defined in the international trade statistics. This gives us more data than exercises that rely on aggregations to industry levels to match data in input-output tables. The drawback is that we are unable to use input-output tables to construct precisely some interaction variables that appear elsewhere in the literature. The transport characteristics of products are best measured/inferred over disaggregated commodities, so we view this as preferable to the industry level aggregations used to test other theories of RCA.

The second difference with much of the existing literature is our choice to use Poisson Pseudo Maximum Likelihood (PPML) estimator rather than the more common log-linear OLS regressions. Our PPML specification is as follows:

$$X_i^k = exp\left[\alpha_i + \alpha^k + \mathbf{C}_i\mathbf{Z}^k\boldsymbol{\Upsilon}\right] + \epsilon_i^k \tag{26}$$

where α_i and α^k are country- and sector- fixed effects, $\mathbf{C}_i$ a vector of country characteristics, $\mathbf{Z}^k$ a vector of product characteristics, $\boldsymbol{\Upsilon}$ a vector of coefficients on the country-sector interactions, and ϵ_i^k is a mean zero error term.

PPML has properties that make it better suited to estimation of RCA than log-linear OLS. First, the PPML specification allows the inclusion of zero values of trade in the estimation. Most studies in the literature on the determinants of comparative advantage estimate with log-linear versions of (26). Log-linear specifications drop products with zero exports from a given country, an outcome that likely generates selection bias. Selection bias may be an especially important problem in RCA estimation because most countries export small numbers of products. Since we estimate at the product rather than the industry level we have many more zeroes, making the selection problem in log-linear OLS much worse.

A second advantage of the PPML estimator is that the fixed effects in the PPML specification impose important adding up conditions, conditions which are quite useful in the specific context of estimating RCA indices. Fally (2015) shows that, in PPML estimation, the estimated coefficients associated with dummy variables take values that impose that the sum of the fitted values associated with the dummy variable equal the sum of the observed data associated with the dummy variable. This property gives the estimated fixed effects important meaning in the context of estimation of RCA. Using (26) as an example, the $\widehat{\alpha}_i$ terms take values such that $\sum_k \widehat{X}_i^k = \overline{X}_i$, and the $\widehat{\alpha}^k$ terms take values such that $\sum_i \widehat{X}_i^k = X_w^k$. In the context of a cross country RCA estimation, these adding up conditions imply that each country's $\widehat{\alpha}_i$ term effectively controls for country i's total participation in exporting $\overline{X}_i$, while each product's $\widehat{\alpha}^k$ efficiently controls for it's role in world exports X_w^k. In this way, the explained variation in the trade flows $\left(exp\left[\mathbf{C}_i\mathbf{Z}^k\widehat{\mathbf{\Upsilon}}\right]\right)$ matches the variation in the RCA index, since the variation due to $\overline{X}_i$ and X_w^k have been exactly purged by estimation that satisfies adding up. These adding-up conditions of the PPML specification do not obtain for the log-linear regressions commonly used in the RCA literature, because of Jensen's inequality.

A related property of PPML estimation recognized in Fally (2015) is that PPML specifications act (implicitly) as weighted regressions, with weights proportional to the dollar values of $\widehat{X}_i^k$. Fally explains that log-linear regressions implicitly apply weights of $\widehat{ln(X)}_i^k$. Log-linear specifications overweight small trade flows (relative to their contribution to world trade), and underweight large flows, generating bias. While any weighting scheme has its advantages and disadvantages, weights that are proportional to the dollar value of trade are particularly useful in estimation of RCA, since our specifications consider both within-product (cross-country) variation and within-country (cross-product) variation. The PPML estimator's consistent US-dollar weighting insures country-size variation and product-size variation are treated consistently in the estimation. This property is important because it weights more heavily the top products in a countries' trade. Chile's fitted

RCA in copper in a PPML regression is stronger than it would be in an OLS regression. A product in which Chile exports much less, but still more than the average product (e.g. the fifth largest export product, cherries), will get less weight in an OLS than in a PPML specification.[22]

4 Data

Our empirical estimates include two types of regressions. We first estimate preliminary regressions that link our transport intensity measure to the associated transport mode. We then estimate a model of export specialization as a function of products' transport characteristics. Here we describe the data we use in each exercise.

4.1 Data for transport intensity/mode regressions

Our preliminary regressions characterizing the relationship between transport mode and transport cost shares is estimated over a sample of 2017 U.S. imports that arrived by air or sea. We exclude import flows from Mexico and Canada and those arriving in Puerto Rico. We take this sample from the U.S. Merchandise Imports database of the U.S. Census.[23] This data reports the value (in USD), weight (in kg.) and freight charges (in USD) incurred on every import flow arriving in the U.S. These data are also available at the combination year, HS10-digit product code, country of origin, U.S. customs district of entry, and transport mode (air or vessel, containerized or not).

In order to be comparable with the international regressions, we aggregate products to the level of HS6-digit product code, country of origin, and U.S. customs district of entry. The LHS variable is the share of freight charges in the value of U.S. imports (inclusive of these charges). We wish to relate this fraction across commodities and the mode by which imports were shipped to the U.S. Our independent variables focus on the mode of transport, including whether or not the import arrived in a shipping container. For each triplet defined by export country, HS6-digit product and U.S. customs district of entry we construct 1) the value share arriving by air; 2) the share of a triplets' imports that arrive in containers; 3) a dummy variable equaling 1 for fuel and related products (a proxy for shipping by liquid vessels and tankers); and 4) a dummy variable equaling to

[22] It is important to note that log weights used elsewhere have no *a priori* rationale, while the PPML weights treat one dollar's worth of exports the same across all products.

[23] We retrieve annual files from U.S. import data from Peter Schott's web page: https://sompks4.github.io/sub_data.html.

