

Entrepôt: Hubs, Scale, and Trade Costs*

Sharat Ganapati
Georgetown University

Woan Foong Wong
University of Oregon

Oren Ziv
Michigan State University

March 2020

Abstract

Entrepôts are hubs that facilitate trade between various origins and destinations. We study the role these hubs, and the networks they form, play in international trade. Using novel data, we trace the paths of containerized goods entering the United States. We show that the majority of trade is indirect and sent through a small number of entrepôts, resulting in lower transport costs through scale economies by using larger ships. We build a model of endogenous entrepôt formation incorporating route choice by exporters within a Ricardian setting. We use the model to estimate trade costs on each shipping leg and develop a geography-based instrument to estimate a leg-level scale elasticity. Counterfactuals opening the Arctic Passage and Brexit quantify the effects of both network spillovers and scale economies. We find that spillovers from the transportation network doubles baseline welfare gains, with scale economies further tripling them.

Keywords: trade costs, scale, hubs, transport costs, transportation networks, international trade, shipping

*Contact: sganapati@georgetown.edu, wfwong@uoregon.edu, and orenziv@msu.edu. We thank Treb Allen, Costas Arkolakis, Johannes Boehm, Kerem Coşar, Anca Cristea, Meredith Crowley, Carsten Eckel, Stefania Garetto, Matthew Grant, Reka Juhász, Ezra Oberfield, Andrés Rodríguez-Clare, Robert Staiger, Meredith Startz, and seminar participants at the Federal Reserve Board, Ludwig-Maximilians-Universität Munich, University of Cambridge, Georgetown University, Stanford University, University of Mannheim, and University of Virginia as well as participants at the NBER Conference on Cities, Labor Markets, and the Global Economy, 2019 Mid-Atlantic International Trade Workshop (Duke), 2020 CeMENT Mentoring Workshop, and other conferences for helpful comments. We apologize for any cancelled presentations due to COVID-19. Andrew Castro, Ray McCormick, Giacomo Romanini, and Philip Valtadoros provided outstanding research assistance.

International trade is generally thought of as a bilateral arrangement between an exporter and an importer. However, the act of exchanging goods over borders involves more than the production and consumption of these goods: shipping, transshipping, and distribution can include multiple agents and additional countries beyond the producers and consumers.

These activities are concentrated at entrepôts, trading hubs where goods travel through—from other origins, and bound for other destinations. The notion that countries stand to gain by exploiting scale economies and becoming hubs for these commercial activities has a long history and continues to be a powerful narrative. Governments and local port authorities invest billions of dollars with the specific aim of becoming or maintaining their role as entrepôts.¹

This paper studies the implications of the transportation network formed by entrepôts for international trade. We seek to answer three questions: (1) How indirect is trade? (2) What drives the formation of entrepôts? and (3) What are the impacts of indirect trade and entrepôts on trade flows and welfare? Our central finding is that by concentrating shipments through a relatively small number of nodes, entrepôts take advantage of scale economies in shipping; furthermore, these hubs generate large trade spillovers through the network and have out-sized impacts on global trade and welfare.

We begin by constructing two new datasets that jointly map individual journeys at the shipping container level in order to document the ubiquity of indirect trade and describe the role of entrepôts in concentrating international trade. We then build a model of international trade where route choice endogenously gives rise to hubs and use the resulting estimation equations to estimate leg-specific trade costs that rationalize observed shipping volumes with minimal assumptions. Next, to assess the role of scale economies and the concentration of commerce at entrepôts in reducing shipping costs, we estimate a scale elasticity and embed our results in a quantitative general equilibrium model to quantify regional and global effects of entrepôts on trade volumes and welfare.

Our novel data sets allow us to uniquely characterize the global trading network: a shipment-level data set of bills of lading for the universe of US container imports, and a global data set on container ships' ports of call. Together, these two data sets enable us to reconstruct the origin

¹Saudi Arabia has implemented a \$7 billion project to expand the container capacity to “be a major east-west marine transshipment location.” ([Financial Times, 2015](#)). India is spending \$4 billion to rival Chinese facilities ([Reuters, 2016](#)). Established entrepôt Singapore is investing \$1.1 billion to boost its capacity to “stay ahead of the curve as a world-class hub port” ([Port Technology, 2018](#)) following a \$3 billion project to construct an automated container yard ([Ship and Bunker, 2012](#)).

to destination journeys taken by individual shipments for nearly all in-bound containerized US shipments. This is the first comprehensive look at how containerized shipments into the United States travel through the global shipping network.

Three stylized facts emerge. First, the majority of trade arrives at US ports indirectly. Figure 1 plots, by country of origin, the average number of stops made by containers. Very few countries export directly to the US and the average shipment stops at two additional countries (*third-party countries*). This significantly adds to the distance goods travel and their shipment time. Second, this indirectness is incredibly concentrated, with a large number of shipments channelled through a small number of entrepôts. Third, increasing ship sizes, which is known to reduce costs, is correlated with both entrepôt activity and indirect shipping patterns. In sum, indirectness is ubiquitous and concentrated at entrepôts, and the shipping network is a hub-and-spoke system where large ships connect global hubs and smaller ships service small local routes. By centralizing shipments, entrepôts appear to reduce shipping costs through scale economies.²

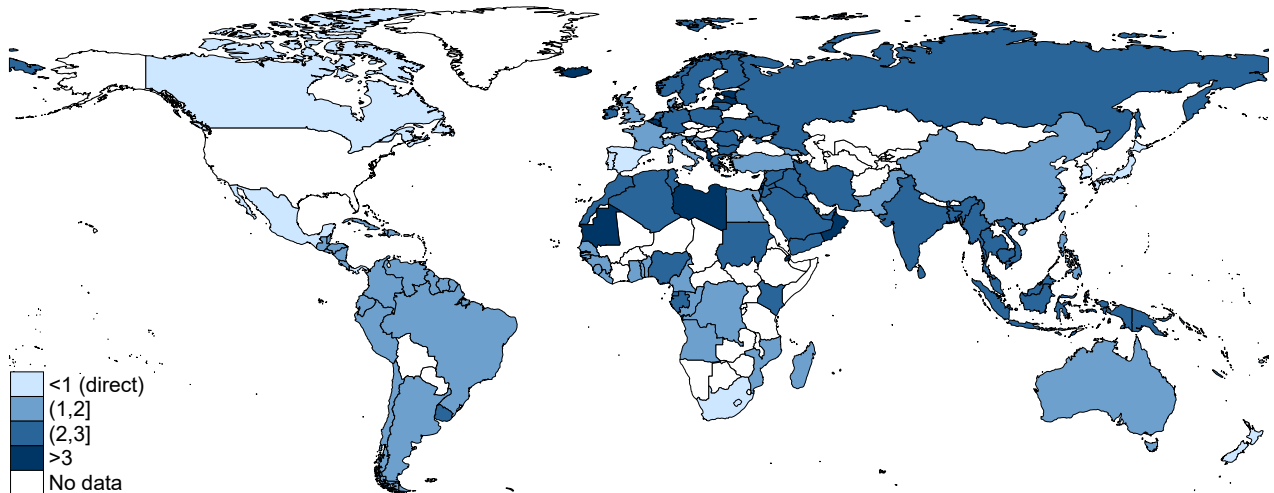


Figure 1: Number of Stops between Origin and US Destination

Notes: Stops are by country and weighted by container volume (Twenty-foot equivalent shipping units - TEU). The destination country US is excluded. Landlocked countries are also excluded, since they would need to stop at a coastal country. 34 of the shipment origin countries are landlocked accounting for 1.6 percent of total TEUs. The missing remaining countries are either due to lack of overall trade with the US (e.g. Somalia) or due to the merge process (e.g. Namibia). China is not direct because we consider Hong Kong, Macau, and Taiwan as separate countries.

Source: Authors' calculations using AIS and Bill of Lading data. ■

To understand the forces behind observed indirectness, we build a general equilibrium model of global trade with entrepôts. Individual firms choose shipping routes and compete for con-

²This network structure is not specific to containerized shipping but is also prevalent in air transport and freight services like UPS or DHL (Rodrigue, Comtois and Slack, 2013).

sumers in destination countries in a generalized Ricardian setting, flexibly accommodating input-output linkages. Low-cost routes can involve shipments through third-party countries, and entrepôts endogenously arise at ports through which shipping costs are lowest. Crucially, scale economies may work to reduce shipping costs on individual links as traffic grows at individual ports or links.

Using global data on shipment flows, we use the model to estimate trade costs for each country pair. An advantage of our modeling approach is that we need to make very few structural assumptions on the production and consumption setting; our model recovers a full trade cost matrix for every origin and destination that best rationalizes the observed traffic given the observed trade flows. Likewise, although we allow for one, we do not take a stand on scale economies in trade cost estimation.

With our cost estimates, we find that scale economies are part of the mechanism driving network concentration and their existence would alter how shocks affect global welfare through entrepôts. To estimate the scale elasticity—the causal effect of quantity on trade costs—we construct an instrument using the geography of the trade network. Embedded in our model is the intuition that some legs are inherently higher traffic (higher demand) routes because they lie closer to the shortest path between large origins and destinations.³ Leveraging this, we construct a novel instrument for demand. For each leg, we compute the distance to and from the leg relative to the shortest distance between each origin and destination, recovering a weighted average of each leg’s proximity to global trade. Using this instrument, we estimate a strong scale elasticity. Our results imply that a 1% increase in traffic on a given leg reduces costs on the same leg by over 0.05%.

Finally, to estimate the impact of the trade network on global shipping and welfare, as well as the localized effects and spillovers from hubs, we adopt further structure to our general model, and use the resulting quantified general equilibrium model to run counterfactual predictions. We run two sets of counterfactuals to quantify and illustrate the effects of the network structure and scale economies on trade volumes and welfare. Our first series of counterfactuals consider the effects of global warming opening up the Arctic Ocean to regular year-round shipping, significantly reducing shipping distances between many Asian, North American and European ports. Our second set of counterfactuals considers the ramifications of worsening relations

³The emergence of entrepôts as hubs in geographically advantageous locations is consistent with the findings of Barjamovic et al. (2019).

between the United Kingdom and its trading partners. The baseline effects of these shocks are first doubled by the network structure of trade and further tripled by the feedback loop imposed by scale economies in container shipping.

This paper characterizes the nature of the global container shipping networks and its implication for international trade. It contributes to the literatures on the presence of networks in trade, on endogenous trade costs, and on trade and transportation technology.

Our main contributions are to a growing quantitative literature investigating the role of trade networks (Fajgelbaum and Schaal, 2017; Redding and Turner, 2015; Allen and Arkolakis, 2019). Our data is a first and systematic observation of indirect trade through the (containerized) shipping network.⁴ We extend the Allen and Arkolakis (2019) Armington framework where route cost shocks are born by consumers to a general Ricardian setting, where traffic volumes reflect both route choice and head-to-head competition on prices at destinations and show how to estimate the model in a multi-industry setting with missing traffic flows. Comparing our estimation to micro-data on US shipment routes, we assess the validity of the Allen and Arkolakis (2019) approach, reporting a tight match between predictions and data.

We also contribute to a growing literature on endogenous transport costs.⁵ In our setting, trade costs are endogenously determined in equilibrium with trade flows as part of the transportation network.⁶ Brancaccio, Kalouptsi and Papageorgiou (2017) estimate a model of endogenous trade costs arising from search frictions between exporters and dry bulk ships carrying homogeneous commodities, where all trade is direct. We model transport costs as part of a global network of container shipping routes, a dramatically different setting which accounts for two-thirds of annual trade moved by sea (World Shipping Council).⁷ Because our model is within a general equilibrium spatial trade framework, our counterfactuals contribute to this literature by addressing the overall welfare impacts of the endogenous transportation

⁴We provide our estimated data on indirect flows which can be used for future empirical work. Previous work have either imputed indirect trade or just used port of call data alone (Wang and Wang, 2011; Kojaku et al., 2019; Lazarou, 2016). They cannot directly observe indirect trade since they do not have information on the loading or unloading of shipments.

⁵See Hummels (2007) and Limao and Venables (2001) for reviews of this literature.

⁶There are a number of papers studying endogenous transport costs in the context of market power, including Hummels, Lugovskyy and Skiba (2009) and Asturias (Forthcoming). We follow Sutton (1991) and allow larger markets to induce entry and competition to lower prices, but do not directly measure market structure.

⁷Like taxis, dry bulk ships depart from destinations without cargo and therefore have to search for it. Containerships, like buses, travel on fixed schedules between many locations. As a result, indirect trade, and the network structure giving rise to systematic trade spillovers due to network linkages, are absent from the bulk carrier setting.

network and its spillovers through entrepôts.⁸

Finally, we contribute to the literature on transportation technology and trade. Several papers investigate the effects of containerization (Coşar and Demir, 2018; Bernhofen, El-Sahli and Kneller, 2016; Wong, 2019). Ducruet et al. (2019) is a complementary study which investigates the aggregate and distributional impact of node-level infrastructure investment—new port technologies from containerization in the 1970s. We document the global network effects of the container shipping technology and of the network formed by these nodes on trade.

One important aspect of transportation technology in our model is the scale economy in shipping, which is empirically important, and in line with findings from the literature on economies of scale in trade.⁹ Our paper shows how scale economies, acting through the global shipping network, can generate shipping hubs and have out-sized impacts on trade volumes and welfare. In this respect, we are also related to a literature in economic geography which considers the role of localized scale economies in the emergence of agglomerations (Allen and Arkolakis, 2014; Allen and Donaldson, 2018). Trade costs typically act as dispersive forces in this literature, we show that with scale economies in transportation this relationship is inverted: the presence scale in trade costs contributes to agglomeration, particularly for entrepôts.

2 Data

We compile and combine two proprietary data sets in this project: global ports of call data for containerships, which allows us to reconstruct the routes taken by specific ships, and US bill of lading data for containerized imports, which gives us shipment-level data on imports into the United States. Independently, these datasets allow us to partially describe the global shipping network. By merging them, we reconstruct nearly the entire journey of shipments entering the United States, from their origin to their US port of entry. To our knowledge, we provide the most comprehensive reconstruction of the global shipping network and routes undertaken by individual shipments into the US.¹⁰

⁸Our estimates provide a matrix of bilateral trade costs and market access measures that can be used for further empirical work.

⁹See especially Alder (2015); Holmes and Singer (2018); Anderson, Vesselovsky and Yotov (2016); Asturias (Forthcoming); Skiba (Forthcoming).

¹⁰Data Appendix A.1 explains both data sets and their merge procedure in detail.

Figure 2: Map of Global Port of Call Network



Notes: Each dot represents a port (total of 1,203 ports). Each line represents a journey between port pairs undertaken by a containership (total of 4,986 ships). Source: Authors' calculations from AIS data. ■

Port of call data Our proprietary ports of call data from Astra Paging captures vessel movements using ship Automatic Identification System (AIS) transponders.¹¹ For each vessel, this captures identifying information, time-stamped ports of call, capacity and height in the water before and after stopping in the port, the latter two jointly indicating the vessel's load. Using these data elements, we can observe the volume shipped between each port pair.

Our sample covers a six months period, from April to October 2014. Over this period, we have information on 4,986 unique container ships with a combined capacity of 18.13 million twenty-foot equivalent shipping units (TEUs)—over 90% of the global container shipping fleet—making 429,868 calls at 1,203 ports.

Figure 2 shows the coverage of the shipping network in our port of call data. Each line represents a journey between port pairs, the dots, undertaken by a containership. Some ports, not necessarily in the largest countries, are more connected than others. The Suez Port and Port of Balboa (due to the Suez and Panama Canals), Singapore and Rotterdam—well-known entrepôts—are particularly connected.

Bill of lading data We put together proprietary bills of lading data, which captures shipment-level information for all containerized imports into the United States. Our data

¹¹Port receivers collect and share AIS transponder information (including ship name, speed, height in water, latitude and longitude). Using geographic AIS variables, we track global port entry and exit data.

captures the foreign location where the shipment originated from, the foreign port where it was loaded on the containership (port of lading) which brings it into the US, and the US port where it was unloaded from the containership (port of unloading). In addition, we know the name and identification number of the containership which transported the shipment as well as the shipment’s weight, number of containers (TEUs), and product information. Over the same six months period, we see a total of 14.8 million TEUs weighting 106 million tons were imported into the US from 227 shipment origin countries, and 144 countries with ports of lading.

Reconstructing shipment routes Using the containership information, port of arrival information, timing of unloading and ports of call at US ports, and port of lading information, we are able to match the bills of lading to the journeys of specific containerships, then use the ports of call between lading and unloading to reconstruct each shipment’s path from its foreign origin to US destination. Over 90% of containerized TEUs entering the US on bills of lading can be matched to routes using this method. Appendix Figure [A.1](#) visualizes this merge.

What remains unobserved is the shipment’s journey between origin and the first stop (port of lading) we observe in our data. In particular, this initial portion of the shipment’s journey could take place overland (by trucks or rail) or by sea on another ship. This information is impossible for us to observe, leading us to under-count the overall level of indirectness. However, this will not affect our model estimation, which is not reliant on observing the full path.

3 Stylized Facts

We use the datasets above to explore the nature of the international shipping network and the routes taken by goods entering the US along that network. Our analysis generates three stylized facts about trade and entrepôts. First, the global container shipping network forms a hub-and-spoke system where ships act as busses on city streets, carrying goods through multiple stops and meeting at central depots or transfer points. Second, this network concentrates shipments through entrepôts, generating large costs for individual shipments, in the form of added distance. Third, the indirectness is optimal, and the resulting concentration allows for the use of supersized cargo ships with lower unit costs.

3.1 The majority of trade is indirect

We begin by asking: how realistic is the bilateral view of trade? Panel (A) in Figure 3 reports the distribution of the number of observed country stops made by each shipment, weighted by TEU. Only about 20 percent of containers are exported to the US directly from the origin country, making stops in no other country along the way. The average TEU entering the US stops at around 2 countries that are not involved as either a producer or a consumer (mean of 1.5 and s.d. of 1.3). The average number of port stops is higher (Figure A.3, mean of 4.6 and standard deviation of 3.5). This result is robust for shipment weight and value (Figure A.7).

Figure 3: Indirect trade distributions, by container and country



Notes: Panel (A) shows the distribution of containers by the number of countries the containers visited. Panel (B) show the distribution of countries, by the share of shipments that are transhipped to the United States. This plot is weighted by the aggregate exported containers (TEU). Source: Authors’ calculations using AIS and Bill of Lading data. ■

This is also true at the country level: the majority of US trading partners export indirectly to it. This can be roughly gleaned from Figure 1 where many countries have more than one stop (i.e. have darker shades of blue). On average, a country’s shipments stop at two other countries before reaching the US (mean of 2.1 and standard deviation of 0.74).¹² Shipments come directly to the US from only 9 countries.¹³

Another aspect of indirectness is *transshipment*, which we define as when the origin country

¹²At the port-level, the average stops are 5.2 with a standard deviation of 1.7.

¹³The 9 direct countries are Canada, Mexico, Panama, Japan, South Korea, Spain, Portugal, South Africa, and New Zealand. China is not included in this list due to Hong Kong, Taiwan, and Macau being considered separate countries in our data set.

of a shipment is not the same as the country where it was loaded onto the containership bound for the US (Stop 1 in Figure A.1). 27% of shipments by volume are transshipped in third-party countries. Moreover, the average good from a majority of non-landlocked US trading partners is transshipped in a third-party country and over 60% of non-landlocked US trading partners transship more than 90% of US-bound goods in third countries (Panel (B), Figure 3).¹⁴

Appendix C explores the high degree of variation in connectivity evident already in Figure 1, showing that that variation is reasonably explained by traditional gravity variables, and further explores variation in routes from unique origins into the US.¹⁵

3.1.1 Indirect trade increases shipping distances

Are the additional country stops simply incidental stops along the way, or do they constitute a trip that is meaningfully distinct from what a “direct” path would look like? One possibility is that the observed indirectness is optimal but only incidental—perhaps additional stops only have small effects on cost, and therefore may be optimal even if the benefit of indirectness is small. However, the significant additional distance and time incurred by indirect travel, documented here, implies this is unlikely to be the case.

