NBER WORKING PAPER SERIES

# ENTREPOT: HUBS, SCALE, AND TRADE COSTS

Sharat Ganapati
Woan Foong Wong
Oren Ziv

Working Paper 29015
http://www.nber.org/papers/w29015

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2021

We thank Treb Allen, Costas Arkolakis, Panle Jia Barwick, Bruce Blonigen, Johannes Boehm, Mark Colas, Kerem Coşar, Anca Cristea, Meredith Crowley, Carsten Eckel, Stefania Garetto, Matthew Grant, Keith Head, Reka Juhasz, Steve Matusz, Ezra Oberfield, Nina Pavcnik, Andres Rodriguez-Clare, Robert Staiger, Meredith Startz, and seminar participants at Stanford University, University of Michigan, University of Virginia, Boston College, Georgetown University, the Federal Reserve Board, Ludwig-Maximilians-Universität Munich, Indiana University, Syracuse University, University of Cambridge, University of Mannheim, and University of Warwick as well as participants at the NBER Conference on Cities, Labor Markets, and the Global Economy, 2019 Mid-Atlantic International Trade Workshop (Duke), 2020 CeMENT Mentoring Workshop, 20th Annual Nordic International Trade Seminars, and other conferences for helpful comments. Andrew Castro, Ray McCormick, Giacomo Romanini, and Philip Valtadoros provided outstanding research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Entrepot: Hubs, Scale, and Trade Costs
Sharat Ganapati, Woan Foong Wong, and Oren Ziv
NBER Working Paper No. 29015
July 2021
JEL No. F10,F12,F14

**ABSTRACT**

Entrepôts are hubs that facilitate trade between multiple origins and destinations. We study these entrepôts, the network they form, and their impact on international trade. We document that the trade network is a hub-and-spoke system, where 80% of trade is shipped indirectly—nearly all via entrepôts. We estimate indirect-shipping consistent trade costs using a model where shipments can be sent indirectly through an endogenous transport network and develop a geography-based instrument to estimate economies of scale in shipping. Counterfactual infrastructure improvements at entrepôts have on average ten times the global welfare impact of improvements at non-entrepôts.

Sharat Ganapati
Department of Economics
Georgetown University
Edward B. Bunn, S.J. Intercultural Center
37th and O Streets, NW
Washington, DC 20057
and CESifo
and also NBER
sg1390@georgetown.edu

Woan Foong Wong
Economics Department
1285 University Of Oregon
Eugene, OR 97403
wongwoanfoong@gmail.com

Oren Ziv
Michigan State University
Department of Economics
orenziv@msu.edu

A data appendix is available at http://www.nber.org/data-appendix/w29015

# 1 Introduction

Exchanging goods over borders involves more than production and consumption: shipping, transshipping, and distribution can include multiple agents and additional countries beyond producers and consumers. These activities are concentrated at entrepôts, trading hubs which goods travel through—from other origins and bound for other destinations. The idea that entrepôts are integral to the trade network and are engines of growth have been the impetus behind many policies aimed at attaining or maintaining entrepôt status (Financial Times (FT) 2015, Reuters 2016, Wall Street Journal (WSJ) 2021).

This paper studies entrepôts, the trade network they form, and their impact on international trade. We seek to answer the following questions: (1) How do goods move from their origins to their destinations and what role do entrepôts play in facilitating this process? (2) What trade costs can explain the observed routes that goods take and the existence of entrepôts? and (3) How does this pattern of trade through entrepôts impact global and regional trade as well as welfare?

Our first contribution documents that trade is indirect and flows from origins to destinations through entrepôts. We construct and merge two new datasets that jointly map the journeys of containerized shipments from their origins to US destinations including stops in other countries. This microdata grants us a unique, comprehensive look at how goods move through the global trading network. Previous work observed origin-destination trade alone or aggregated ship movements, and was unable to examine indirect trade, which we define as trade journeys that make stops with the shipment either on-board or transshipped—transferred onto a ship—at additional countries (*third-party countries*).

We document two stylized facts. First, the majority of trade—80%—is shipped indirectly.[1] The median shipment stops at two additional countries before its destination. On average, most countries trade with the US indirectly. We further show that indirect trade increases shipping time and distance by 30%. Second, this indirectness is incredibly concentrated, with over 90% of indirect trade channelled through a small number of entrepôts. This establishes that international trade takes place over a hub-and-spoke network. These facts highlight an inherent trade-off. Indirectness increases the distance and time costs of trade, but by revealed preference it lowers costs, especially for the spokes of the network which disproportionately choose to ship via entrepôts.

[1] The majority of trade is also transshipped via a third-party country before its destination.

Our second contribution estimates the trade costs that rationalize the documented direct and indirect trade through the global trading network. To achieve this, we build a general equilibrium model of trade with entrepôts and endogenous trade costs which flexibly accommodates input-output linkages. Producers choose shipping routes and compete for consumers in destination countries in a generalized Ricardian setting. Low-cost routes can involve shipments through third-party countries, and entrepôts endogenously arise at locations through which shipping costs are lowest. Crucially, we allow for both scale economies and dis-economies to govern shipping costs on these network links.

Expanding beyond the US to include global data on shipment flows, we estimate these trade costs for each leg of the network, generating a new set of origin-destination trade costs that is consistent with the global trade network. An advantage of our model is that we need to make very few assumptions on the production and consumption settings; we recover a trade cost matrix that best rationalizes the observed link-level traffic given the observed origin-destination-level trade flows. An important contribution is that we establish the validity of both our estimates and modeling approach by finding a tight match between our estimated trade costs and external freight rate data, as well as between our predicted network flows and microdata on US-bound shipment journeys. Our measures of indirectness, estimates of leg- and origin-destination-level trade costs, as well as resulting market access measures are available online to researchers.