1 for finished transport equipment (a proxy for movement by roll-on/roll-off (or RoRo) ships.[24] Our estimates also include common cost shifters for freight charges: 1) the log of the weight-to-value ratio (WV) for each imported HS6 product, country-of-origin, U.S. customs-district triplet, and 2) the log of the great circle distance from a country of origin to a given U.S. customs district.[25]

4.2 Data for export composition regressions

Our empirical model of RCA uses three sorts of data: 1) global export flows at the country-product level; 2) country-characteristics; and 3) data on product characteristics. The value and weight of each country's exports are calculated at the HS6-digit level.

Global export flows are taken from the BACI database, which we source from the BACI database - sourced from CEPII (2024). We use data from 2017. This sample contains data from 156 regions/countries and 4,620 HS6-digit products.[26] In order to estimate without missing data we use the set of regions/countries and products for which data from all country and product control variables are available.[27] We impute zero trade values for country product-pairs in which no exports are observed.

[24] In particular, the dummy variable "*Dummy Tankers*^{k}" equals 1 for products with HS4-digit codes equal to *2709 - Petroleum oils and oils obtained from bituminous minerals, crude*, *2710 - Petroleum oils and oils obtained from bituminous minerals, other than crude;* and *2711 - Petroleum gases and other gaseous hydrocarbons*. The dummy variable "*Dummy RoRo*^{k}", equals 1 for finished wheeled vehicles for which both the air share and the share shipped in containers is less than 10 percent. In our database, these products have HS4 digit codes *8429 - Self-propelled bulldozers, angledozers, graders, levelers, scrapers, mechanical shovels, excavators, shovel loaders, tamping machines and road rollers*, *8602 - Other rail locomotives*, *8604 - Railway or tramway maintenance or service vehicles*, *8701 - Tractors*, *8702 - Motor vehicles for the transport of ten or more persons*, *8703 - Motor cars and other motor vehicles principally designed for the transport of persons*, *8704 - Motor vehicles for the transport of goods*, and *8705 - Special purpose motor vehicles*. Additionally, these products have HS4 digit code *430 - Other moving, grading, leveling, scraping, excavating, tamping, compacting, extracting or boring machinery*, *8601 - Rail locomotives powered from an external source of electricity or by electric accumulators*, *8603 - Self-propelled railway or tramway coaches, vans and trucks*, and *8605 - Railway or tramway passenger coaches, not self- propelled.* This latter set of products are similar to those above after checking by hand the Harmonized System of products.

[25] To calculate these distances, we use the GPS coordinates of every place that we retrieved from https://simplemaps.com/data/world-cities.

[26] Because of our interest in characterizing how products' transport intensity and countries' geographical location explain countries' trade specialization, where possible, we include data for regions such as Greenland that are not independent states but BACI reports trade data. We also exclude data from regions/countries that the BACI database seems to report incompletely their trade data. These are regions/countries for which we estimate the BACI database reports an export and import trade share of less than 5 percent with their parent country; a level of trade we consider reasonable between countries with a close relationship. We cannot correctly estimate all models with a large share of trade missing for some countries.

[27] We estimate over a fuller sample to ensure that the main relationships we estimate in this model are not affected by our use of this limited sample. Most of the missing data are from countries with low values of exports (which means that they have low weight in the PPML regressions). The larger sample includes data from 202 regions/countries and 5,146 HS-digit products.

Our primary country characteristics of interest are logged market access, population and per capita income. We construct our market access variable following Redding and Venables (2004), but use the (PPML) estimator:

$$X_{ij} = \exp\left(\omega_i + \omega_j + \phi_1 \ln(\mathrm{DIST}_{ij}) + \phi_2 \mathrm{CONT}_{ij}\right) + \epsilon_{ij} \tag{27}$$

where X_{ij} is the (aggregate) value of bilateral trade exports from i to j, ω_i and ω_j are origin- and destination- fixed effects capturing the scale of region i's exports and region j's, respectively, DIST_{ij} the distance between i and j, CONT_{ij} a dummy variable taking the value of 1 when countries i and j are contiguous, ϕ_1 is the distance elasticity of trade, ϕ_2 the geographical contiguous effect of trade, and ϵ_{ij} is a mean zero error term. Our estimates indicate that $\widehat{\phi_1} = -0.633$ and $\widehat{\phi_2} = 0.803$. Market access of exporting country i is calculated as $MA_i = \sum_j \exp(\widehat{\omega_j} + \widehat{\phi_1} \ln(\mathrm{DIST}_{ij}) + \widehat{\phi_2}\mathrm{CONT}_{ij})$. Country-level data on GDP per capita and population are taken in most cases from CEPII, but we replace missing observations with data from the World Development Indicators (WDI), International Monetary Fund (IMF) or United Nations (UN), as available. We use 2015 data for both measures, so that they may be taken as exogenous for our 2017 trade flows.

The product-characteristic data that are our primary interest are the transport modes most commonly used to deliver each product to the U.S. (exclusive of shipments from Canada or Mexico). The underlying data source is the U.S. Merchandise Import database of the U.S. Census. The "Air share" and "Container share" variables we calculate are quite similar to those calculated for our preliminary regressions, except that in this case we collapse over U.S. customs district and origin countries to create aggregate statistics for each product. We take these statistics as product characteristics that we map onto non-U.S. exports.

In a robustness check, we replace our transport mode share variables with bin dummies in which products' bin membership is based on their weight-to-value (WV) ratio. We use U.S. import data to calculate aggregate WV. We focus on separating products with extreme WV values from the others. Products in the top 10 percent of the distribution of WV ratios in U.S. data are coded with a high-WV dummy; products in the bottom tenth percentile are coded with a dummy indicating low WV. These high- and low-WV bins are proxy variables meant to loosely predict products that are likely to be shipped by bulk ships and air shipment, respectively.

