# Entrepôt: Hubs, Scale, and Trade Costs*

Sharat Ganapati
Georgetown University

Woan Foong Wong
University of Oregon

Oren Ziv
Michigan State University

Preliminary

February 2020

Most Recent Version Here

**Abstract**

Entrepôts are trading hubs where goods travel through, from other origins and bound for other destinations. We study the role these hubs and the networks they form play in international trade. Using novel data, we retrace the paths of goods entering the US and show that the majority of trade is indirect, sent through a small number of entrepôts, and that entrepôts justify larger fleets. We build a model of endogenous entrepôt formation which incorporates route choice by exporters within a Ricardian setting and economies of scale in shipping, and use the model to estimate trade costs on each leg. We develop a geography-based IV to estimate a leg-level scale elasticity. We conduct counterfactuals on opening the Arctic Passage and Brexit to quantify the effects of both spillovers and scale economies. Considering changes to underlying transportation costs, we find that spillovers from the trade network doubles baseline welfare gains, with scale economies further tripling them.

*Contact: sganapati@georgetown.edu, wfwong@uoregon.edu, and orenziv@msu.edu. We thank Meredith Crowley, Carsten Eckel, Reka Juhasz, and seminar participants at the Federal Reserve Board, Ludwig-Maximilians-Universität Munich, University of Cambridge, Georgetown University, and University of Mannheim as well as participants at the NBER Conference on Cities, Labor Markets, and the Global Economy, 2019 Mid-Atlantic International Trade Workshop (Duke), and 2019 Society for Advancement of Economic Theory Conference for very helpful comments. Andrew Castro 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 distributing these goods can include multiple agents and additional countries beyond the producers and consumers.

These activities are often concentrated at entrepôts, trading hubs where goods travel through—from other origins, and bound for other destination. 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 continue to invest billions of dollars with the specific aim of becoming or maintaining their role as entrepôts.[1]

This paper studies the role of the transportation network formed by entrepôts in international trade. We seek to answer three questions: (1) How indirect is trade? (2) What drives the formation of trade hubs? and (3) What are the impacts of indirect trade and entrepôt on trade volumes 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; 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 to map individual journeys at the container level in order to document the ubiquity of indirect trade and describe the role of entrepôts in concentrating international trade, 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 the volumes of shipments observed. Next, we estimate the role of scale economies and the concentration of commerce at entrepôts in reducing shipping costs. Lastly, we 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

[1]Saudi Arabia is implementing 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 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 over 90% of in-bound US shipments. This is the first comprehensive look at how shipments into the United States travel through the global shipping network.

Three stylized facts emerge from our data. First, the majority of trade indirect arrives at US ports. This significantly adds to the distance goods travel as they come to the US. Figure 1 plots the average number of stops made by containers by their country of origin. Second, this indirectness is incredibly concentrated, with a large number of shipments channelled through a small number of entrepôts. Third, we show that ship size is correlated with both entrepôt activity and indirect shipping patterns. Increasing ship size is known to reduce costs, and entrepôts can justify large ship fleets. We find that 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. Entrepôts appear to play a central role in reducing shipping costs through economies of scale.

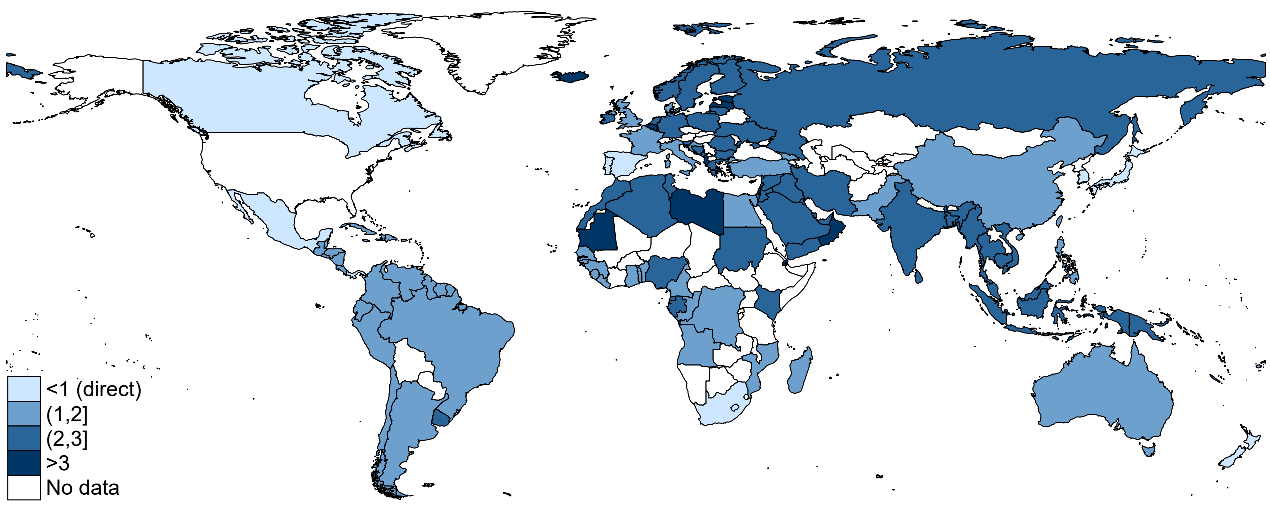



**Figure 1:** Number of Stops between Origin and US Destination

Notes: Stops are by country and weighted by container volume (TEU). 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 from Bill of Lading data. ■

To understand how the indirectness and concentration we observe occurs in general equilibrium and to explore its role in global trade, we build a model of global trade with entrepôts. Individual firms choose shipping routes and compete for consumers in destination countries in a generalized Ricardian setting which is flexible enough to accommodate a broad range of input-output linkages. Lowest-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 data on shipment flows through entrepôts, we use the model to estimate trade costs for each country pair in a subset of our data. 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.

Our model recovers an endogenous price-quantity relationship, but to estimate the causal effect of quantity on price, we need a demand shifter, which we construct using the inherent 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 many large origins and destinations. We leverage this to construct an instrument for demand; for a given leg, we compute the distance to and from the leg relative to the straight-line distance between each origin and destination. Summing across pairs, we recover a weighted average of each leg's proximity to global trade. Using this instrument we estimate, to a given trade elasticity, a strong scale elasticity. For a trade elasticity of 4.5 (Simonovska and Waugh, 2014), our results imply that a 10% 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 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 a series 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 effect of global warming opening up the Arctic Ocean to regular year-round shipping, significantly reduce shipping distances between many Asian, North American and European ports. Our second series of counterfactuals considers the ramifications of a worsening of relations between the United Kingdom and its trading partners. In both cases, the baseline effects of these shocks are first doubles by the network structure of trade and further tripled by the feedback loop imposed by scale economies in container shipping.

This paper makes contributions to three separate literatures: on the presence of networks in trade, on the endogenous trade costs, and on trade and transportation technology. First, this paper characterizes the nature of the global container shipping networks and its implication for international trade. A key contribution towards that goal is the direct observation of indirect trade. To our knowledge, we are the first to do so.[2,3]

More generally, a growing quantitative literature investigates the role of trade networks (Fajgelbaum and Schaal, 2017; Redding and Turner, 2015). Our model is most closely related to Allen and Arkolakis (2019). We extend their Armington framework where route cost shocks are born by consumers to a general Ricardian setting, where link-specific trade volumes reflect not only choice of routes but head-to-head competition and selection on prices at destinations. By mapping our estimation of the model to our micro-data on US shipment routes, our paper acts as check to the validity of the Allen and Arkolakis (2019) approach, showing a tight match between model estimates and the micro-data.

Second, we contribute to a growing literature on endogenous transport costs. In our paper, trade costs and volumes are are endogenously determined in equilibrium with trade flows, similar to Behrens and Picard (2011), Hummels (2007), and Limao and Venables (2001).[4] Our estimates provide a matrix of bilateral trade costs that can be used for further empirical work. Brancaccio, Kalouptsidi and Papageorgiou (2017) estimate a model of endogenous trade costs arising in a very different context, from search frictions between exporters and specialied ships carrying homogeneous commodities. We model transport costs as part of a global network of container shipping routes, which accounts for two-thirds of annual trade moved by sea (World Shipping Council), and where the nature and consequences of the trade network differ dramatically from bulk shipping.[5] Because our model is within a spatial general equilibrium trade

[2]Without information on the loading or unloading of shipments, previous work which has used ship data alone cannot directly observe indirect trade. This work, mostly in the network science and transportation literature, document these networks by solely using ports of call data (Kojaku et al., 2019; Wang and Wang, 2011). By contrast, we reconstruct the journeys that individual shipments take on these ships, providing direct evidence on indirectness and how global shipping network informs international trade flow patterns.

[3]Lazarou (2016) is a complementary study which utilizes network theory to study transshipment hubs within a monopolistic competition framework. As mentioned in his paper, the lack of actual observed indirect route data in his empirical exercise leads to “ambiguity as to the existence, followed by the identification, and therefore selection, of good-specific routes and their respective measurement of distance.” Our dataset is a contribution in that we observe exactly this. We provide a set of tables of model-based results so that future studies using AIS data can estimate indirect traffic.

[4]There are also a number of papers studying endogenous transport costs in the context of market power, including Hummels, Lugovskyy and Skiba (2009), Asturias (Forthcoming), and Francois and Wooton (2001).

[5]The main difference between dry bulk shipping and container shipping is similar to the distinction between taxis and buses. Dry bulk ships likely have to depart from their destination without cargo which gives rise to the search for cargo, whereas this would be less true for containerships by virtue of containers being able to

framework, our counterfactuals are able to contribute to this literature by addressing the overall welfare impacts of the network effects of these endogenous transport costs and their spillover impact of entrepôts on trade and welfare.

One important aspect of endogenous trade costs in our model is the existence of scale economies, which we findare empirically important, in line with findings from the literature on economies of scale in trade (Alder, 2015; Holmes and Singer, 2018; Anderson, Vesselovsky and Yotov, 2016; Skiba, Forthcoming; Asturias, Forthcoming). 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 large literature in economic geography which considers the role of localized scale economies in the emergence of agglomerations, where trade costs acting as dispersive forces (Allen and Arkolakis, 2014; Allen and Donaldson, 2018). Our story inverts this relationship: hubs emerge and are shaped by scale economies in transportation costs, rather than at locations. We are unaware of other work which draws the same connection.

Third, we contribute to the literature on containerization and trade (Coşar and Demir, 2018; Bernhofen, El-Sahli and Kneller, 2016; Rua, 2014; Wong, 2019; Blonigen and Wilson, 2008) by documenting and quantifying the global network effects of the container shipping technology on international trade. In particular, 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. Our contribution here is to examine the effects of the network formed by these nodes on international trade.

The rest of this paper is organized as follows: Section 1 describes both our novel data and how we recover the journeys of individual shipments by combining them. Section 2 uses the data to highlight four stylized facts characterizing the indirectness of containerized trade into the US. Section 3 lays out our general model which shows how the trade network can be analyzed under a wide possible set of functional forms. Section 4 uses the model to estimate trade costs and Section 5 uses the results to estimate scale economies. Section 6 estimates a quantitative general equilibrium model as a special case of the general model and uses it to generate counterfactuals to illustrate the differential effects of changes to the trade network.

accommodate a lot of different products. Containerships are more like buses in the regard that they travel on fixed schedules between a few locations (the "round trip effect", (Wong, 2019)). As a result, dry bulk shipping does not have the same hub-and-spoke network structure which would not give rise to systematic trade spillovers due to network linkages like our container shipping framework would.

Section 7 concludes.

# 1 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, 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 US.

**Figure 2:** Map of Global Port of Call Network

Notes: Each dot represents a port. Each line represents a journey between port pairs undertaken by a containership.
Source: Authors' calculations from AIS data. ■

**Port of call data** Our proprietary ports of call data from Astra Paging captures vessel movements using the transponders on these ships (known as the automatic identification system, AIS). [6] For each vessel, this data captures identifying information, time-stamped ports of call, and the vessel's height in the water before and after stopping in the port, which indicates the

[6]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, our data marks entry and exit into a number of ports all over the world.

vessel's load. Using these data elements, we are able to calculate an estimated shipment volume between each port pair.

Figure 2 shows the coverage of the shipping network in our port of call data.[7]Our sample covers the same six months period, from April to October 2014. Over this period, we have information on 4986 unique container ships with a combined capacity of 18.13 million TEU. This is over 90% of the global container shipping fleet. These ships make 429,868 calls at 1203 ports.[8]

**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.[9] 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 identification number of the containership which transported the shipment as well as the shipment's weight, number of containers (TEUs), and product information.[10] 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 194 countries with ports of lading.[11][12]

[7]Each line represents a journey between port pairs, the dots, undertaken by a containership. The data in the figure is unweighted by volume or number of journeys.

[8]Ports with no AIS receivers or where information is not shared will not show up in our data. In addition, if transponders are turned off or transmissions are not recorded, ports of call can be missed.

[9]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 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 via the Freedom of Information Act from Panjiva.

[10]For a subset of the shipments, we observe value information.

[11]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).

[12]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.

**Reconstructing shipment routes** Using the containership information,[13] port of arrival information,[14] timing of unlading and ports of call at US ports, and port of lading information,[15] we are able to match the bills of lading to the journeys of specific containerships, then use the ports of call between lading and unlading to reconstruct each shipment's path from its foreign origin to US destination. Over 90% of containerized TEU entering the US are on bills of lading can be matched to routes using this method.[16] Appendix Figure A.2 visualizes this merge.

What remains unobserved is the shipment's journey between its Origin and the first stop 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.