Our estimation finds the presence of bilateral scale economies—the causal effect of increasing shipping volumes on decreasing trade cost—using an instrumental variable approach. Embedded in our model is the intuition that some legs have inherently higher traffic (higher demand) because they are geographically closer to the shortest path between origins and destinations. For example, Singapore lies along the Straits of Malacca, close to the lowest-distance route between many European and Asian countries.[2] We use this variation to construct an instrument for shipping quantities: for each leg, we compute the distance to and from the leg relative to the shortest distance between each origin and destination, recovering a weighted average of each leg's proximity to global trade. We find that increasing traffic volume on a leg by 1% would reduce costs by 0.06%. As the median journey in our microdata has 3 legs, a 10% increase in overall origin-destination trade translates into a 0.17% decrease in trade costs.

Our third contribution is to quantify the global trade and welfare impact of the trade

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

network. Our main counterfactual quantifies the regional and global trade, and welfare benefits of transport infrastructure improvements for each country in our sample. We show that entrepôts are pivotal to the global trade network: welfare impacts of infrastructure investment are on average ten times higher at entrepôts than non-entrepôts. Conflating transportation and non-transportation trade costs, or failing to account for the endogenous response of the network to changes in non-transportation trade costs, impact estimated effects by an order of magnitude on average. This is especially true for entrepôts, which differentially concentrate the benefits of infrastructure improvements regionally relative to non-entrepôts. We find that scale economies amplify the concentration of benefits, highlighting how scale economies in transportation can be a source of agglomeration. We establish that Egypt (and the Suez Canal) is the most pivotal location in the trade network, as reflected by the strain in global supply chains when it was blocked in March 2021 (Wall Street Journal, Financial Times, AP News, 2021).

Our next two counterfactuals illustrate the magnitude and distributional effects of transportation network changes, first through the network's endogenous response to non-transportation trade cost changes and then through direct changes to the transportation network. Our second counterfactual investigates the ramifications of worsening trade relations between one hub, the United Kingdom (UK), and its trading partners—Brexit. In our baselines analysis which only considers the direct impact of increased non-transportation trade costs, we find that outcomes are affected through direct trade with the UK or multilateral resistance. When our analysis accounts for the interaction of network trade costs and scale economies, we find that smaller countries like Ireland and Iceland that use the UK as an entrepôt to access all other trading partners are disproportionately hurt (as recognized in Financial Times, 2020), illustrating how such interactions can lead to different distributional outcomes even when the initial changes are unrelated to transport. Our third counterfactual studies the effects of opening up the Arctic Ocean to regular year-round shipping, connecting countries in East Asia and Europe. We find that the network structure of trade distributes gains beyond directly impacted countries with pre-existing shipping routes. Network spillovers are an order of magnitude above those observed from input-output linkages, and global welfare impacts are further tripled by the feedback loop imposed by scale economies.

This paper ties two broad literature together, combining detailed microdata on the flow of goods through the trade network with a structural model of trade and transportation. The first dives deeply into the technology underpinning the fundamentals of international trade, such as container shipping and infrastructure investment (Coşar and Demir, 2018). The second considers the geography and cost structures of transportation networks (Allen and Arkolakis, 2019). After documenting novel aspects of the trade network by tracing the path of shipments, we estimate the cost structures and scale economies that are consistent with the technologies underpinning these fundamentals in order to quantify the global and local welfare effects of the network.

With regards to the technologies underpinning trade, a wide literature shows how both containerization and infrastructure investments have local outcomes (Heiland et al., 2019; Ducruet et al., 2019; Wong, 2020; Coşar and Demir, 2018; Bernhofen, El-Sahli and Kneller, 2016; Rua, 2014).[3] We model transport costs as part of a global network of container shipping routes, a setting which accounts for two-thirds of annual trade moved by sea (World Shipping Council).[4] Using our general equilibrium spatial trade framework, our counterfactuals show how endogenous changes in trade costs propagate via the network and through entrepôts as well as quantify their trade and welfare impacts.

Our modeling embeds trade networks within a class of gravity models (Head and Mayer, 2014). We provide empirical evidence for a growing quantitative literature investigating the role of trade networks (Allen and Arkolakis, 2019; Fajgelbaum and Schaal, 2017; Redding and Turner, 2015).[5] We extend the Armington framework in Allen and Arkolakis (2019)—where route cost shocks are born by consumers—to a general Ricardian setting—where traffic volumes reflect both route choice and head-to-head competition on prices at destinations and demonstrate how to estimate the model in a multi-industry setting in the presence of unobserved traffic flows. Finding a tight match between our model predictions and external cost estimates, ship sizes, and a sample of observed trade routes, we establish that our estimates reflect actual costs and indirect flows in the trade network in addition to serving as a check to the validity of our modelling approach as

[3] Hummels, Lugovskyy and Skiba (2009),Grant and Startz (2020), and Asturias (2020) study transport costs in the context of market power. While container shipping firms may hold market power, we generalize away from the profits of the shipping companies. Models allowing for leg-level oligopoly, fixed costs and endogenous entry competition fit within our framework (Sutton, 1991), but we leave the study of how market power works through the hub-and-spoke network for future study.

[4] Brancaccio, Kalouptsidi and Papageorgiou (2017) estimate endogenous trade costs arising from search frictions for dry bulk ships carrying homogeneous commodities, where all trade is direct.

[5] We provide the first and systematic documentation of indirect trade through the containerized shipping network. Previous work have either imputed indirect trade or just used port of call data alone (Heiland et al., 2019; Wang and Wang, 2011; Kojaku et al., 2019; Lazarou, 2016). Relative to these papers, we also quantify the welfare and trade impacts of the trade network using a general equilibrium model with endogeneous trade costs.

well as the Allen and Arkolakis (2019) framework.

One important aspect of transportation technology in our model is the scale economy in shipping. For the median route into the US, our leg-level scale economy implies that a 10% increase in volume leads to a 1.7% decrease in costs, which is about three-quarters of the estimates in Asturias (2020) and Skiba (2017).[6] While Lashkaripour and Lugovskyy (2019) and Bartelme et al. (2019) both consider the trade consequences of production scale economies, we consider scale in transportation. Our paper shows how scale economies in transportation can interact with the global trade network to concentrate economic activity. In this respect, we are also related to a literature in economic geography which considers the role of localized scale economies in the emergence of agglomerations (Allen and Arkolakis, 2014; Allen and Donaldson, 2018). Scale economies typically generate agglomerations by acting on the volume of economic activity at locations. Our counterfactuals show that scale economies can concentrate trade and welfare gains at entrepôts by acting on transportation costs over a network.