5 Results

5.1 Summary Statistics

5.1.1 Transport cost shares by mode

The preliminary regressions that we estimate to illustrate different values of transport intensities by transport mode use data from 5,202 HS6-digit products, 230 countries of origin, and 41 U.S. customs districts. Panel A in Table 1 reports the summary statistics for these regressions. We leave inspection of these to the reader, but note some key outcomes.

1. The transport share of delivered cost is only 0.14 at the 90th percentile of the distribution. While the maximum value is much bigger (0.94), the mass of the data is clearly at the low end of the distribution. These data motivate the parametrization of our numerical simulations above, where the highest transport intensity parameter is 0.2.

2. The mean air share of shipment is 0.49, while the mean container share is 0.46. These two modes account for most of the transportation in U.S. imports.

3. Our indicators for the likely use of Roll-on-Roll-off vessels and tanker vessels each apply to less than one percent of the sample.

5.1.2 Summary statistics for export composition regressions

Summary statistics for the export composition regressions are in Panel B of Table 1. We largely leave inspection of these data to the reader. The key country characteristics of interest are (logged) market access $ln(MA_i)$, population $ln(Pop_i)$ and per capita income $ln(GDPpc_i)$. The latter two are familiar, but it may be useful to know that countries with high market access scores are Canada, Belgium, Luxembourg, the Czech Republic and France, while countries with the least market access are Fiji, New Zealand, Sao Tome and Principe, Australia and Comoros.

In this section, we report the regression results from our two empirical exercises. We first conduct preliminary regressions with the aim of describing the relationship between transport mode and a measure corresponding to γ^k in our model, the freight share of cost in delivered value. We subsequently report estimates for our primary specification of interest, equation (26).

5.2 Preliminary regression results

We estimate preliminary regressions in order to establish that transport intensities of products are related to the transport mode in which the product usually travels (on intercontinental voyages). Specifically we want to show that the cost share of transport in delivered cost (which corresponds to γ^k in our theory) varies systematically with mode of transport. We establish this both in levels, and with interactions of transport mode with bilateral distances. The results of these regressions are reported in Table 2.

Columns 1 and 2 show that a) air-shipped goods have higher transport cost shares than container-shipped goods, b) container-shipped goods have higher cost shares than goods not shipped in containers, and c) longer distance shipments have higher transport cost shares. Column 2 includes our dummy variables for other forms on non-containerized freight, and the cost shares for these shipments are not statistically different than those of other non-containerized freight. Including these dummies does not affect the lessons from Column 1.

Column 3 shows that all forms on non-bulk freight see transport cost shares rise more rapidly in distance than the excluded category of bulk freight. These effects are moderated when we include a WV control in the regression in column 4. This result reflects strong selection effects of commodities into mode according to WV ratios. Column 5 includes an unreported interaction between WV and transport mode share as a further control. Products shipped by air, container vessels and tanker vessels continue to see their transport cost share rise faster with distance than the cost share of the excluded category, bulk vessels.

Our lesson from these regressions is that an interpretation of our model in which different forms of transport capital imply different transport intensities of the products they carry. In the next set of regressions we show that these differences are manifested in global patterns of export specialization.

5.3 Transport intensity and RCA

We estimate (26) to establish that the product composition of countries' exports is determined by an intersection of a product characteristic - the products' most appropriate mode of intra-continental shipping - and a vector of country characteristics. We are particularly interested in the interaction of products' transport characteristic proxies with our market access variable, but we include population size and per capita income because our model also generates predictions with respect to these variables. We estimate with the three mode shares available in the data, with a broader set of shipping types, and then with a series of other controls for comparative advantage. We need to split the results into two tables to fully illustrate the many interesting results of these regressions. We report our main estimates of interest in Table 3, and the results for our control variables in Table 4.

Column 1 of Table 3 shows the results of a regression that only includes our primary variables of interest. Non-containerized maritime freight is the excluded category. Products that tend to be exported by air or container vessel are more likely to be exported by countries that are large or geographically central. In addition, products most suitable for air shipment are more likely to be exported by high-income countries.

Column 2 includes interactions with dummy variables that indicate products associated with particular kinds of non-containerized - but also non-bulk - maritime transport. These regressions reveal that products commonly shipped in tanker vessels are most commonly exported by countries with low populations and low per capita incomes. Motorized vehicles - those most likely to be transported by RoRo vessels - are most commonly exported by geographically central, large-population, and high-income countries. Coefficients on our main variables of interest maintain the same sign pattern.

Column 3 includes interactions of countries' WTO member status with products' transport characteristic proxies. The inclusion of the WTO controls maintains the same sign patterns on our variables of interest, but per capita income becomes less important as a country-characteristic. Results in Table 4 show that WTO members export conditionally higher values of products commonly transported by air, and conditionally lower values of products shipped in tanker vessels.

Column 4 includes interactions of the Upstreamness measure with country characteristics. The inclusion of this control is intended as an acknowledgement that production-line position matters for RCA, as demonstrated by Antràs and de Gortari (2020). the inclusion of the Upstreamness interactions leave coefficients on the main variables of interest largely unaffected, though levels of statistical significance are marginally reduced in some cases. Table 4 results confirm Antràs and de Gortari (2020)'s theory that more remote countries specialize in upstream activities. The results also show that low population countries export relatively more upstream products. The evidence confirming the Antràs and de Gortari (2020) result, together with the consistency of our main results, offers fairly persuasive evidence that the von Thünen mechanism operates even after controlling for the effects of products vertical position in supply chains.

In column 5 we include interactions that have proven important elsewhere in the literature: interactions of relative factor intensities with relative factor abundances, and interactions between the rule of law and measures of contract intensity. The coefficients on the main variables of interest do not change in their statistical significance. Results in Table 4 show that comparative advantage based on factor endowments and institutional quality is evident in these data too. The von Thünen channel remains significant even in the presence of these controls.