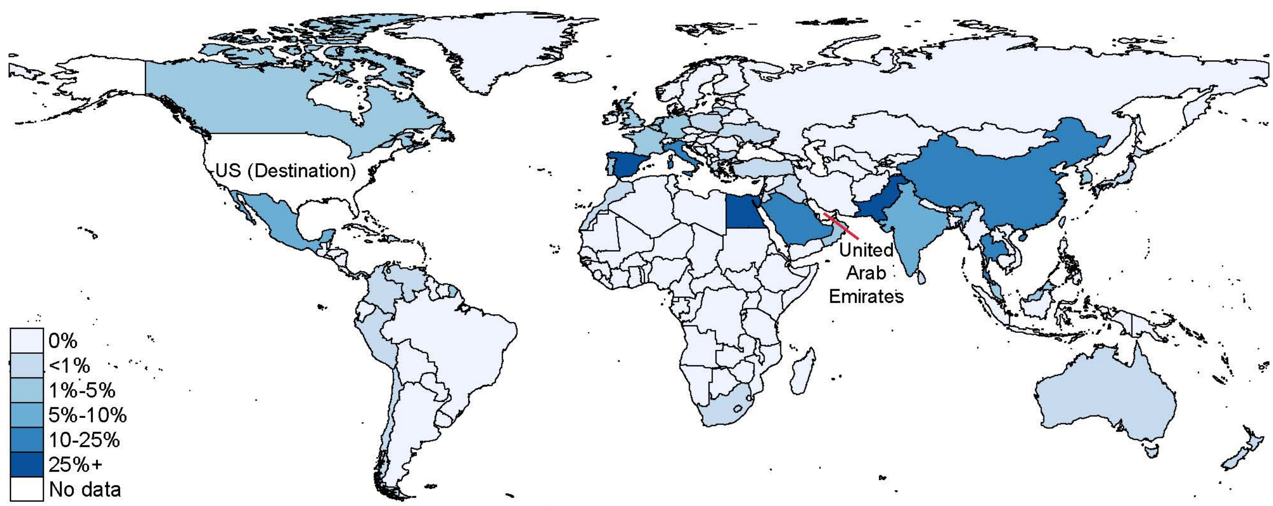



**Figure 3:** Percent of UAE-US trade that stops in each country

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.

[13]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.

[14]Ports of arrival are recording using UNLOCODEs in the port of call 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.

[15]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.

[16]Unmatched shipments may have missing and unrecoverable ship information, or ports of call that do not match lading and unlading records on bills of lading.

As an example, Figure 3 plots for all containerized trade from the United Arab Emirates, 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 U.S. Many of the most popular are regional neighbor hubs, including Egypt, Pakistan, but Spain and China also facilitate UAE-US trade.

# 2 Stylized Facts

In this section, we use the datasets above to explore the nature of the international shipping network and the routes goods entering the US take along that network. Our analysis generates three stylized facts about trade and entrepôts. First, we find indirectness is both ubiquitous and that indirectness results in significantly longer trips. Second, the global trade network is very concentrated, sending a large number of shipments through a small number of entrepôts. Third, shipments flowing through entrepôts are more likely to arrive at US ports on larger, more efficient ships.

The picture that emerges is that 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. This network concentrates shipments through entrepôts, generating large costs for individual shipments, in the form of added distance. Nevertheless, the indirectness is optimal, and the resulting concentration appears to allow for the use of lower-marginal-cost vessels.

## 2.1 The majority of trade is indirect

We begin by asking: how realistic is the bilateral view of trade? Panel (A) in Figure 4 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 standard deviation of 1.3). The average number of port stops is higher at 5 stops (Figure A.3, mean of 4.6 and standard deviation of 3.5).[17]

This is also true at the country level: the majority of US trading partners export indirectly

[17] This result is robust for shipment weight and value as well (Figure A.7).

**Figure 4:** 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. ■

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).[18] Shipments from only 9 countries are direct more than not.[19]

Another aspect of indirectness is *transshipment*, which we define as when the origin country 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.2). 27% of shipments by volume are transshipped in a third country before arriving in the US. Moreover, the average good from a majority of non-landlocked US trading partners is transshipped in another country (Panel (B), Figure 4) and about 60 percent of non-landlocked US trading partners transship all US-bound goods in third countries.[20] [21]

Appendix A.4 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

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

[19]The 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.

[20]Examples include Denmark, Bangladesh, Cambodia, and Ecuador.

[21]Appendix Figure A.4 maps the percentage of goods transshipped for each country of origin.

further explores variation in routes from unique origins into the US.[22]

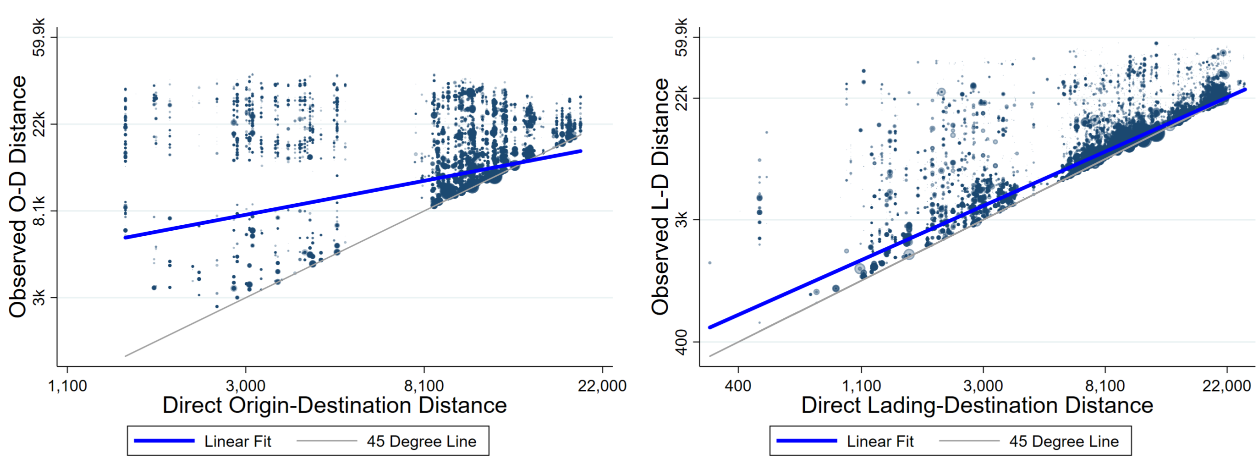
### 2.1.1 Indirect trade increases shipping distances

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 distance 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 5). 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. Appendix Table A.1 further evaluates the relationship between indirectness and journey length, we find that doubling the number of stops adds 10-15% to distance travelled, even after controlling for direct journey length or origin-by-destination fixed effects.

**Figure 5:** Difference between Traveled Distance and Direct Distance for Indirect Shipments

(A) Shipment Origin to Destination (B) Place of Lading (Stop 1) to Destination




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

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

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 to us that it carries with it a cost reduction over and above these costs.

From these results, we can summarize our first stylized fact:

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

## 2.2 Indirect trade takes place through entrepôts

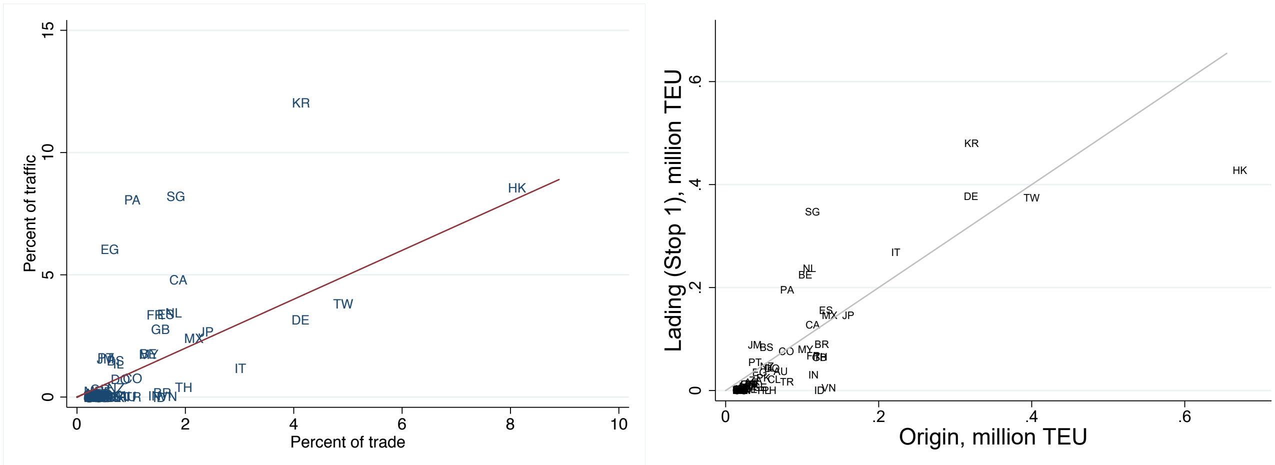
When shipments stop in third countries, where are these countries located? In this section, we show that the stops along indirect shipping routes are not arbitrarily distributed throughout the world. Shipments in our data are channelled through well-known entrepôts. These locations disproportionately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

Panel (A) of Figure 6 scatters each country's percent of total stops against percent of total trade. Here we see where the above concentration is taking place. Some countries 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-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 (SGP), Belgium (BEL), Netherlands (NLD), Korea (KOR), and Panama (PAN), 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. A key takeaway here is that the countries with the largest concentrations of shipments are also more likely to be disproportionately used as entrepôts.[23] Appendix Figure A.5 tabulates the percent of all goods entering the US stopping in a country, broken into goods originated there and elsewhere, for each of the top ten countries.

[23]Since 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 like Jamaica and the Bahamas. Appendix Figure A.6 repeats the exercise in Panel (B) using all our AIS data (no longer merged with US bills of lading), scattering percent of global trade against percent of global AIS traffic.

**Figure 6:** Direct vs Indirect shipments

(A) Percent Originated vs Transit Volume (B) Percent Originated vs Laded




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-country stops (the Y-axes)? Appendix Table A.4 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-country for almost 400 times the number of shipments (by TEU) as the median country, Iraq. This ratio is high by most standards.[24] Transhipsment is similarly concentrated. A natural benchmark might be the same concentration ratio in trade. The same ratio for trade is just over 76, and at all reported ratios, trade is significantly less concentrated than third-country stops.

These relationships can be summarized in our second stylized fact:

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

## 2.3 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

[24]The same ratio for IT-employment in US cities is 300 (Moretti (2019).

unobserved cost reduction at entrepot and may be at work similtaneously.[25] 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.[26] 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 7 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 7 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) that confirm these findings.[27] Column one 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 two adds the log of quantity laded at each shipment's port of lading. Both coefficients are positive but the coefficient on origin volumes are reduced, indicating much of the impact of origin volumes on ship size acts through the size of the lading port. The impact of lading port volumes on ship size is robust to these controls. While shipments' ship sizes are correlated with their origin country volumes, shipments laded in larger ports disproportionately lade on larger ships.

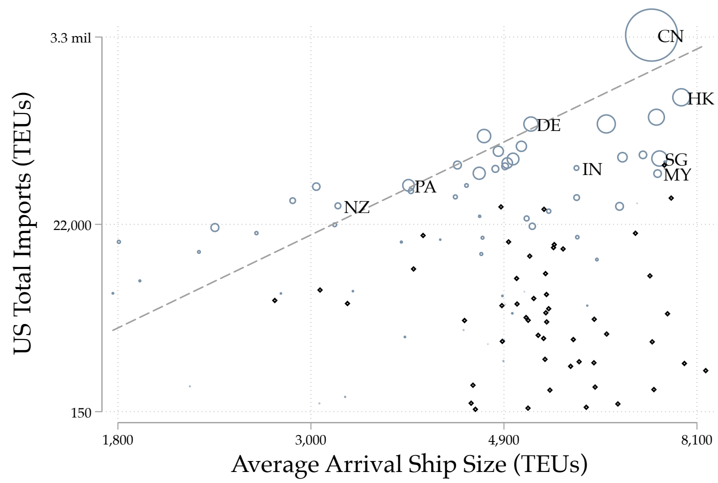
Furthermore, goods that use third countries as transshipment points come on larger ships than would be predicted by their own country's trade volumes. On the figure, the dark highlighted points indicate countries where all shipments coming into the US are always laded elsewhere. These countries, who disproportionately use entrepôts, appear to be outliers, hav-

[25] High-traffic routes are served by many carriers, using ships capable of carrying 25,000 containers with automated lading and unlading 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.

[26] Cullinane and Khanna (2000) find that shipping costs decrease as ship size increases, with this effect being more pronounced on longer trans-oceanic routes.

[27] Here, again, we list the full specifications in the table in the appendix for brevity.

**Figure 7:** 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. ■

ing shipments arriving in the US on ships much larger than what would be expected given their overall trading volumes.

Turning again to the shipment-level data, column three of Table 1 fully interacts variables the specification in column two fully with an indicator variable for shipments that are laded in their origin countries, effectively splitting the specification into two samples. For those shipments whose origin country differs from lading country, which is indicated by a value of 0 for the indicator, the coefficient on lading port volume is considerably higher. Furthermore, as suggested by the figure, shipments' ship sizes are not strongly correlated with origin country volumes when they lade in third countries.

Finally, we note that goods lading at smaller transshipment points along major routes are also on larger ships. On the graph, 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. One explanation may be that lading on ships bound for or coming from much larger ports, shipments at such smaller ports benefit from disproportionately larger ship size.