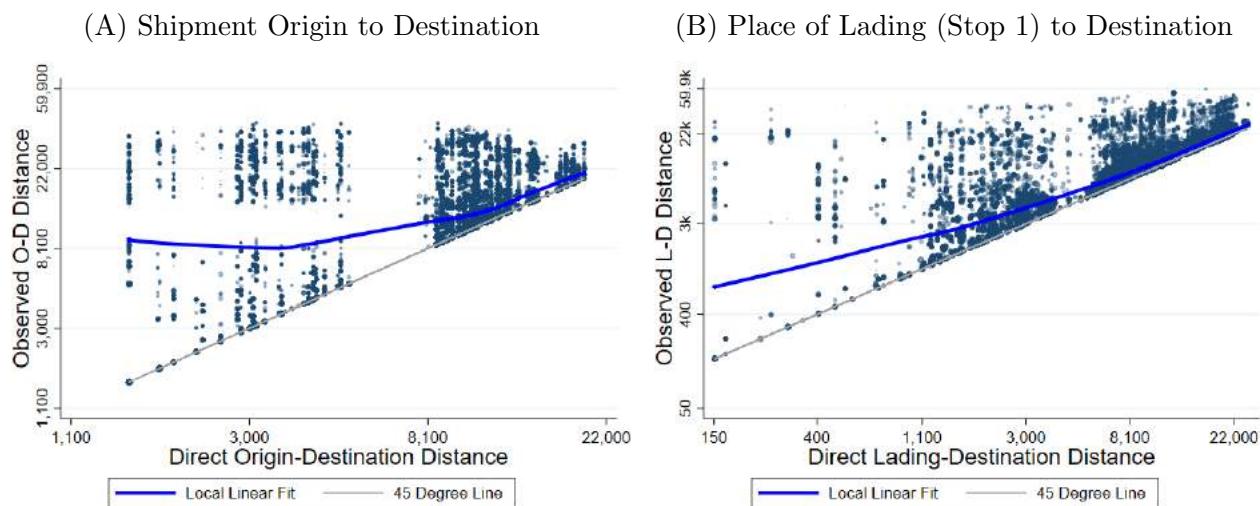
On average, the actual traveled distance between a shipment’s origin and its US destination is 31 percent more than its direct ocean distance (Panel (A) in Figure 4). Panel (B) shows the actual traveled distance between a shipment’s lading location and its final destination and here we see that the gap is smaller at 14 percent. Table A.1 further evaluates the relationship between indirectness and journey length, finding that doubling the number of stops adds 10% to distance travelled and 33% to time travelled, even after controlling for direct journey length or origin-by-destination fixed effects.

Furthermore, beyond the distance costs we document, pecuniary costs of transshipment and time costs of stopping at additional ports will increase the cost of indirectness. We conclude that indirectness is meaningful in the sense that it is costly. The fact that this organizational structure remains optimal implies that it carries with it a cost reduction over and above these costs. From these results, we can summarize our first stylized fact:

¹⁴Examples include Denmark, Bangladesh, Cambodia, and Ecuador. See Figure A.4 for a map of the percentage of goods transshipped for each country of origin.

¹⁵The existence of within-origin route variation is an important assumption in our model and is used in our validity checks.

Figure 4: Difference between Traveled Distance and Direct Distance for Indirect Shipments



Notes: These figures show only indirect shipments, with different direct and travelled distances. Figure (A) compares the direct shipping distance from the country of origin to the US to the actual route travelled. Figure (B) compares the direct shipping distance from the place a container was last loaded on a ship before arrival to the US to the actual route travelled. The local linear fit line is a locally weighted regression of the observed on direct pair-wise distance.

Source: Authors' calculations using AIS and Bill of Lading data. ■

Stylized Fact 1. *The majority of containerized trade into the US is indirect and results in a significant increase in shipping distance and time.*

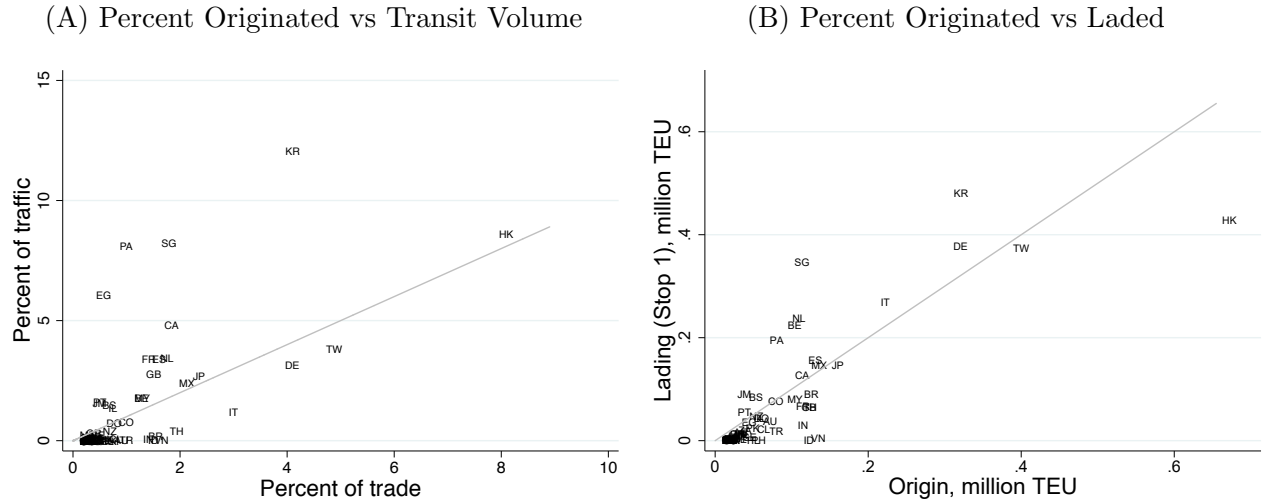
3.2 Indirect trade is routed through entrepôts

When shipments stop in third-party countries, how are they routed? In this section, we show that the stops along indirect shipping routes are not arbitrarily distributed throughout the world. Shipments in our data are channelled through well-known entrepôts. These locations disproportionately service shipments originating in other countries.

Panel (A) of Figure 5 scatters each country's percent of total stops against percent of total trade. Here we see where the above concentration is taking place. Some locations like Taiwan, are both popular stopping points but also major countries of origin for goods. A few key countries, especially Korea, Singapore, Panama, and Egypt, are not only high on the Y-axis, denoting they are especially important third-party countries, but they are above the 45° line, indicated that they disproportionately participate in trade as a third-country in US-bound shipments. This metric captures countries long-associated with entrepôt, including Singapore (SG), Belgium (BL), Netherlands (NL), Korea (KR), and Panama (PA), as well as others. Panel (B) plots each country's total volume as a first stop (lading volume) against their origin volume.

A similar set of countries lies above the 45° line.¹⁶ Countries with the largest concentrations of lading shipments are also more likely to be disproportionately used as entrepôts.¹⁷

Figure 5: Direct vs Indirect Shipments



Notes: Figure (A) compares the share of world container that originated in the country versus the share that passed through that country, but did not originate there. Figure (B) compares the share of shipments to the United States that originated in a country to the share of shipments that were last loaded onto a ship in that country. For scale, China is omitted. Source: Authors' calculations using AIS and Bill of Lading data. ■

Overall, how concentrated are third-party country stops (the Y-axes)? Table A.2 reports 99-50, 95-50, and 90-50 concentration ratios for all third-party shipments, transshipments, and trade volumes. For shipments, the 99th-percentile country, Korea, acts as a third-party country for almost 400 times the number of shipments (by TEU) compared to the median country, Iraq (99-50 ratio). This ratio is high by most standards.¹⁸ Transshipment is similarly concentrated (more than 400 for the 99-50 ratio, for example). A natural benchmark to compare these ratios to might be the concentration ratio in trade. The same 99-50 ratio for trade is only 96, and at all reported ratios, trade is significantly less concentrated than third-party country stops and transshipment. These relationships can be summarized in our second stylized fact:

Stylized Fact 2. *Indirect shipping routes are concentrated through well-known entrepôts.*

¹⁶For top 10 countries by country stops and lading volumes, Figure A.5 tabulates the percent of all goods entering the US stopping in that country, broken into goods originated there and elsewhere.

¹⁷Since this dataset is only for US imports, this relative ranking may miss entrepôts that deal mainly in non-US-bound traffic or may overemphasize the importance of US-proximate locations. Figure A.6 repeats the exercise in Panel (B) using all our global AIS data (no longer merged with US bills of lading). We find that many of the entrepôts who play a major role in US trade continue to do so in global trade

¹⁸The same ratio in employment in the highly concentrated IT sector across US cities is 300 (Moretti, 2019).

3.3 Indirect trade increases ship sizes

In the standard gravity model, trade costs are a function of geographic characteristics like distance. By revealed preference, shipping through entrepôt appears to generate a cost reduction over and above the costs incurred by travelling indirectly. What is the nature of the cost reductions attracting indirect trade to entrepôts? A number of mechanisms may well account for the unobserved cost reduction at entrepot and may be at work simultaneously.¹⁹ In this subsection, we focus on the relationship between indirect shipping and ship size, which we directly observe. A well-documented inverse relationship exists between unit shipping costs and ship size.²⁰ In what follows, one might interpret ship size as a mechanism that lowers shipping cost or equally as a proxy variable for (unobserved) shipping cost.

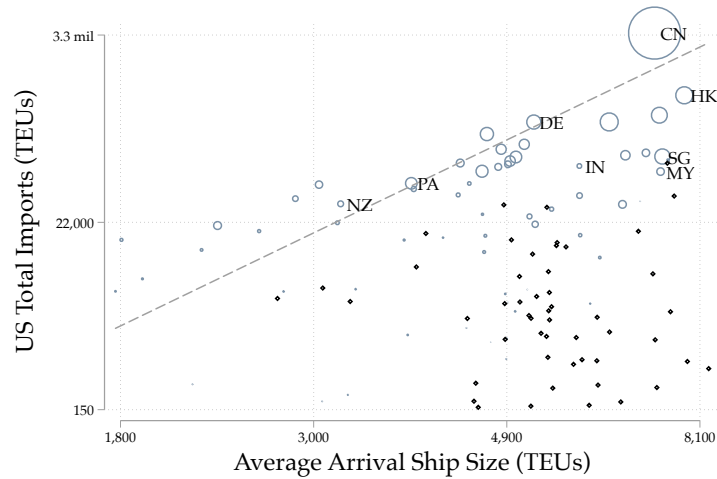
Ports with higher volumes of transshipments send goods on larger ships. The regression line in Figure 6 shows a positive country-level relationship between the volume of US imports and the average size of the incoming ships; unsurprisingly, larger trade volumes require larger ships to transport. However, some of the smaller US import partners also arrive to the US in similarly large ships like Hong Kong (HK) and Singapore (SG). Many of these origin points are entrepôts, where originated shipments are able consolidated onto larger ships filled with goods transshipped through the entrepôt. The blue circles in Figure 6 shows the lading sizes of these countries. The skew of larger vs smaller circles relative to the regression line implies that countries with larger volumes of lading ship on larger ships.

Table 1 displays regression on the same data at the shipment level (weighted by shipment TEU) which confirms these findings. Column 1 regresses, for our sample of shipments, the log of ship size against the log of total origin origin country volumes shipped (TEUs), confirming a positive relationship at the shipment level. Column 2 adds the log of quantity laded at each shipment's port of lading. Both coefficients are positive but the coefficient on origin volumes is almost halved (0.084 in Column 1 compared to 0.043 in Column 2). This indicates that much of the correlation between origin volumes and ship size acts through the size of the lading port. While shipments' ship sizes are correlated with their origin country volumes, shipments laded in larger ports disproportionately lade on larger ships.

¹⁹High-traffic routes are served by many carriers, using ships capable of carrying 25,000 containers with automated lading and unloading technologies. Internal and external scale economies in shipping, and competition among shippers could all generate a negative relationship between volume and costs, as could factors such as port infrastructure.

²⁰Cullinane and Khanna (2000) find that shipping costs decrease as ship size increases.

Figure 6: Link between US trade and ship size



Notes: Each circle represents an exporting country to the United States. The y-axis shows the total exports from that country to the United States. The x-axis shows the average size of a ship a good from that exporter arrives at the United States. The size of the circle represents the total volume of exports that was last loaded onto a ship in that country. Countries without direct shipments to the United States are denoted with solid black dots.

Source: Authors' calculations using AIS and Bill of Lading data. ■

Furthermore, goods that use third countries as transshipment points come on larger ships than would be predicted by their own country's trade volumes. The dark highlighted points in Figure 6 indicate countries where all shipments coming into the US are always laded elsewhere. These countries, who disproportionately use entrepôts, are outliers, having shipments arriving on ships much larger than what would be expected given their trading volumes.

Turning again to the shipment-level data, column 3 of Table 1 fully interacts variables in column 2 with an indicator variable for shipments that are laded in their origin countries ($\mathbb{1}(\text{Lading is Origin})$), effectively splitting the specification into two samples. For those shipments whose origin country differs from lading country, which have an indicator value of 0, the correlation between ship size and lading volume is considerably and significantly higher (0.13, Column 3). Furthermore, as suggested by the figure, shipments' ship sizes are not strongly correlated with origin country volumes when they lade in third countries (0.009, Column 3).

Finally, we note that goods lading at smaller transshipment points along major routes are also on larger ships. Figure 6 shows that goods from Malaysia appear to arrive on ships disproportionate to both its origination or lading volumes. In our data, a significant number of ships stopping in Malaysia also stop in Singapore. Shipment laded at small ports on ships bound for or coming from much larger ports, may also benefit from disproportionately larger

Table 1: Determinants of ship size

	(1)	(2)	(3)	(4)
	ln Ship Size	ln Ship Size	ln Ship Size	ln Ship Size
ln Volume at Origin	0.0843 (0.0163)	0.0432 (0.0179)	0.00925 (0.0121)	
ln Volume at Lading		0.0803 (0.0202)	0.127 (0.0230)	0.0282 (0.0182)
$\mathbb{1}(\text{Lading is Origin})$			-0.0220 (0.300)	
$\mathbb{1}(\text{Lading is Origin}) \times \ln \text{Volume at Lading}$			-0.0937 (0.0295)	
$\mathbb{1}(\text{Lading is Origin}) \times \ln \text{Volume at Origin}$			0.0861 (0.0220)	
ln Largest Port Stop				0.121 (0.0250)
Observations	215,656	215,656	215,656	215,656
R^2	.124	.174	.199	.21
F-stat	26.82	14.66	13.51	26.73

Notes: Observations represent all matched imported containers to the United States. Observations weighted by the size of shipment (TEU). Standard errors clustered by lading and destination ports.

Source: Authors' calculations using AIS and Bill of Lading data. ■

ship size.

In column 4 we investigate this possibility by regressing shipments' log ship size against the log volume laded at their port of lading and the log volume laded at the largest port at which we observe the shipment making a port call. The effect of the max-port-size variable is large, positive, and overall stronger than the effect of lading port volumes alone. Over and above the size of their port of lading, shipments lade onto larger ships when, on route to the US, those shipments also stop at entrepôt, and indirectness appears to facilitate larger ship size beyond transshipment alone. These relationships are summarized in the following stylized fact:

Stylized Fact 3. *Goods from and through entrepôts are loaded onto larger ships.*

These facts outline an inherent trade-off: indirectness increases the distance and time costs of trade, but the resulting concentration appears to lower costs. The level of indirectness and concentration we have documented in the data are shaped by this trade-off, and the goal of our empirical estimation is to understand the forces underlying this trade-off. We first present our theoretical framework, which uses Allen and Arkolakis (2019) to recover trade costs from observed global trade and shipping patterns. We implement this estimation, then estimate the role of scale using underlying geography as an instrument for shipping volumes.

4 Theoretical Framework

In this section, we present a model of global trade where shipments are sent indirectly through an endogenously formed transport network. We embed the Allen and Arkolakis (2019) route selection model in a generalized Eaton and Kortum (2002) framework where production technologies in each industry and country are non-stochastic, but idiosyncratic variation in a product's optimal route generates random variation in the price of each product-origin pair.

Entrepôts emerge as ports through which goods flow but which are neither the goods' origins nor their destinations. Throughout, we maintain a production and consumption setting that is as general as possible, allowing for any number of goods, industries, and input-output linkages. This model is agnostic to scale economies or dis-economies in transportation costs, which could work to either amplify or attenuate shipments through entrepôts. Restrictions on route cost heterogeneity generate moment conditions that can be matched to the data to yield estimates of leg-specific shipping costs.

4.1 Setup

Consumption and Production In each country j , consumers consume goods $\omega_n \in \Omega_n$ from each of N industries n according to some function $U_j = U_j(C_j)$, where U_j is a continuous, twice differentiable function and C_j is a matrix of quantities of an arbitrarily large number of goods ω_n in industry $n \in N$ in country j .²¹ Within each industry and product category, goods are homogeneous and normal.²²

Goods can be produced using a variety of traded and non-traded inputs including labor, capital, and traded and non-traded varieties from any industry. The production technology for good ω is common for all goods in the same industry n , and includes a vector of factor inputs L , as well as inputs of other goods.²³ Production functions can vary across industries and countries. Crucially, cost minimization leads the competitive fringe of firms in each country

²¹The utility function itself is allowed to vary across destinations, and the number of goods in each industry need not be a continuum, but can be.

²²The model and empirics can accommodate arbitrarily fine industry classifications in order to ensure this assumption holds.

²³The production function is given by $q_{in}(\omega) = f_{in}(z_{in}, L_{in}, Q_{in})$ where f_{in} is a continuous, twice differentiable country-industry-specific production function, z_{in} is a production technology common to industry n and country i , L_{in} is a vector of non-tradable factor inputs, and Q_{in} is a country-industry specific matrix of inputs of other goods ω from all industries. All inputs are treated as homogeneous.

and industry to have the same marginal cost of production. Marginal cost of a good ω is

$$c_{in} \equiv c_{in}(z_{in}, W_i, P_i),$$

where P_i is the matrix of prices of all goods ω in industries n in i and W_i is the vector of factor prices in country i . Because producers in the same industry and country share the same input prices and production function, costs are shared within country-industries. These costs correspond to the classic Ricardian comparative advantage.

Pricing In order to sell goods abroad at any destination $j \in J$, a firm producing product ω in industry n must pay tariffs κ_{ijn} and iceberg transport costs $\tau_{ijn}(\omega)$ after optimally choosing the route r between i and j to minimize the shipping costs incurred. Competitive firms selling from i to j price their goods at marginal cost. Observed prices for these products at j are

$$p_{ijn}(\omega) = c_{in} \kappa_{ijn} \tau_{ijnr}(\omega).$$

where purchasers of good ω in industry n at j source the lowest cost supplier globally.²⁴

Shipping Producers seek to minimize shipping costs by choosing the lowest cost shipping route available. A shipping route r is comprised of a series of K_r legs of a journey with $K_r - 1$ stops along the way between the origin, i , (or $k = 1$) and destination j , (or $k = K_r$).

Following Allen and Arkolakis (2019), we assume that moving from stop to stop involves iceberg transport costs as well as product- and route-specific idiosyncratic cost shocks $\epsilon_{ijnr}(\omega)$.²⁵ This shock is drawn from the Fréchet distribution such that $F_{ijn}(\epsilon)$, the cumulative distribution function of the idiosyncratic draws is the following.²⁶

$$F_{ijn}(\epsilon) \equiv \Pr\{\epsilon_{ijnr}(\omega) \leq \epsilon\} = \exp\{-\epsilon^{-\theta}\},$$

where shape parameter $\theta > 0$ captures the randomness or dispersion in the choice of routes from i to j . A higher $\epsilon_{ijnr}(\omega)$ draw means that industry n has a lower cost for route r .

Accordingly, product ω 's shipping cost along route r from country i to country j is:

$$\tau_{ijnr}(\omega) = \frac{1}{\epsilon_{ijnr}(\omega)} \prod_{k=1}^{K_r} t_{k_r-1, k_r} \equiv \frac{1}{\epsilon_{ijnr}(\omega)} \tilde{\tau}_{ijr}, \quad (1)$$

²⁴Here, non-tradable goods can be assumed to have infinite transport costs.

²⁵Because of the max-stable property of the Fréchet distribution, an isomorphic specification would have firm-specific cost shocks with a finite mass of potential competitive firms in each country. This would affect the interpretation of the source of idiosyncratic variation (firm variation or product variation) and of shape parameter θ .

²⁶This distribution is the same for each product across industries so product-industry subscripts n is dropped.

where $\tilde{\tau}_{ijr}$ is the product of all leg-specific costs t_{k_{r-1},k_r} and is common to all products taking the same route r . While transport costs are usually a function of distance, we place no structure on these leg trade costs, allowing them to be a flexible function of exogenous and endogenous variables: $t = f(X_{exog}, X_{end})$. Our estimation will first maintain maximally flexible in estimates of t , then estimate a scale economy, where t is a function of distance and volume.

This structure is consistent with a host of mechanisms, including but not limited to port-level effects and leg-level scale economies.²⁷ In terms of market power, we do not directly model the decision of shipping firms, but rather consider an overall industry equilibrium within a Sutton (1991) framework, where larger markets induce more entrants and lower marginal costs, with profits being absorbed by fixed costs.²⁸ As discussed further in the text, differences between these mechanisms will not impact the model estimation but will manifest in the interpretation of scale economies and for counterfactual predictions.