Since most of the extant literature on RCA are estimated solely with manufacturing data, we estimate with that limited sample. The results - reported in column 6 - are much weaker than in the full sample. The exclusion of agricultural and mining products from the sample gives sharply reduces the number of products shipped by bulk vessels, and so limits the ability of the other transport characteristic variables to appear significant relative to the excluded variable. The location of products shipped in containers in large and geographically central countries holds up nonetheless. Due to the lack of most bulk products from this regression, we do not view these results as particularly important for our thesis.

The link between our PPML specification and RCA relies on the adding up properties of PPML noticed by Fally (2015). The results reported so far do not satisfy these properties because a) the properties only strictly apply for a balanced data set that includes a square matrix of countries, and b) we have excluded imports into the U.S. because our product characteristics were calculated in U.S. data. In unreported results we find that estimating the model with U.S. imports returns the same sign patterns as we observe here. Remote and low population countries appear to have RCA in products shipped in non-containerized maritime vessels, even after our battery of controls for other explanations.

5.3.1 Export composition and WV

Our regressions so far proxy products' transport intensity with the shares of particular transport modes used in each products' U.S. imports (exclusive of imports coming from Canada and Mexico). Some readers might find that assignment of transport intensity unintuitive, because overland transport is an important part of most countries' trade. Our assignment is not a statement about actual mode choices made in each country, we are asserting that the mode choices made on U.S. imports by air and sea are indicative of product characteristics that affect their transport intensity. In the context of our model, the idea is that stocks of different forms of transport capital are fixed at any given moment of time, and production will be organized so that remote countries produce the goods that will be shipped by modes that provide relatively less service at lower transport costs per unit of distance. If the specific mode choices made in U.S. import data are not representative of products' transport intensity, then the assignment of specific U.S. mode shares as product characteristics may not be informative.

Our alternative is to use a product characteristic that is directly measurable: the product's weight-to-value ratio product. Dispensing with mode shares, we take WV (again from US import data) and group products into bins (as described in the Section 4). We interact these bins with the country characteristics of interest, and observe how these relationships hold up when our vector of interactive controls are added to the regression. These results are reported in Tables 5 and 6.

The results in Table 5 focus on WV interactions with our country characteristics of interest. The "Light products" dummy represents products likely to move by air shipment. The "Mid Products" dummy is a proxy category for container shipping, and the (excluded) "Heavy Products" dummy is associated with bulk shipping. The results are largely consistent with the results that use mode shares. Physically light products are exported from large-population and high-income countries, middle weight products are exported from geographically central, large-population and high-income countries. The residual category - heavy products - are exported from small, low income and peripheral countries. The viability of heavy products from remote countries depends on modes with lower transport costs, modes that offer fewer transport-related services. These relationships hold up when all of our control variables are added. Column 6 regressions that exclude agricultural and mining products produce coefficients that are less often statistically significant, most likely because agricultural and mining products are such an important component of export composition in the periphery.

The sign patterns on the vector of control interactions reported in Table 6 are almost exactly identical to those estimated when transport mode shares were the variables of interest. We also estimate the full WV model with U.S. imports included, making the link between the PPML regression and RCA more explicit. There is little effect at all of including the U.S. data.

6 Conclusion

A prominent and long-standing feature of the global trade pattern is the specialization of remote countries in the exports of primary commodities or lightly processed versions thereof. Key questions for policy are if, and how, countries that are specialized in this way can develop by moving through the product space, as many countries in East Asia have successfully done. In an era where industrial policy is again *en vogue*, how easily can remote commodity exporters develop this way? In our view it is likely that the scope of the potential product space for export specialization is potentially limited in geographically disadvantaged countries.

While other channels are also important, we propose a model that explains limits on the ability of remote countries to specialize in transport-intensive goods. The model follows the logic of von Thünen (1826) and more recently Venables and Limao (2002), two models that link revealed comparative advantage to interaction between country-level variation in relative distance of countries to destination markets and product-level variation in the costs of transport. We propose a fully general equilibrium model in which products vary in their transport intensity. Monopolistically competitive firms choose where to produce their variety of each good. When the cost shares of transport are sufficiently low, we replicate the strong specialization results of von Thünen (1826). For high average transport intensities, the model produces less-than-complete specialization, though the model's intuition remains the same: more transport intensive products are exported from geographically central countries, while geographically remote countries specialize in less transport-intensive goods.

The model has similar predictions about the role of country-size. Other things equal, large countries are closer to global demand, and therefore specialize in transport-intensive activities. Small countries are left to produce less transport-intensive products. These predictions mimic those of the home market effect literature, and our model produces them in a similar way. Our model illustrates the link between HMEs and specialization a la von Thünen.

In our empirical exercises we employ a PPML estimating model to explain cross-country variation in RCA. We argue that products have physical and/or market characteristics that largely determine their mode of shipment in long-distance transport. We measure the degree to which products are shipped by air transport, shipping containers or other kinds maritime transport in long-distance U.S. imports. We take these are proxies for the importance of transport intensity for each product, and map these measures onto national export bundles. We find evidence that small/remote countries tend to specialize in products that are not shipped by air transport or in shipping containers. We include a number of explanations for comparative advantage that have been proposed in this literature. The inclusion of these control variables largely confirms other evidence presented in the literature, even as the transport characteristic variables remain important for explaining variation in RCA. We redo the exercises with bins based on products weight-to-value ratios in place of the transport mode measures, and find that our evidence is robust to that change.