In column four of Table 1 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

**Table 1:** Determinants of ship size

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Ln ship size | | | |
| Ln volume at lading | | 0.0922*** | 0.148*** | 0.0414** |
| | | (0.0210) | (0.0245) | (0.0188) |
| Lading is origin | | | 0.104 | |
| | | | (0.267) | |
| Lading is origin x Ln vol lading | | | -0.106** | |
| | | | (0.0431) | |
| Ln volume origin country | 0.0943*** | 0.0435** | 0.00695 | |
| | (0.0183) | (0.0204) | (0.00654) | |
| Lading is origin x Ln vol origin | | | 0.0885*** | |
| | | | (0.0261) | |
| Ln largest stop | | | | 0.117*** |
| | | | | (0.0184) |
| Constant | 7.420*** | 6.951*** | 6.754*** | 6.682*** |
| | (0.232) | (0.311) | (0.276) | (0.281) |
| Observations | 178,396 | 178,396 | 178,396 | 178,396 |
| R-squared | 0.142 | 0.200 | 0.223 | 0.227 |

Notes: Observations represent all matched imported containers to the United States. Observations weighted by the size of shipment (TEU). Standard errors clustered by exporting country.
Source: Authors' calculations using AIS and Bill of Lading data. ■

port at which we observe the shipment making a port call after having been laded on the ship. 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 larger ports, 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 estimate 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. Once we implement this estimation, we estimate the role of scale using a historical data and underlying geography as an instrument for modern shipping volumes.

# 3 Theoretical Framework

In this section, we develop a model of global trade where shipments are sent indirectly, through an endogenously formed transport networks. Entrepôts emerge as ports through which goods flow but which are neither goods' origin or destination. Scale economies at these ports or along routes can further reduce costs at, and concentrate shipments through, entrepôts.

We embed the Allen and Arkolakis (2019) route selection model in a generalized Eaton and Kortum (2002) framework, allowing for a complex production function and trade and consumption in intermediates, as well as final goods.

Unlike Eaton and Kortum (2002), in our setting, the technologies available to firms in each industry and country is non-stochastic. Instead, firms additionally solve an optimal route choice problem, where idiosyncratic variation in a products' optimal route generates random variation in the price of each intermediate product-origin pair.

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. Restrictions on route cost heterogeneity generate moment conditions that can be matched to the data to yield estimates of leg-specific shipping costs.

## 3.1 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 $\omega_n$ in industry $n \in N$ in country $j$.[28] Within each industry and product category, goods are homogeneous and normal.[29]

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 $\omega$ is common for all goods in the same industry $n$, and includes a vector of factor inputs

[28] The 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.

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

$L$, as well as inputs of other goods.[30]

Crucially, cost minimization leads the competitive fringe of firms in each country and industry to have the same marginal cost of production. However, the production process itself can be arbitrarily different across industries and countries. Marginal cost of a good $\omega$ is

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

where $P_i$ is the matrix of prices of all goods $\omega$ 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 $\omega$ in industry $n$ must pay tariffs $\kappa_{ijn}$ and iceberg transport costs $\tau_{ijn}(\omega)$ after optimally choosing the route between $i$ and $j$ to minimize the shipping costs incurred. Because markets are competitive, firms selling from $i$ to $j$ price their goods at marginal cost. Observed prices for these products at $j$ are therefore

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

Producers and consumers purchasing a good $\omega$ in industry $n$ at location $j$ source the lowest cost supplier globally.[31]

**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 $\epsilon_{ijnr}(\omega)$,[32]

[30]The 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 $\omega$ from all industries. All inputs are treated as homogeneous.

[31]Here, non-tradable goods can be assumed to have infinite transport costs.

[32]Because of the max-stable property of the Frechét 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 $\theta$.

drawn from the Fréchet distribution such that $F_{ijn}(\epsilon)$, the cumulative distribution function of the idiosyncratic draws is:[33]

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

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 $r$ when shipping from $i$ to $j$.

Accordingly, product $\omega$'s shipping cost along route $r$ from country $i$ to country $j$ is

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

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 iceberg transport costs are usually a function of distance (including in (Allen and Arkolakis, 2019)), we need place no structure on the leg-specific trade costs. Instead, we allow trade costs to be a flexible function of exogenous and endogenous variables: $t = f(X_{exog}, X_{end})$. In our estimation these will later include distance and scale or route volume, which implies an economy of scale in shipping.

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 $\omega$ in industry $n$:[34]

## 3.2 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, including transshipments. We then close the model by imposing 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

[33] This distribution is the same for each product across industries so we drop the product-industry subscripts $n$

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

$\omega$ being shipped on route $r$ from $i$ to $j$ only if the final price of $\omega$, 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 $\omega$ from all other origin country-route combinations.

We can then use the properties of the Frechét distribution to consider the probability that a given country and route $r'$ will be selected as the lowest cost route-supplier combination for good $\omega$ 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 $\omega$ and that country $i$ is the lowest-cost supplier of good $\omega$ to $j$ as:

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

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 $r$.[35][36] We define two matrices following (Allen and Arkolakis, 2019). First $A = [a_{ijn} \equiv t_{ijn}^{-\theta}]$ , where each element is a function of the leg-specific trade cost $a_{ijn} \equiv t_{ij}^{-\theta}$. Second, we define the matrix $B$ as $B \equiv (I - A)^{-1}$. 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

$$\pi_{ijn}^{kl} = \left[(c_{in}\kappa_{ijn})^{-\theta} \cdot b_{nik}a_{nkl}b_{nlj}\right] \cdot \Phi_{jn}^{-1}, \tag{4}$$

where $\Phi_{jn} = \sum_{i'}(c_{i'n}\kappa_{i'jn})^{-\theta} \cdot b_{ni'j}$ is a multilateral resistence 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 $\Phi_{jn}$,

[35] Recall the number of goods in each industry is set arbitrarily large so that the law of large numbers will hold.

[36] The unconditional (pre-selection) average transport cost from $i$ to destination $j$:

$$\tau_{ijn} = \gamma^{-1/\theta}\left(\sum_{r\in R_{ij}} \tilde{\tau}_{ijr}^{-\theta}\right)^{-1/\theta} \tag{3}$$

where $\gamma$ is the function $\Gamma(t) = \int_0^{\infty} x^{t-1}\exp^{-x} dx$ evaluated at $\left(\frac{1+\theta}{\theta}\right)^{-\theta}$.

which accounts for the Ricardian selection of lowest-price sources for each good $\omega$. Intuitively, the traffic flowing on a given leg responds both to that leg's effectiveness in reducing route costs as well as to comptitive forces which make trades increasingly less likely to be pursued for more expensive routes. However, multilateral resistence 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.

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

$$\pi^{kl} = \sum_n a_{nkl} \cdot \sum_j b_{nlj} \frac{\Phi_{kn}}{\Phi_{jn}}. \tag{5}$$

Equations (4) and (5) 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:

$$\pi_{ijn} \equiv \sum_r \frac{\left[c_{in}\kappa_{ijn} \cdot \tilde{\tau}_{ijr}\right]^{-\theta}}{\sum_{i' \in I} \left[(c_{i'n}\kappa_{i'jn})^{-\theta} \cdot \sum_{r' \in R_{i'j}} \tilde{\tau}_{i'jr'}^{-\theta}\right]} = \frac{\left(c_{in}\kappa_{ijn}\tau_{ijn}\right)^{-\theta}}{\Phi_{jn}}. \tag{6}$$

**Closing the model** Factor and goods market clearing and balanced trade conditions close the model. Because they are unnecessary for this stage, we refrain from writing these until we impose further structure.

## 3.3 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 $t_k l$ 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.

On the other hand, when there is no endogenous scale response, only the final term that 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.

This scale-free case is instructive for understanding the way the trade cost matrix interacts with Ricardian competitive forces in the model. In this case, the relationship is positive if and only if the following condition is true:

$$\epsilon_{X_{jn},t_{kl}} + \epsilon_{c_{in},t_{kl}} + \epsilon_{\tau_{in},t_{kl}} > -\epsilon_{\Phi_j,t_{kl}} \tag{7}$$

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 the following holds:

$$\epsilon_{X_{jn},t_{kl}} + \left[\epsilon_{c_{in},t_{kl}} + \epsilon_{\tau_{in},t_{kl}}\right](1 - \pi_{ijn}) > \sum_{i' \not{i}} (\epsilon_{c_{i'n},t_{kl}} + \epsilon_{\tau_{i'jn},\tau_{kl}})\pi_{ijn} \tag{8}$$

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 $t_{kl}$.

# 4 Trade Cost Estimation

Using equations (4) and (6) we can calculate the probability of any good traveling through leg $k, l$ conditional on being sold from $i$ to $j$. 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

$$\Xi_n^{kl} \equiv \sum_i \sum_j X_{ijn} \cdot b_{ikn} a_{kln} b_{ljn} b_{ijn}^{-1} \tag{9}$$

Note this equation is identical to what is established in Allen and Arkolakis (2019). Conditional on the observed trade values $X_{ijn}$, the contribution of trade between $i$ and $j$ to traffic between $k$ and $l$ is invariant to multilateral resistance, tariffs, or technology. This is despite the significant differences between our economic frameworks. In particular, expensive trade routes here suffer from Ricardian selection at destination markets, where the route's impact on prices make them noncompetitive. 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 volume of trade, but do not differentially favor one route from an origin $i$ to a destination $j$. Put differently, any change to a cost in one country will affect trade from that country and others proportionally on all routes.

Equation (9) gives us a relationship between trade volumes, 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 (9) over industries $n \in \bar{N}$ yields

$$\Xi_{\bar{N}}^{kl} = \sum_i \sum_j X_{i\bar{N}} \cdot b_{\bar{N}ik} a_{\bar{N}kl} b_{\bar{N}lj} b_{\bar{N}ij}^{-1} \tag{10}$$

Equation (9) tells us that to accurately measure transport costs, we need only data on transportation and traffic for all goods in an industry. Equation (10) 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.

Conversely, we need not observe all traffic nor account for all trade to estimate transport costs.

We will operationalize equation (10) in our estimation by using observed total containerized traffic volumes, isolating the set of industries that are containerized and assuming that all trade in those industries are in fact by container, no other trade is by container, and all transport costs are identical across containerized industries.

## 4.1 Recovering Trade Costs

To estimate the impact of counterfactuals that alters the geography of container trade, we need estimates for point-to-point container trade costs. These costs are fully determined by the matrix A. Our goal in this section is not to show what causes $A$, but rather simply recover it using all available observable information. Our estimating equation requires two objects linked to observables: trade flows and traffic flows.[37]

We use 2014 US Customs data on containerized and non-containerized shipments to construct the share of each HS 4-digit commodity code that travels by container. All commodities with a containerized share above 80% are labeled as a containerized good.[38] With this data, we then collapse the 2014 EORA International Input-Output database at the country level, segregating containerized and non-containerized commodities.[39] We can then extract out a trade matrix for containerized commodities, $X$, for all countries in the database.

In an ideal world, estimation would recover the trade costs that rationalize observed bilateral containerized traffic flows. While we directly observe ocean traffic, this omits land based trade and internal within-country trade. This is a problem for network links between geographically contiguous countries.[40]

[37]It bears repeating that this procedure is agnostic to the exact specification of any particular trade model that generates trade 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.

[38]In practice we find a bimodal distribution, with some commodities being never containerized (oil and iron ore) and other always containerized (washing machines and children's toys).

[39]We are implicitly shutting down the substitution between containerized and non-containerized trade. This is supported by our bi-modal distribution of goods. Furthermore, while in the past, some ports did not handle containerized trade, by 2014, nearly all ports had some ability to handle containerized trade, relegating break-bulk trade to a small portion of all trade volumes.

[40]For example, containerized trade from Canada to the US can occur via truck or train, and will not be observed in our traffic matrix $\Xi$, even though it appears in the trade matrix $X$. Furthermore, trade between China and Mexico could first take a containership from China to the US, before being loaded on a truck between the US and Mexico.

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:[41]

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

where $a_{ij}$ is an element of the trade cost matrix $A$ and the matrix $Z$ is a vector of covariates defined according to

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

where $\beta_0$ is an intercept, $\beta_1$ considers sea distance between the nearest principal port,[42] $\beta_2$ considers port-to-port traffic and $\beta_3$ and $\beta_4$ consider the total incoming and outgoing traffic at ports $i$ and $j$ respectively. $\beta_5$ considers the role of the backhaul problem from Wong (2019), where ship capacity is fixed by the shipping direction with the higher demand. The indicator variable simply consider if the traffic from $i$ to $j$ is more or less than the traffic from $j$ to $i$. Finally $\beta_6$ consider an indicator variable for two countries that share a land border.[43]

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.[44] However, at this stage these relationships are not of interest to us. Our objective is not the vector $\beta$ of coefficients, but the resulting predictions for $a_{ij}$.[45] In effect, we accept that elements of the vector $\beta$ will be endogenous, and seek 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.

Secondly, note that parameters for $\beta$ will yield estimates of every trade cost $a_{ij}$, but it is not necessary to discipline $\beta$ by comparing traffic on every link. Specifically, we can omit

[41]This functional form is useful as it maps from the real numbers to the unit interval, as is required by our theoretical foundation.

[42]This variable is largely constant within country, with the exception of a small set of countries. For example, in the United States, we use Los Angeles for Asia, New York for Europe and Africa, and Houston for parts of Latin American and the Caribbean.

[43]We 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. See Allen and Arkolakis (2019) for estimates of internal road transport trade costs.

[44]This methodology follows the spirit of Allen and Arkolakis (2019).

[45]While we eventually seek to understand the role played by both scale and traditional distance in determining trade costs, this stage simply aims to estimate trade costs $A$ that rationalize the data, we postpone casual effects to the next section.

within-country traffic as well as traffic between countries that share overland routes and still recover estimates of $a_{ij}$.

Our estimation there will seek to find the vector $\beta$ that match observed container traffic with expected container traffic. We denote this expected traffic, $\hat{\Xi}\left(A\left(\beta\right);X\right)$, which is constructed from estimates of $\beta$ as well as trade data $X$.

We then find estimates of parameters that minimize the differences of expected traffic and observed traffic for countries that do not share a land border:

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

As noted before, sharing a land border indicates that we may not fully observe traffic of containerizable goods on that given leg. We thus cannot match traffic data from these countries with our expected traffic. However, we still need to rationalize observed trade flows $X$, the land border element of $Z$ accounts for this.