Each product's ultimate shipping cost from i to j will be the minimum transport cost route over a set of all other routes for origin i , destination j , and product ω in industry n .²⁹

4.2 Equilibrium

We use the properties of the Fréchet distribution to find expressions for two observables: (1) the equilibrium mass of products that will be shipped from any origin to any destination through a specific leg and (2) the total volume of trade between any country pair. Closing the model, which we defer to Section 7, imposes goods and labor market clearing conditions, the former generating a within-industry gravity equation, the latter generating an expression for wages.

Route volume Firms from origin country i select the lowest-cost route before consumers in j select the lowest-cost intermediate good supplier across all the origin countries. We observe ω being shipped on route r from i to j only if the final price of ω , which includes both the marginal cost of production and shipping cost on route r from i to j ($p_{ijnr}(\omega)$), is lower than all other prices of good ω from all other origin country-route combinations.

We can then use the properties of the Fréchet distribution to consider the probability that a given country and route r' will be selected as the lowest cost route-supplier combination for

²⁷It also allows for spatial correlation in link costs, say between t_{kl} and t_{lm} .

²⁸We omit discussion of the optimal shipping network from the perspectives of a firm with market power, and focus on leg-level scale instead.

²⁹The price of a product ω in industry n from i to j conditional on route r is $p_{ijnr}(\omega) = c_{in}w_i\tau_{ijnr}(\omega)$.

good ω conditional on price p :

$$G_{jn\omega}(p) \equiv \Pr \left\{ \min_{i \in I, r \in R_{ij} \setminus r'} p_{ijnr}(\omega) > p \right\} = 1 - \exp \left\{ -p^\theta \cdot \sum_i \left[(c_{in} \kappa_{ijn})^{-\theta} \cdot \sum_{r \in R_{ij}} \tilde{\tau}_{ijr}^{-\theta} \right] \right\}.$$

We can define the joint probability that a route r is the lowest-cost route from i to j for good ω and that country i is the lowest-cost supplier of good ω to j as:

$$\pi_{ijnr\omega} \equiv \Pr \left\{ p_{ijnr\omega} \leq \min_{i' \in I \setminus i, r' \in R_{ij} \setminus r} p_{i'jn r' \omega} \right\} = \frac{[c_{in} \kappa_{ijn} \cdot \tilde{\tau}_{ijr}]^{-\theta}}{\sum_{i' \in I} [(c_{i'n} \kappa_{i'jn})^{-\theta} \cdot \sum_{r' \in R_{ij}} \tilde{\tau}_{i'jr'}^{-\theta}]}. \quad (2)$$

By the law of large numbers this is also the share of all goods sold in j in industry n that come from i and take route r .³⁰ We define two matrices following Allen and Arkolakis (2019). First,

$$A_n = [a_{ijn} \equiv t_{ijn}^{-\theta}], \quad (3)$$

where each element is a function of the leg-specific trade cost $a_{ijn} \equiv t_{ijn}^{-\theta}$. Second, we define the matrix B , the expected trade cost matrix, as,

$$B_n = [b_{ijn}] \equiv (I - A)^{-1}. \quad (4)$$

Using these and substituting for the definition of $\tilde{\tau}_{ij}$ (equation (1)) and summing across routes r that pass between leg k to l , we can express the share of imports in industry n in destination j that come from origin i which passes through leg k, l as:

$$\pi_{ijn}^{kl} = \left[(c_{in} \kappa_{ijn})^{-\theta} \cdot b_{nik} a_{nkl} b_{nlj} \right] \cdot \Phi_{jn}^{-1}, \quad (5)$$

where $\Phi_{jn} = \sum_{i'} (c_{i'n} \kappa_{i'jn})^{-\theta} \cdot b_{ni'j}$ is a multilateral resistance term that accounts for the average costs, openness, and connectivity of competitors from all other countries i' . This equation is the direct analogue to equation (7) in Allen and Arkolakis (2019). The key distinction here is Φ_{jn} , which accounts for the Ricardian selection of lowest-price sources for each good ω . Intuitively, the traffic flowing on a given leg responds both to that leg's effectiveness in reducing route costs as well as to competitive forces which make trades increasingly less likely to be pursued for more expensive routes. However, multilateral resistance is j, i -level, and therefore enters proportionately into traffic flows for all k, l -pairs – a fact that will be crucial for our estimation.

³⁰Recall the number of goods in each industry is set arbitrarily large so that the law of large numbers will hold. The unconditional (pre-selection) average transport cost from i to destination j : $\tau_{ijn} = \gamma^{-1/\theta} \left(\sum_{r \in R_{ij}} \tilde{\tau}_{ijr}^{-\theta} \right)^{-1/\theta}$ where γ is the function $\Gamma(t) = \int_0^\infty x^{t-1} \exp^{-x} dx$ evaluated at $\left(\frac{1+\theta}{\theta} \right)^{-\theta}$.

Furthermore from summing across industries, origins, and destinations, we can recover the share of observed global shipping that passed through leg k, l :

$$\pi^{kl} = \sum_n a_{nkl} \cdot \sum_j b_{nlj} \frac{\Phi_{kn}}{\Phi_{jn}}. \quad (6)$$

Equations (5) and (6) correspond the shares of goods passing through leg k to l , including shipments bound for l and those continuing onward to other destinations. Because they account both for optimal route selection and competition on price, they correspond to observable volumes after route selection and competition among producers.

The sum of products sold in j in industry n from country i is equal to the share of all products sold in j in industry n that come from i and take route r , summed across all r routes:

$$\pi_{ijn} \equiv \sum_r \frac{[c_{in} \kappa_{ijn} \cdot \tilde{\tau}_{ijr}]^{-\theta}}{\sum_{i' \in I} [(c_{i'n} \kappa_{i'jn})^{-\theta} \cdot \sum_{r' \in R_{i'j}} \tilde{\tau}_{i'jr'}^{-\theta}]} = \frac{(c_{in} \kappa_{ijn} \tau_{ijn})^{-\theta}}{\Phi_{jn}}. \quad (7)$$

Closing the model Factor and goods market clearing and balanced trade conditions close the model. Unnecessary for estimation, we defer them to Section 7 when we conduct counterfactuals.

4.3 The Network Effect of Adjustments on Trade

A change in the leg cost between k and l (t_{kl}) can affect trade volumes between an origin i and destination j through the trade network. However, Ricardian competition can interact with the trade network to generate unexpected effects. For any change to the cost t_{kl} , trade volumes between i and j will adjust according to the following equation:

$$\frac{dX_{ijn}}{dt_{kl}} = \frac{\partial X_{jn}}{\partial t_{kl}} \cdot \pi_{ijn} + X_{jn} \cdot \left[\frac{\partial c_{in}^{-\theta}}{\partial t_{kl}} \cdot \frac{\pi_{ijn}}{c_{in}^{-\theta}} + \frac{\partial \tau_{ijn}^{-\theta}}{\partial t_{kl}} \cdot \frac{\pi_{ijn}}{\tau_{ijn}^{-\theta}} + \frac{\partial \Phi_{jn}^{-\theta}}{\partial t_{kl}} \cdot \frac{\pi_{ijn}}{\Phi_{jn}^{-\theta}} \right].$$

The first term on the right is the effect of t_{kl} on trade with i through a change in the total volume consumed at j in industry n . The first term in parentheses is the effect through any changes to the production costs at i , which can happen if the price of inputs changes or through a change in wages. The second term in parentheses the effect through trade costs between i and j in industry n , and the final term is the effect through multilateral resistance. This final term is the effect of a change in t_{kl} on trade volumes from i via multilateral resistance.

What can we say about the signs on these terms? When the trade cost matrix is endogenous to trade volumes, as it would be in the presence of scale economies, these terms are ambiguous,

as a change in t_{kl} , by changing trade volumes, changes traffic volumes at each leg, and therefore equilibrium effects on the full matrix of trade costs.

When there is no endogenous scale response, only the final term can be negative. Intuitively, a reduction in trade costs between k and l can increase consumption at j , reduce expected trade costs between i and j , and reduce production costs at i . All of these result in an increase in trade volumes between i and j . However, a reduction in trade costs between k and l also stiffens competitions at j . If this last effect is large enough, it can overturn the sign of the first three. In this scale-free case, the total effect is positive if and only if the following condition is true:

$$\epsilon_{X_{jn}, t_{kl}} + \epsilon_{c_{in}, t_{kl}} + \epsilon_{\tau_{in}, t_{kl}} > -\epsilon_{\Phi_j, t_{kl}}. \quad (8)$$

That is, if the elasticities of consumption at j ($\epsilon_{X_{jn}, t_{kl}}$), production costs at i ($\epsilon_{c_{in}, t_{kl}}$), and trade costs between i and j ($\epsilon_{\tau_{in}, t_{kl}}$) with respect to t_{kl} are larger than the elasticity of multilateral resistance at j with respect to t_{kl} ($\epsilon_{\Phi_j, t_{kl}}$). Rearranging terms, we have $\frac{\partial X_{ijn}}{\partial t_{kl}} > 0$ if and only if:

$$\epsilon_{X_{jn}, t_{kl}} + [\epsilon_{c_{in}, t_{kl}} + \epsilon_{\tau_{in}, t_{kl}}] (1 - \pi_{ijn}) > \sum_{i'j} (\epsilon_{c_{i'n}, t_{kl}} + \epsilon_{\tau_{i'jn}, \tau_{kl}}) \pi_{i'jn}. \quad (9)$$

The sum of the effects on production and transport costs between all other countries i' (other than i) and j has to be less than a function of the effects on production and transport cost at i and the overall propensity of consumption at j to grow. This last expression shows most clearly that the effect of a decline in trade costs between k and l has the potential to negatively affect trade flows between i and j . In particular, if the shift differentially favors trade and production costs from other countries to j , trade volumes from i to j will suffer.

Finally, because the elasticity $\epsilon_{\tau_{in}, t_{kl}}$ is equivalent to the proportion of trade from i to j that goes through k, l , a decline in t_{kl} is more likely to positively affect X_{ij} the more k, l is used in proportion to other routes and the higher that proportion is relative to the same proportion for other countries i' . The higher the same proportion is in other countries, the more likely trade volumes between X_{ij} will fall with a fall in trade costs t_{kl} .

5 Trade Cost Estimation

To estimate counterfactuals that alter the geography of container trade, we need point-to-point container trade costs, the matrix A . In this section we estimate A using observable data.

5.1 Estimation Equation

Using equations (5) and (7) we can calculate the probability of any good traveling through leg k, l conditional on being sold from origin i to destination j . If X_{ijn} is the total value of trade between i and j in industry n , we can express the total volume of traffic between k and l in a given industry n as:

$$\Xi_n^{kl} \equiv \sum_i \sum_j X_{ijn} \cdot b_{ikn} a_{kln} b_{ljn} b_{ijn}^{-1}. \quad (10)$$

This equation is identical to Allen and Arkolakis (2019). Conditional on the observed trade values X_{ijn} , the contribution of trade between i and j to the traffic between legs k and l is invariant to multilateral resistance, tariffs, or technology. This is despite the significant differences between our frameworks. In particular, expensive trade routes here suffer from Ricardian selection at destination markets, where the route's impact on prices make them less competitive. Yet, this does not impact the estimation of trade costs.

The intuition for this result is that Ricardian selection, tariffs, and multilateral resistance all operate through adjusting the total value of trade, but do not differentially favor one route from an origin i to a destination j . Put differently, any change to a non-transportation related costs in one country will affect trade from that country and others proportionally on all routes.

Equation (10) gives us a relationship between trade values, trade costs, and traffic for a given industry. To map our model into the data we make one final assumption: there is a set of industries \bar{N} for which trade costs are identical and all trade ($X_{\bar{N}} \equiv \sum_{n \in \bar{N}} X_n$) and traffic ($\Xi_{\bar{N}} \equiv \sum_{n \in \bar{N}} \Xi_n^{kl}$) are observable. Summing equation (10) over industries $n \in \bar{N}$ yields

$$\Xi_{\bar{N}}^{kl} = \sum_i \sum_j X_{i\bar{N}} \cdot b_{\bar{N}ik} a_{\bar{N}kl} b_{\bar{N}lj} b_{\bar{N}ij}^{-1}. \quad (11)$$

Equation (10) tells us that to accurately measure transport costs, we only need data on transportation and traffic for all goods in an industry. Equation (11) tells us that we can use traffic across multiple industries so long as we have the correct trade aggregate, we see all traffic for those industries, and we can assume transport costs are identical in those industries. We implement equation (11) using observed total containerized traffic, isolating containerized industries, and assuming that transport costs are similar across containerized industries.

5.2 Recovering Trade Costs

Our estimating equation requires two observable objects: trade values and traffic volumes.³¹ We use 2014 US Customs data on containerized and non-containerized shipments to construct the share of each HS 4-digit commodity code that is transported by container. All commodities with a containerized share above 80% are labeled as containerized.³² We then collapse the 2014 EORA International Input-Output database at the country level, segregating containerized and non-containerized commodities (Lenzen et al., 2012).³³ We can then extract a country-level trade matrix for containerized commodities, X , for all countries in the database. For clarity, we drop industry subscript n , only considering containerized industries.

In an ideal world, estimation would recover the trade costs that directly rationalize observed bilateral containerized traffic flows—a just identified case. While we directly observe ocean traffic, our data omits land based trade and internal within-country trade. This is an issue for network links between geographically contiguous countries.³⁴

We overcome this limitation by assuming a functional form that allows for estimation without requiring the direct observation of overland links. We consider the exponential mapping:³⁵

$$a_{ij} = t_{ij}^{-\theta} = \frac{1}{1 + \exp(Z\beta)} \in [0, 1],$$

where a_{ij} is an element of the matrix A and the matrix Z is a vector of covariates defined as

$$\begin{aligned} Z\beta = & \beta_0 + \beta_1 \log \text{sea distance}_{ij} + \beta_2 \log \text{traffic}_{ij} + \beta_3 \log \text{traffic}_i \\ & + \beta_4 \log \text{traffic}_j + \beta_5 \mathbb{1}_{backhaul} + \beta_6 \mathbb{1}_{\{i, j \in \text{Land Borders}\}}, \end{aligned}$$

where β_0 is an intercept, β_1 considers sea distance between the nearest principal port,³⁶ and

³¹This procedure is agnostic to the exact specification of any particular trade model that generates trade value flows X . By conditioning estimation on these flows X , all origin, destination, and origin-destination factors are controlled for. In particular, items such as all origin-destination tariffs and non-tariff barriers are all accounted for. This does not mean that we can disentangle the two, rather we can directly account for these factors collectively.

³²In practice we find a bimodal distribution, with some commodities being never containerized (e.g. oil and iron ore) and others always containerized (e.g. washing machines and children’s toys).

³³We are implicitly shutting down the substitution between containerized and non-containerized transport. This is supported by our bi-modal distribution of goods. While in the past, some ports did not handle containerized trade, by 2014, nearly all ports had some ability to handle containerized trade.

³⁴For example, containerized trade from Canada to the US can occur via truck or train, and will not be observed in our traffic matrix Ξ , even though it appears in the trade matrix X .

³⁵This functional form maps from the real numbers to the unit interval, as is required by our theory.

³⁶For each country pair, we calculate the volume-weighted mean sea distance across all port pairs. These data are available for download from our websites.

β_2 considers port-to-port traffic. β_3 and β_4 consider the total incoming and outgoing traffic at ports i and j respectively. β_5 considers the role of the backhaul problem from Wong (2019), where ship capacity is fixed by the shipping direction with the higher demand. Finally β_6 considers an indicator variable for two countries that share a land border.³⁷

It is crucial to note two things about this strategy. First, the above equations posit relationships between observables, such as distance and traffic flows to trade costs. However, at this stage these relationships are not of interest to us. Our objective is not the vector β of coefficients, but the resulting predictions for a_{ij} . We accept that elements of the vector β will be endogenous, seeking instead to fully saturate the variation in the data in order to generate the closest prediction the data can yield for the matrix A relative to the just-identified case. This recovers the trade costs while being agnostic to their underlying determinants, including potential market power as well as possible geographic indicators. Secondly, note that parameters for β will yield estimates of every trade cost a_{ij} , but we need not discipline β by comparing traffic on every link. We can omit within-country traffic as well as traffic between countries that share overland routes and still recover estimates of a_{ij} .

Our estimation finds the vector β that that minimize the differences of expected traffic, $\hat{\Xi}(A(\beta); X)$, which is constructed from estimates of β as well as trade data X and observed traffic for countries that do not share a land border:

$$\arg_{\beta} \min \sum_{ij \neq \text{land borders}} \left| \Xi^{kl} - \hat{\Xi}^{kl}(A(\beta); X) \right|.$$

As noted before, sharing a land border indicates that we may not fully observe traffic of containerizable goods on that given leg, and we do not match traffic data from these countries with our expected traffic. Yet, disciplined by non-land border traffic flows, guesses for a_{kl} for overland link k to l are still generated from any guess of β , and these along with observed trade flows X generate predicted traffic flows for sea-only legs where traffic is observed.

Appendix Figure A.12 graphs our resulting matrix of pairwise trade costs. We present the vector β estimates in the Appendix Table A.5 as purely predictive parameters, not fundamentals that we can alter in the counterfactuals. Instead, we simply need to know if our β estimates produce containerized ship traffic that reflects the world. With a full A matrix, we also can

³⁷We do not estimate the diagonals a_{ii} , the cost of shipping from one's own port back to itself. We assume that these costs do not change in the counterfactual and we only estimate our data on international trade data, abstracting away from domestic trade. Allen and Arkolakis (2019) provide estimates of internal trade costs.

generate a full B matrix, or average bilateral transport cost between points. Appendix Table A.6 compares these bilateral trade costs to distance measures more commonly used.³⁸

Alternative Data Definitions Estimates of a_{ij} are at the country-level. With port level traffic, estimation of a port-level cost matrix is possible. However, that estimation requires sub-national trade data X , which is not broadly available. Using port traffic and imputed trade data, we can guess bilateral port trade data and run a version of the above estimation. Results from the port-level estimation are broadly in line with results of main estimation and later scale elasticity estimation, with the correlation between weighted port-pair costs and country-pair costs of 0.6. However, due to the highly speculative assumptions required to estimate sub-national trade data, we view country-level estimates as more accurate.

5.3 Model Fit and External Validity

First, we check our estimation results by comparing our model-predicted traffic and trade values against their observed counterparts in the data (Figure 7). In Panel (A), we compare actual observed global container traffic shares with the our model-predicted shares using our β variables and estimated trade costs. We include both a best fit line and a 45 degree line. As we achieve the ideal just identified case, all observations would line up along the 45 degree line. In general, we fit the data extremely well, with a correlation between the observed and predicted shares (in logs) of 0.97. In Panel (B), we compared actual observed trade shares to our estimated trade shares.³⁹ Even though our model does not explicitly target trade values, we still fit the data well with a high correlation (in logs) of 0.73.

To assess the model’s ability to capture the actual paths of goods, we compare model predictions for traffic for US-bound shipments to our US microdata. Our estimation, which uses global traffic data rather than US microdata, delivers predictions for how US-bound shipments travel through the shipping network—the routes US imports take. Specifically

$$\widehat{\pi}_{iUS}^{kl} = b_{ik} a_{kl} b_{lj} b_{ij}^{-1},$$

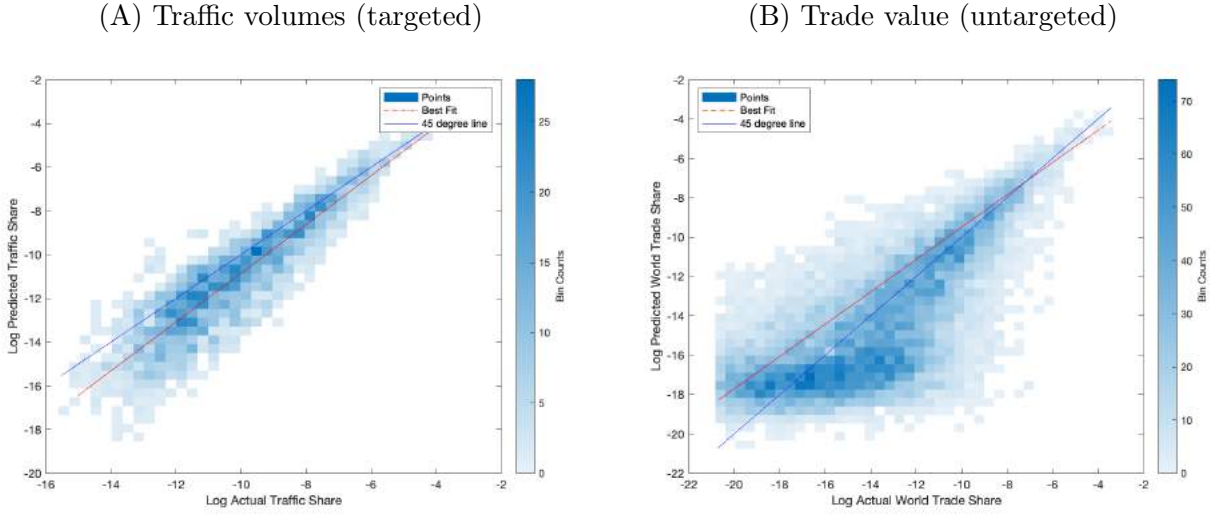
is the ratio of all shipments from i to the US that are observed flowing through leg k, l .