Our evidence suggests that small and remote countries face important constraints on their ability to develop via manufacturing successively more complex goods, a pathway suggested by Hausmann and Klinger (2006). Our framework points to grounds for hope that peripheral countries may have opportunities to develop via increased exports of internet-related services, where transport margins are non-existent or virtually so. This would be more useful if there are significant economies of scale in services exports. In our framework, growing incomes in the periphery would increase the share of transport capital that arrive there. The arrival of more transport vessels should provide greater opportunities to export more complicated goods, a related implication that is not in the model because we do not model back-hauling.

References

Antràs, P. and D. Chor (2013). Organizing the Global Value Chain. *Econometrica 81*(6), 2127–2204.

Antràs, P. and A. de Gortari (2020). On the Geography of Global Value Chains. *Econometrica 88*(4), 1553–1598.

Balassa, B. (1965). Trade Liberalisation and "Revealed" Comparative Advantage. *The Manchester School 33*(2), 99–123.

Balistreri, E. J. and T. F. Rutherford (2013). Computing general equilibrium theories of monopolistic competition and heterogeneous firms. In *Handbook of computable general equilibrium modeling*, Volume 1, pp. 1513–1570. Elsevier.

CEPII (2024). BACI: International Trade Database at the Product-Level at HS2002. Centre d'Etudes Prospectives et d'Informations Internationales. Available from https://tinyurl.com/ycpycu9k accessed in February, 2024.

Debaere, P. (2014). The Global Economics of Water: Is Water a Source of Comparative Advantage? *American Economic Journal: Applied Economics 6*(2), 32–48.

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

Evans, C. L. and J. Harrigan (2005). Distance, time, and specialization: Lean retailing in general equilibrium. *American Economic Review 95*(1), 292–313.

Fally, T. (2015). Structural Gravity and Fixed Effects. *Journal of International Economics 97*(1), 76–85.

Francois, J. F. and I. Wooton (2001). Trade in international transport services: the role of competition. *Review of International Economics 9*(2), 249–261.

Hanson, G. H. and C. Xiang (2004). The home-market effect and bilateral trade patterns. *American Economic Review 94*(4), 1108–1129.

Hausmann, R., J. Hwang, and D. Rodrik (2007). What You Export Matters. *Journal of Economic Growth 12*, 1–25.

Hausmann, R. and B. Klinger (2006). Structural transformation and patterns of comparative advantage in the product space. *Center for International Development and KSG Faculty Research Working Paper Series (CID-128 and RWP06-041), September*.

Hidalgo, C. A. and R. Hausmann (2009). The Building Blocks of Economic Complexity. *Proceedings of the National Academy of Sciences 106*(26), 10570–10575.

Kalouptsidi, M. (2018). Detection and impact of industrial subsidies: The case of chinese shipbuilding. *The Review of Economic Studies 85*(2), 1111–1158.

Krugman, P. (1980). Scale Economies, Product Differentiation, and the Pattern of Trade. *The American Economic Review 70*(5), 950–959.

Krugman, P. and A. J. Venables (1995). Globalization and the Inequality of Nations. *The Quarterly Journal of Economics 110*(4), 857–880.

Krugman, P. and A. J. Venables (1996). Integration, Specialization, and Adjustment. *European Economic Review 40*(3-5), 959–967.

Levchenko, A. A. (2007). Institutional Quality and International Trade. *The Review of Economic Studies 74*(3), 791–819.

Markusen, J. R. (2021). Global comparative statics in general equilibrium: model building from theoretical foundations. *Journal of Global Economic Analysis 6*(2), 86–123.

Mathiesen, L. (1985). Computation of economic equilibria by a sequence of linear complementarity problems. *Economic equilibrium: model formulation and solution*, 144–162.

Norris, J. J. (2025). A many-location home market effect and a home biased geography. *Journal of International Economics 154*, 104057.

Nunn, N. (2007). Relationship-specificity, Incomplete Contracts, and the Pattern of Trade. *The Quarterly Journal of Economics 122*(2), 569–600.

Redding, S. and P. K. Schott (2003). Distance, skill deepening and development: will peripheral countries ever get rich? *Journal of development economics 72*(2), 515–541.

Redding, S. and A. J. Venables (2004). Economic Geography and International Inequality. *Journal of International Economics 62*(1), 53–82.

Reed, T. (2024, November). Export-led industrial policy for developing countries: Is there a way to pick winners? *Journal of Economic Perspectives 38*(4), 3–26.

Rodrik, D. (2006). What's so Special About China's Exports? *China & World Economy 14*(5), 1–19.

Romalis, J. (2004). Factor Proportions and the Structure of Commodity Trade. *American Economic Review 94*(1), 67–97.

Rutherford, T. F. (1995). Extension of gams for complementarity problems arising in applied economic analysis. *Journal of Economic Dynamics and Control 19*(8), 1299–1324.

Venables, A. J. and N. Limao (2002). Geographical Disadvantage: a Heckscher–Ohlin–von Thünen Model of International Specialisation. *Journal of International Economics 58*(2), 239–263.

von Thünen, J. (1826). *Der Isolierte Staaat in Beziehung auf Landtschaft und Nationalokonomie.* Hamburg.