Appendix Figure A.10 graphs our resulting matrix of pairwise trade costs. We present the estimates for the vector $\beta$ in the Appendix as they are simply incidental parameters, not fundamentals that we can alter in counterfactuals. Instead, we simply need to know if our $\beta$ estimates produce containerized ship traffic that reflects the world.

## 4.2 Model Fit and External Validity

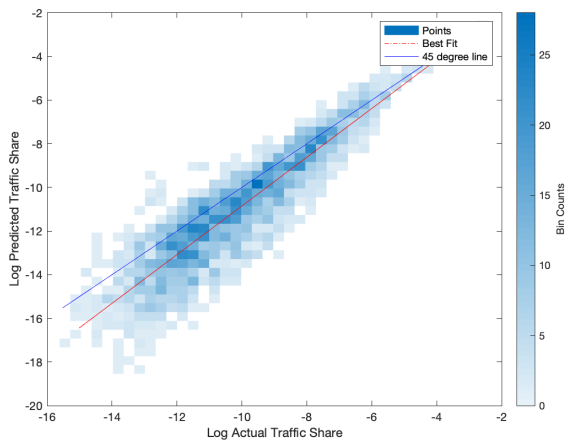
One way to check our estimation is to take our model and compare our expected traffic and trade value flows to actual flows (Figure 8). In Panel (A), we compare actual global observed container traffic shares with the container traffic predicted with our $\beta$ variables. We include both a best fit line and a 45 degree line. If our model perfectly fit the world, all observations would lie on the 45 degree line. In general, we fit the data extremely well, with a correlation between the expected and expected shares (in logs) of .97. In Panel (B), we compared actual observed trade flows to our estimated flows.[46] Even though our model does not target trade value flows, we still fit the data well with a high correlation (in logs) of .73.

We turn now to ask: how good are our model's estimates of leg transportation costs? Using our estimation, which uses global traffic data, the model delivers predictions for how volumes

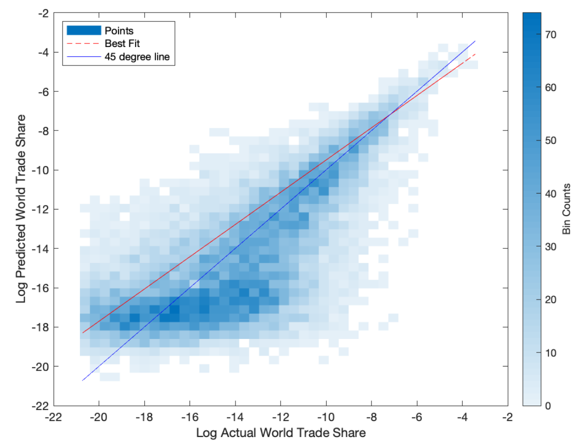
[46]To generate trade flows, we close the model using the full setup in 6.

**Figure 8:** Model Fit Comparisons

(A) Traffic volumes (targeted)




(B) Trade value (untargeted)




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 6. These moments are not targeted by the trade cost estimation.
Source: Authors' calculations using AIS and Bill of Lading data. ■

of US-bound shipments travel through the shipping network – the routes US imports take,

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

as the ratio of all shipments from $i$ to $j$ which can be observed flowing from $k$ to $l$. For goods coming into the US, this is the proportion of goods coming from a specific origin country $i$ that flow through leg $k, l$.

This ratio can be matched to the data in Sections 2 and 3, shipment-level observations of individual routes, which are not involved in the estimation. (The model is estimated using global AIS shipping data.) We can compare the above value from the estimation to the proportion of goods coming into the U.S. from any origin $i$ on leg $k, l$ in our route data, which we call $\pi^*_{ijkl,obs}$

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 for US trade. Our model predicts that there should be some small amount of volume on every leg, which is not the case, as demonstrated by the 30-fold jump in observations. However, because the model predicts extremely low volumes of trade on these legs, including these links does not significantly change the estimates or model fit.

**Table 2:** External Validity Checks

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | $\pi^*_{ijkl}$ | $\pi^*_{ijkl}$ | $\Xi_{kl}$ | $\Xi_{kl}$ |
| $\pi^*_{ijkl,US}$ | 0.846 | 0.872 | | |
| | (0.119) | (0.121) | | |
| $\Xi_{kl,US}$ | | | 1.224 | 1.240 |
| | | | (0.128) | (0.126) |
| Observations | 13813 | 366010 | 652 | 2153 |
| Data | US | Global | US | 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.
Source: Authors' calculations using AIS and Bill of Lading data. ■

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

$$\Xi_{kl} = \sum_i X_{ij}\pi^*_{ijkl}. \tag{11}$$

where $X_{ij}$ is the total volume of trade from origin $i$ to the US. Once again we can compare this to the total volume of shipments in the data moving between a given leg, which we call $Vol_{kl,obs}$.

Column 3 and 4 report these univariate regressions, again adding back in legs with zero observed volumes in 4. Here, the coefficient is strong and positive and still 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.

# 5 Scale Economy Estimation

Understanding the role of scale economies in shipping is critical to understanding the role of entrepôts in trade. Moreover, 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 section, we estimate the causal relationship between scale, or volume, on a given leg and leg cost.

We begin by estimating a "reduced form" relationship between leg costs obtained from our estimation and shipping volumes. Column 1 of Table 3 reports the following specification

$$ln(c_{kl}) = \alpha_0 + \alpha_1 \cdot ln(Vol)_{kl} + \alpha_2 \cdot d_{kl} + \epsilon_{kl}, \tag{12}$$

where $\alpha_0$ is a constant, $\alpha_1$ is the ordinary least squares (OLS) relationship between price and quantity, $\alpha_2 \cdot d_{kl}$ is the coefficient and measure of sea-distance from $k$ to $l$ respectively, and $c_{kl} = a_{kl}^{-1} - 1$, which allows us to interpret $\alpha_1$ as the elasticity between cost and volume to a trade elasticity $\theta$. That is, to interpret our results as elasticities, they must be deflated by $\theta$, which we do not recover directly.

Column 1 of Table 3 reports the OLS relationship between prices and quantities. Unsurprisingly, the strong negative coefficient echoes the strong negative relationship between costs and volumes 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 volume are associated with 0.18% lower transport costs. Although we are estimating a leg-traffic-volumes leg-cost elasticity, this is in the ballpark of what has been established in the ocean shipping and scale literature.[47]

**Table 3:** Scale 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 | | | -0.267 | -0.423 |
| | (0.0113) | | | (0.128) | (0.0919) |
| $\ln(z_{kl})$ | | -0.0419 | 0.157 | | |
| | | (0.0274) | (0.0372) | | |
| $\ln(d_{kl})$ | 0.495 | 0.652 | -0.272 | 0.580 | 0.649 |
| | (0.0329) | (0.0607) | (0.0769) | (0.0499) | (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 node k and node l
Source: Authors' calculations using AIS and Bill of Lading data. ■

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

[47] Asturias (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.

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.

We build such a shifter by constructing a geography-based instrument for demand. We hypothesize that 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 Europe to Korea that include the leg Singapore-to-China, for example, are close to the direct route between Europe and Korea, and more Europe-Korean trade should flow through the Singapore-China link than the Singapore-Australia link, which would involve a longer detour before eventually arriving in Korea.

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 destination to the distance of two legs of a three-leg journey, where the omitted middle leg is the object of interest. For each $k, l$ pair, we calculate the instrument $z_{kl}$ as the sum

$$z_{kl} = \sum_{i \backslash k,l} P_i \sum_{j \backslash k,l} P_j \frac{d_{ij}^2}{(d_{ik} + d_{lj})^2}, \tag{13}$$

where $d_{ij}$ is the sea distance between $i$ and $j$, and the square of the relative excess distance is weighted by the 1960 population $P$ at each origin $i$ and destination $j$.[48]

Links including Singapore scores favorably on this measure, likely because of its strategic location by the Straits of Malacca make it "on the way" for many origin-destination pairs. In Column 2 of Table 3, we report estimates from the reduced-form relationship between our instrument:

$$ln(c_{kl}) = \alpha_0 + \alpha_1 \cdot ln(z_{kl}) + \alpha_2 \cdot ln(d_{kl}) + \epsilon_{kl}. \tag{14}$$

Results show that higher values of the instrument indeed are associated with lower costs as we predict. Column 3 reports that first stage, given by the equation

$$ln(Vol_{kl}) = \alpha_0 + \alpha_1 \cdot ln(z_{kl}) + \alpha_2 \cdot ln(d_{kl}) + \epsilon_{kl} \tag{15}$$

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, legs which are in more distance-minimizing journeys are on average higher traffic legs.

[48] 1960 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.

Column 4 reports our instrumented scale elasticity. As expected, 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.

While the elasticity continues to appear large, it's important to reiterate that these elasticities are only interpretable up to $\theta$. Thus, 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%.[49] This is more modest but broadly consistent with the strong scale economies from ship size in Cullinane and Khanna (2000) as well as our leg-volume-ship-size results in Section 2. Furthermore, these result lend support to our initial hypothesis that a major role of entrepôts are their facilitation of scale through concentration of shipments.

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 the direction of change we expect if omitted port-level information includes cost reduces that are correlated with our demand shifter, but is the direction we expect if port-level congestion is being accounted for. Column 6 includes origin and destination port fixed effects. The instrument is further weakened and the scale elasticity moves closer to the OLS result. As a result, we rely on our column-4 elasticity, our most conservative estimate of scale economies, in counterfactuals to account for the effects of higher volumes on route costs.

# 6 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 then consider two

[49] For this scale economy to be interpretable as the true, externally valid scale economy of shipping, we need the following restriction to hold:

$$Cov(\epsilon_{c,kl}, ln(z_{kl})) = 0$$

where $\epsilon_{c,kl}$ is the error or difference between the model-implied cost $c_{kl}$ and the true leg price. Appendix E further discusses this condition and evaluates its validity using external measures of freight costs.

counterfactuals. The first considers the trade cost effects of global warming, with the Arctic opening up to trade 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.

## 6.1 Closing the Model

To close our model, we adopt the Caliendo and Parro (2014) framework. A continuum of intermediate goods $\omega_n$ 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 $\omega_{nc}$ and $\omega_c$ goods are sourced internationally. Trade routes are modeled for all three sectors but will only be estimated for intermediates $\omega_c$.

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

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

where $\eta_n$ is the Cobb-Douglas industry share, $\sum_n \eta_n = 1$. Because each product $\omega$ 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 $\omega$: 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 $\omega$ is

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

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

The marginal cost of production for firms is

$$c_{in} \equiv \frac{\Upsilon_{in} w_i^{\gamma_{in}} \prod_{n'}^{N} P_{in'}^{\gamma_{in}^{n'}}}{z_{in}}, \tag{16}$$

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

**Composite goods production** In each country $i$, composite goods in industry $n$ are produced using a CES aggregate of intermediates $\Omega_n$, 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

$$p_{jn}(\omega) = min_{i,r} \left\{ c_i \kappa_{ijn} \tilde{\tau}_{ijn} / \epsilon_{ijnr}(\omega) \right\}, \tag{17}$$

where $\tilde{\tau}_{ijn}$ is the expected trade cost from $i$ to $j$ in industry $n$ and $\epsilon_{ijnr}(\omega)$ is the Frechét draw for a given route.

When compared to the relevent equation in Caliendo and Parro (2014), 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

$$P_{jn} = A_n \left[ \sum_{i=1}^{I} c_i^{-\theta_n} \kappa_{ijn}^{-\theta_n} \tilde{\tau}_{ijn}^{-\theta_n} \right], \tag{18}$$

where $A_n$ is a constant.

### 6.1.1 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 $\hat{\tau}_{ijn} = \tau'_{ijn}/\tau_{ijn}$. We follow the the logic of Caliendo and Parro (2014) to show how changes in tariffs, trade costs, and/or productivity change in the cost of production, price index, wage levels, trade flows, and welfare.[50] See Appendix H for full details. Formally we solve for how wages and prices change $\left\{\hat{w}_i, \hat{P}_i\right\}$ as a function of changes of our model primitives, $\left\{\hat{\tau}_{ijn}, \hat{z}_{in}, \hat{\kappa}_{ijn}\right\}$. We compute welfare as the change in real wages, $\left\{\hat{w}_i, \hat{P}_i\right\}$ as a function of

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

changes $\left\{\hat{\tau}_{ijn}, \hat{z}_{in}, \hat{\kappa}_{ijn}\right\}$. Furthermore, we can also compute changes in marginal costs $\hat{c_{in}}$ and trade volumes $\hat{X_{ij}}$.

## 6.2 Counterfactual Methodology

**Algorithm 1** Scale Counterfactual Algorithm

**procedure** WELFARE CHANGE$(X_0, \Xi_0, \hat{t})$ ▷ Find a new equilibrium
  Initialize current trade flows $X_0$ and traffic $\Xi_0$
  Initialize changes in cost fundamentals $\hat{\tau}$ ▷ Example: shipping distances changes
  Compute $A_0 = A(\Xi_0; \hat{\tau})$ ▷ Following equation 12
  Compute $B_0 = (I - A_0)^{-1}$
  Initialize difference $= \infty$, tolerance $= \epsilon$
  **while** $difference < tolerance$ **do**
    Update trade flows $X_1 = X(B_0)$ ▷ Solving 6.1.1
    Update traffic $\Xi_1 = \Xi(X_1, A_0, B_0)$ ▷ Following equation 10
    Update leg costs $A_1 = A(\Xi_1)$
    Update trade costs $B_1 = (I - A_1)^{-1}$
    Compute $difference = \Sigma_{ij}(B_1 - B_0)^2$
    Update $A_0 = A_1$ and $B_0 = B_1$
  Return final trade flows $X_1$
  Compare welfare and price index changes between $X_1$ and $X_0$ ▷ Solving 6.1.1

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 $\eta$ and $\gamma$.[51] We follow the literature Simonovska and Waugh (2014) as conservatively set $\theta = 4$. For any change in trade costs $\hat{\tau}$, we can calculate changes in any country's price index $\hat{P}$ and trade flows $\hat{X}$.