This ratio can be matched to our compiled microdata on shipment-level observations of

³⁸Using the B -matrix, we make our bilateral transport cost estimates and country-level measures of market access available for future research.

³⁹To generate trade flows, we close the model using the full setup in 7.

Figure 7: Model Fit Comparisons



Notes: Panel (A) compares our predicted container traffic volumes from any two ports to the actual container traffic volumes (normalized as a share to total world container traffic). Panel (B) compares aggregate trade shares versus predicted trade shares for containerized traffic. This computation uses the full model described in Section 7. These moments are not targeted by the trade cost estimation. Source: Authors' calculations using AIS and Bill of Lading data. ■

individual routes for US-bound trade in Sections 2 and 3. Note that this microdata is not used in the estimation. The model is instead estimated using containership port of call data at the global level. We can compare the above value from the estimation to the proportion of goods coming into the US from any origin i on leg k, l in our microdata, which we call $\pi_{iUS,Data}^{kl}$.

Table 2: External Validity Checks

	(1)	(2)	(3)	(4)
	$\widehat{\pi}_{iUS}^{kl}$	$\widehat{\pi}_{iUS}^{kl}$	$\widehat{\Xi}_{kl}$	$\widehat{\Xi}_{kl}$
$\pi_{iUS,Data}^{kl}$	0.846 (0.119)	0.872 (0.121)		
$\Xi_{kl,Data}$			1.224 (0.128)	1.240 (0.126)
Observations	13813	366010	652	2153
Data	Nonzero	Global	Nonzero	Global
R^2	0.513	0.513	0.659	0.669
F	50.58	51.79	91.75	97.04

Notes: Standard errors clustered by origin and destination countries. $\widehat{\pi}_{iUS}^{kl}$ is the model-predicted share of goods from origin i to US destination flowing through leg k, l and $\widehat{\Xi}_{kl}$ is the model-predicted total US-bound traffic on a given leg k, l . Their corresponding variables observed in the compiled microdata are indicated with subscript "Data", $\pi_{iUS,Data}^{kl}$ and $\Xi_{kl,Data}$. Columns (1) and (3) are restricted to nonzero traffic volumes in the US microdata (Nonzero Data). Columns (2) and (4) includes journeys with zero traffic volumes in the US microdata (Global Data).

Source: Authors' calculations using AIS and Bill of Lading data. ■

Column 1 of Table 2 reports the univariate regression, weighted by total origin TEU. The coefficient is strong and positive, and we cannot exclude a 1-to-1 relationship. Over half of the variation in the observed distribution can be explained using the predicted probabilities.

In Column 2, we add back in legs for which there are no observed journeys, or zero traffic volumes, in the microdata. Our model predicts that there should be some small amount of traffic volume on every leg as demonstrated by the 30-fold jump in observations. However, because our model predicts extremely low volumes of trade on these legs, including these links does not significantly change our estimates or model fit.

Next, summing predicted probabilities across origins, the model delivers a prediction for the total amount of US-bound traffic on a given leg:

$$\widehat{\Xi}_{kl} = \sum_i X_{iUS} \pi_{ijkl}^* \tag{12}$$

where X_{iUS} is the total volume of trade from origin i to the US. Again, we can compare this to the total volume of shipments in the data moving between a given leg, which we call $\Xi_{kl,Data}$.

Column 3 and 4 report these univariate regressions, again adding back in legs with zero observed volumes in 4. Here as well, the coefficient is significantly positive and not different from 1. The R-squared is now higher – close to 0.70. These results are also robust to tobit specifications which allow for lower and upper censoring limits.

By showing that there is a tight fit between our model estimates to our micro-data on US shipment routes, our paper serves as a check to the validity of the Allen and Arkolakis (2019) approach. Allen and Arkolakis (2019) imputed traffic and trade flows within the US highway system for their estimation since this data is not readily available.⁴⁰

6 Scale Economy Estimation

Results from Section 3 imply that scale economies in shipping are crucial to understanding the role of entrepôts play in global trade. As we turn to counterfactual estimation, the existence of such a scale economy implies that perturbations to the global shipping network that change trade volumes will in turn impact the leg cost matrix estimated in the previous section. Such effects must be accounted for in order to correctly estimate counterfactual adjustments. In this

⁴⁰They assume that the observed traffic for a link is proportional to the underlying value of trade on that link. This assumption is later on verified by comparing their predicted trade flows to actual flows from the Commodity Flow Survey.

section, we estimate the causal impact of scale, or traffic volume, on leg-level trade costs.

6.1 Naive Scale Elasticity

As discussed in Section 4, scale economies, a negative relationship between the volume of goods traveling on a leg and the cost on that leg, could result from a number of different mechanisms. We begin by estimating a “reduced form” relationship between leg costs obtained from our estimation and traffic volumes. Column 1 of Table 3 reports the following specification

$$\ln c_{kl} = \alpha_0 + \alpha_1 \cdot \ln \Xi_{kl} + \alpha_2 \cdot \ln d_{kl} + \epsilon_{kl}, \quad (13)$$

where α_0 is a constant, α_1 is the ordinary least squares (OLS) relationship between price and quantity, $\alpha_2 \cdot \ln d_{kl}$ is the coefficient and measure of log sea-distance from k to l respectively, and $c_{kl} = a_{kl}^{-1} - 1$, which allows us to interpret α_1 as the elasticity between cost and traffic volumes to a trade elasticity θ . That is, to interpret our results as elasticities, they must be deflated by trade elasticity θ , which we do not recover directly.

Column 1 of Table 3 reports the OLS relationship between leg-level transport costs and traffic. Unsurprisingly, the strong negative coefficient echoes the strong negative relationship between costs and traffic found in our GMM estimation. For a trade elasticity value of $\theta = 4.5$ (Simonovska and Waugh, 2014), this relationship implies legs with 1% more traffic volume are associated with 0.18% lower transport costs. Although we are estimating a leg-level elasticity, this is similar to what has been established in the ocean shipping and scale literature.⁴¹

Of course, this OLS relationship between price and quantity cannot be taken as causal. Lower cost legs may face larger demand precisely because unobserved cost-reducers induce higher levels of demand on those legs. Essentially, we wish to observe the supply elasticity, but we have only market-clearing prices and quantities. We therefore need a demand shifter.

6.2 Geography-Based Instrumental Variables

We build such a shifter by using the intuition of our model to construct a geography-based instrument for demand. Demand for a given leg will be higher if the leg lies along an otherwise lower-cost route between an origin and a destination. Routes from South Korea to the Netherlands that include the leg China-Singapore, for example, are closer to the direct route between Korea and the Netherlands compared to the leg China-Australia. As such, more Korea-

⁴¹Asturias (Forthcoming) reports an origin-destination country trade-volume trade-cost elasticity of 0.23 while Skiba (Forthcoming) reports an elasticity of 0.26 using product-level import data from Latin America.

Table 3: Scale Elasticity Estimates

	(1)	(2)	(3)	(4)	(5)
	$\ln c_{kl}$	$\ln c_{kl}$	$\ln \Xi_{kl}$	$\ln c_{kl}$	$\ln c_{kl}$
$\ln \Xi_{kl}$	-0.814 (0.0113)			-0.267 (0.128)	-0.423 (0.0919)
$\ln z_{kl}$		-0.0419 (0.0274)	0.157 (0.0372)		
$\ln d_{kl}$	0.495 (0.0329)	0.652 (0.0607)	-0.272 (0.0769)	0.580 (0.0499)	0.649 (0.0732)
k-level FE					Y
Specification	OLS	RF	1st St	IV	IV
Observations	1947	1947	1947	1947	1947
R^2	.89	.14		.55	.74
KP F-stat				17.79	15.27

Notes: Robust standard errors in parentheses clustered two-ways by nodes k and l .
Source: Authors' calculations using AIS and Bill of Lading data. ■

Netherlands trade should flow through the China-Singapore link than the China-Australia link, which would involve a longer detour before eventually arriving in the Netherlands.

Effectively, we wish to calculate how far out of the way a given leg would be for most journeys. To operationalize this intuition, we relate the direct sea-distance between an origin and a destination to the distance of two legs as part of a four-leg journey, where the omitted middle leg is the object of interest. For each kl pair, we calculate the instrument z_{kl} as:

$$z_{kl} = \sum_{i \setminus \{k,l\}} Pop_{i,1960} \sum_{j \setminus \{k,l\}} Pop_{j,1960} \frac{d_{ij}^2}{(d_{ik} + d_{lj})^2}, \quad (14)$$

where d_{ij} is the sea distance between origin i and destination j , and the square of the relative excess distance between legs $i - k$ and $l - j$ ($d_{ik} + d_{lj}$) is weighted by the year 1960 population at each origin i and destination j , $Pop_{i,1960}$ and $Pop_{j,1960}$.⁴²

Links including Singapore score favorably on this measure; its strategic location by the Straits of Malacca makes it “on the way” for many key country pairs. Column 2 of Table 3 reports estimates from the reduced-form relationship between our instrument and link costs:

$$\ln c_{kl} = \beta_0 + \beta_1 \cdot \ln z_{kl} + \beta_2 \cdot \ln d_{kl} + \varepsilon_{kl}. \quad (15)$$

We find that higher values of the instrument—more strategic locations—are indeed associated

⁴²1960 Population here stands in place of GDP, which may be endogenous to the trade costs in our model. The year is chosen both because immigration and populations prior to 1960 could not plausibly be impacted by 2014 containerized shipping costs.

with lower costs as we predict. Our first stage results are given by the following (Column 3):

$$\ln \Xi_{kl} = \gamma_0 + \gamma_1 \cdot \ln z_{kl} + \gamma_2 \cdot \ln d_{kl} + v_{kl}. \quad (16)$$

The F-statistic confirms the instrument is strong. Both reduced-form and first-stage show relatively strong variation in the direction we expect. On average, strategic legs along more distance-minimizing journeys exhibit higher traffic volumes.

Column 4 reports our instrumented scale elasticity from our second stage regression (equation (13)). Our demand shifter reduces the scale elasticity in Column 1, roughly halving the result. This is consistent with the direction of bias we would expect from reverse causation. For the widely used value $\theta = 4.5$ (Simonovska and Waugh, 2014), the interpretation of our causal estimate is that increasing volume on a route by 1% would reduce costs by 0.05%.⁴³ These results are consistent with our third stylized fact linking trade and ship-size in Section 2. and lend support to our initial hypothesis that a major role of entrepôts are their facilitation of scale through concentration of shipments.

An important note is that many specific mechanisms may be at work generating this observed scale economy, including ship size and market power among others. Indeed, because a multitude of such mechanisms may be at work simultaneously, we choose a model-consistent, agnostic approach in our estimation of scale. Yet different mechanisms may generate different out of sample results, and further work should be done to isolate and test for these.

6.3 Validity of Identification Strategy

In order for the IV strategy to be valid, our demand shifter instrument has to be generally uncorrelated with unobserved changes in cost determinants for a particular leg ($\text{corr}(\epsilon_{kl}, \ln z_{kl}) = 0$). We include leg-level sea-distance to directly control for the physical leg-level costs of shipping.

What are the possible threats to identification? We are chiefly concerned with port-level omitted variables that could put a downward bias on our estimates, rather than differences in at-sea costs. For example, port level infrastructure investments could reduce costs and increase demand, showing up as a scale economy where none exists.

To address this concern we add origin-port fixed effects to the regression in column 5. The coefficient becomes larger in magnitude, closer to the OLS. This is not what we should expect

⁴³This leg-level elasticity is more modest, but broadly consistent with the strong scale economies from ship size in Cullinane and Khanna (2000), which measure origin-destination elasticities that would compound, on average, three leg level elasticities.

if omitted port-level cost reduces are correlated with our demand shifter, but is the direction we expect if port-level congestion is being accounted for. Accordingly, we rely on our column 4 elasticity, our most conservative and model-consistent estimate, to account for the effects of higher volumes on route costs in counterfactuals.

A second threat to identification could be that locations that are strategically close to each other in sea distance are also close to each other in land distance and have easier access to other modes of transportation like road or rail. This availability of other modes along such kl legs could be potentially correlated with any unobserved cost determinants ϵ_{kl} . To address this concern, we recalculate our instrument in equation (14) by omitting the shortest 10 percentile distances for each origin i and destination j respectively.⁴⁴ Our results are robust to this setting—our coefficients retain the same signs and stay within a standard error of the baseline results (Table A.3). We find a stronger negative correlation (higher scale elasticity) between costs and more strategic locations that are further away. This may imply that strategic locations play a bigger role in reducing trade costs through scale when they are further away.

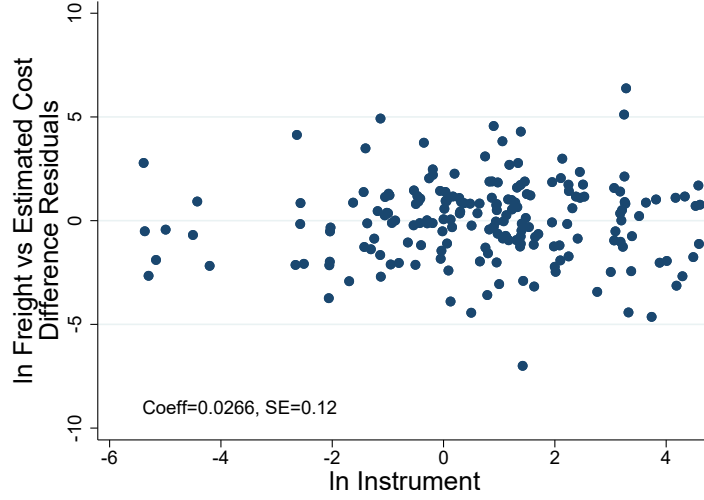
While it is not possible to directly test the validity of our exclusion restriction, we can show the lack of correlation between our instrument and an approximation of ϵ_{kl} . We can interpret ϵ_{kl} as the error or difference between the model-implied cost c_{kl} and the true leg transport cost. Using external measures of freight costs from Wong (2019), we can calculate the difference between these external costs to our model-implied costs for a subset of legs. This difference should approximate ϵ_{kl} and be correlated with ϵ_{kl} . Figure 8 shows a weak and insignificant correlation between this residualized approximation of ϵ_{kl} , after controlling for sea-distance, and our instrument. This is insufficient to concretely validate our instrument, but performs the same role as a balancing test, showing an absence of evidence of exclusion restriction violations.

7 Counterfactuals

In order to estimate our counterfactuals, we now introduce structural assumptions into our general theory model to deliver a quantifiable general equilibrium framework. With a fully specified model and estimates of both trade costs and scale elasticizes, we consider two counterfactuals. The first considers the trade cost effects of global warming, with the Arctic opening up to trade

⁴⁴To use our previous example, this means the set of routes from South Korea to the Netherlands would exclude the shortest legs from South Korea (like Japan) as well as the shortest legs to the Netherlands (like Belgium or France).

Figure 8: Balancing Test: Residualized plot of correlation between instrument and approximation of unobserved cost determinants



Notes: The approximation of unobserved cost determinants on the y-axis is the difference between external measures of freight costs and the model-implied costs. This difference is residualized after controlling for sea-distance. Robust standard errors are clustered two-ways by nodes k and l .
 Source: Container freight rates from Wong (2019), Authors' calculations using AIS and Bill of Lading data. ■

between the Pacific and Atlantic Oceans, bypassing the Suez and Panama canals. The second considers the role of a negative trade shock, the United Kingdom leaving the European Union.

7.1 Closing the Model

To close our model, we adopt the Caliendo and Parro (2015) framework. A continuum of intermediate goods ω_n are used in the production of composite goods that are in turn used domestically both as final goods and as materials for intermediate production by firms in each industry n . We assume there are three sectors ($N = 3$): containerized tradables c , non-containerized tradables nc , and nontradables nt ($n \in [c, nc, nt]$). Intermediates in the nt sector are only sourced domestically while ω_{nc} and ω_c goods are sourced internationally. Trade routes are modeled for all three sectors but will only be estimated for intermediates ω_c .

Consumption In each country i , consumers consume composite goods m_{in} from each sector n , maximizing Cobb-Douglas utility.

$$U_i = \prod_n m_{in}^{\eta_n}; \text{ where } \sum \eta_n = 1,$$

where η_n is the Cobb-Douglas industry share, $\sum_n \eta_n = 1$. Since each product ω in sector n is homogeneous, consumers in country j will choose the lowest-cost provider of each such product.

Intermediate goods production The traded goods are intermediates, which are used in each country as building blocks for industry composite goods. In each country i and industry n , firms produce a continuum of intermediate goods, indexed in each industry by $\omega_n \in \Omega_n$. There are two types of input required for the production of ω : labor and composite goods. The production of intermediate goods across countries differs in their efficiency by a country-industry specific constant z_{in} , a Ricardian technology. The production technology for intermediate ω is

$$q_{in}(\omega) = z_{in} [l_{in}]^{\gamma_{in}} \prod_{n'}^N [m_{in}^{n'}]^{\gamma_{in}^{n'}},$$

where l_{in} is labor. $\gamma_{in}^{n'}$ is share of materials from sector n' used in production of intermediate good ω , γ_{in} is share of value added, with $\sum_{n'}^N \gamma_{in}^{n'} = 1 - \gamma_{in}$.

The marginal cost of production for firms is

$$c_{in} \equiv \frac{\Upsilon_{in} w_i^{\gamma_{in}} \prod_{n'}^N P_{in'}^{\gamma_{in}^{n'}}}{z_{in}}, \quad (17)$$

where w_i is the wage in country i , $P_{in'}$ is the price of a composite good from sector n' , and constant $\Upsilon_{in} = \prod_{n'}^N (\gamma_{in}^{n'})^{\gamma_{in}^{n'}} (\gamma_{in})^{\gamma_{in}}$.

Composite goods production In each country i , composite goods in industry n are produced using a CES aggregate of intermediates Ω_n , purchased and sold domestically at marginal cost. In traded industries, intermediates are sourced internationally from lowest-cost suppliers. For a given product, the price available to composite producers in j is

$$p_{jn}(\omega) = \min_{i,r} \{ c_i \kappa_{ijn} \tau_{ijnr} \}, \quad (18)$$

where τ_{ijnr} is the realized trade cost from i to j in industry n .

When compared to the relevant equation in Caliendo and Parro (2015), it's clear that idiosyncratic route draws are generating the stochastic price dispersion usually assumed to be idiosyncratic TFP. Using the standard aggregation, the resulting price at j of the composite in industry n is expected to be the following (where A_n is a constant):

$$P_{jn} = A_n \left[\sum_{i=1}^I c_i^{-\theta_n} \kappa_{ijn}^{-\theta_n} \tilde{\tau}_{ijn}^{-\theta_n} \right] \quad (19)$$

7.1.1 Equilibrium in changes

The general equilibrium can be defined using hat algebra which dictate how equilibrium-set parameters adjust to a change in trade costs $\hat{t}_{kl} = t'_{kl}/t_{kl}$ through a change in expected trade costs $\hat{\tau}_{ijn} = \tau'_{ijn}/\tau_{ijn}$. Changes in tariffs, trade costs, and/or productivity alter the endogenous costs of production, price indices, wage levels, trade flows, and welfare.⁴⁵ See Appendix F for full details. Formally we solve for how wages and prices change $\{\hat{w}_i, \hat{P}_i\}$ as a function of changes of our model primitives, $\{\hat{\tau}_{ijn}, \hat{z}_{in}, \hat{\kappa}_{ijn}\}$. We compute welfare as the change in real wages, $\{\hat{w}_i, \hat{P}_i\}$ as a function of changes $\{\hat{\tau}_{ijn}, \hat{z}_{in}, \hat{\kappa}_{ijn}\}$. Furthermore, we can also compute changes in marginal costs \hat{c}_{in} and trade volumes \hat{X}_{ij} .