7 Tables

Table 1: Summary statistics

	Panel A: Transport intensity by mode							
	# Obs	Mean	Std. Dev.	Min	Perc. 10	Median	Perc. 90	Max
f_{ij}^k/M_{ij}^k	678,293	0.06	0.07	0.00	0.01	0.04	0.14	0.94
Air Share_{ij}^k	678,293	0.49	0.47	0.00	0.00	0.40	1.00	1.00
Ctnr Share_{ij}^k	678,293	0.46	0.47	0.00	0.00	0.29	1.00	1.00
Dummy Tankers^k	678,293	0.00	0.04	0.00	0.00	0.00	0.00	1.00
Dummy RoRo^k	678,293	0.00	0.07	0.00	0.00	0.00	0.00	1.00
$\ln(\text{WV}_{ij}^k)$	678,293	-3.19	1.94	-16.75	-5.74	-3.09	-0.81	7.04
$\ln(\text{DIST}_{ij})$	678,293	9.09	0.37	3.32	8.73	9.10	9.52	9.85

	Panel B: Product and Country Characteristics							
	# Obs	Mean	Std. Dev.	Min	Perc. 10	Median	Perc. 90	Max
Air Share^k	720,720	0.20	0.29	0.00	0.00	0.06	0.73	1.00
Ctnr Share^k	720,720	0.70	0.32	0.00	0.12	0.84	0.99	1.00
Dummy Tankers^k	720,720	0.00	0.04	0.00	0.00	0.00	0.00	1.00
Dummy RoRo^k	720,720	0.01	0.11	0.00	0.00	0.00	0.00	1.00
Dummy Light Products^k	720,720	0.10	0.30	0.00	0.00	0.00	0.00	1.00
Dummy Mid Products^k	720,720	0.80	0.40	0.00	0.00	1.00	1.00	1.00
$\ln(\text{MA}_i)$	720,720	-1.51	0.29	-2.04	-1.83	-1.55	-1.12	-0.62
$\ln(\text{Pop}_i)$	720,720	16.07	1.73	11.60	13.58	16.12	18.19	21.04
$\ln(\text{GDPpc}_i)$	720,720	8.58	1.46	5.75	6.50	8.54	10.66	11.57
WTO_i	720,720	0.90	0.29	0.00	1.00	1.00	1.00	1.00
Upstreamness^k	720,720	2.35	0.88	1.00	1.28	2.41	3.55	4.65
Capital_i	720,720	98.78	96.31	1.82	6.97	63.49	263.66	419.57
Land_i	720,720	0.50	0.50	0.00	0.06	0.34	1.06	3.21
Fuels_i	720,720	5.52	13.08	0.00	0.00	0.33	17.58	74.78
Minerals_i	720,720	3.93	12.55	0.00	0.00	0.24	7.69	96.23
1 - Labor Intensity^k	720,720	0.46	0.19	0.03	0.25	0.43	0.73	0.95
Rule of Law_i	720,720	-0.05	0.96	-2.26	-1.11	-0.28	1.57	2.02
Dummy DIF^k	720,720	0.61	0.49	0.00	0.00	1.00	1.00	1.00
Dummy REF^k	720,720	0.32	0.47	0.00	0.00	0.00	1.00	1.00

Note: All statistics in Panel A. are calculated to all variables (except to $\ln(\text{DIST}_{ij})$) over a sample of U.S. imports shipped either by air or sea in 2017 (excluding flows from Mexico and Canada and arriving in Puerto Rico) at the combination product-country of origin and U.S. Customs District. k indicates a HS6-digit product code, j a U.S. Customs District, and i a country of origin. f_{ij}^k/M_{ij}^k is calculated as the ratio of shipping freight charges to the value of U.S. imports including shipping freight charges. $\ln(\text{DIST}_{ij})$ is calculated as the natural logarithm of the Great Circle distance between every U.S. Customs District and every country of origin. All statistics in Panel B. are calculated over a sample of world export flow data in 2017 at the combination product-country. As in Panel A., k indicates a HS6-digit product code, and i a country of origin.

Table 2: Transport cost shares by mode regressions

VARIABLES	f^k_{ij}/M^k_{ij}				
	(1)	(2)	(3)	(4)	(5)
Air Share^k_{ij}	0.00851***	0.00856***	-0.172***	0.00783	-0.0102
	(0.000523)	(0.000531)	(0.0126)	(0.0113)	(0.0114)
Ctnr Share^k_{ij}	0.00249***	0.00254***	-0.119***	-0.000535	-0.0228**
	(0.000519)	(0.000526)	(0.0126)	(0.0113)	(0.0113)
Dummy Tankers^k		-0.000352	-0.231***	-0.254***	-0.162***
		(0.00221)	(0.0406)	(0.0366)	(0.0368)
Dummy RoRo^k		0.00138	-0.0616	0.00415	0.0328
		(0.00128)	(0.0379)	(0.0341)	(0.0345)
$\ln(\text{DIST}_{ij})$	0.0101***	0.0101***	-0.00546***	0.00809***	0.00400***
	(0.000743)	(0.000743)	(0.00148)	(0.00133)	(0.00134)
$\ln(\text{DIST}_{ij})\times$ Air Share^k_{ij}			0.0198***	0.00579***	0.00990***
			(0.00138)	(0.00124)	(0.00125)
$\ln(\text{DIST}_{ij})\times$ Ctnr Share^k_{ij}			0.0134***	-0.000113	0.00418***
			(0.00138)	(0.00124)	(0.00124)
$\ln(\text{DIST}_{ij})\times$ Tankers^k			0.0255***	0.0222***	0.0136***
			(0.00453)	(0.00407)	(0.00410)
$\ln(\text{DIST}_{ij})\times$ RoRo^k			0.00689*	-0.000272	-0.00135
			(0.00419)	(0.00377)	(0.00378)
$\ln(\text{WV}^k_{ij})$				0.0223***	0.0142***
				(5.60e-05)	(0.000260)
Constant	-0.0337***	-0.0338***	0.108***	0.0318***	0.0524***
	(0.00677)	(0.00677)	(0.0135)	(0.0122)	(0.0122)
$\ln(\text{WV}^k_{ij})\times$ transport mode	NO	NO	NO	NO	YES
Observations	678,774	678,774	678,774	678,293	678,293
R-squared	0.031	0.031	0.032	0.216	0.217

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimates over a sample of U.S. import data in 2017 shipped either by air or sea (excluding flows from Mexico and Canada and arriving in Puerto Rico) at the combination product k, country of origin i and U.S. Customs of District j. The LHS variable of all models (f^k_{ij}/M^k_{ij}) is the ratio of shipping freight charges to the value of U.S. imports including shipping freight charges. All models are estimated using the OLS estimator, including fixed effects at the U.S. Customs District and country of origin levels.