To decompose the effects of (1) bilateral distance or trade costs, (2) the global container transportation network, and (3) scale economies on modeled trade costs $\hat{\tau}$, we run three variants. First, we directly only allow for changes in trade costs only on routes that are directly affected by our counterfactual policy. This captures the direct effect of distance or trade policy. Second, we then allow for these trade costs to affect all trade, including indirect trade. The 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.

[51] We hold trade deficits constant in all our counterfactuals.

We map the change in distance to trade costs through Equation (12), with $\alpha_2$ 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.[52]

**Table 4:** Aggregate Counterfactual Outcomes

| | Directly Affected Routes | Full Trade Network Effects | Allowing Scale Economies |
|---|---|---|---|
| Panel A: Arctic Passage | | | |
| $\Delta$ Average Global Welfare | 0.015% | 0.033% | 0.089% |
| $\Delta$ Container Trade Volumes | 0.174% | 0.382% | 1.018% |
| Panel B: Brexit | | | |
| $\Delta$ Average Global Welfare | | -0.023% | -0.100% |
| $\Delta$ Container Trade Volumes | | -0.245% | -1.127% |

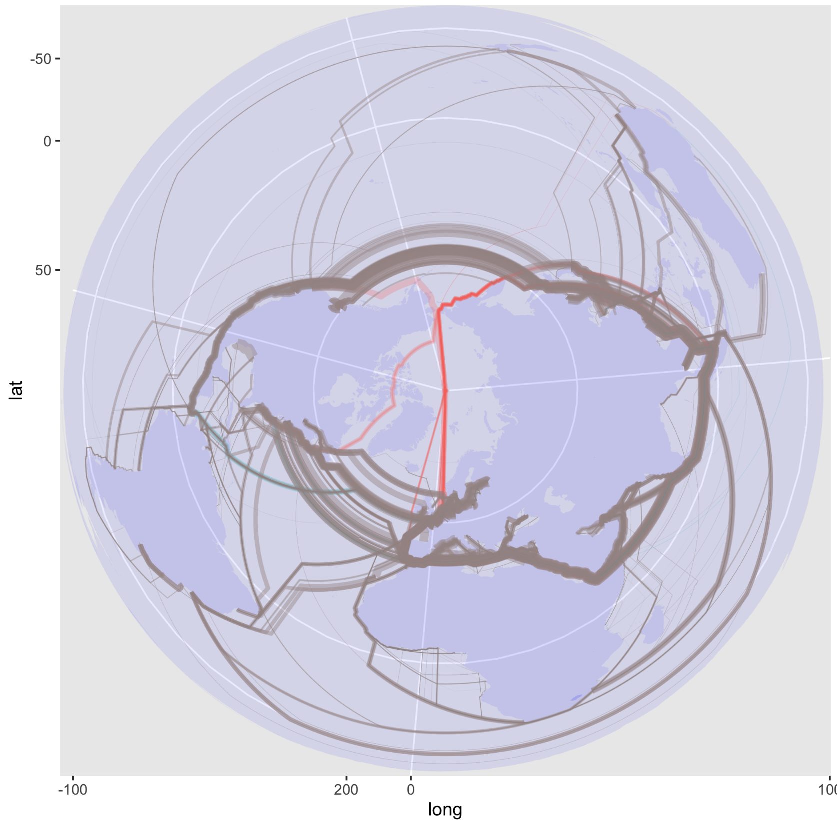
Notes: Notes: In Panel A, we first only reduce the trade costs between country pairs, whose distance is shortened by crossing the Arctic Ocean. Second, we allow for the full network structure $B$ to allow for indirect shipping to also cross the Arctic Ocean. Third, we allow for scale economies to create 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. ■

## 6.3 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 (the Economist, 2018). Today, a small number of ships make this journey, but within the next

[52]In the appendix, we consider the robustness of these equilibria.

**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. ■

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.[53] 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

[53] We measure the mean ocean-going distance between all bi-lateral 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.

without arctic ice caps.[54]

We simulate three different counterfactuals. First, we 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.

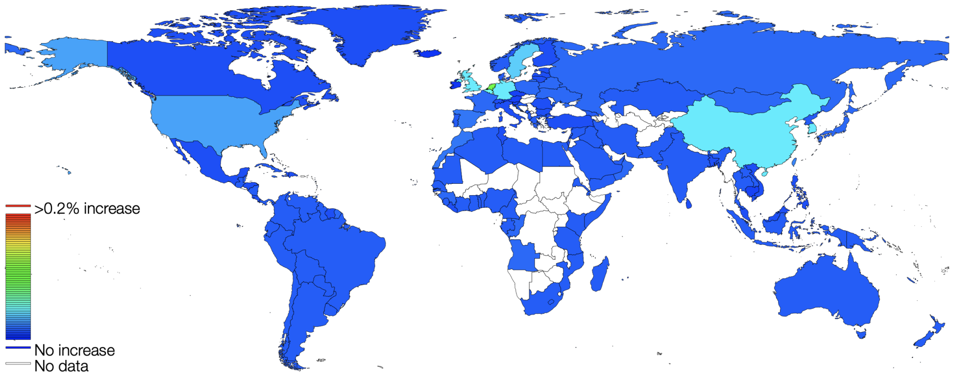
We first show the aggregate gains, averaging across global welfare, before decomposing our heterogeneity. We display the summary statistics for all three scenarios in Panel A of Table 4. The first column, shows that with our simply input-output structure, the direct effects of the Arctic Passage are 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 container traffic.

However these relatively small global effects mask significant heterogeneity across countries. Figure A.13 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.13 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 amply 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.

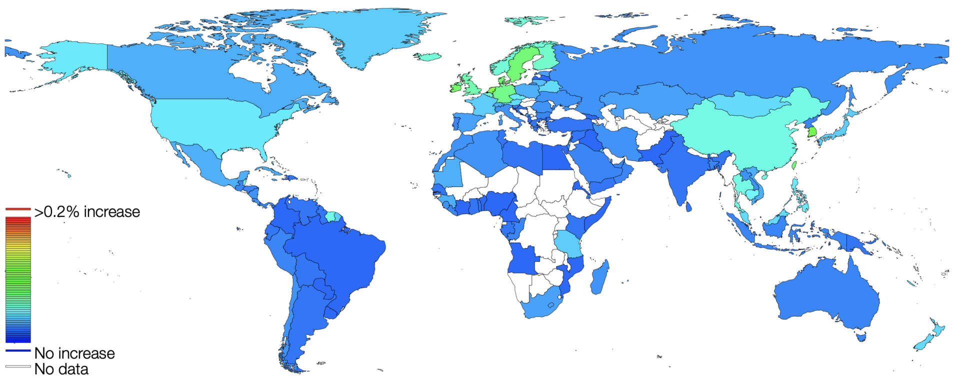
[54] For more information, see the 'gdistance' package.

**Figure 10:** Welfare Changes - Arctic Passage

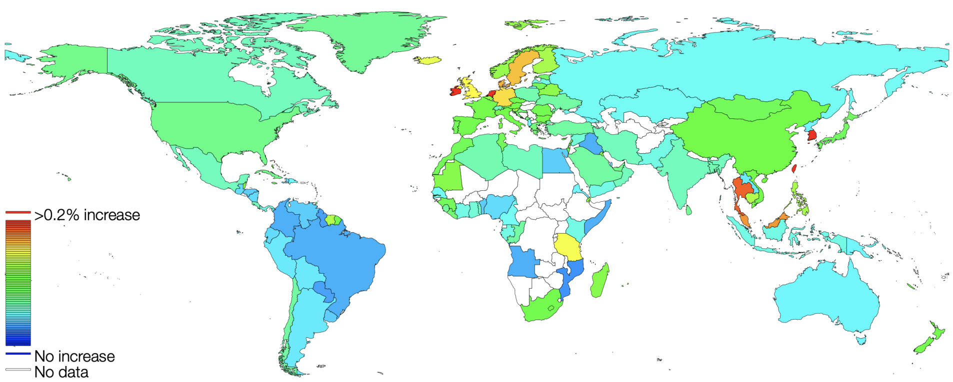
(A) Only Directly Affected Routes




(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) from all countries 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. ■

## 6.4 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.[55] We assume that these tariffs and related costs will

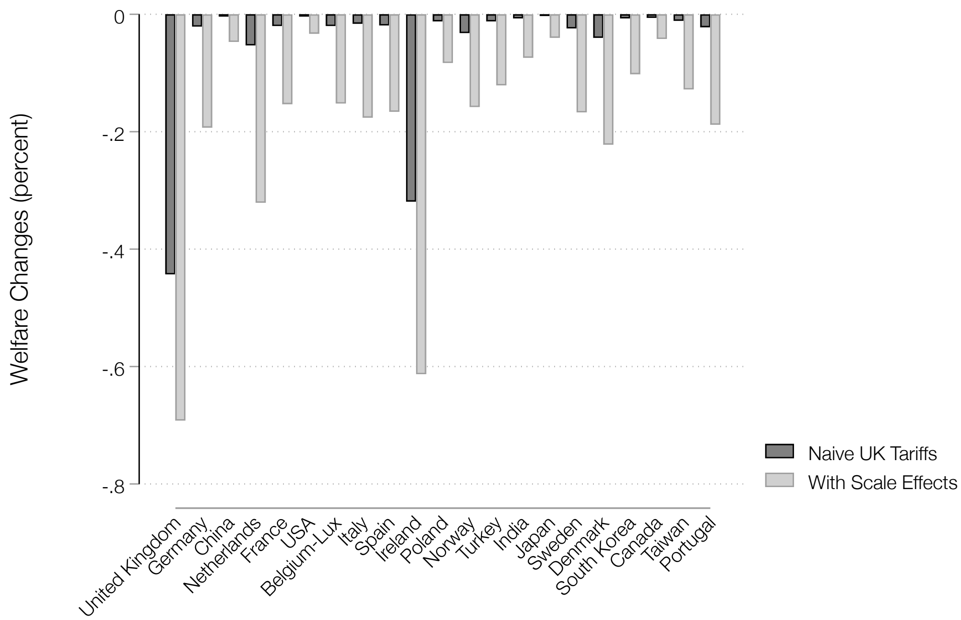
[55]Alternatively, we can model the effect of port infrastructure improvements either between bilateral pairs or for any single nation.

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

We thus 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%. For example 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. Thus 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 12 highlights the distributional effects in terms of welfare (see Appendix Table A.14 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.

**Figure 11:** 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. ■

# 7 Conclusion

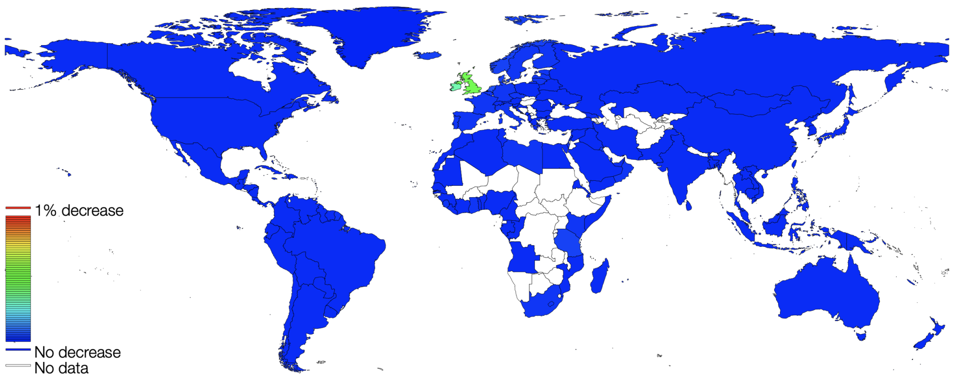
World trade does not all get shipped directly from an origin to a destination. It often takes meandering routes, aided by scale economies that consolidates trade in hubs, or entrepôts. Guided by a series of salient facts, we construct a model world trade with endogenous trade costs, estimating both the underlying trade costs on all container-ship routes, as well as scale economies.

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.

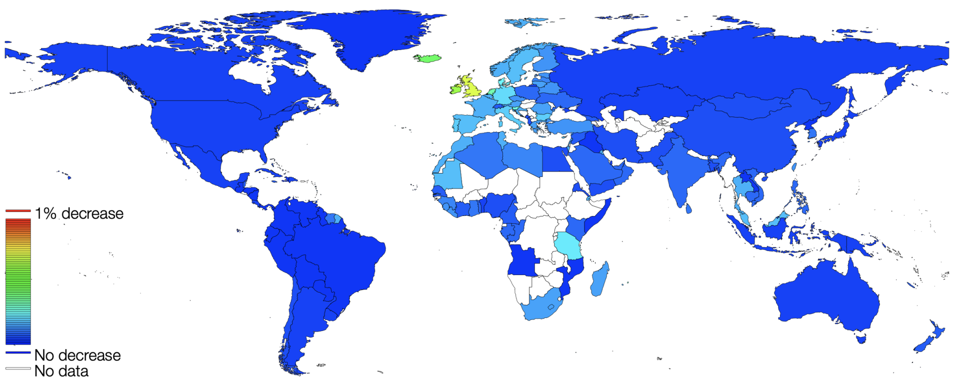
While such hub and spoke networks have been studied in contexts such as airline travel, we take take our trade costs and embed them into a tractable general equilibrium framework. However, our analysis is inherently limited. We only consider the role played by containerized shipping. Do similar patterns hold across other traded goods? What are the roles of fixed costs in enabling these scale economies?

**Figure 12:** 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 reflects 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. ■

# 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."

**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.*

**Behrens, Kristian, and Pierre M Picard.** 2011. "Transportation, freight rates, and economic geography." *Journal of International Economics*, 85(2): 280–291.