7.2 Counterfactual Methodology

Algorithm 1 Scale Counterfactual Algorithm

- 1: **procedure** WELFARE CHANGE(X_0, Ξ_0, \hat{t}) ▷ Find a new equilibrium
 - 2: Initialize current trade flows X_0 and traffic Ξ_0
 - 3: Initialize changes in cost fundamentals $\hat{\tau}$ ▷ Example: shipping distances changes
 - 4: Compute $A_0 = A(\Xi_0; \hat{\tau})$ ▷ Following equation 13
 - 5: Compute $B_0 = (I - A_0)^{-1}$
 - 6: Initialize difference = ∞ , tolerance = ϵ
 - 7: **while** *difference* < *tolerance* **do**
 - 8: Update trade flows $X_1 = X(B_0)$ ▷ Solving 7.1.1
 - 9: Update traffic $\Xi_1 = \Xi(X_1, A_0, B_0)$ ▷ Following equation 11
 - 10: Update leg costs $A_1 = A(\Xi_1)$
 - 11: Update trade costs $B_1 = (I - A_1)^{-1}$
 - 12: Compute *difference* = $\sum_{ij} (B_1 - B_0)^2$
 - 13: Update $A_0 = A_1$ and $B_0 = B_1$
 - 14: Return final trade flows X_1
 - 15: Compare welfare and price index changes between X_1 and X_0 ▷ Solving 7.1.1
-

We combine our trade volume data with country level input-output data from the EORA database aggregating over three sectors: non-traded goods, container-shipped traded goods and non-container traded goods and use country level consumption and production data to compute Cobb-Douglas shares η and γ .⁴⁶ We follow the literature and conservatively set $\theta = 4$ (Simonovska and Waugh, 2014). For any change in trade costs $\hat{\tau}$, we can calculate changes in any country's price index \hat{P} and trade flows \hat{X} .

⁴⁵As in the literature we assume that trade is balanced up to a constant deficit shifter.

⁴⁶We hold trade deficits constant in all our counterfactuals.

We map the change in distance to trade costs through equation (13), with α_2 relating the change in distance to our trade costs. We use this, translated through the model, to reflect changes in the realized trade cost matrix B between every bilateral trading in our data - even those that are not directly connected with each other. For scale, we model a new equilibrium in the short-to-medium run, by following an iterated procedure in Algorithm 1. In this procedure, we start at today’s equilibrium and allow all shippers optimize their transportation patterns. We then allow trade costs to shift due to scale economies for all origin-destination pairs. We then iterate allowing re-optimization until a new stable equilibrium is reached. By construction, there may be many alternative equilibrium, however we focus on the unique equilibrium from our current starting point - the world today.⁴⁷

Table 4: Aggregate Counterfactual Outcomes

	Directly Affected Routes	Full Trade Network Effects	Allowing Scale Economies
<u>Panel A: Arctic Passage</u>			
Δ Average Global Welfare	0.015%	0.033%	0.089%
Δ Container Trade Volumes	0.174%	0.382%	1.018%
<u>Panel B: Brexit</u>			
Δ Average Global Welfare		-0.023%	-0.100%
Δ Container Trade Volumes		-0.245%	-1.127%

Notes: Notes: In Panel A, we first only reduce the trade costs between country pairs, whose distance is shortened by crossing the Arctic Ocean. Second, we allow for the full network structure B to allow for indirect shipping to also cross the Arctic Ocean. Third, we allow for scale economies to create a feedback loop. In Panel B, we model Brexit as an effective 5% tariff increases for all trade originating or destined for the United Kingdom. The three columns reflect the naive effect only on British trading partners, the full network effect over all trade partners, and finally include the full effects of scale economies. See text for full details. ■

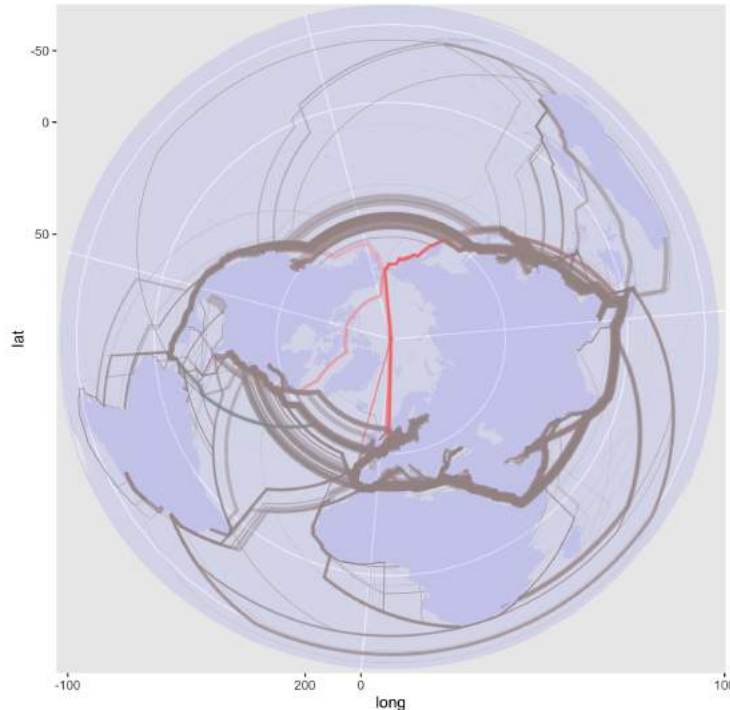
7.3 The Arctic Passage

We model the opening of the once-fabled Northeast and Northwest Passages through the Arctic Ocean between North America, Northern Europe and East Asia as a viable shipping route due to global warming. As an example, a ship traveling from South Korea to Germany would take roughly 34 days via the Suez Canal but only 23 days via the Northeast and Northwest Passages

⁴⁷Kucheryavy, Lyn and Rodríguez-Clare (2019) establishes a common mathematical structure that characterizes the unique equilibrium in multi-industry gravity trade models with industry-level external economies of scale. Their structure requires that the product of the trade and scale elasticities to be not higher than one, which is satisfied in our case.

(the Economist, 2018). Today, a small number of ships make this journey, but within the next 5-10 years, regular trips are expected (Reuters, 2019). For every bilateral pair, we consider the change trade cost for containerized trade due to changes in shipping distance using the shortest ocean-going distance between ports.⁴⁸ We find a more than 30% reduction in average shipping distances for the top 15 routes. This has a direct effect on the underlying trade cost matrix A for these Northern European, North American, and Asian countries, with no difference for all other routes. Figure 9 compares the top 150 existing shipping routes today and shortest ocean-going distance of these routes after the Arctic sea passage is viable. Existing shipping routes are highlighted in blue, new routes going through the Arctic passage are in red, and non-changing routes are in brown. We compute the change in distance using Dijkstra’s algorithm on a world map with and without arctic ice caps (Appendix A.2).

Figure 9: Comparison of shipping routes: existing and after the opening of the Arctic sea



Notes: Blue lines indicate existing shipping routes, red lines indicate the Arctic sea routes, and brown lines indicate routes that do not change. The width of each route reflects the total number of containers (TEU) on that route. Source: Authors’ calculations based on AIS and Bill of Lading data. ■

To decompose the effects of (1) bilateral distance, (2) the global container transportation network, and (3) scale economies on modeled trade costs $\hat{\tau}$, we run three variants. First, we

⁴⁸We measure the mean ocean-going distance between all bilateral ports within each origin-destination pair. The opening of Arctic sea routes, allows for ships to pass through the Bering sea from the Pacific to the Arctic Ocean and through the Labrador/Norwegian seas from the Atlantic to the Arctic.

directly only allow for changes in trade costs only on routes that are directly affected by our counterfactual scenario. This captures the direct effect of distance. Second, we then allow for these trade costs to affect all trade, including indirect trade. This reflects the notion that even countries that do not directly ship between each other - for example between China and landlocked Ukraine - have changes in trade costs. Third, as trade costs change, trade volumes change, causing a feedback loop through our estimated scale elasticity.

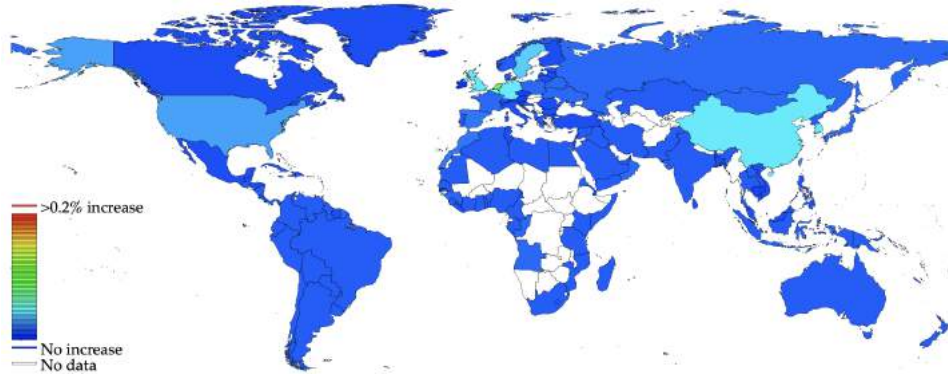
In each of the three simulations, we first show the aggregate gains, averaging across global welfare, before decomposing our heterogeneity. We display the summary statistics for all three scenarios in Panel A of Table 4. The first column, shows that with our simply input-output structure, the direct effects of the Arctic Passage are positive, with aggregate welfare increasing 0.015%, and container trade volumes increasing 0.2%. Allowing for the full trade network, including indirect shipping, doubles the aggregate welfare effect to 0.03% and increases worldwide container volumes 0.4%. Finally, allowing for scale economies increases welfare six-fold and trade volumes five-fold, with a 0.09% welfare gain and a full 1% increase in global traffic.

However these relatively small global effects mask significant heterogeneity across countries. Figure A.8 show changes in the relative wage-adjusted price index (interpreted as national welfare, if we omit the costs of climate change) across our three scenarios. (Appendix Figure A.8 shows related changes in country-by-country containerized exports.) In the baseline scenario in Panel (A), we see increases from trade between countries on the Northeast passage, but very little spillover effects - only those reflecting the classic multilateral resistance term in trade models and cascading effects from value chains. These direct changes in trade costs due to the Northeast passage only have moderate affects on the relative price index, which the biggest effects seen in trade between East Asian and North European countries. Panel (B) allows for indirect trade, allowing the benefits of trade to pass on to nearby countries, including those without direct transcontinental trade routes. Panel (C) allows for scale economies to amplify effects, with the gains from trade being particularly pronounced in East Asia, who are able to cheaply ship a large amount of products for consumption in Europe. While we find welfare improvements in general, countries near the Panama and Suez Canals see smaller increases in welfare, highlighting the effects from route diversion towards the Arctic Passage.

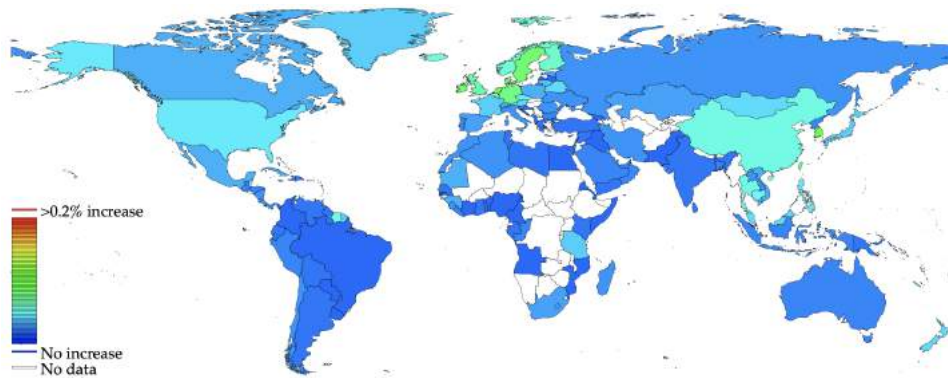
Figure 11 zooms in on the welfare changes of Singapore, Hong Kong, and Taiwan as well as their surrounding countries as a result of the opening of the Arctic Passage. In the baseline scenario in Panel (A), we see that these entrepôts have a direct welfare increase from the passage

Figure 10: Welfare Changes - Arctic Passage

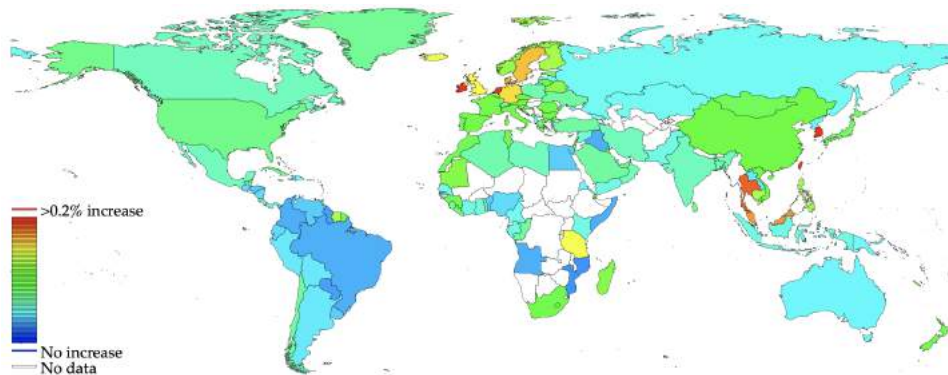
(A) Only Directly Affected Routes



(B) Full Trade Network Effects



(C) Full Trade Network Effects and Scale Economies

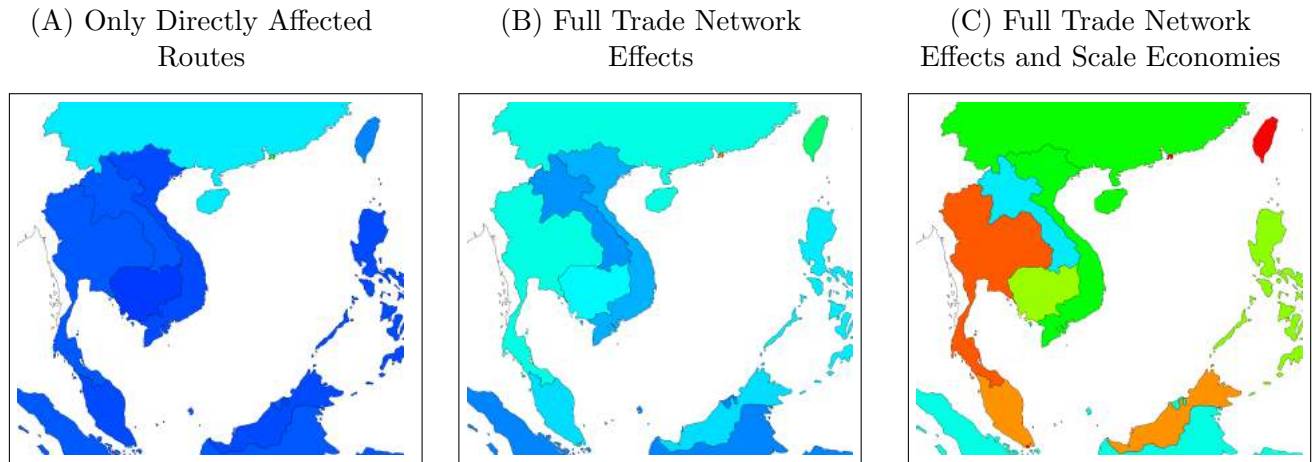


Notes: These three plots show the percent change in welfare (the relative price index) for all countries in our dataset. Darker reds reflect a greater increase and blue represents no change. White represents omitted countries. Panel (A) reflects changes if we only allow trade costs to decrease on routes whose distance is directly reduced to the Arctic Passage. Panel (B) reflects changes if we allow all countries to indirectly access the Arctic Passage through the trade network. Panel (C) allows for scale economies and allows for a feedback loop for all countries. ■

opening since they have direct routes to Northern European countries and North America. When allowing for indirect trade in Panel (B), the neighboring countries of these entrepôts see an increase in welfare because they are now able to benefit from using these entrepôts to trade

with the Northern European countries and North America. As a result of this indirect trade, entrepôts are going to benefit further. When allowing for scale economies to amplify effects in Panel (C), the entrepôts and their neighboring countries are going to benefit even further. The concentration of welfare gains in entrepôts from this counterfactual highlights a novel source of agglomeration—scale economies in transportation and transport networks can help contribute to and shape entrepôts. This is in contrast to the general literature in economic geography where transport costs act as a dispersive force on agglomeration.

Figure 11: Welfare Changes on Asian Entrepôts - Arctic Passage



Notes: These three plots are a magnified part of figure 10 to show the percent change in welfare (the relative price index) for a subset of Asian Entrepôts in our dataset. Darker reds reflects a greater increase and blue represents no change. White represents omitted countries. Panel (A) reflects changes if we only allow trade costs to decrease on routes whose distance is directly reduced to the Arctic Passage. Panel (B) reflects changes if we allow all countries to indirectly access the Arctic Passage through the trade network. Panel (C) allows for scale economies and allows for a feedback loop for all countries. ■

7.4 Hard Brexit

We model one potential issue of a “Hard” Brexit, increases in the costs for only goods that originate or are destined for British use.⁴⁹ We assume that these tariffs and related costs will not spill over to goods that are only transhipped or temporarily stop at British ports on their way to far flung destinations. In a traditional setup, the affects of Brexit will only be felt through changes in costs through either direct trade relationship or through the traditional multilateral resistance term (including the direct effect on global value changes). However, with scale economies, a decrease in the viability of British trade will have a spillover effect. Potential tariffs will decrease trade volumes, increasing trade costs, not only affecting Britain,

⁴⁹Alternatively, we can model the effect of port infrastructure improvements either between bilateral pairs or for any single nation.

but countries that currently find the United Kingdom as a preferred entrepot.

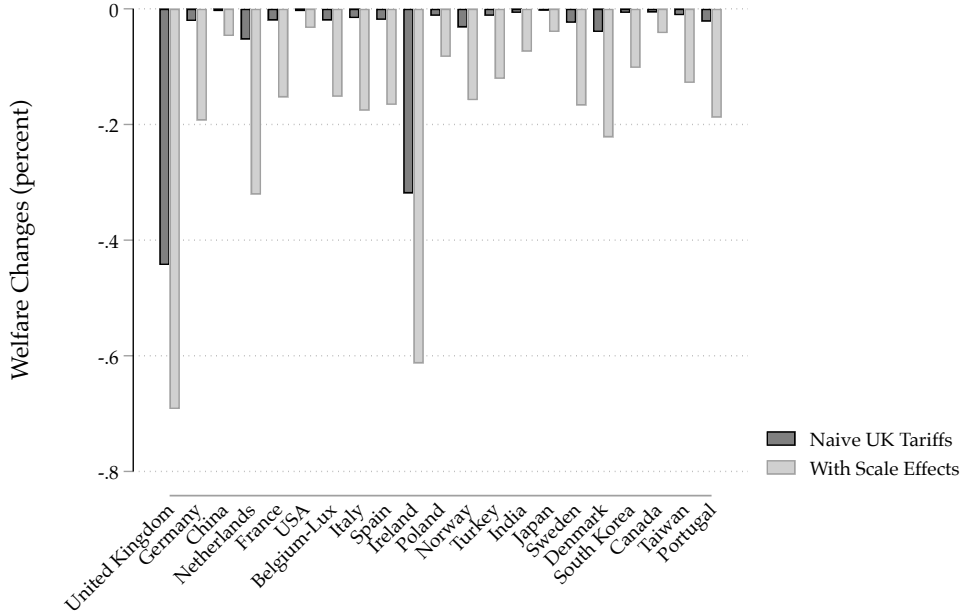
We model this scenario in two steps. First, we replicate a naive Brexit counterfactual, where the cost of shipping doesn't increase, but the cost of entering/exiting the British market increases by 5%. Irish exports destined for Britain will face an increase tariff cost, while Irish exports destined for the United States will not be directly affected - even if the good stops in the British port of Felixstowe first. Second, we additionally allow for scale economies. Now, as British trade volumes fall, trade costs increase. Irish exports to the United States will be more costly, as they will have to pay either the increased costs of travelling through Britian, find an alternative entrepot (perhaps Le Havre, France), or take a low-volume and costly direct trip.

Panel B in Table 4 highlights our results. The direct effect decreases global welfare by 0.04% and scale economies decrease global welfare by 0.16%. Trade volumes follow a similar pattern. Figure 13 highlights the distributional effects in terms of welfare (see Appendix Table A.9 for trade volumes). Direct effects are only significantly felt in the United Kingdom and Ireland, but the structure of global trade and intermediate good use spreads out welfare effects to much of the European Union. Finally scale economies amplify effects, significantly impacting the rest of Europe. Significant effects are also seen in Iceland and in other Nordic countries. Many of these small countries rely on local United Kingdom feeder routes to get goods to large vessels that ply transoceanic trade with Shanghai and New York. Some countries see small effects, such as Norway and Sweden, as they can substitute through Rotterdam, Netherlands and Bremerhaven, Germany. However, Irish and Icelandic exporters and importers suffer, as their trade is not as easily routed through alternatives.