# 2 Data

We 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 United States bill of lading data for containerized imports, which gives us shipment-level information on US imports. Independently, these datasets partially describe the global shipping network. Merged, they reconstruct the journey of individual shipments as they navigate the trade network, from their origin to their US port of entry. To our knowledge, we provide the most comprehensive reconstruction of the global trading network and routes undertaken by individual shipments into the US.[7]

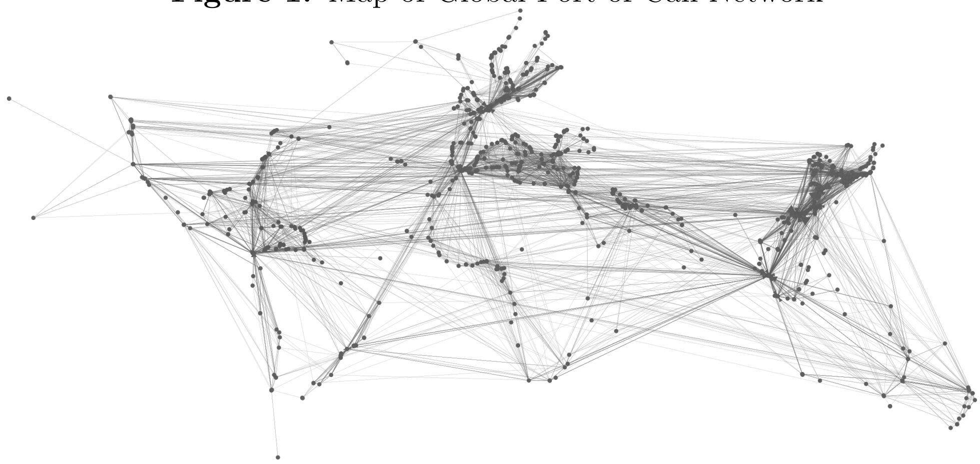
Our ports of call data captures vessel movements using Automatic Identification System (AIS) transponders.[8] For each vessel, this data captures the vessel's characteristics, time-stamped ports of call, capacity, and height in the water before and after stopping at each port. The latter two pieces of information indicates the vessel's load at these ports, allowing us to observe volumes shipped between port pairs.

[6] Asturias (2020) reports an origin-destination country trade-volume trade-cost elasticity of 0.23 while Skiba (2017) reports an elasticity of 0.26 using product-level import data from Latin America. See also Alder (2015); Holmes and Singer (2018); Anderson, Vesselovsky and Yotov (2016).

[7] Data Appendix A.1 explains both data sets and their merge procedure in detail.

[8] Port receivers collect and share AIS transponder information (including ship name, speed, height in water, latitude and longitude). Using Astra Paging data, we track global port entry and exit data.

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



Notes: Each dot represents a port (total of 1,203 ports). Each line represents a journey between port pairs undertaken by a containership (total of 4,986 ships).

Our sample covers 4,986 unique container ships with a combined capacity of 18.1 million twenty-foot equivalent shipping units (TEUs)—over 90% of the global container shipping fleet—making 429,868 calls at 1,203 ports from April to October 2014. Figure 1 shows the coverage of the shipping network in our port of call data. Each line represents a containership journey. We use this global data along with CEPII global trade data when estimating our model in Section 5.

With this port of call data alone, shipment journeys within the trading network remain unobserved. To remedy this, we merge the port of call data with US bills of lading data, which captures shipment-level information for all containerized imports. We observe each shipment's origin country, the port where they are loaded onto containerships (also known as port of lading), and the US port where they are unloaded (port of unlading). We observe the name and identification number of the containership which transported the shipment as well as the shipment's weight, number of containers (TEUs), and product information. Over the same six months period, we see a total of 14.8 million TEUs weighting 106 million tons were imported into the US from 227 origin countries and loaded onto US-bound containerships (laded) in 144 countries.

Using details on containerships, ports, and arrival times, we reconstruct each shipment's journey from its foreign origin to US destination by matching each shipment to the containership that it was transported on. Over 90% of containerized TEUs entering the US can be matched to routes using this method (Appendix Figure A.1 visualizes this merge).[9] While the shipments' exact journey between origin and the first stop (the port

[9]See Appendix A.1 for further details on each of these datasets as well as the merge process.

where they are loaded onto containerships) remain unobserved, this initial portion can either take place overland (by trucks or rail) or by sea on another containership because they are containerized. Not observing this portion in fact leads us to under-count the overall level of indirectness. We empirically deal with unobserved transit in Section 5.

# 3 Stylized Facts

We analyze the international trade network and the routes taken by goods entering the US along that network. We find that the majority of trade takes place indirectly in a manner which is costly—increasing both shipping time and distance travelled. We further show that the global trade network is a hub and spoke system, concentrating a large number of shipments through a small number of entrepôts.

## 3.1 The Majority of Trade is Indirect

Panel (A) in Figure 2 reports the distribution of the number of observed country stops made by each shipment, weighted by TEU containers. Only 20% of containers are exported to the US directly from their origin countries—making no stops in third-party-countries. The average container entering the US stops at around two third-party-countries.[10] The map in Figure 2, Panel (B) shows that this is also true at the country level: the majority of US trading partners export to it indirectly. Only shipments from 9 countries typically enter the US directly.[11] Similarly, the average shipment from a majority of US trading partners is transshipped in a third-party country—over 60% of US trading partners transship more than 90% of their US-bound goods.[12] Figure A.5 reports the percent of goods that are transshipped at third-party countries.