**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.

**Blondel, Vincent D, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre.** 2008. "Fast unfolding of communities in large networks." *Journal of statistical mechanics: theory and experiment*, 2008(10): P10008.

**Blonigen, Bruce A, and Wesley W Wilson.** 2008. "Port efficiency and trade flows." *Review of International Economics*, 16(1): 21–36.

**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.** 2014. "Estimates of the Trade and Welfare Effects of NAFTA." *The Review of Economic Studies*, rdu035.

**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.

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

**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.

**Kojaku, Sadamori, Mengqiao Xu, Haoxiang Xia, and Naoki Masuda.** 2019. "Multiscale core-periphery structure in a global liner shipping network." *Scientific reports*, 9(1): 404.

**Lazarou, Nicholas-Joseph.** 2016. "Endogenous Hub Formations in International Trade." Forum for Research in Empirical International Trade Working Paper 1144.

**Limao, Nuno, and Anthony J Venables.** 2001. "Infrastructure, geographical disadvantage, transport costs, and trade." *The World Bank Economic Review*, 15(3): 451–479.

**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.

**Rua, Gisela.** 2014. "Diffusion of containerization." *Finance and Economics Discussion Series Staff working paper 2014–88. Federal Reserve Board, Washington, DC.*

**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.*

**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.

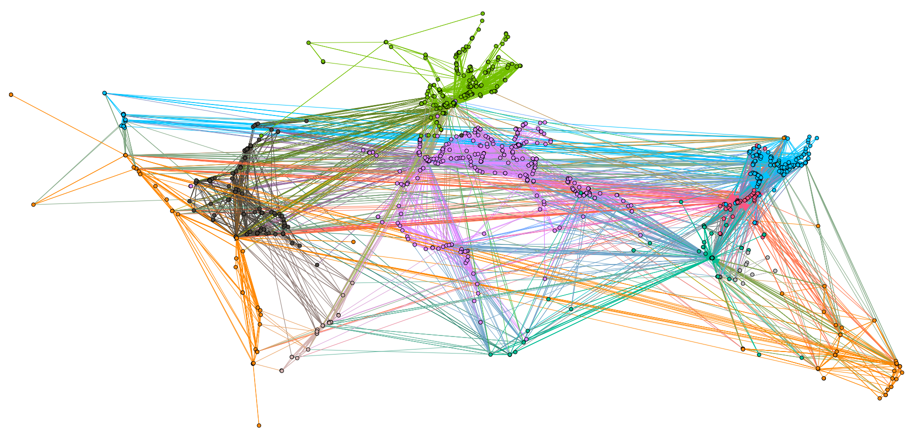
**Wong, Woan Foong.** 2019. "The Round Trip Effect: Endogenous Transport Costs and International Trade." Working Paper.

# Appendix A Additional Figures and Tables

# Appendix B Data Appendix

In figure A.1 below, we use the Louvain method created by (Blondel et al., 2008) to identify the communities—groups of nodes that are more densely connected to one another than to other nodes—within the shipping network.

**Figure A.1:** Map of Global Port of Call Network



Notes: Notes: Coloring uses the Louvain method (Blondel et al., 2008) to identify clusters. Source: Authors' calculations of AIS data. ■

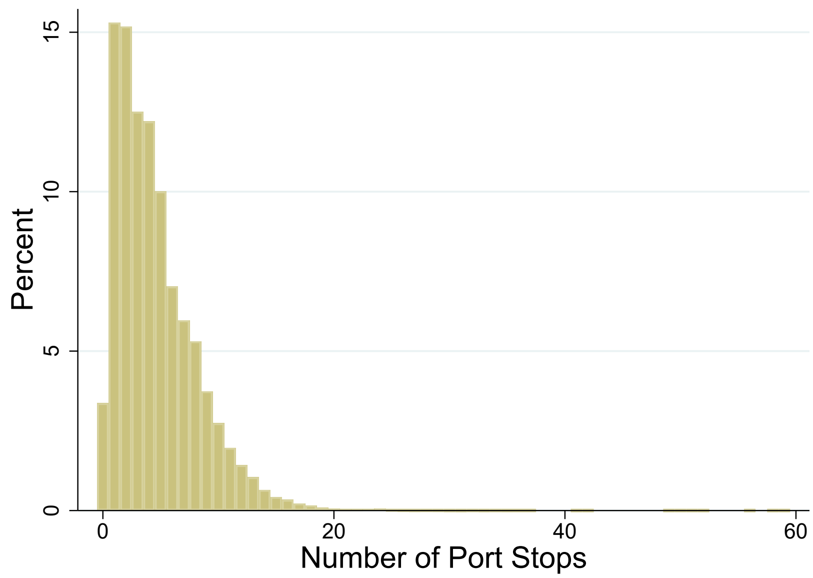
# Appendix C Additional Figures

Panel (A) of Figure A.5 tabulates, for each of the top ten countries, the percent of all goods entering the US stopping in that country. The share of shipments accounted for by shipment origination is in blue while shipments observed stopping in the country but not originated in the country is in red. Unsurprisingly, many recognizable entrepôts are listed, including Korea,

**Figure A.2:** Combined Dataset: Routes Undertaken by Shipments into the US

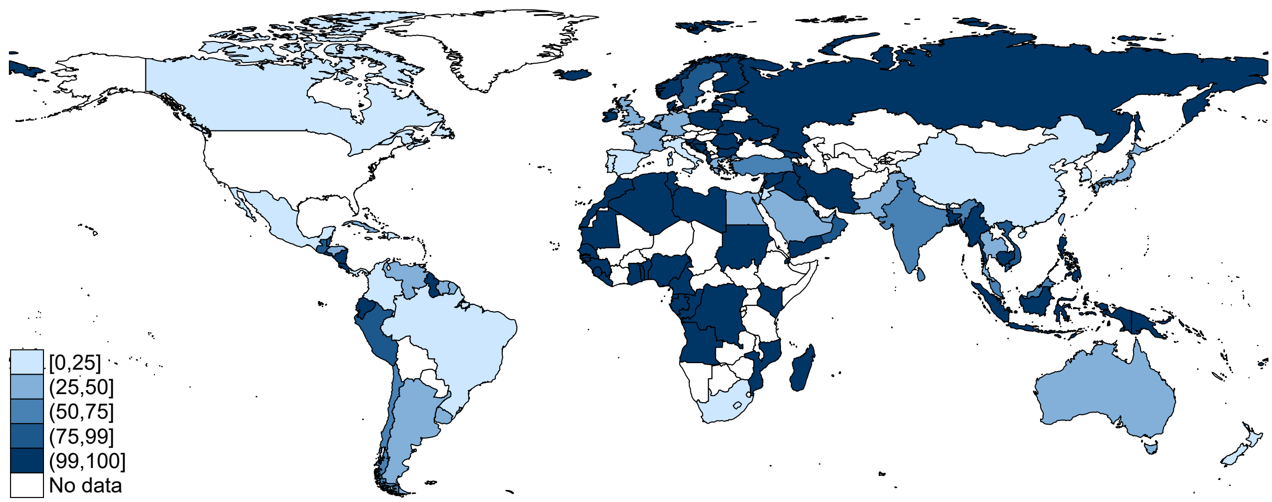
$$\begin{array}{c}\text{Origin}\\\text{(Foreign)}\end{array} \longrightarrow \underbrace{\begin{array}{c}\text{Stop 1}\\\text{(Foreign)}\end{array} \longrightarrow \begin{array}{c}\text{Stop 2}\\\text{(Foreign)}\end{array} \dashrightarrow \begin{array}{c}\text{Stop X}\\\text{(Foreign)}\end{array} \longrightarrow \begin{array}{c}\text{Destination}\\\text{(US)}\end{array}}_{\text{Containership}}$$

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




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

**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. ∎

Panama, Singapore, and Egypt. Perhaps more surprisingly, more than 50% of the containers entering into the US stop in China. While this panel sums to over 1, since each container stops in more than one country, over 80% of shipments to the US stop in at least 1 of 5 countries:

**Table A.1:** Number of stops vs distance

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Log observed distance | | | |
| Log stops | 0.159*** | 0.164*** | 0.150*** | 0.147** |
| | (0.0415) | (0.0477) | (0.0496) | (0.0585) |
| Log direct dist_direct | 0.829*** | 0.821*** | 0.852*** | |
| | (0.0400) | (0.0778) | (0.0356) | |
| Constant | 1.438*** | 1.506** | 1.248*** | 8.304*** |
| | (0.337) | (0.656) | (0.279) | (3.074) |
| Lading port FEs | No | Yes | No | No |
| Unlading port FEs | No | No | Yes | No |
| Lading-unlading pair FEs | No | No | No | Yes |
| Observations | 178,395 | 178,395 | 178,395 | 178,395 |
| R-squared | 0.911 | 0.931 | 0.916 | 0.953 |

Note. 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.
*** p<0.01, ** p<0.05, * p<0.1

China, Panama Singapore, Korea, or Egypt.[56]

Panel (B) replicates Panel (A) but for country of lading. Here the total of all bars (including those not graphed) sum to 1, and 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 important as entrepôt rather than as countries of origin.

# Appendix D Variation in Connectivity

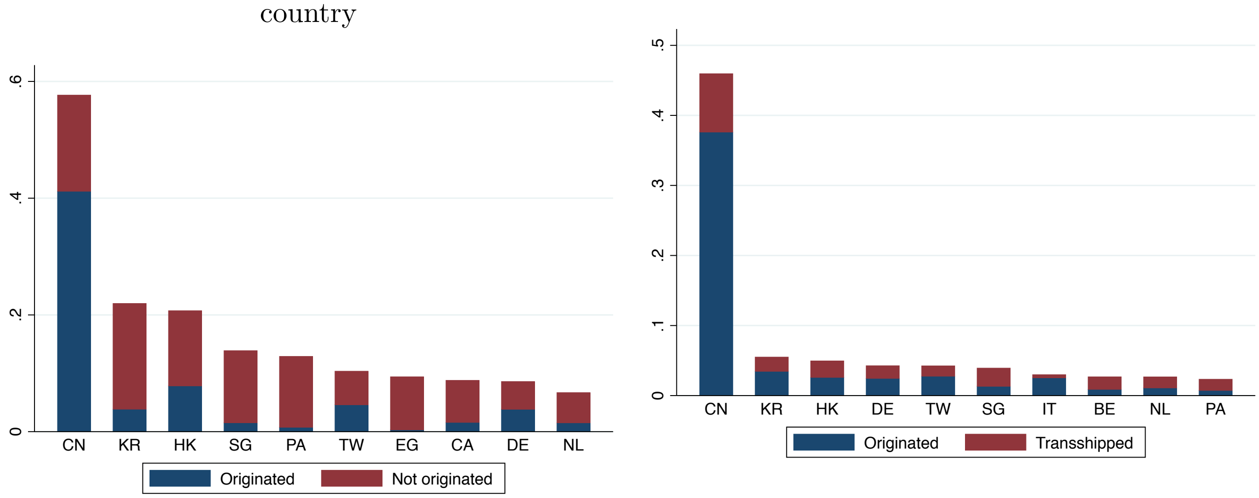
As is evident in Figures 1 and A.4, there is a high degree of variance in indirectness across countries. This variation is reasonable explained by traditional gravity variables. In Panel (A) of figure A.8, we plot number of stops against country GDP and find that countries with higher GDPs are more likely to have less stops on their export journeys to the US. In Panel (B), we plot number of stops against distance instead and find that countries which are larger in size and closer have more direct trade with the US. These results are robust to using port stops

[56] Of course, the sum of these five bars is greater than 80% because the average shipment makes multiple stops).

**Figure A.5:** Most Used Countries

(A) Percent of shipments making stops, by country

(B) Percent of shipments laded, by country

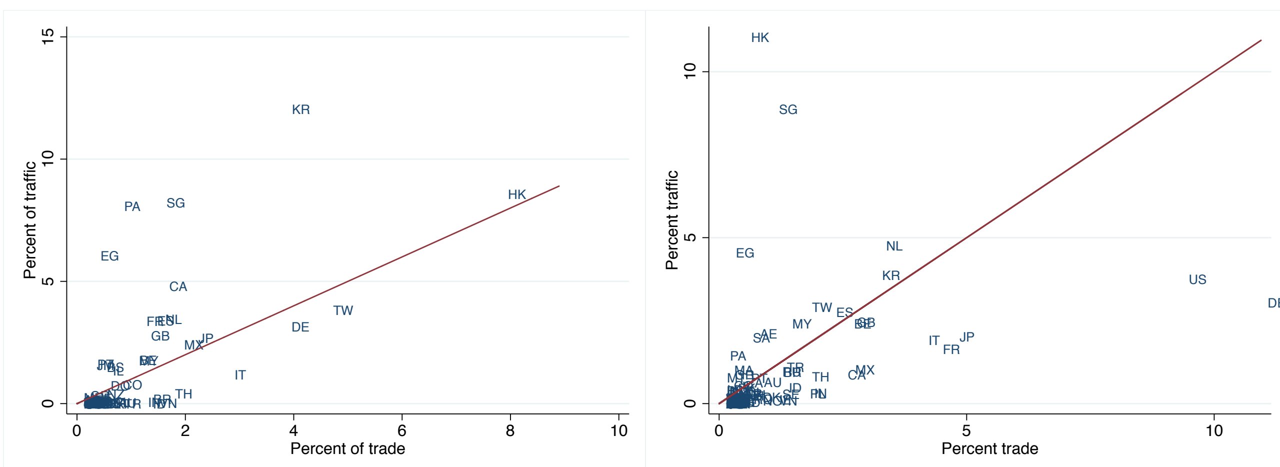



Notes: Source: Authors' calculations of AIS and Bill of Lading Data. ■

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

(A) Traffic v. Trade, US Shipments

(B)Traffic v. Trade, Global Data



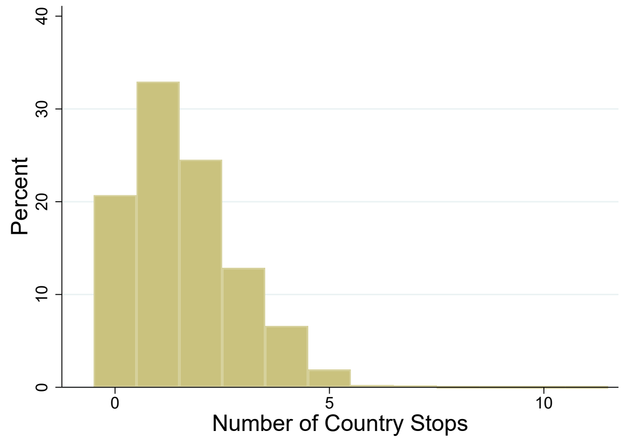

Notes: Note: We omit China as it is out of the data frame.
Source: Authors' calculations of AIS and Bill of Lading Data. ■

instead of country stops (table A.2) 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 availablility of direct trade to the US could in principle reverse the causality.