8 Conclusion

World trade does not all get shipped directly from an origin to a destination. It often takes meandering routes, aided by scale economies that consolidates trade in hubs, or entrepôts. We characterize this global container shipping network and its implications for international trade. Guided by a series of novel and salient facts, we model world trade with endogenous trade costs, estimating both the underlying trade costs on all containership routes, as well as scale economies. We develop a novel geography-based IV in order to causally identify the scale impact of traffic on trade costs. Our results imply that a 1% increase in traffic flows on a leg would reduce its corresponding trade costs by 0.05%.

Figure 12: Welfare Changes - Brexit - Largest Trading Partners



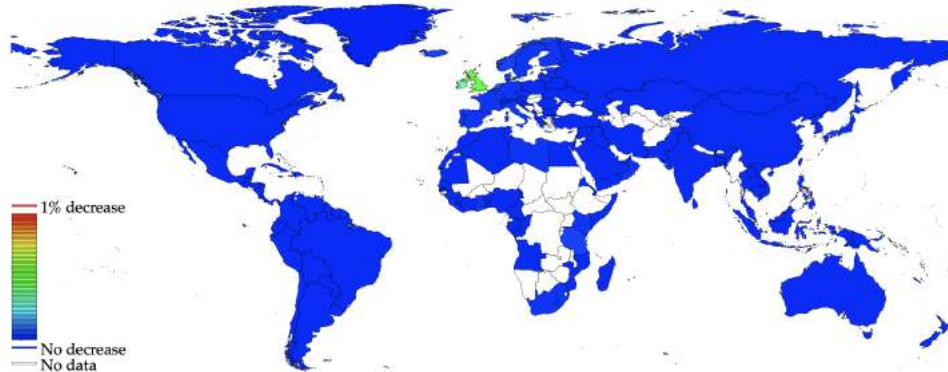
Notes: This show the percent change in welfare (the relative price index) of a simulated 5% increase in trading costs with the United Kingdom the largest 15 trading partners. The first bar reflects changes if shipping costs remain constant, reflecting only welfare changes due to changes in prices. The second bar allows for scale economies and allows for a feedback loop for all countries. ■

Armed with estimates, we show that accounting for both the network effects of the full trade network and their associated scale economies have quantitatively significant welfare and trade volume effects. Combined, they globally increase the effect of trade shocks more than five-fold. The concentration of welfare gains in entrepôts from our Arctic Passage counterfactual highlights a novel source of agglomeration—scale economies in transportation and transport networks can help contribute to and shape entrepôts. This is in contrast to the general literature in economic geography where trade cost typically act as a dispersive force on agglomeration.

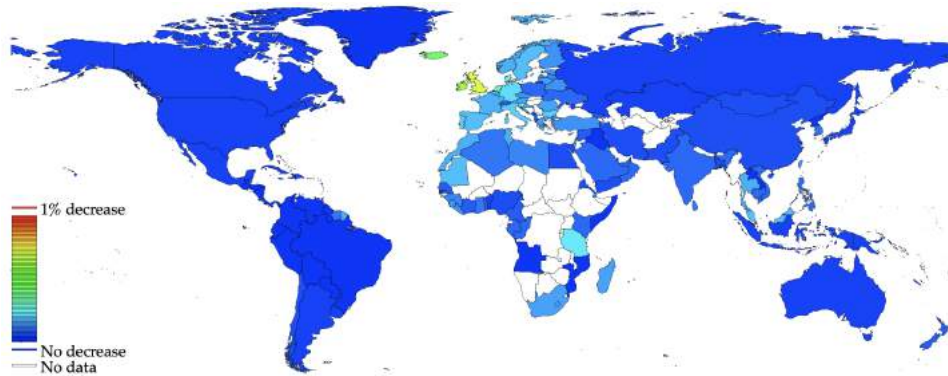
While such hub and spoke networks have been studied in contexts such as airline travel, we take our trade costs and embed them into a tractable general equilibrium framework to be able to quantify welfare effects. We are singularly focused on containerized shipping in our setting as containerized trade accounts for the majority of global seaborne trade. It is worth emphasizing, however, that the hub and spoke network and its implications for trade is not specific to just containerized trade. Such networks are also prevalent in freight services like UPS or DHL in addition to air transport (Rodrigue, Comtois and Slack, 2013). Our treatment of scale economies in this paper is intentionally agnostic to the multitude of potential underlying mechanisms that are likely at work. Future work should especially be done to

Figure 13: Welfare Changes - Brexit

(A) Tariff Change, No Network Scale Effects



(B) Full Trade Network Effects and Scale Economies



Notes: These two plots show the percent change in welfare (the relative price index) of a simulated 5% increase in trading costs with the United Kingdom for all countries in our dataset. Darker reds reflect a greater increase and blue represents no change. White represents omitted countries. Panel (A) reflects changes if shipping costs remain constant, reflecting only welfare changes due to changes in prices. Panel (B) allows for scale economies and allows for a feedback loop for all countries. ■

consider mechanisms the roles of fixed costs in enabling the scale economies in containerized shipping, such as the costs incurred by potential oligopolies in setting shipping networks.

References

- Alder, Simon.** 2015. “Chinese roads in India: The effect of transport infrastructure on economic development.” Working Paper.
- Allen, Treb, and Costas Arkolakis.** 2014. “Trade and the Topography of the Spatial Economy.” *The Quarterly Journal of Economics*, 129(3): 1085–1140.
- Allen, Treb, and Costas Arkolakis.** 2019. “The welfare effects of transportation infrastructure improvements.” National Bureau of Economic Research.
- Allen, Treb, and Dave Donaldson.** 2018. “The geography of path dependence.” *Working Paper*.
- Anderson, James E, Mykyta Vesselovsky, and Yoto V Yotov.** 2016. “Gravity with scale effects.” *Journal of International Economics*, 100: 174–193.

- Asturias, Jose.** Forthcoming. “Endogenous Transportation Costs.” *European Economic Review*.
- Barjamovic, Gojko, Thomas Chaney, Kerem Coşar, and Ali Hortaçsu.** 2019. “Trade, merchants, and the lost cities of the bronze age.” *The Quarterly Journal of Economics*, 134(3): 1455–1503.
- Bernhofen, Daniel M, Zouheir El-Sahli, and Richard Kneller.** 2016. “Estimating the effects of the container revolution on world trade.” *Journal of International Economics*, 98: 36–50.
- Bertoli, Simone, Michaël Goujon, and Olivier Santoni.** 2016. “The CERDI-seadistance database.” *FERDI*.
- Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou.** 2017. “Geography, Search Frictions and Endogenous Trade Costs.” National Bureau of Economic Research.
- Caliendo, Lorenzo, and Fernando Parro.** 2015. “Estimates of the Trade and Welfare Effects of NAFTA.” *The Review of Economic Studies*, 82(1): 1–44.
- Coşar, A Kerem, and Banu Demir.** 2018. “Shipping inside the box: Containerization and trade.” *Journal of International Economics*, 114: 331–345.
- Cullinane, Kevin, and Mahim Khanna.** 2000. “Economies of scale in large containerships: optimal size and geographical implications.” *Journal of transport geography*, 8(3): 181–195.
- Ducruet, César, Réka Juhász, Dávid Krisztián Nagy, and Claudia Steinwender.** 2019. “All aboard: The aggregate effects of port development.” Working paper.
- Eaton, Jonathan, and Samuel Kortum.** 2002. “Technology, geography, and trade.” *Econometrica*, 70(5): 1741–1779.
- Fajgelbaum, Pablo D, and Edouard Schaal.** 2017. “Optimal transport networks in spatial equilibrium.” National Bureau of Economic Research.
- Holmes, Thomas J, and Ethan Singer.** 2018. “Indivisibilities in Distribution.” National Bureau of Economic Research.
- Hummels, David.** 2007. “Transportation costs and international trade in the second era of globalization.” *The Journal of Economic Perspectives*, 21(3): 131–154.
- Hummels, David, Volodymyr Lugovskyy, and Alexandre Skiba.** 2009. “The trade reducing effects of market power in international shipping.” *Journal of Development Economics*, 89(1): 84–97.
- International Maritime Authority.** 2003. “AIS Guidelines.”
- Kojaku, Sadamori, Mengqiao Xu, Haoxiang Xia, and Naoki Masuda.** 2019. “Multi-scale core-periphery structure in a global liner shipping network.” *Scientific reports*, 9(1): 404.
- Kucheryavy, Konstantin, Gary Lyn, and Andrés Rodríguez-Clare.** 2019. “Grounded by gravity: A well-behaved trade model with external economies.” Working Paper.
- Lazarou, Nicholas-Joseph.** 2016. “Endogenous Hub Formations in International Trade.” Forum for Research in Empirical International Trade Working Paper 1144.
- Lenzen, Manfred, Keiichiro Kanemoto, Daniel Moran, and Arne Geschke.** 2012. “Mapping the structure of the world economy.” *Environmental science & technology*, 46(15): 8374–8381.
- Limao, Nuno, and Anthony J Venables.** 2001. “Infrastructure, geographical disadvantage, transport costs, and trade.” *The World Bank Economic Review*, 15(3): 451–479.
- Moretti, Enrico.** 2019. “The effect of high-tech clusters on the productivity of top inventors.” National Bureau of Economic Research.

- Redding, Stephen J, and Matthew A Turner.** 2015. "Transportation costs and the spatial organization of economic activity." *Handbook of regional and urban economics*, 5: 1339–1398.
- Rodrigue, Jean-Paul, Claude Comtois, and Brian Slack.** 2013. *The geography of transport systems*. Routledge.
- Schott, Peter K.** 2008. "The relative sophistication of Chinese exports." *Economic policy*, 23(53): 6–49.
- Simonovska, Ina, and Michael E Waugh.** 2014. "The elasticity of trade: Estimates and evidence." *Journal of international Economics*, 92(1): 34–50.
- Skiba, A.** Forthcoming. "Regional economies of scale and regional welfare." *Review of International Economics*.
- Sutton, John.** 1991. *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration*. MIT press.
- Wang, Chengjin, and Jiaoe Wang.** 2011. "Spatial pattern of the global shipping network and its hub-and-spoke system." *Research in Transportation Economics*, 32(1): 54–63.
- Wong, Woan Foong.** 2019. "The Round Trip Effect: Endogenous Transport Costs and International Trade." Working Paper.

Appendix (For Online Publication Only)

Appendix A Data Construction

A.1 Shipment Microdata

We compile and combine two proprietary micro-data sets in this project: global ports of call data for all containerships, which allows us to reconstruct the routes taken by specific ships, and United States bill of lading data for containerized imports, which gives us shipment-level data on imports into the United States. Independently, each of these datasets allow us to partially describe the global shipping network. By merging them, we are able to reconstruct nearly the entire journey most shipments entering the United States take, from their initial origin point or place of receipt, to the port of entry into the United States. To our knowledge, we provide the most comprehensive reconstruction of the global shipping network and routes undertaken by individual shipments into the United States.

Port of call data We partner with Astra Paging, which provides us with port of call data for containerships. Astra Paging’s data captures vessel movements using the transponders on these ships (known as the automatic identification system, AIS). A network of receivers at ports collects and shares AIS transponder information (including ship name, speed, height in water, latitude, and longitude). Using the geographic variables in the AIS data, Astra Paging marks entry and exit into a number of ports all over the world and provides us with a dataset of ships’ entry and exit from ports of call, timestamps, and ships’ height in the water, or draft. Using these data elements, we are able to calculate an estimated shipment volume between each port pair by taking the observed draft relative to maximum observed draft and multiplying by total ship capacity.

Our sample covers the a six months period, from April to October 2014. Over this period, we have information on 4,986 unique container ships with a combined capacity of 18.13 million TEU. This represents over 90% of the global container shipping fleet. These ships make 429,868 calls at 1,203 ports. Ports with no AIS receivers or where information is not shared do not show up in our data. In addition, if transponders are turned off or transmissions are not recorded, ports of call can be missed. However transponders are required to be operational by the International Maritime Organization on ships engaging in international voyages 300 gross

tons, applying to all containerships in our sample International Maritime Authority (2003).

Bill of lading data We partner with Panjiva Inc. (Now a division of Standard and Poor’s) to acquire bill of lading information for all seaborne US imports from April to October 2014. Panjiva cleans this data to standardize the names of the ports, ships, companies, and container volumes. We subset this data to only consider goods that arrive on seaborne container ships.

We put together proprietary bills of lading data, which captures shipment-level information for all containerized imports into the United States. International shipping relies on an industry-standardized system of bills of lading, which act as receipts of shipment, recording all information on the shipment, all the parties involved in the shipping process. The US Customs and Border Patrol (CBP) agency collects these bills in addition to customs information at all ports of entry into the US and this data is obtained from the agency by Panjiva.¹

Our data captures the foreign location where the shipment originated from, the foreign port where it was loaded on the containership which brings it into the US, and the US port where it was unloaded from the containership. In addition, we know the name and International Maritime Organization identification number of the containership (IMOs) which transported the shipment as well as the shipment’s weight, number of containers (TEUs), and product information. For a subset of the shipments, we observe value information.

Over a six months of US imports from April to October 2014, we see a total of 14.8 million TEUs weighting 106 million tons were imported into the US from 227 shipment origin countries, 225 place of receipt countries, and 144 countries with ports of lading. This accounts for about three quarters of the 2014 TEU and tonnage imports, 77 percent and 74 percent respectively (Maritime Administration, US Department of Transportation).² Non-containerized goods, including goods on roll-ons (vehicle carriers), bulk cargo liners (for commodities), and non-containerized cargo ships are not observed in our data.

Specifically, for the purposes of this study, our data captures the following location information for each shipment into the US: the foreign location where the shipment originated from (*shipment origin*), the foreign port where it was loaded on the containership which brings it into the US (*port of lading*), and the US port where it was unloaded from the containership (*port of*

¹US Bill of Lading data is immediately available for direct purchase from the Department of Homeland Security or through a lag using a Freedom of Information Act. However, this raw data requires substantial computing resources for processing and needs to be standardized over time.

²In particular, we miss containers that arrive on trucks and trains from either Mexico or Canada. Our estimation strategy explicitly accounts for this unobserved data.

unlading). In addition, we know the name and identification number of the containership which transported the shipment as well as the shipment's weight, number of containers (TEUs), and product information. For a smaller subset of the shipments, we observe value information.

This data set allows us to start tracing the journey of a shipment from its origin to its destination US port, in particular we can determine whether this shipment was loaded at its origin location onto the vessel that brings it directly to its final US destination, or if it went through at least one other location during its journey. When matched with the port of call data, we can reconstruct most of the remaining journey after its port of lading.

Reconstructing shipment routes Using the containership information, port of arrival information, timing of unloading and ports of call at US ports, and port of lading information, we are able to match the bills of lading to the journeys of specific containerships, then use the ports of call between lading and unloading to reconstruct each shipment's path from its foreign origin to US destination.

We use Vessel IMOs which are identifiers that are unique to containership vessels and stay with vessel hulls for the lifetime of their operation. Only about 4000 ships are identified in Bills of lading by IMO. An additional (roughly) 2,000 ships are matched to IMOs using a fuzzy string match, after which matches are made with the help of excellent undergraduate research assistants.

Ports of arrival are recording using UNLOCODEs in the AIS data and US Census Schedule D codes in the Bill of Lading data. We construct a crosswalk with the excellent help of undergraduate research assistants.

Port of lading are recorded using CBP's Schedule K Foreign ports on Bill of Lading and UNLOCODES in the port of call data. We construct a crosswalk between these with the help of our excellent undergraduate research assistants.

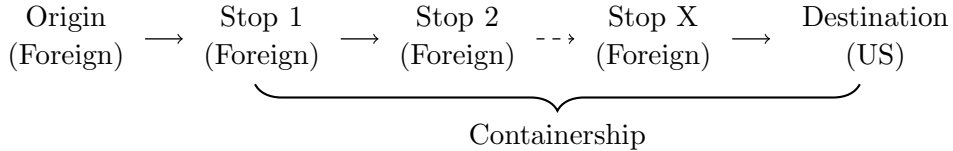
What remains unobserved is the shipment's journey between its Origin and the first stop (port of lading location) we observe in our data. In particular, this initial portion of the shipment's journey could take place overland (by trucks or rail) or by sea on another ship. This information is not recorded by both our datasets and therefore is impossible for us to observe. This will under-count overall indirectness overall, but will not affect our model estimation.

For each bill of lading, we match ship, date of unloading and port of unloading to the AIS data on ships' port of call. Once we match shipments to ships, we record each port of call in

the AIS data before the port of unloading as a stop the shipment makes, then remove all stops observed before the ship stopped at the port of lading. If the port of lading is not observed, the route is discarded and the shipment remains unmatched. Furthermore, any routes that include the port of unloading before the date of unloading are discarded, as they represent loops where the port of call for the port of lading is missing.

Over 90% of containerized TEU entering the US are on Bills of Lading can be matched to routes using this method. Appendix Figure A.1 visualizes this merge. Unmatched shipments may have missing and unrecoverable ship information, or ports of call that do not match lading and unloading records on bills of lading. In addition, a small number of reconstructed routes have implicit voyage speeds above 50KPH, and are discarded.

Figure A.1: Combined Dataset: Routes Undertaken by Shipments into the US



Notes: Origin is the foreign location where shipment originated from, Stop 1 is the location where the shipment was loaded on its US-bound containership (also known as the port/location of lading), Stop 2 to Stop X are the subsequent stops that the US-bound containership made while the shipment remains on the ship, and Destination is the US port where the shipment was unloaded from containership.

Source: Authors' calculations of AIS and Bill of Lading Data. ■

As an example, Figure A.2 plots for all containerized trade from the United Arab Emirates (UAE), the proportion that stops in each country. This illustrates the paths shipments take when being transported from the UAE on to the US. Shipments from the UAE collectively stop in many countries before continuing onto the US. Many of the most popular are regional neighbor hubs, including Egypt, Pakistan, but Spain and China also facilitate UAE-US trade.

A.2 Geographic Distance Data

Geographic distance data is computed using two rasterized (with pixels) world maps. One map consists of the all the navigable oceans and large seas, with a polar ice cap, as well as the Suez and Panama canals. The second map, assumes that the Arctic ice sheet melts away due to anthropogenic climate change. In both maps we compute the sea distance between ports of call, and aggregate to the national level using using port-to-port container flows. We do this computation in R using using Dijkstra's algorithm on a world map with and without Arctic ice

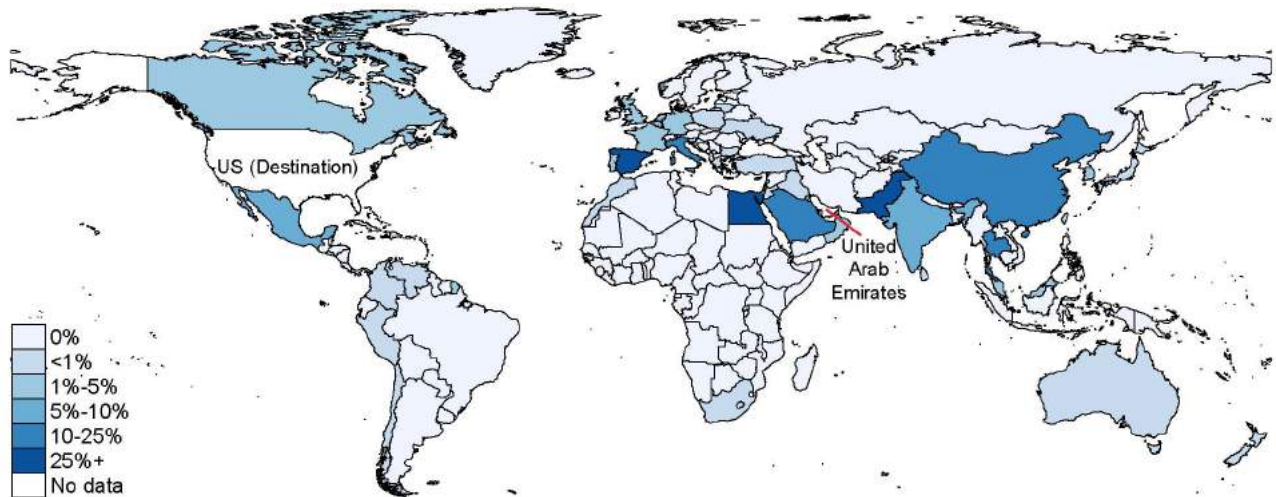


Figure A.2: Percent of UAE-US trade that stops in each country

Notes: Each country’s color represents the share of shipments from the UAE to the United States that stop in that country. Stops computed at the country level and weighted by total container volume (TEU).

Source: Authors’ calculations from AIS data and US Bill of Lading data. ■

caps.³

A.3 Aggregate Economic and Trade Statistics

For our main estimation, we also require data on the value of containerized trade between countries. We use aggregate trade data from Centre d’études Prospectives et d’Informations Internationales (CEPII) and their BACI international database for 2014. This database aggregate data from the UN Comtrade Database, aligning data from origin and destination countries. This provides us data on trade volumes from origin to destinations by industry using Harmonized System (HS) codes.