We explore the high degree of variation in connectivity in Appendix B.4, showing that this variation is in part explained by traditional gravity variables. We show that there is substantial variation in routes from unique origins into the US, which is an important assumption in our model and is used in our validity checks (Figure A.9, Panel (B)).

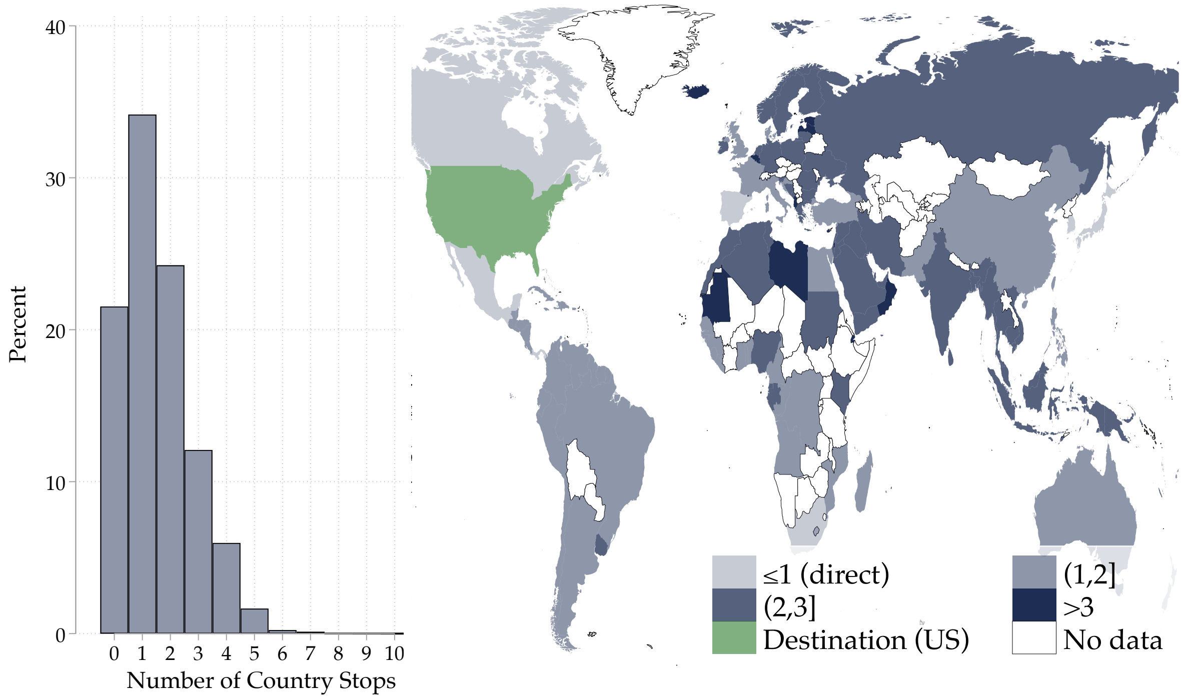
[10] Mean of 1.5 and s.d. of 1.3. Landlocked countries are excluded. The average number of port stops is higher (Figure A.3, mean of 4.6 and standard deviation of 3.5). This result is robust for shipment weight and value (Figure A.4). Multiple stops at the same third-party country are not counted.

[11] These countries are Canada, Mexico, Panama, Japan, South Korea, Spain, Portugal, South Africa, and New Zealand. We treat Mainland China, Hong Kong, Taiwan, and Macau as separate locations.

[12] Both on-board stops and transshipment are important measures of indirect trade. For completeness, all results are broken out here or in the appendix using transshipment only. Examples of countries transshipping more than 90% of goods include Denmark, Bangladesh, Cambodia, and Ecuador.

**Figure 2:** Indirect Trade Distributions, by Container and Country

(A) Country Stops per Container (B) Average Stops between Origin and the US



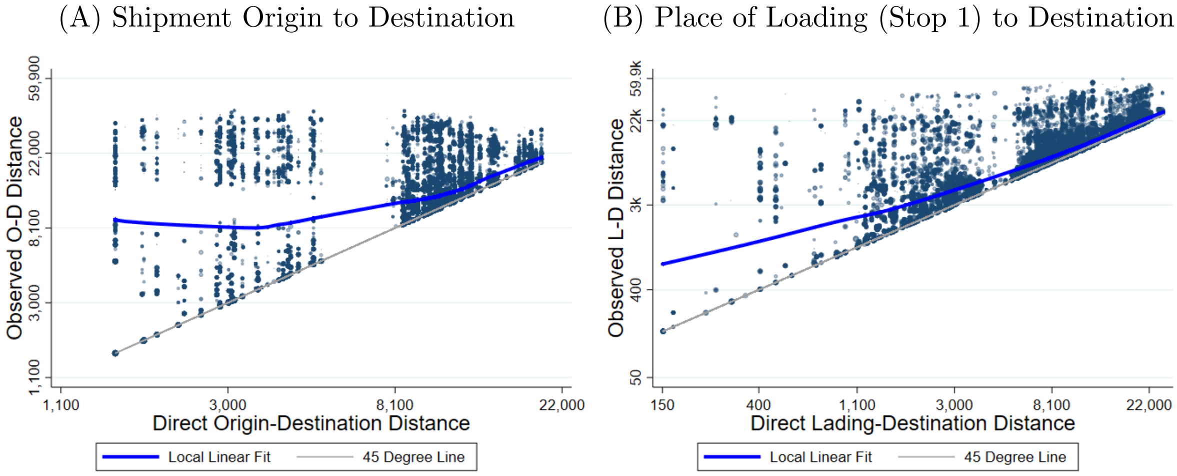

Notes: Panel (A) shows the distribution of containers by the number of unique third-party countries the containers visited. In Panel (B), for each origin country, we calculate the average number of third-party country. The destination country (US) is excluded (in white). Plots are at the shipment level and weighted by the aggregate exported containers (TEU). Landlocked countries are also excluded (in white), since they would mechanically 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 excluded either due to lack of overall trade with the US (e.g. Somalia) or due to the merge process (e.g. Namibia). ■

**Indirect trade increases shipping distances and time.** Are the additional country stops simply incidental stops along the way, or do they constitute a trip that is distinct from a "direct" path? As an example, do goods transiting the Straits of Malacca stop at Singapore since it is "on-the-way," or are goods making lengthier detours during their stops? One possibility is that the observed indirectness is optimal but only incidental—perhaps additional stops only have small effects on costs, and so may be optimal even if the benefit of indirectness is small. However, the significant additional distance and time incurred by indirect travel, documented here, implies this is unlikely to be the case.