Table 3: Export composition by transport intensity and country type

VARIABLES	X_i^k					
	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(\text{MA}_i) \times$ Air Share^k	0.838**	1.668***	1.666***	0.969**	0.842**	0.335
	(0.382)	(0.497)	(0.497)	(0.450)	(0.426)	(0.359)
$\ln(\text{Pop}_i) \times$ Air Share^k	0.336***	0.242***	0.229***	0.139**	0.143**	0.108*
	(0.0743)	(0.0633)	(0.0628)	(0.0606)	(0.0579)	(0.0566)
$\ln(\text{GDPpc}_i) \times$ Air Share^k	0.439***	0.310**	0.278*	0.267*	0.184	0.185
	(0.126)	(0.146)	(0.147)	(0.139)	(0.142)	(0.113)
$\ln(\text{MA}_i) \times$ Ctnr Share^k	1.323***	2.207***	2.204***	1.636***	1.573***	1.250***
	(0.336)	(0.470)	(0.471)	(0.411)	(0.396)	(0.283)
$\ln(\text{Pop}_i) \times$ Ctnr Share^k	0.265***	0.163***	0.158***	0.0855**	0.0815**	0.108***
	(0.0568)	(0.0382)	(0.0381)	(0.0394)	(0.0391)	(0.0382)
$\ln(\text{GDPpc}_i) \times$ Ctnr Share^k	0.0442	-0.100	-0.113	-0.119	-0.151	-0.0893
	(0.0991)	(0.124)	(0.126)	(0.115)	(0.125)	(0.0818)
$\ln(\text{MA}_i) \times$ Dummy Tankers^k		0.841	0.769	0.642	0.636	
		(0.608)	(0.624)	(0.606)	(0.549)	
$\ln(\text{Pop}_i) \times$ Dummy Tankers^k		-0.240***	-0.201**	-0.218***	-0.200***	
		(0.0798)	(0.0824)	(0.0813)	(0.0761)	
$\ln(\text{GDPpc}_i) \times$ Dummy Tankers^k		-0.376**	-0.237	-0.236	-0.199	
		(0.161)	(0.164)	(0.161)	(0.184)	
$\ln(\text{MA}_i) \times$ Dummy RoRo^k		1.551***	1.546***	0.431	0.307	0.0694
		(0.487)	(0.488)	(0.443)	(0.441)	(0.388)
$\ln(\text{Pop}_i) \times$ Dummy RoRo^k		0.190***	0.184**	0.0456	0.0550	0.0252
		(0.0717)	(0.0719)	(0.0776)	(0.0754)	(0.0710)
$\ln(\text{GDPpc}_i) \times$ Dummy RoRo^k		0.486***	0.467***	0.461***	0.367**	0.364**
		(0.176)	(0.179)	(0.173)	(0.170)	(0.155)
Only Manufactures	NO	NO	NO	NO	NO	YES
WTO$_i$ interactions	NO	NO	YES	YES	YES	YES
Upstreamness^k interactions	NO	NO	NO	YES	YES	YES
Factor abundance$_i \times$ Intensities^k	NO	NO	NO	NO	YES	YES
Rule of Law$_i \times$ Type of Product^k	NO	NO	NO	NO	YES	YES
Observations	716,196	716,196	716,196	716,196	716,196	626,964
Pseudo R2	0.760	0.764	0.767	0.769	0.781	0.802

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimates over the sample of world export flow data (excluding U.S. flows) in 2017 at the combination product k and country of origin i. The LHS variable of all models X_i^k is the exported value from every country of origin i of every product k. All models are estimated using the PPML estimator, including fixed effects at the product and country of origin level and excluding singleton observations.

Table 4: Control variables for export composition regressions

VARIABLES	X_i^k					
	(1)	(2)	(3)	(4)	(5)	(6)
$WTO_i \times$ Air Share^k			2.494***	2.472***	1.937***	2.803***
			(0.367)	(0.362)	(0.374)	(0.407)
$WTO_i \times$ Ctnr Share^k			0.251	0.266	-0.111	0.323
			(0.296)	(0.293)	(0.316)	(0.320)
$WTO_i \times$ Dummy Tankers^k			-2.009***	-1.996***	-1.805***	
			(0.369)	(0.363)	(0.379)	
$WTO_i \times$ Dummy RoRo^k			1.023*	1.022*	0.670	1.051*
			(0.549)	(0.545)	(0.549)	(0.551)
$\ln(MA_i) \times$ Upstreamness^k				-0.558***	-0.576***	-0.480***
				(0.101)	(0.0949)	(0.0901)
$\ln(Pop_i) \times$ Upstreamness^k				-0.0695***	-0.0621***	-0.0628***
				(0.0187)	(0.0182)	(0.0184)
$\ln(GDPpc_i) \times$ Upstreamness^k				-0.00735	-0.0162	-0.0195
				(0.0293)	(0.0331)	(0.0335)
$Capital_i \times$ (1- Labor Intensity^k)					0.00790***	0.00970***
					(0.00130)	(0.00129)
$Land_i \times$ (1- Labor Intensity^k)					1.294***	
					(0.311)	
$Minerals_i \times$ (1- Labor Intensity^k)					-0.0100	
					(0.0126)	
$Fuels_i \times$ (1- Labor Intensity^k)					0.0660***	
					(0.0121)	
Rule of $Law_i \times$ Dummy DIF^k					0.636***	0.459***
					(0.110)	(0.102)
Rule of $Law_i \times$ Dummy REF^k					0.546***	0.418***
					(0.113)	(0.0971)
Constant	15.89***	19.63***	19.03***	21.01***	19.84***	18.14***
	(1.321)	(1.885)	(1.853)	(2.119)	(2.158)	(1.933)
Only Manufactures	NO	NO	NO	NO	NO	YES
Observations	716,196	716,196	716,196	716,196	716,196	626,964
Pseudo R2	0.760	0.764	0.767	0.769	0.781	0.802

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimates over the sample of world export flow data (excluding U.S. flows) in 2017 at the combination product k and country of origin i. The LHS variable of all models X_i^k is the exported value from every country of origin i of every product k. All models are estimated using the PPML estimator, including fixed effects at the product and country of origin level and excluding singleton observations.