To aggregate industry trade to industries that use container shipments versus trade that does not, we use aggregate data from 2014 from the United State Customs, as disseminated by Schott (2008).⁴ This data reports the share of shipments by HS Codes that arrive by containerships. We consider 4-digit HS Codes as a consistent level of aggregation. The distribution of containership share by HS code is bi-modal, with one peak around 0% and another around 100%. We use a cutoff of 80%. So HS codes that are shipped by containership to the US over 80% of the time are classified as “containerizable” trade.

For aggregate trade and economic statistics for using in the counterfactual, we use the

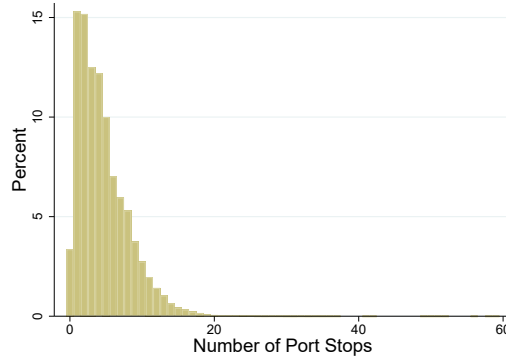
³For more information, see the ‘[gdistance](#)’ package.

⁴This data has been continually updated by the author following the initial publication

Eora global supply chain database with a multi-region input-output table (EORA-MRIO).⁵ We collapse all world trade into three categories; those that are non-tradable, those that typically traded over oceans by containerized vessels, and those that are not typically traded over oceans by containerized vessels.⁶ We again classify industries using the methods of Schott (2008).

Appendix B Additional Figures and Tables

Figure A.3: Distribution of Port Stops per Container (TEU)



Notes: Landlocked countries are excluded. Mean of 4.6 with standard deviation of 3.5.
Source: Authors' calculations of AIS and Bill of Lading Data. ■

Table A.1: Relationship between Indirectness vs Distance and Time

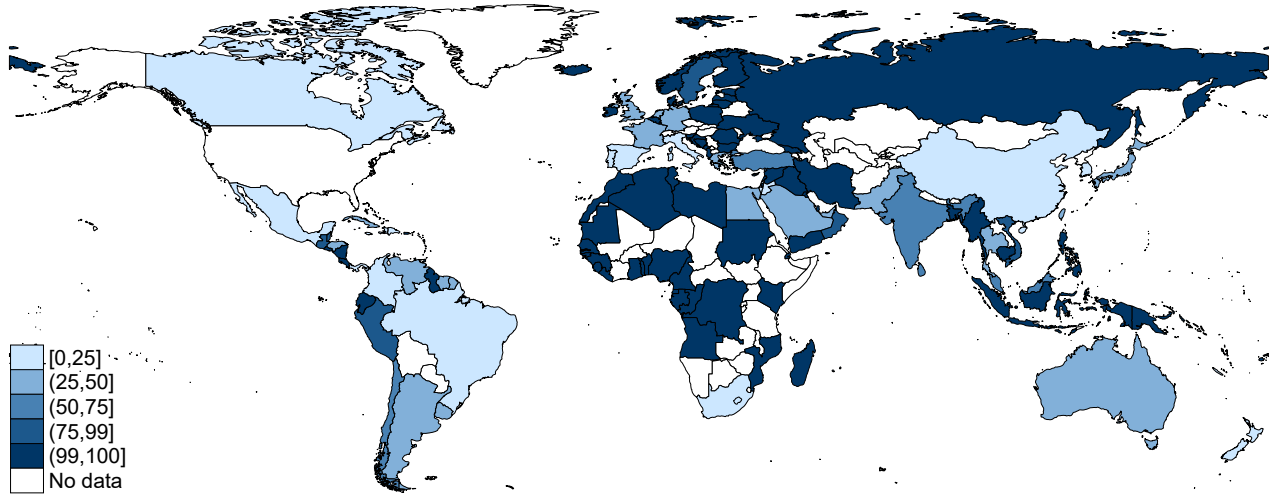
	(1)	(2)	(3)	(4)	(5)
	ln Observed Dist	ln Observed Dist	ln Observed Dist	ln Observed Dist	ln Time Travelled
ln Country Stops	0.112 (0.0223)	0.109 (0.0237)	0.101 (0.0270)	0.104 (0.0300)	0.333 (0.0819)
ln Direct Dist	0.881 (0.0276)	0.918 (0.0347)	0.896 (0.0282)		
Lading Port FE		Y			
Unlading Port FE			Y		
Lading-Unlading Ports FE				Y	Y
Observations	215,655	215,655	215,655	215,656	215,656
R^2	.942	.954	.945	.966	.774
F-stat	1360.62	1818.20	1242.46	12.11	16.49

Notes: Indirectness is measured in number of country stops between foreign origin and US destination. Distance is measured in kilometers while time is measured in hours. Shipment-level observations are weighted by TEU. Shipments originating in landlocked countries are omitted. Standard errors in parentheses are clustered at the port of lading and port of unlading levels.
Source: Authors' calculations of AIS and Bill of Lading Data. ■

⁵Freely available for academic use from <https://worldmrio.com/>.

⁶This includes bulk shipping, roll-on roll-off ships, as well as air freight.

Figure A.4: Transshipped Trade Share between Origin and US Destination



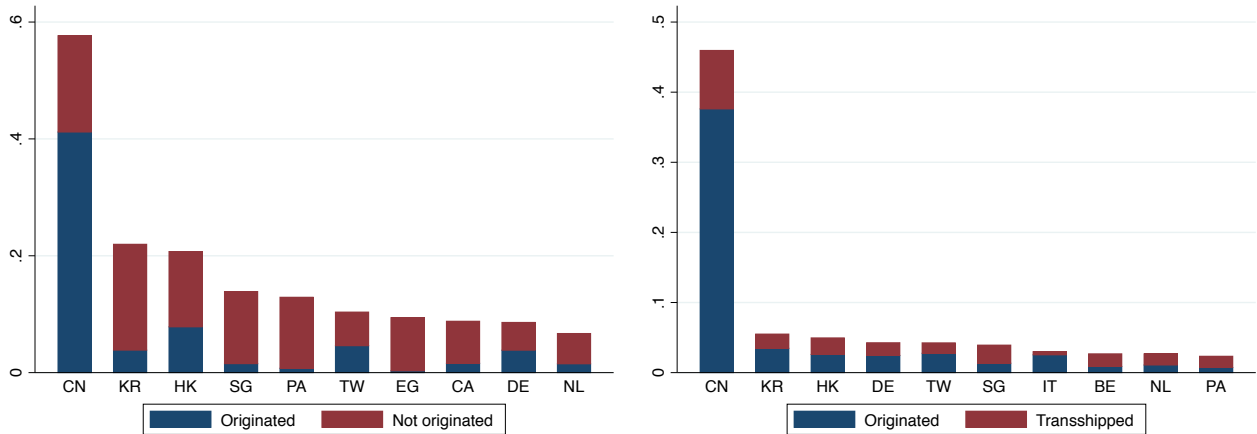
Notes: Lighter colors indicated lower levels of transshipped trade share (ie. more direct trade). The US is not included since it is the destination country. Landlocked countries are also not included, since by definition they would need to stop at a coastal country. 34 of the shipment origin countries are landlocked accounting for 1.6 percent of total TEUs. The missing remaining countries are either due to lack of overall trade with the US (e.g. Somalia) or due to the merge process (e.g. Namibia).

Source: Authors' calculations of AIS and Bill of Lading Data. ■

Figure A.5: Roles of Countries in Bilateral Trade: Origin vs Entrepôts

(A) Share of shipments stopping in country, for top ten countries

(B) Share of shipments loaded in country, for top ten countries



Notes: The blue portion in Panel (A) highlights shipment shares that originate in that country while the red accounts for shipments stopping in that country (not originated). Many entrepôts are listed among the top ten countries, including Korea, Panama, Singapore, and Egypt. Also, more than 50% of the containers entering into the US stop in China. Panel (B) replicates Panel (A) but for country of lading. China again dominates as a source of lading. A few of these top countries, like Germany in (A) and Italy in (B) are majority blue, implying they are important to the US because of their role as an origination country. Other countries, like Singapore, are differentially red, and appear to play a bigger role as an entrepôt rather than as countries of origin.

Source: Authors' calculations of AIS and Bill of Lading Data. ■

Figure A.6: Percent Originated vs Percent Third-Country Volume

(A) US Shipment Imports Data

(B) Global Data

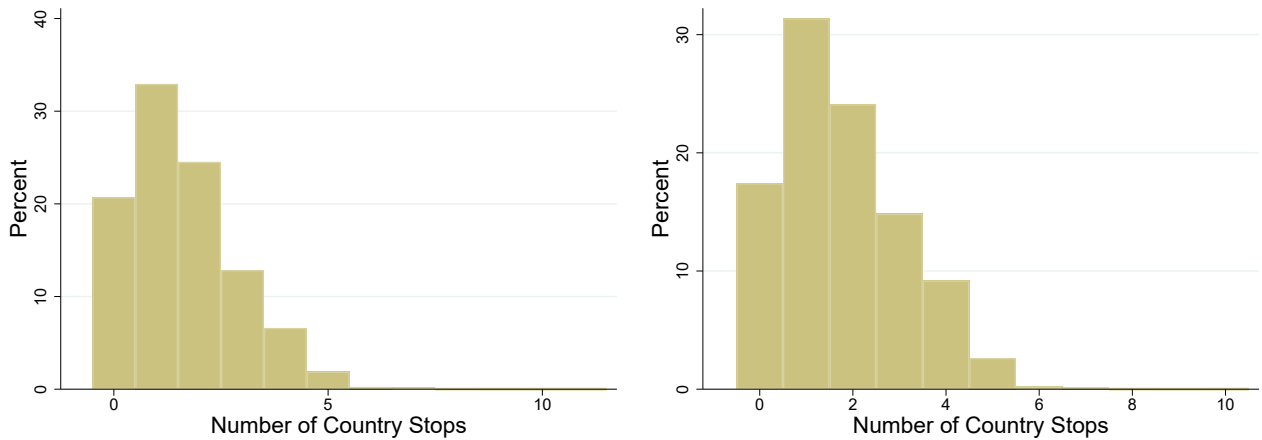


Notes: China is omitted as it is out of the data frame. Panel (A) is reproduced from Panel (A) in Figure 5. Panel (B) repeats this exercise in the previous panel using our global AIS data (no longer merged with US bills of lading), scattering percent of global trade against percent of global AIS traffic. Many of the entrepôts who play a major role in US trade continue to do so in global trade, like Hong Kong, South Korea, and Singapore. Source: CEPII, Authors' calculations of AIS and Bill of Lading Data. ■

Figure A.7: Distribution of Third-Party Countries Involved in Bilateral Trade by Weight and Value

Panel (A) Number of Countries per Ton

Panel (B) Number of Countries per USD Value



Notes: Landlocked countries are excluded. We find that only about 20% by weight (measured in tons) and less than 20% by value are transported between an exporting origin and the US directly with no stops in between. Source: Authors' calculations of AIS and Bill of Lading Data. ■

Table A.2: Concentration Ratios

	Third-Party Stops	Transshipment	Trade
Max/50	426	476	400
99/50	398	476	96
95/50	213	135	27
90/50	112	91	15

Notes: Data represent the ratio between a numerator and denominator, where ports are ranked by percentile. For example, the largest port has 426 times the number of third-party stops at the median (50th-percentile) port.

Source: Authors' calculations of AIS and Bill of Lading Data. ■

Table A.3: Robustness check of scale estimates:
Removal of shortest 10 percentile distances for each origin and destination

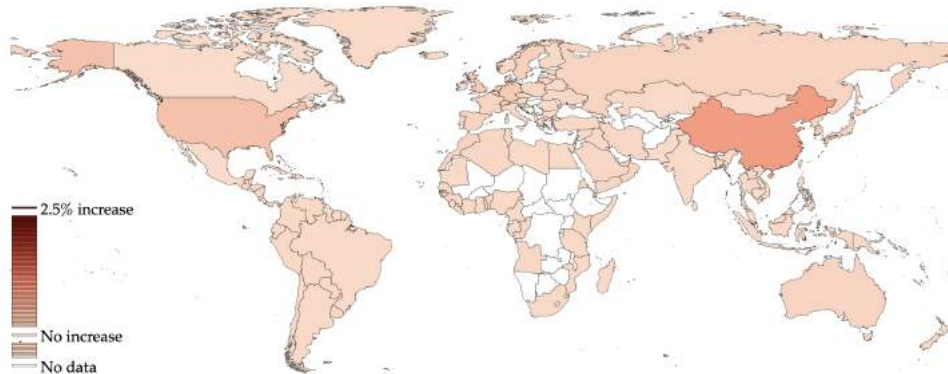
	(1)	(2)	(3)	(4)	(5)
	$\ln(c_{kl})$	$\ln(c_{kl})$	$\ln(\Xi_{kl})$	$\ln(c_{kl})$	$\ln(c_{kl})$
$\ln(\Xi_{kl})$	-0.814 (0.0113)			-0.328 (0.122)	-0.459 (0.0896)
$\ln(z_{kl})$		-0.0522 (0.0293)	0.159 (0.0414)		
$\ln(d_{kl})$	0.495 (0.0329)	0.647 (0.0595)	-0.233 (0.0752)	0.570 (0.0464)	0.629 (0.0710)
k-level FE					Y
Specification	OLS	RF	1st St	IV	IV
Observations	1947	1947	1947	1947	1947
R^2	.89	.14		.62	.77
KP F-stat				14.82	12.05

Notes: Robust standard errors in parentheses clustered two-ways by node k and node l . Instrument z_{kl} in equation (14) is recalculated by omitting the shortest 10 percentile distances for each origin i and destination j respectively.

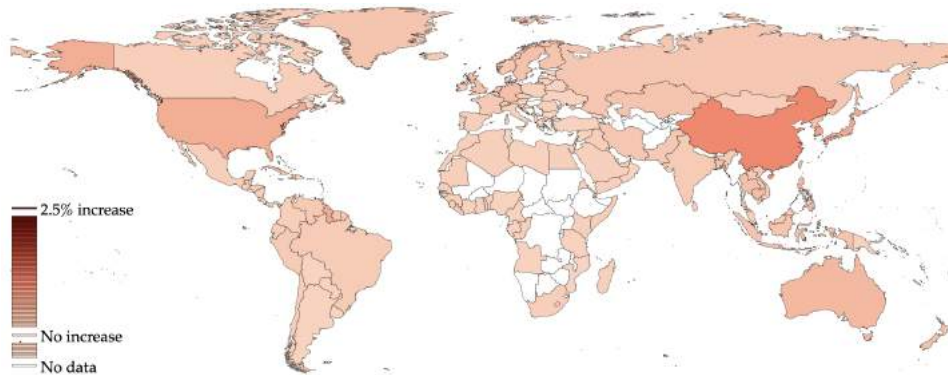
Source: Authors' calculations using AIS and Bill of Lading data. ■

Figure A.8: Export Volume Changes - Arctic Passage

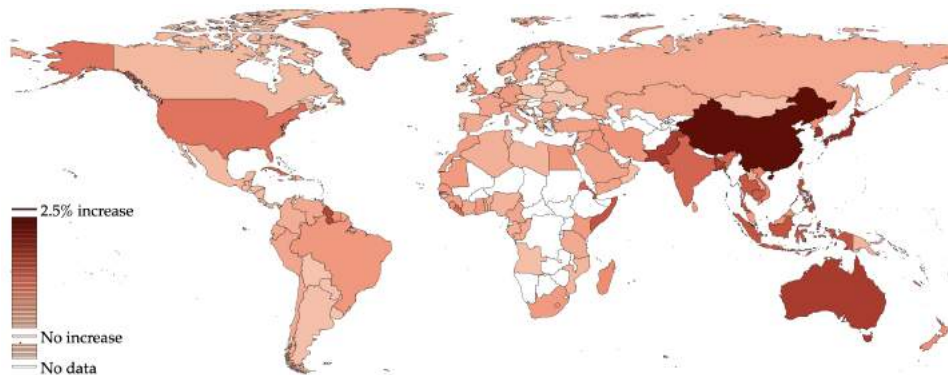
(A) Only Directly Affected Routes



(B) Full Trade Network Effects



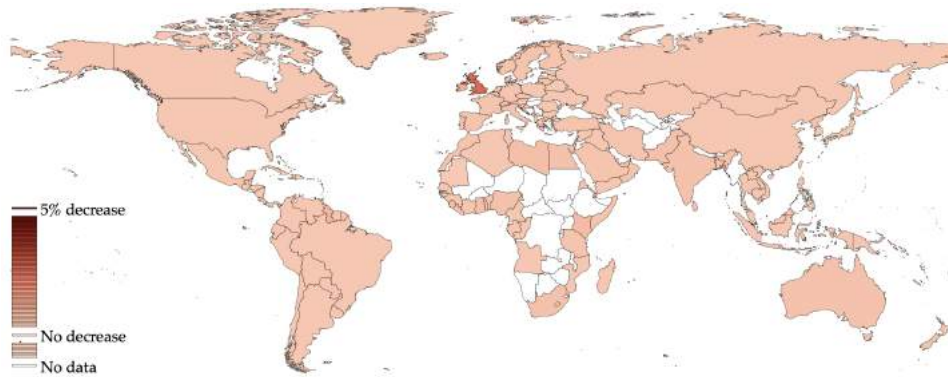
(C) Full Trade Network Effects and Scale Economies



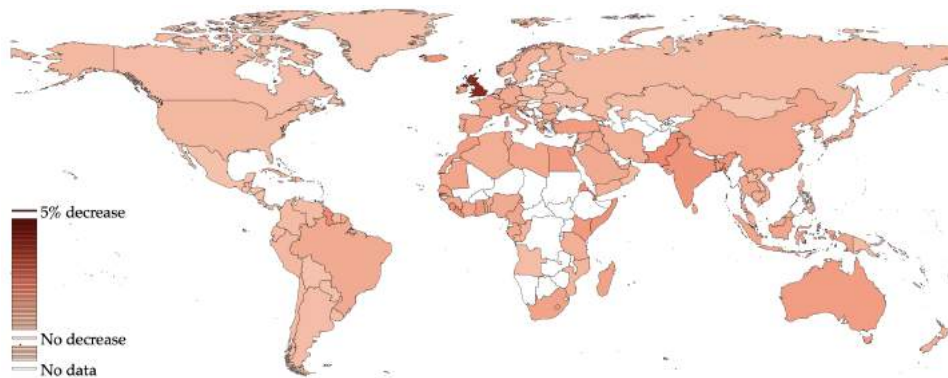
Notes: These three plots show the percent change in exports from all countries in our dataset. Darker reds reflect a greater increase in exports. White represents omitted countries. Panel (A) reflects changes if we only allow trade costs to decrease on routes whose distance is directly reduced to the Arctic Passage. Panel (B) reflects changes if we allow all countries to indirectly access the Arctic Passage through the trade network. Panel (C) allows for scale economies and allows for a feedback loop for all countries. ■

Figure A.9: Export Volume Changes - Brexit

(A) Tariff Change, No Network Scale Effects



(B) Full Trade Network Effects and Scale Economies



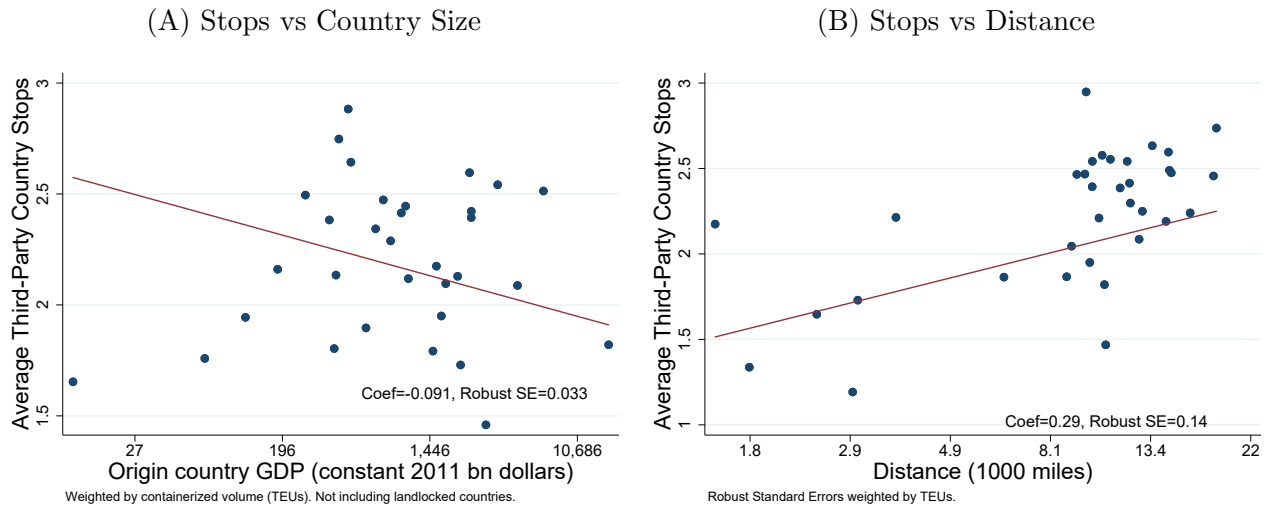
Notes: These two plots show the percent change in exports of a simulated 5% increase in trading costs with the United Kingdom for all countries in our dataset. Darker reds reflect a greater increase. White represents omitted countries. Panel (A) reflects changes if shipping costs remain constant, reflecting only trade changes due to changes in prices. Panel (B) allows for scale economies and allows for a feedback loop for all countries.