On average, the actual traveled distance between a shipment's origin and its US destination is 31% more than its direct ocean distance (Panel (A) in Figure 3). Panel (B) shows the actual traveled distance between the location where the shipment was last loaded onto a ship and its final destination. Here the remaining gap is still substantial at 14%. Table A.1 further evaluates the relationship between indirectness and journey length. Controlling for direct journey length or origin-by-destination fixed effects, doubling the number of stops adds 10% to distance travelled and 33% to time travelled. These distance and time costs do not include pecuniary costs of transshipment. Consequently, this indirectness is meaningful in the sense that it is costly. These longer shipping routes imply a cost reduction from indirectness that is over and above the additional time and distance 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 and time.*

**Figure 3:** Difference Between Traveled Distance and Direct Distance




Notes: These figures show only indirect shipments, with different direct and observed distances. Dots are shipments, shaded by TEU. Panel (A) compares the direct shipping distance from the shipment's origin country to the US, to the actual route travelled. Panel (B) compares the direct distance from the place a shipment was last loaded onto a US-bound ship (Stop 1 in Appendix Figure A.1), to the actual route travelled. Sea distances for observed and direct routes are calculated using Dijkstra's algorithm. The local linear fit line is a locally weighted regression of the observed on direct pair-wise distance. ■

## 3.2 Indirect Trade Is Routed Through Entrepôts

When shipments stop in third-party-countries, how are they routed? We show that the stops along indirect shipping routes are not arbitrarily distributed throughout the world. Instead, they are channelled through a small number of hubs, which disproportionately service shipments originating in other countries.

Panel (A) of Figure 4 plots each country's share of total third-party-country stops against its share of total US trade. Some locations are both popular stopping points and major countries of origin for goods like China, Germany, and Japan. Key countries like Korea, Singapore, Panama, and Egypt disproportionately participate as third-party-countries in US-bound shipments.[13] This leads to our measure of entrepôt activity:

$$Entrep\hat{o}t_{l,j} \equiv \pi_j^l - \pi_{l,j} \tag{1}$$

where country $j$'s usage of entrepôt $l$ for its imports is the difference between $\pi_j^l$, the share of $j$'s imports flowing through $l$, and $\pi_{l,j}$ the share of $j$'s imports originating at $l$. This captures the use of location $l$ above and beyond its role as an exporter to $j$.[14]

Panel (B) of Figure 4 repeats the exercise in (A) using *global* traffic minus trade shares.[15] Results are broadly consistent with the microdata in Panel (A), while some countries such as Canada and Panama which are specifically integral to the US network are now below or closer to the 45 degree line. We define the top 15 countries as entrepôts.[16] In the microdata, 72% of all shipments pass through at least one entrepôt. Of indirect shipments, 90% pass through an entrepôt. In both panels, third-party-country stops (the Y-axes) are significantly more concentrated than trade (the X-axis).[17]

Which countries disproportionately use entrepôts? We find that smaller origin countries are simultaneously more likely to go indirectly and more likely to use entrepôts (see Appendix B.3 for further details). Jointly, these figures confirm that smaller countries are spokes which disproportionately use entrepôts for their trade.[18] These relationships can be summarized in our second stylized fact:

**Stylized Fact 2.** *Indirect shipping routes are concentrated through entrepôts. International trade occurs over a hub-and-spoke network.*

Our two facts outline an inherent trade-off: indirectness increases distance and time

[13]Figure A.6 tabulates the percent of all goods entering the US stopping in that country, broken into goods originated there and elsewhere.

[14]$Entrep\hat{o}t_{l,j}$ is directly proportional to the total volume of goods moving through $l$ that do not originate at $l$. Appendix C shows how this measure arises from our model as the difference between $l$'s on-board marginal cost selling to $j$ and its network relation to $j$, and that lowering location $l$'s leg-level transport costs to other origins increases $Entrep\hat{o}t_{l,j}$. Our results here and throughout are be robust to other functional forms—for example log differences.

[15]We subtract country $l$'s share of observed global containerized trade $\pi_l$ from its observed share of global container traffic $\pi^l$, with an adjustment for unobserved overland traffic as described in Section 5. Appendix C clarifies how this is a consistent aggregation of the country-level measure in Equation 1.

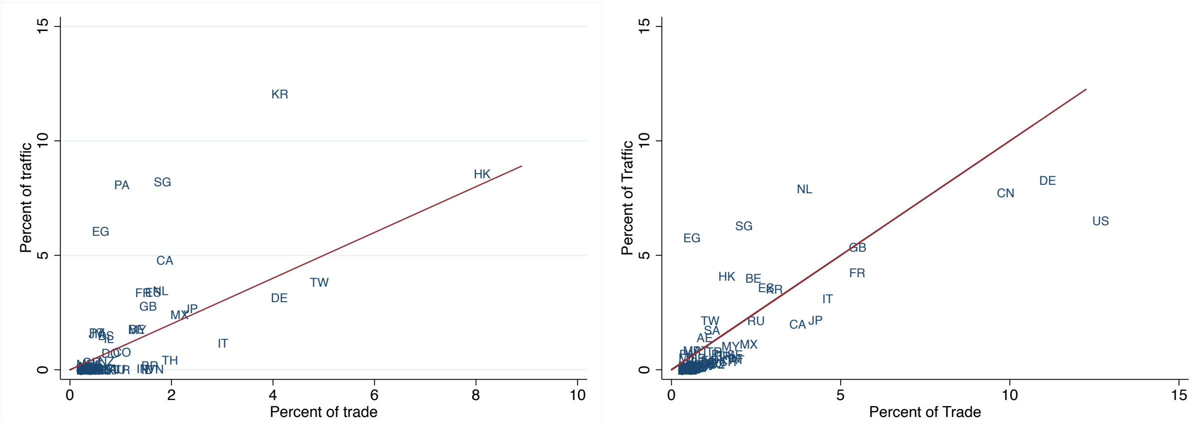
[16]Our set of global entrepôts are: Egypt, Singapore, Netherlands, Hong Kong, Belgium, Taiwan, Spain, Saudi Arabia, South Korea, the United Arab Emirates, Morocco, Panama, Malta, Portugal, and the United Kingdom.