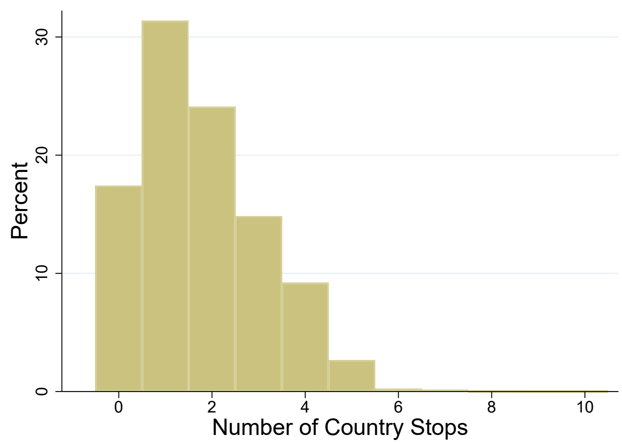
Do shipments from a given origin follow a unique path to the US? Appendix C shows that even from a single origin, observed routes are indeed varied. The existence of this within-origin route variation will be a particularly important assumption in our model and external validity

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

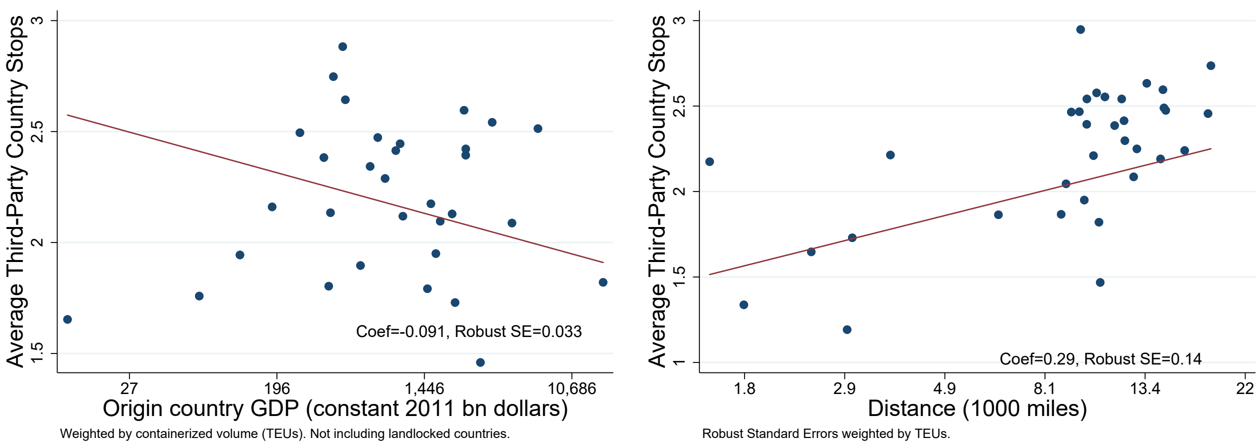
checks.

Panel (A) in Figure A.9 shows the distribution in the number of unique routes to the US by origin country. The average is about 397 routes with wide variation (sd 681). 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.

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.9 shows that almost 70 percent of origin countries have fairly low concentration of routes—HHI less than 1500 which is the low concentration threshold determined by the US

**Figure A.8:** Larger and closer countries have lower number of average stops

(A) Stops vs Country Size (B) Stops vs Distance

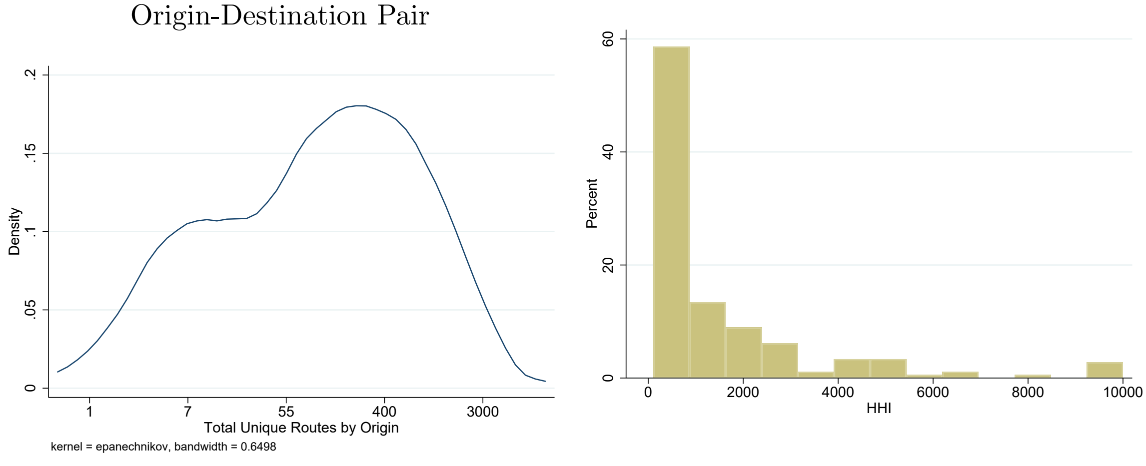



Notes: Binned scatter plot with observation at the origin level. Robust standard errors weighted by total container TEU trade by origin.
Source: World Bank WITS, Seadistance.org, and author's calculations of AIS and Bill of Lading Data. ■

Department of Justice. 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.9:** Variation in trade indirectness

(A) Number of Unique Routes by Origin-Destination Pair (B) Distribution of Route Concentration




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

| | (1) Ctry Stops | (2) Ctry Stops | (3) Ctry Stops | (4) Port Stops | (5) Port Stops | (6) Port Stops |
|---|---|---|---|---|---|---|
| ln GDP | -0.0915** | | -0.114*** | -0.0283 | | -0.0659+ |
| | (0.0329) | | (0.0253) | (0.0514) | | (0.0393) |
| ln Dist | | 0.294* | 0.401* | | 0.619*** | 0.681*** |
| | | (0.143) | (0.162) | | (0.174) | (0.195) |
| Obs | 133 | 133 | 133 | 133 | 133 | 133 |
| F-stat | 7.752 | 4.261 | 12.27 | 0.303 | 12.71 | 6.183 |
| $R^2$ | 0.193 | 0.117 | 0.399 | 0.00995 | 0.280 | 0.331 |

Robust standard errors in parentheses. Weighted by TEUs.
Not including landlocked countries.

**Table A.2:** Relationship between stops and country size as well as distance

| | (1) ln HHI | (2) ln HHI | (3) ln HHI |
|---|---|---|---|
| ln GDP | -0.189*** | | -0.159*** |
| | (0.0387) | | (0.0445) |
| ln Dist | | -0.501*** | -0.341* |
| | | (0.126) | (0.165) |
| Obs | 123 | 123 | 123 |
| F-stat | 23.79 | 15.83 | 16.95 |
| $R^2$ | 0.164 | 0.0997 | 0.206 |

Robust standard errors in parentheses.

**Table A.3:** Relationship between stops and country size as well as distance

**Table A.4:** Concentration Ratios

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

# Appendix E External-validity of Model-Derived Scale Elasticity

We join several other papers that derive prices from shares using the Eaton Kortum (2002) framework and then consider the relationship between model-derived price and quantity. Our scale economy is consistent with the model-derived prices. However, an issue with the external validity of this estimation is the contamination of our price estimates with quantity itself.

In particular, if the true relationship between price and quantity is

$$P_{kl}^{true} = \beta_1 Vol_{kl} + \epsilon_{kl} \tag{19}$$

we are estimating instead

$$P_{kl}^{model} = \beta_2 Vol_{kl} + \nu_{kl} \tag{20}$$

we can further define

$$dif_{kl}^1 \equiv P_{kl}^{true} - P_{kl}^{model}$$

where $\beta_2$ is an an unbiased estimate of $\beta_1$ if $Cov(dif_{kl}, Vol_{kl})$ is zero. Unfortunately, we cannot observe the true price, and cannot varify that assumption. Moreover, the reader may be concerned that because our model nonlinearly relates arrives at Prices by minimizing model-based implied shares from actual shares, we may not satisfy this condition.

Our approach to dealing with this hinges on introducing separate measures of prices. Using freight costs from (Wong, 2019), we can consider the relationship

$$P_{kl}^{Wong} = \beta_2 Vol_{kl} + \kappa_{kl}$$

along with the difference

$$dif_{kl}^2 \equiv P_{kl}^{Wong} - P_{kl}^{model}.$$

At this point, we must make the further assumption that $Cov(\kappa_{kl}, Vol_{kl}) = 0$. As these measures of prices are not model or share-related, this is a more standard assumption to make. Under that assumption, it should be clear that $Cov(dif_{kl}^2, Vol_{kl}) = 0$ if and only if $Cov(dif_{kl}^1, Vol_{kl}) = 0$. We test the former equation using both volume and our instrument in place of volume in table A.5 below.

Column one shows there is only a low, negative correlation between volume and differences to begin with. Column two shows a negative relationship with our instrument. Because the difference is larger when we underestimate price more relative to Wong (2019), this implies the direction of bias from the instrumental variable is negative, and that our scale economy is

**Table A.5:** Over- or Under-estimate of Cost vs Volume

| | Log difference, freight vs estimated costs |
|---|---|
| Log instrument | -0.0499 |
| | (0.0272) |
| Log distance | 0.5054 |
| | -0.0742 |
| Constant | 1.9350 |
| | (0.8798) |
| Observations | 211 |
| R-squared | 0.42 |

Robust standard errors in parentheses clustered by node k.

**Table A.6:** Trade Cost Estimates

| Coefficient | Estimate |
|---|---|
| $\beta_0$ (intercept) | 5.36 |
| $\beta_1$ (log distance) | 0.10 |
| $\beta_2$ (log route traffic) | -0.97 |
| $\beta_3$ (log outgoing port traffic) | 0.27 |
| $\beta_4$ (log incoming port traffic) | 0.38 |
| $\beta_5$ (land borders) | -.4272 |

Notes: Results are highly preliminary and subject to change. Positive coefficients indicate higher trade costs.

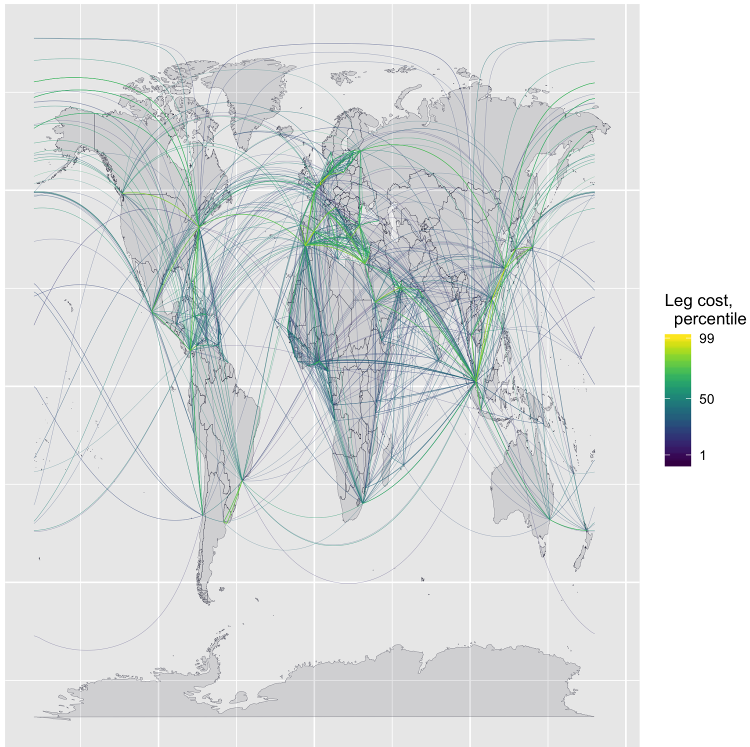
conservatively estimated.

# Appendix F Recovery of Predicted Trade Costs

Table A.6 shows the results of our estimation. Positive values for $\beta$ indicates 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 volumes in $X$.

Notably, we find container trade exhibits significant scale effects on the route-level. This result is very different from Allen and Arkolakis (2019), who find significant congestion costs in domestic road trade. This likely reflects the true underlying nature of the shipping industry versus the trucking industry. While shipping lanes can get crowded or "full", they do not experience anywhere near the same level of congestion as highway shipping. Furthermore, while it is difficult for one truck driver to control more than one or two container loads, the largest sea vessels can have capacities of 25,000 containers, without significant increases in

**Figure A.10:** Trade Cost Estimates, All Legs




Notes: Notes: This map display the recovered trade cost between all origins and destinations for containership legs in the AIS data. Source: Authors' calculations of AIS and Bill of Lading Data. ■

operating costs. On the other hand, port level trade exhibit significant congestion costs, there are costs to having to many incoming and outgoing ships.[57]

# Appendix G Additional Estimation Results

Below, estimated route costs are drawn. Thicker and ligher colors are lower-cost routes. Shorter and more heavily trafficed 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.