■

Appendix C Variation in Connectivity

There is a high degree of variance in indirectness across countries, as shown in Figures 1 and A.4. This variation is reasonable explained by traditional gravity variables. In Panel (A) of Figure A.10, we plot number of stops against country GDP and find that countries with higher GDPs are more likely to have less stops on their journeys to the US. In Panel (B), we plot number of stops against distance instead and find that countries which are closer have more direct trade with the US. These results are robust to using port stops instead of country stops (table A.4) as well as to weighting by containers, tons, and value. One natural interpretation of this would be the endogenous response of shippers to the scale of shipments from these countries. Of course, the availability of direct trade to the US could in principle reverse the causality.

Figure A.10: Larger and closer countries have lower number of average stops



Notes: Binned scatter plot with observation at the origin level. Robust standard errors weighted by total container TEU trade by origin.

Source: World Bank WITS, Seadistance.org, and author's calculations of AIS and Bill of Lading Data. ■

Do shipments from a given origin follow a unique path to the US? Panel (A) in Figure A.11 shows the distribution in the number of unique routes to the US by origin country. With an average of about 397 routes with wide variation (sd 681), observed routes from a single origin are indeed varied. The countries with the highest number of unique routes are big trading partners like China, the United Kingdom, Germany, and well-established entrepôts like Hong Kong. Countries with the lowest unique routes are smaller trading partners like American Samoa, Nauru, Tonga, and Montserrat. The existence of this within-origin route variation will

Table A.4: Relationship between stops and country size as well as distance

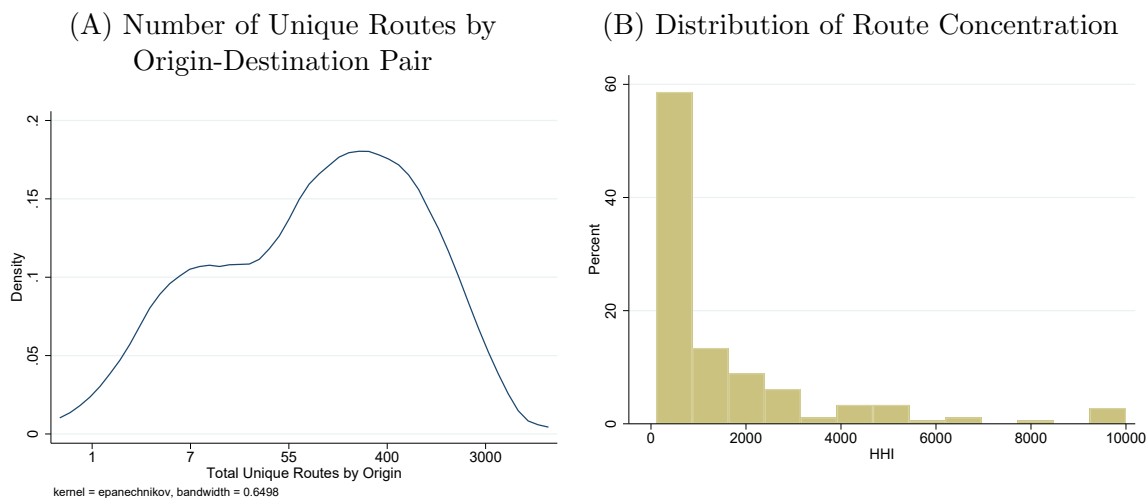
	(1)	(2)	(3)	(4)	(5)	(6)
	ln Ctry Stops	ln Ctry Stops	ln Ctry Stops	ln Port Stops	ln Port Stops	ln Port Stops
ln GDP	-0.0371 (0.0187)		-0.0488 (0.0140)	-0.00226 (0.00966)		-0.00935 (0.00719)
ln Distance		0.166 (0.0851)	0.212 (0.0934)		0.119 (0.0352)	0.128 (0.0384)
Observations	133	133	133	133	133	133
F-stat	3.933	3.795	8.878	0.0546	11.50	5.644
R^2	0.120	0.142	0.339	0.00185	0.305	0.335

Notes: Robust standard errors in parentheses. Weighted by TEUs and excluding landlocked countries.
 Source: Penn World Table, CERDI-SeaDistance (Bertoli, Goujon and Santoni, 2016), Authors' calculations of AIS and Bill of Lading Data. ■

be a particularly important assumption in our model and external validity checks.

We can measure the concentration of these unique routes by constructing a Herfindahl-Hirschman Index (HHI) for each origin country using the container shares of each route. Panel (B) in Figure A.11 shows that almost 70 percent of origin countries have fairly low concentration of routes (HHI less than 1500). The average HHI overall is 1475 (sd 1974). Examples of countries with high levels of concentration are like Vanuatu, Cuba, and Liberia while countries with low levels of concentration are Macau, Hong Kong, and Belgium-Luxembourg.

Figure A.11: Variation in trade indirectness



Notes: Source: Authors' calculations of AIS and Bill of Lading Data. ■

Table A.5: Predictive Trade Cost Estimates

Coefficient	Estimate
β_0 (intercept)	7.95
β_1 (log distance)	0.17
β_2 (log route traffic)	-1.04
β_3 (log outgoing port traffic)	0.28
β_4 (log incoming port traffic)	0.28
β_5 (land borders)	-0.39
β_6 (back-haul)	-0.05

Notes: These results are not causal, and cannot be used for either inference or counterfactuals. They represent the predictive power of various (possibly endogenous) variables in predicting a trade cost matrix that rationalizes leg-level containerized traffic flow. ■

Appendix D Recovery of Predicted Trade Costs

Table A.5 shows the results of our estimation that predicts leg-level trade costs. Positive values for β indicate increases in trade costs and negative values indicate decreases in trade cost. We find that distance increases trade cost and increased shipping traffic decreases trade costs, after fully accounting for the role of total trade values in X .

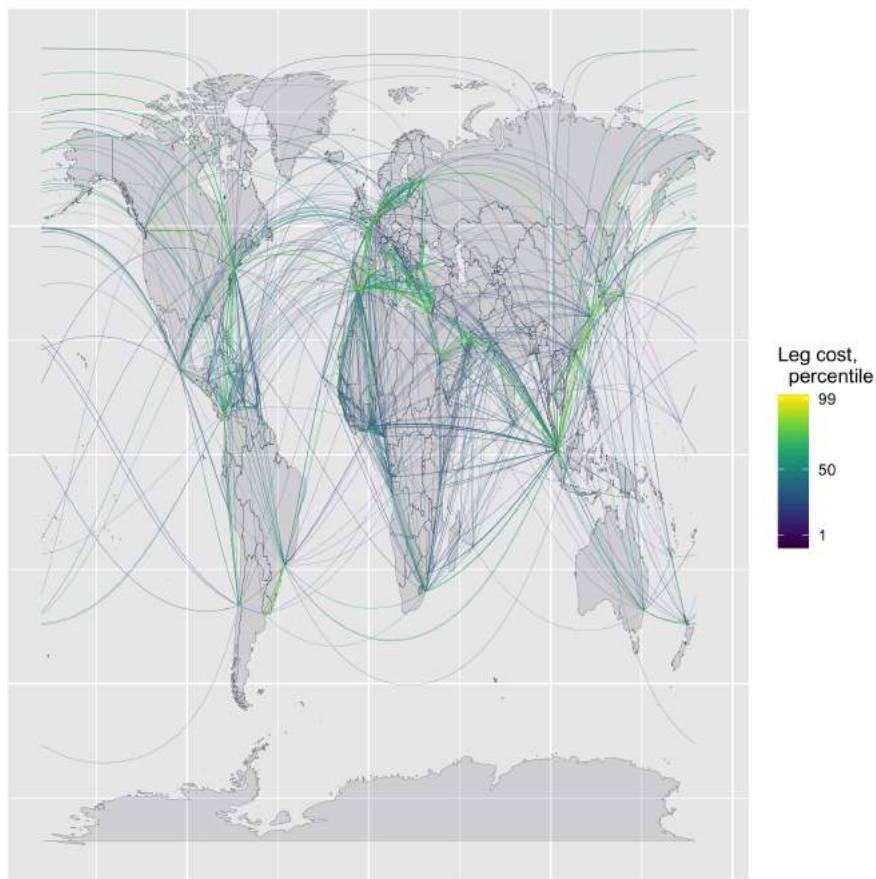
However, these estimates are not casual, and cannot be used for either inference or counterfactuals. They represent the power of various (including highly endogenous) variables in predicting a trade cost matrix that rationalizes leg-level containerized traffic flow. The fit of our predictions is highlighted in Figure 7.⁷

In spirit, such analysis reflects the spirit of pure prediction and cannot satisfy the “Lucas Critique”, as they are purely observational and do not reflect fundamental economic parameters, forces, or relationships. In Section 6 we address endogeneity and causality, using an instrument to find the relationship between route-level volume and trade costs.⁸

⁷If we had more possible useful predictive variables, we would use a machine learning technique to tease out the best basis of variables to predict model-consistent trade costs.

⁸This estimation follows Allen and Arkolakis (2019) and abstracts away from endogeneity and model misspecification concerns.

Figure A.12: Trade Cost Estimates, All Legs



Notes: This map display the recovered trade cost between all origins and destinations for containership legs in the AIS data. Lighter colors indicate lower trade costs.

Source: Authors' calculations of AIS and Bill of Lading Data. ■

Appendix E Additional Estimation Results

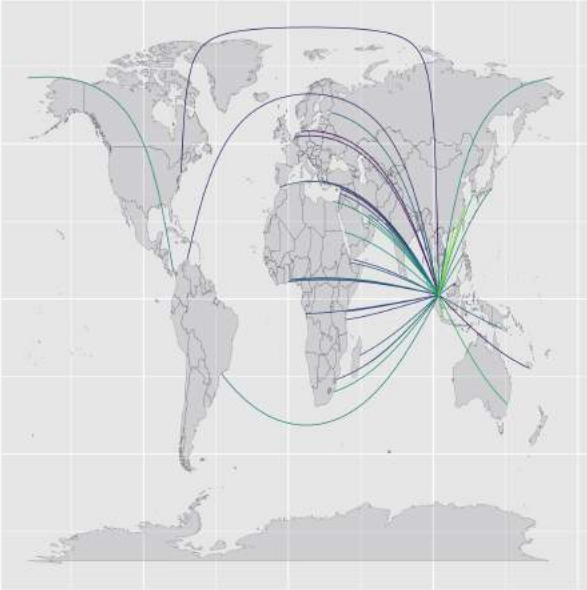
Below, estimated route costs are drawn. Thicker and lighter colors are lower-cost routes. Shorter and more heavily trafficked routes are the cheapest. The effect of scale is observable here: Syria to France is one of the highest cost legs, significantly higher than Singapore to Gibraltar, a much longer distance. Even among the subset of bilateral pairs for which we observe traffic, the triangle inequality is violated 280 times.

Figure A.13 plots bilateral incoming and outgoing trade costs for Singapore and Lebanon separately. Singapore is not only well-connected but both as an origin and destination has some of the cheapest legs. Singapore ships to Lebanon which has both fewer and shorter connections.

Figure A.14 plots country-level averages of the expected trade cost (from the B-matrix). Entrepôts such as Egypt, Panama, and (not visible) Singapore and Gibraltar have generally

Figure A.13: Trade Costs by Country

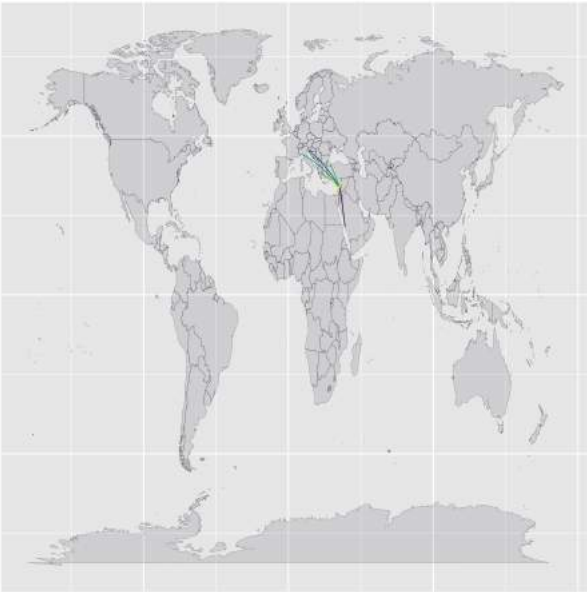
(A): Singapore, Origin



(B): Singapore, Destination



(C): Lebanon, Origin



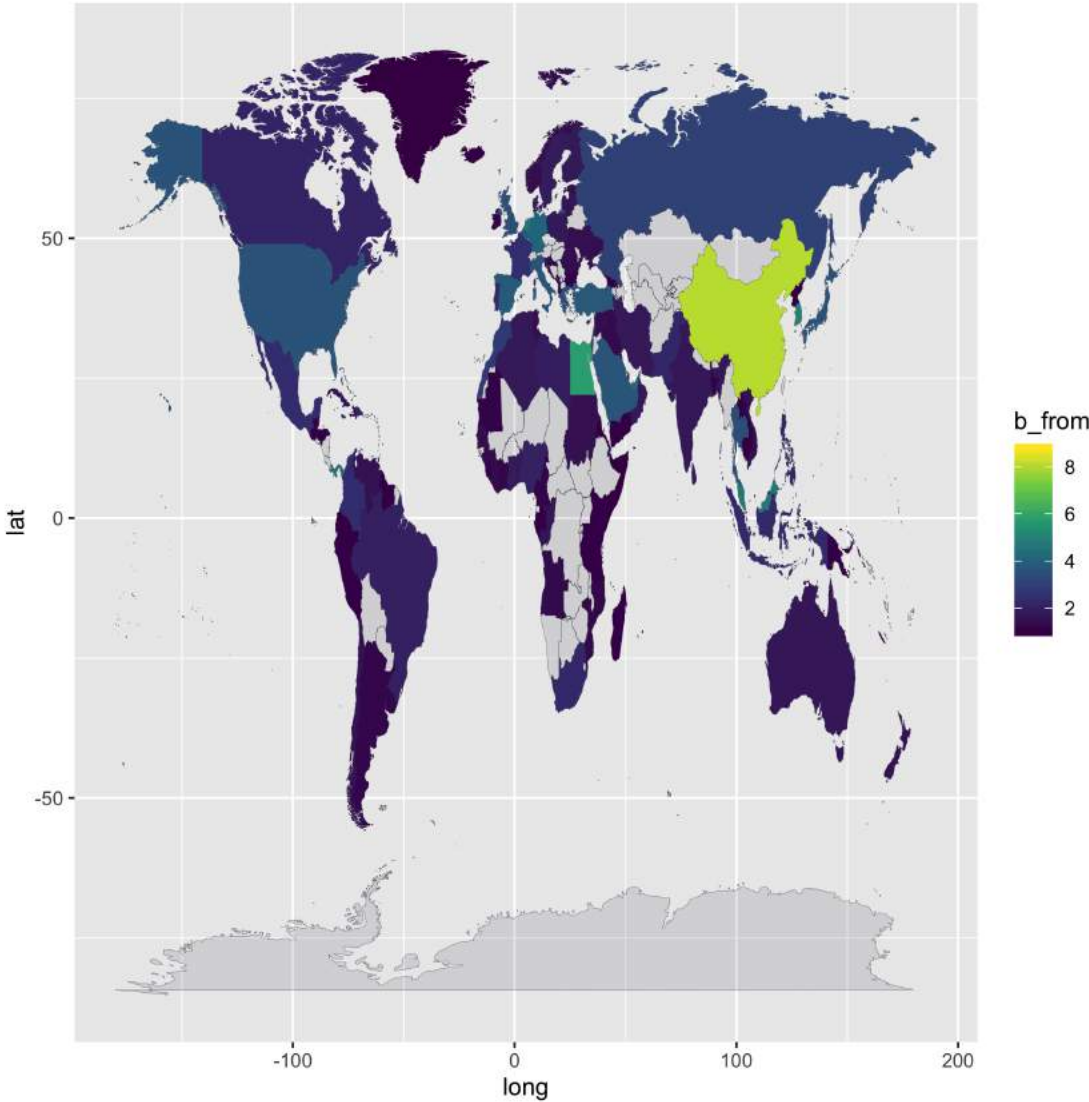
(D): Lebanon, Destination



Notes: Lighter colors indicate lower trade costs.

Source: Authors' calculations of AIS and Bill of Lading Data. ■

Figure A.14: Expected Trade Costs, Country Average



Notes: Lighter colors indicate lower expected trade costs.
Source: Authors' calculations of AIS and Bill of Lading Data.

■

cheaper trade costs, as does China, due to the scale of shipping as well as access to nearby low-cost entrepôt (Korea, Singapore, and Japan).

Table A.6 reflects the log-linear relationship between our estimated trade cost τ , aggregate bilateral trade values, and distance. These results highlight the reduced form relationships between these three variables, as well as the predictive power of our computed trade costs. Without origin or destination fixed effects, our trade costs alone can explain 29% variation of global trade. The logarithm of distance can account for less than 3% . However, we caution against reading too far into this; as we do use trade values on operational sea routes to predict trade costs.

Table A.6: τ vs Distance

	(1)	(2)	(3)	(4)	(5)	(6)
			Log trade volumes			
Log $\tau_{ij}^{-\theta}$	0.462 (0.0296)		0.449 (0.0308)	0.758 (0.0313)		0.524 (0.0287)
Log dist		-0.752 (0.0980)	-0.406 (0.0909)		-1.325 (0.0621)	-0.652 (0.0599)
Constant	12.67 (0.352)	14.76 (0.892)	16.16 (0.746)	15.63 (0.312)	19.87 (0.554)	19.11 (0.435)
Orig, Dest FEs	No	No	No	Yes	Yes	Yes
Observations	22,985	22,985	22,985	22,985	22,985	22,985
R-squared	0.286	0.029	0.294	0.760	0.751	0.770

Robust standard errors in parentheses

Appendix F General Equilibrium Model in Changes

The production costs in country i and industry n respond to a shock to a given t_{kl} according to the equation:

$$\hat{c}_{in} = \hat{w}_i^{\gamma_{in}} \prod_{k=1}^N \hat{P}_{ik}^{\gamma_{ink}}. \quad (20)$$

The change in the price of the composite intermediate good in country i and industry n relative to shock to t_{kl} is:

$$\hat{P}_{in} = \left[\sum_{i=1}^J \pi_{ijn} [\hat{\tau}_{ijn} \hat{c}_{in}]^{-\theta_n} \right]^{-1/\theta_n}. \quad (21)$$

Bilateral trade shares between i and j in industry n will change according to standard changes through production and transport costs:

$$\hat{\pi}_{ijn} = \left[\frac{\hat{c}_{in} \hat{\tau}_{ijn}}{\hat{P}_{in}} \right]^{-\theta_n}. \quad (22)$$

Trade volumes similarly adjust:

$$X'_{in} = \sum_{k=1}^N \gamma_{ink} \sum_{j=1}^I \frac{\pi'_{ijn}}{1 + \kappa_{ijn}} X'_{jk} + \alpha_{in} I'_i. \quad (23)$$

Lastly, trade is balanced to a deficit shifter such that:

$$\sum_{n=1}^N \sum_{i=1}^I \frac{\pi'_{ijn}}{1 + \kappa_{ijn}} X'_{in} - D_i = \sum_{n=1}^N \sum_{i=1}^I \frac{\pi'_{jin}}{1 + \kappa_{jin}} X'_{jn}, \quad (24)$$

where $I'_i = \hat{w}_i w_i L_i + \sum_{n=1}^N \sum_{i=1}^I \tau'_{ijn} \frac{\pi'_{ijn}}{1 + \kappa_{ijn}} X'_{in} + D_i$.