[17]Table A.2 reports the concentration ratios for trade, transshipment, and third-party-country stops, which are high by most standards. The 99-50 ratio is 400 for third-country stops, 480 for transshipment, and 96 for trade. For comparison, the same ratio for employment in the concentrated IT-sector across US cities is 300 (Moretti, 2019).

[18]Section 5 addresses the extent to which exogenous characteristics like geography are responsible for lower costs at, hence higher concentration of shipments through, entrepôts.

**Figure 4:** Concentration of Indirect Shipments

(A) US Microdata: Transit Volume vs Percent Originated (B) Global Data: Percent Transit Volume vs Percent Originated




Notes: Panel (A) uses US microdata to compare, for each country, the share of shipments to the US that originated in a country (x-axis) to the the share that passed through that country (y-axis), weighted by TEU. For readability, China is omitted in Panel (A). Panel (B) replicates Panel (A) using global port of call and trade data with adjustments made for unobserved overland traffic as discussed in Section 5. ■

costs of trade, but by revealed preference should imply lower costs, especially for the spokes of the network which disproportionately choose to send goods indirectly through entrepôts.[19] The goal of our empirical estimation is to measure this trade-off within the context of the full global trading network by finding a set of node-to-node costs which describe the shipping network and is consistent with the indirect trade we observe.

These facts also trace the existence of a size-cost relationship: shipment along high-concentration entrepôts routes appears by revealed preference to be cost-reducing. As with any scale-cost relationship, both directions of causation may be operational. We model the shipping decision in a way which allows for but does not impose a reduced-form scale economy, and in our estimation, identify the causal impact of scale on costs.

# 4 Theoretical Framework

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

[19] While some entrepôts lie along lowest-cost routes, routes stopping at entrepôts are 3-8% longer. This is true even when comparing shipments sent from the same origin, to the same destination, and using the same total number of stops, and comparing total distance travelled as well as distance from port of lading to US destination.

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

## 4.1 Setup

**Consumption and Production** In each country $j$, consumers consume goods $\omega_n \in \Omega_n$ from each $n$ of $N$ industries according to function $U_j = U_j(C_j)$, where $U_j(\cdot)$ 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 $j$.[20] Within each industry and product category, goods are homogeneous and normal.[21]

Goods are 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 $L$, as well as inputs of other goods.[22] Production functions can vary across industries and countries. Cost minimization results in identical production costs among competitive firms within an industry in each country. The 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 country $i$ 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** To sell goods abroad at any destination $j \in J$, a firm producing product $\omega$ in industry $n$ must pay non-transport trade costs $\kappa_{ijn}$ and iceberg transport costs

[20] We allow for the utility function to vary across destinations, and the number of goods in each industry need not be a continuum but can be.

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

[22] The production function is given by $q_{in}(\omega) = f_{in}(z_{in}, L_{in}, Q_{in})$ where $f_{in}(\cdot)$ is a continuous and twice differentiable country-industry-specific production function, $z_{in}$ is the 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.

$\tau_{ijnr}(\omega)$ after optimally choosing the route $r$ between $i$ and $j$ to minimize the shipping costs incurred. Competitive firms in $i$ selling to $j$ price their goods at marginal cost. The observed prices for these products at $j$ are

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

where purchasers of good $\omega$ in industry $n$ at $j$ source the lowest cost supplier globally.

**Shipping** Producers seek to minimize shipping costs, choosing the lowest cost shipping route available. Shipping route $r$ is comprised 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), moving from stop to stop involves iceberg transport costs as well as product- and route-level idiosyncratic cost shocks $\epsilon_{ijnr}(\omega)$.[23] We place minimal structure on these direct leg-level costs $t_{k_{r-1},k_r}(\cdot)$ between locations $k_{r-1}$ and $k_r$, allowing them to be a function of exogenous and endogenous variables:

$$t_{k_{r-1},k_r} = f(\Xi, \varepsilon_{k_{r-1},k_r}) \tag{2}$$

where $\Xi$ is a matrix of endogenous containerized traffic over the entire network and $\varepsilon_{k_{r-1},k_r}$ reflects exogenous transportation cost elements such as distance.

Route-specific idiosyncratic shocks are drawn from the Fréchet distribution such that $F_{ijn}(\epsilon)$, the cumulative distribution function of the idiosyncratic draws is as follows:[24]

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

where shape parameter $\theta > 0$ captures the randomness or dispersion in the choice of routes from $i$ to $j$.[25] Higher $\epsilon_{ijnr}(\omega)$ draws mean industry $n$ has lower costs for route $r$.

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

$$\tau_{ijnr}(\omega) = \frac{1}{\epsilon_{ijnr}(\omega)} \prod_{k=1}^{K_r} t_{k_{r-1},k_r}(\Xi, \varepsilon_{k_{r-1},k_r}) \equiv \frac{1}{\epsilon_{ijnr}(\omega)} \tilde{\tau}_{ijr}, \tag{3}$$

where $\tilde{\tau}_{ijr}$ is the product of all leg-specific costs $t_{k_{r-1},k_r}(\Xi, \varepsilon_{k_r-1,k_r})$ and is common to all products taking route $r$. Product $\omega$ in industry $n$'s realized shipping cost from $i$ to $j$ is that of the transport-cost minimizing route from the set of all routes from $i$ to $j$.[26]

[23]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 shape parameter $\theta$.