Table 5: Export composition with weight-to-value ratio bins

VARIABLES	X_i^k				
	(1)	(2)	(3)	(4)	(5)
$\ln(MA_i)$ × Dummy Light Products	1.140***	1.178***	0.598	0.532	-0.140
	(0.393)	(0.394)	(0.382)	(0.342)	(0.341)
$\ln(Pop_i)$ × Dummy Light Products	0.371***	0.332***	0.274***	0.255***	0.135***
	(0.0756)	(0.0752)	(0.0703)	(0.0640)	(0.0517)
$\ln(GDPpc_i)$ × Dummy Light Products	0.448***	0.337***	0.309***	0.267**	0.299***
	(0.118)	(0.115)	(0.114)	(0.104)	(0.0902)
$\ln(MA_i)$ × Dummy Mid Products	1.286***	1.322***	0.830**	0.821***	0.157
	(0.345)	(0.348)	(0.324)	(0.281)	(0.247)
$\ln(Pop_i)$ × Dummy Mid Products	0.334***	0.301***	0.251***	0.223***	0.136***
	(0.0519)	(0.0516)	(0.0512)	(0.0461)	(0.0279)
$\ln(GDPpc_i)$ × Dummy Mid Products	0.340***	0.242***	0.217***	0.218***	0.301***
	(0.0894)	(0.0868)	(0.0827)	(0.0748)	(0.0474)
Only Manufactures	NO	NO	NO	NO	YES
WTO_i interactions	NO	YES	YES	YES	YES
$Upstreamness^k$ interactions	NO	NO	YES	YES	YES
Factor $abundance_i$× $Intensities^k$	NO	NO	NO	YES	YES
Rule of Law_i× Type of $Product^k$	NO	NO	NO	YES	YES
Observations	716,196	716,196	716,196	716,196	626,964
Pseudo R2	0.762	0.764	0.766	0.779	0.801

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimates over the sample of world export flow data (excluding U.S. flows) in 2017 at the combination product k and country of origin i. The LHS variable of all models X_i^k is the exported value from every country of origin i of every product k. All models are estimated using the PPML estimator, including fixed effects at the product and country of origin level and excluding singleton observations.

Table 6: Control variables for weight-to-value regressions.

VARIABLES			X_i^k		
	(1)	(2)	(3)	(4)	(5)
WTO$_i \times$ Dummy Light Products		3.782***	3.780***	3.292***	2.475***
		(0.421)	(0.420)	(0.362)	(0.308)
WTO$_i \times$ Dummy Mid Products		1.982***	1.978***	1.506***	0.544**
		(0.358)	(0.354)	(0.292)	(0.220)
ln(MA$_i$) $\times$ Upstreamness^k			-0.523***	-0.528***	-0.449***
			(0.109)	(0.106)	(0.101)
ln(Pop$_i$) $\times$ Upstreamness^k			-0.0530***	-0.0491***	-0.0528***
			(0.0182)	(0.0178)	(0.0180)
ln(GDPpc$_i$) $\times$ Upstreamness^k			-0.0258	-0.0273	-0.0458
			(0.0327)	(0.0372)	(0.0336)
Capital$_i \times$ (1- Labor Intensity^k)				0.00938***	0.0112***
				(0.00138)	(0.00130)
Land$_i \times$ (1- Labor Intensity^k)				1.316***	
				(0.316)	
Minerals$_i \times$ (1- Labor Intensity^k)				-0.0130	
				(0.0135)	
Fuels$_i \times$ (1- Labor Intensity^k)				0.0674***	
				(0.0120)	
Rule of Law$_i \times$ Dummy DIF^k				0.648***	0.461***
				(0.101)	(0.102)
Rule of Law$_i \times$ Dummy REF^k				0.533***	0.398***
				(0.104)	(0.0974)
Constant	13.39***	12.75***	14.39***	13.32***	14.48***
	(1.313)	(1.257)	(1.782)	(1.759)	(1.684)
Only Manufactures	NO	NO	NO	NO	YES
Observations	716,196	716,196	716,196	716,196	626,964
Pseudo R2	0.762	0.764	0.766	0.779	0.801

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Estimates over the sample of world export flow data (excluding U.S. flows) in 2017 at the combination product k and country of origin i. The LHS variable of all models X_i^k is the exported value from every country of origin i of every product k. All models are estimated using the PPML estimator, including fixed effects at the product and country of origin level and excluding singleton observations.

Appendixes

1 Countries in CEPII database with WTO membership updated

Table 1: Countries with updated WTO membership and related WTO countries

Country with updated WTO membership	Related WTO Country
BES	NLD
COK	NZL
CUW	NLD
GRL	DNK
MSR	GBR
NCL	FRA
PYF	FRA
SHN	GBR

Note: All codes correspond to the ISO 3-digit country codes. The column titled "Country with updated WTO membership" lists all countries not completely independent from a WTO member classified in the CEPII database as a non-WTO member. The WTO membership thus is updated. The column titled "Related WTO Country" reports the WTO member to which every country in the column titled "Country with updated WTO membership" is related.