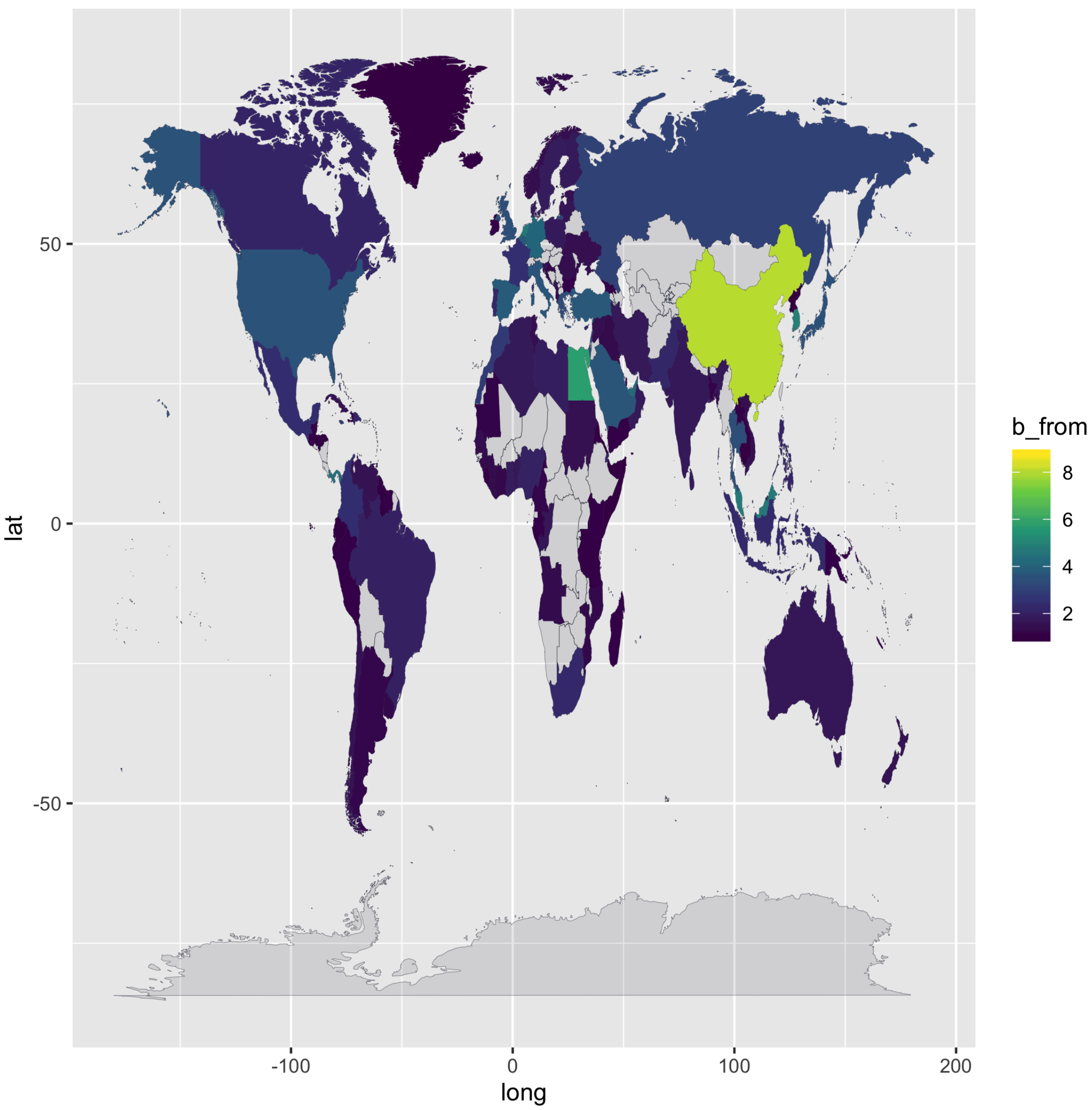
Figure A.11 plots bilateral incoming and outgoing trade costs for Singapore and Lebanon separately. Singapore is not only well-connected but both as an origin and destination has some of the cheapest legs. Singapore ships to Lebanon which has both fewer and shorter connections.

[57]Current analysis follows Allen and Arkolakis (2019) and abstracts away from endogeneity and model misspecification concerns. Forthcoming analysis adopts an instrumental variables approach, instrumenting using expected trade conditional of 1960 economic fundamentals. Results are quantitatively similar.

**Figure A.11:** Trade Costs by Country


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

**Figure A.12:** Expected Trade Costs, Country Average




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

Figure A.12 plots country-level averages of the expected trade cost (from the B-matrix). Entrepôts such as Egpyt, Panama, and (not visible) Singapore and Gibraltar have generally cheaper trade costs, as does China, due to the scale of shipping as well as acecss to nearby low-cost entrepôt (Korea, Singapore, and Japan).

**Table A.7:** $\tau$ vs Distance

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Log trade volumes | | | | | |
| Log $\tau_{ij}^{-\theta}$ | 0.462*** | | 0.449*** | 0.758*** | | 0.524*** |
| | (0.0296) | | (0.0308) | (0.0313) | | (0.0287) |
| Log dist | | -0.752*** | -0.406*** | | -1.325*** | -0.652*** |
| | | (0.0980) | (0.0909) | | (0.0621) | (0.0599) |
| Constant | 12.67*** | 14.76*** | 16.16*** | 15.63*** | 19.87*** | 19.11*** |
| | (0.352) | (0.892) | (0.746) | (0.312) | (0.554) | (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
*** p<0.01, ** p<0.05, * p<0.1

# Appendix H Quantitative General Equilibrium Model in Changes

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

$$\hat{c}_{in} = \hat{w}_i^{\gamma_{in}} \prod_{k=1}^{N} \hat{P}_{ik}^{\gamma_{ink}} \tag{21}$$

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

$$\hat{P}_{in} = \left[ \sum_{i=1}^{J} \pi_{ijn} [\hat{\tau}_{ijn} \hat{c}_{in}]^{-\theta_n} \right]^{-1/\theta_n} \tag{22}$$

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

$$\hat{\pi}_{ijn} = \left[ \frac{\hat{c}_{in} \hat{\tau}_{ijn}}{\hat{P}_{in}} \right]^{-\theta_n} \tag{23}$$

Trade volumes similarly adjust:

$$X'_{in} = \sum_{k=1}^{N} \gamma_{ink} \sum_{j=1}^{I} \frac{\pi'_{ijn}}{1+\kappa_{ijn}} X'_{jk} + \alpha_{in} I'_i \tag{24}$$

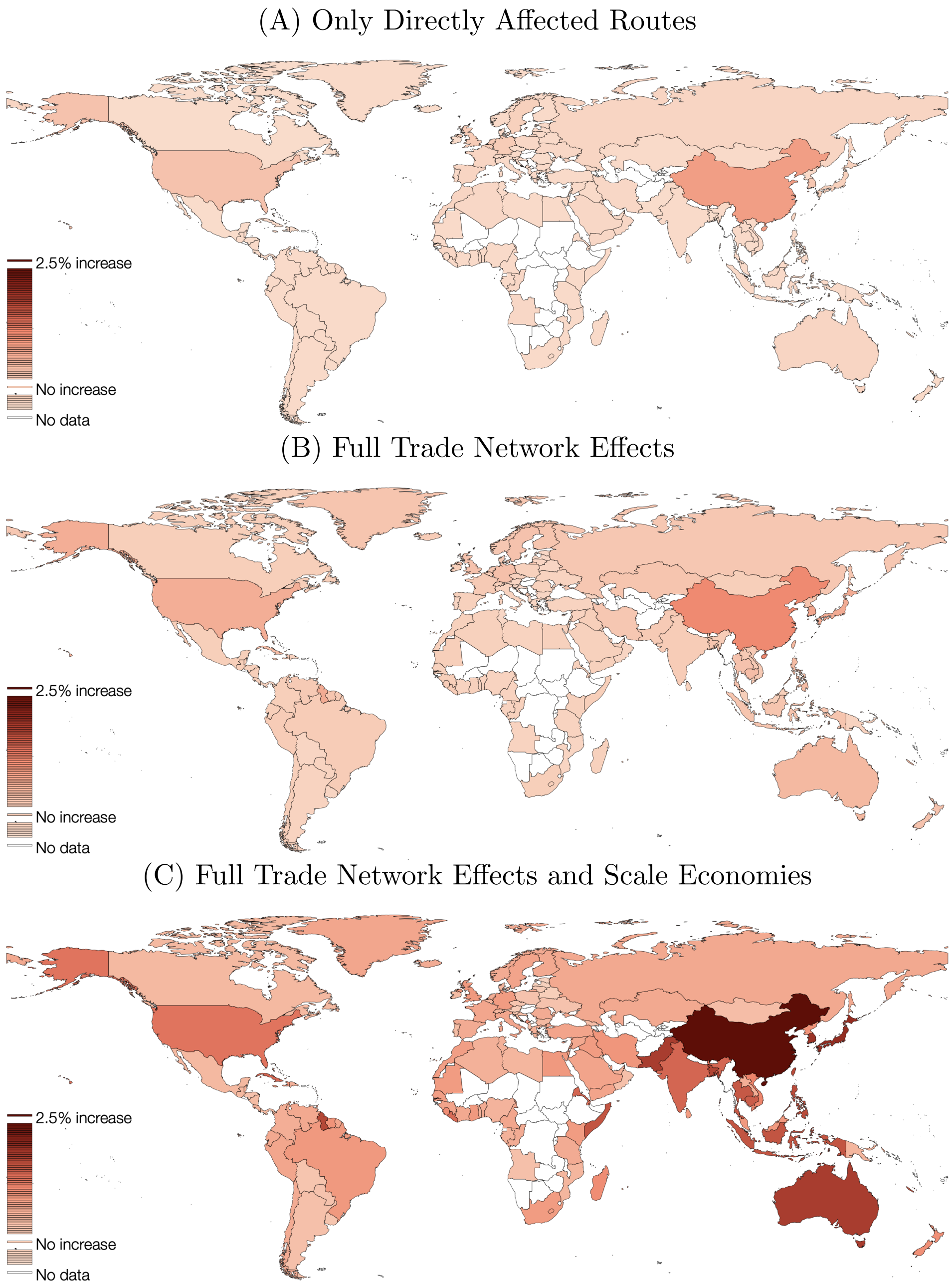
Lastly, trade is balanced to a deficit shifter such that:

$$\sum_{n=1}^{N} \sum_{i=1}^{I} \frac{\pi'_{ijn}}{1+\kappa_{ijn}} X_{in} - D_i = \sum_{n=1}^{N} \sum_{i=1}^{I} \frac{\pi'_{jin}}{1+\kappa_{jin}} X_{jn} \tag{25}$$

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

# Appendix I Additional Counterfactual Results

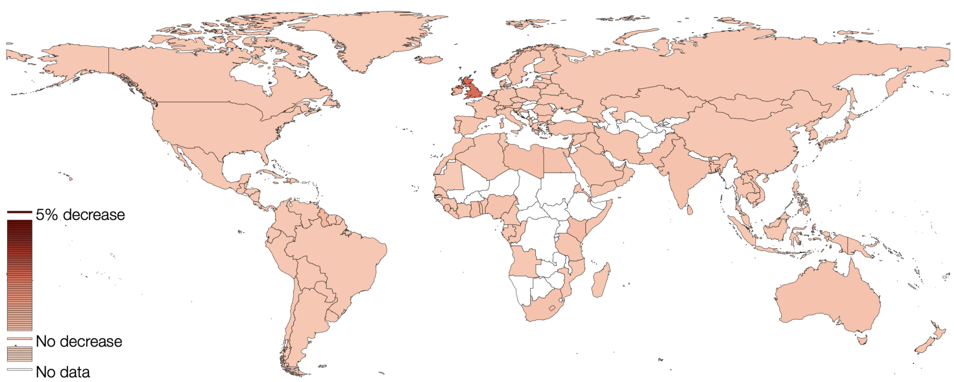
**Figure A.13:** Export Volume Changes - Arctic Passage



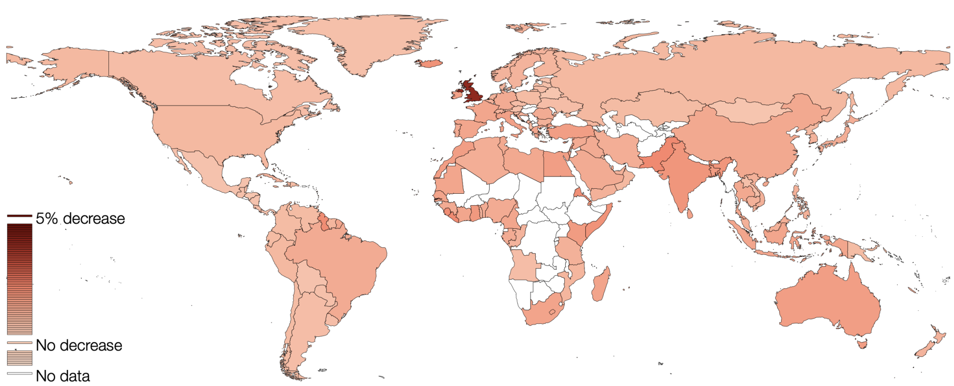

Notes: These three plots show the percent change in exports from all countries in our dataset. Darker reds reflects a greater increase in exports. White represents omitted countries. Panel (A) reflects changes if we only allow trade costs to decrease on routes whose distance is directly reduced to the 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.14:** 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 reflects 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. ■