[24]This distribution is identical across industries so product-industry subscript $n$ is dropped.

[25]This dispersion assumption is reflected in our microdata (Panel (B) in Figure A.9, Appendix B.4) Almost 70 percent of origin countries have fairly low concentration of routes (HHI less than 1500).

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

The multiplicative functional form for $t_{k_{r-1},k_r}$ in Equation (3) allows for an analytical solution to the routing problem. Section 7 subsequently establishes a tight fit between our estimates and two sets of external data, helping alleviate misspecification concerns.

This structure is consistent with a host of mechanisms, including but not limited to port-level effects and leg-level scale economies.[27] With regards to market power, we do not directly model the decision of shipping firms. Instead, our equilibrium can be considered as an overall industry equilibrium within a Sutton (1991) framework, where larger markets induce more entrants and lower marginal costs, with profits being absorbed by fixed costs.[28] Differences between these mechanisms will not impact the model estimation but will manifest in the interpretation of scale economies and for counterfactual predictions.

## 4.2 Equilibrium

**Route volume** Firms from origin $i$ select the lowest-cost route before consumers in $j$ select the lowest-cost intermediate good supplier across all the origins countries. We observe $\omega$ 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 then 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:

$$\pi_{ijnr\omega} \equiv \Pr\left\{p_{ijnr\omega} \le \min_{i'\in I\setminus i,\ r'\in R_{ij}\setminus r} p_{i'jnr'\omega}\right\} = \frac{\left[c_{in}\kappa_{ijn}\cdot\tilde{\tau}_{ijr}\right]^{-\theta}}{\sum_{i'\in I}\left[(c_{i'n}\kappa_{i'jn})^{-\theta}\cdot\sum_{r'\in R_{i'j}}\tilde{\tau}_{i'jr'}^{-\theta}\right]}. \quad (4)$$

By the law of large numbers, this is also the share of goods sold in $j$ in industry $n$ coming from $i$ and taking route $r$.[29] Introducing auxiliary matrix $A_n = [t_{ijn}^{-\theta}(\Xi, \varepsilon_{ij})]$ where each element is a function of the leg-specific transport cost, we define the expected transport

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

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

[29] Recall that the number of goods in each industry is set so the law of large numbers holds.

cost matrix as

$$\left[\tau_{ijn}\right] \equiv \left[\left(I - A_n\left(\Xi, \varepsilon\right)\right)^{-1}\right]^{\circ(-\theta)}, \tag{5}$$

where $\circ$ is the element-by-element Hadamard power.[30] Substituting the definition of $\tilde{\tau}_{ijr}$ (Equation (3)) into Equation (4) 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 $kl$ as:

$$\pi_{ijn}^{kl} = \left[c_{in}\kappa_{ijn} \cdot \tau_{ikn}\left(\Xi, \varepsilon\right) \cdot t_{kln}\left(\Xi, \varepsilon\right) \cdot \tau_{ljn}\left(\Xi, \varepsilon\right)\right]^{-\theta} \Phi_{jn}^{-1}, \tag{6}$$

where $\Phi_{jn} = \sum_{i'}\left[c_{i'n}\kappa_{i'jn} \cdot \tau_{i'jn}\left(\Xi, \varepsilon\right)\right]^{-\theta}$ is the key distinction from Allen and Arkolakis (2019)—a multilateral resistance term that accounts for average costs, openness, and connectivity of competitors from all other countries $i'$. With optimal route selection and competition on price both accounted for, Equation (6) is the realized and observable share of traffic that flows through leg $kl$ from $i$ to $j$.

Next, the model yields a gravity equation. The sum of products sold in $j$ in industry $n$ from country $i$ equals the share of products sold in $j$ in industry $n$ coming from $i$ and taking route $r$, summed across all $r$ routes:

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

Equations (6) and (7) will jointly generate our estimation equation in Section 5.

Finally, we derive an expression for the share of global shipping passing through $kl$:

$$\pi^{kl} = \sum_n \sum_j \sum_i \pi_{ijn}^{kl} = \sum_n t_{kln}\left(\Xi, \varepsilon\right)^{-\theta} \cdot \sum_j \Theta_{jn}\tau_{ljn}\left(\Xi, \varepsilon\right)^{-\theta} \cdot \frac{\Phi_{kn}}{\Phi_{jn}}, \tag{8}$$

where $\Theta_{jn}$ is $j$'s global consumption share of industry $n$. Because optimal route selection and competition on price are both accounted for, Equation (8) corresponds to the observable shares of all goods passing through leg $kl$, including shipments bound for $l$ and those continuing onward to other destinations. In Section 7, we compare our model-implied leg-level trade flows to those observed in the US microdata. We find high correlations which also hold true for higher levels of aggregation across origins and levels as well.

In Appendix C.2, we show how a change in the leg cost between $k$ and $l$ $(t_{kl}(\Xi, \varepsilon_{kl}))$

[30] The expected transport cost from $i$ to destination $j$ is also $\tau_{ijn} = \gamma^{-1/\theta}\left(\sum_{r \in R_{ij}} \tilde{\tau}_{ijr}^{-\theta}\right)^{-1/\theta}$ where $\gamma$ is the function $\Gamma(t) = \int_0^{\infty} x^{t-1} \exp^{-x} dx$ evaluated at $\left(\left(1+\theta\right)/\theta\right)^{-\theta}$.

can affect trade volumes between an origin $i$ and destination $j$ through the trade network.

**Closing the model** In order to close the model, we require markets to clear for factors and goods as well as the balanced trade condition. Unnecessary for estimation, we defer them to Section 8 when we conduct counterfactuals.

# 5 Estimation

We now show how to link our model to real world data, use the model to recover the trade costs underlying the global shipping network, and estimate a scale elasticity in shipping.

## 5.1 Linking the Model with Data

Using equations (6) and (7) we can calculate the probability of any good traveling through leg $kl$ conditional on being sold from origin $i$ to destination $j$. With the total value of trade between $i$ and $j$ in industry $n$ ($X_{ijn}$), we can express the total volume of traffic between $k$ and $l$ in a given industry $n$ as:

$$\Xi_n^{kl} \equiv \sum_i \sum_j X_{ijn} \cdot \left(\tau_{ikn} t_{kln} \tau_{ljn} {\tau_{ijn}}^{-1}\right)^{-\theta}. \tag{9}$$

Equation (9) is identical to Allen and Arkolakis (2019), despite differences in framework. In particular, expensive trade routes here suffer from Ricardian selection at destination markets, where the route's impact on prices make them less competitive. Yet, this does not impact the trade cost estimation. Intuitively, while Ricardian selection, non-transportation trade costs such as tariffs, and multilateral resistance all reduce total trade, they do not differentially favor one route from an origin $i$ to a destination $j$, instead they reduce traffic flows proportionally along all leg-level $kl$-pairs. Accordingly, conditioning on the observed trade values $X_{ijn}$, the contribution of trade between $i$ and $j$ to the traffic between $k$ and $l$ is invariant to multilateral resistance, tariffs, or technology.

Mapping our model into the data we make one final assumption: for a set of industries $\bar{N}$, trade costs are identical and all trade $\left(X_{\bar{N}} \equiv \sum_{n \in \bar{N}} X_n\right)$ and traffic $\left(\Xi_{\bar{N}} \equiv \sum_{n \in \bar{N}} \Xi_n^{kl}\right)$ are observable. Summing Equation (9) over industries $n \in \bar{N}$ yields:

$$\Xi_{\bar{N}}^{kl} \equiv \sum_i \sum_j X_{ij\bar{N}} \cdot \left(\tau_{ik\bar{N}} t_{kln} \tau_{lj\bar{N}} {\tau_{ij\bar{N}}}^{-1}\right)^{-\theta}. \tag{10}$$

Equation (9) tells us that to accurately measure transport costs, we only need data on trade 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. We implement equation (10) using observed total containerized traffic and trade in containerized industries, where transportation costs are likely similar, and apply it in estimation only to legs where all traffic is observed.

## 5.2 Recovering Scale Elasticities

**The cost–scale relationship** The existence of a scale economy in shipping implies that perturbations to the global shipping network that affect traffic volumes will in turn impact the leg cost matrix estimated in the next section. Such effects must be accounted for in order to correctly estimate counterfactual adjustments.

Using leg-level trade costs from Equations (5) and (10), we consider the regression:

$$\ln(t_{kl}^{\theta} - 1) = \alpha_0 + \alpha_1 \cdot \ln \Xi_{kl} + \alpha_2 \cdot \ln d_{kl} + \varepsilon_{kl}, \tag{11}$$

where $\alpha_0$ is a constant, $\alpha_1$ is the relationship between price and quantity, $\alpha_2 \cdot \ln d_{kl}$ is the coefficient and measure of log sea-distance from $k$ to $l$ respectively. $(t_{kl}^{-\theta} - 1)$ allows us to interpret $\alpha_1$ as the elasticity between cost and traffic volumes to a trade elasticity $\theta$. That is, to interpret results from Equation (11) as elasticities, they must be deflated by trade elasticity $\theta$.

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

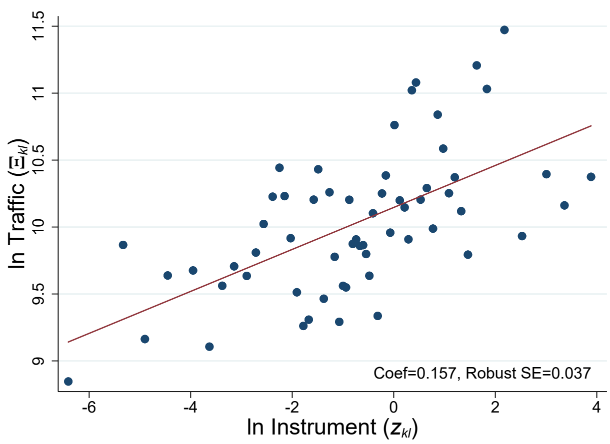
**Geography-Based Instrument** We use the intuition of our model to construct a geography-based instrument for demand. Demand for a given leg will be higher, all else equal, if the leg lies along the most direct route between an origin and a destination. For example, consider routes from origin South Korea to destination the Netherlands. Routes that include the leg China-Singapore are closer to the direct Korea-Netherlands route compared to routes that include the leg China-Australia. As such, more Korea-Netherlands trade should flow through the China-Singapore leg than the China-Australia leg, which would involve a longer detour. Links that are effectively out of the way for most journeys should, all else equal, face lower demand, such as Singapore vs Australia on routes between East Asia and Europe.

Operationalizing this intuition, we relate the direct sea-distance between an origin and a destination to the distance of two legs as part of a three-leg journey, where the omitted middle leg is the object of interest. We calculate the instrument $z_{kl}$ as:

$$z_{kl} = \sum_{i \backslash k,l} Pop_{i,1960} \sum_{j \backslash \{k,l\}} Pop_{j,1960} \frac{d_{ij}^2}{(d_{ik} + d_{lj})^2}, \tag{12}$$

where $d_{ij}$ is the sea distance between origin $i$ and destination $j$, and the square of the relative excess distance between legs $ik$ and $lj$ $(d_{ik} + d_{lj})$ is weighted by the year 1960 population at each origin $i$ and destination $j$, $Pop_{i,1960}$ and $Pop_{j,1960}$.[31] Figure 5 shows the robust first-stage relationship between our instrument and traffic.

**Figure 5:** Residualized Plot of Correlation Between Instrument and Traffic




Notes: The figure shows a binned scatter plot of 1,947 observations of link $kl$ with the natural log of sea distance between $k$ and $l$ is included as a control. The x-axis is the natural log of the instrument $z_{kl}$. The y-axis is the natural log of traffic on leg $kl$. The standard error printed is clustered two ways by nodes $k$ and $l$. ■

For plausible identification, our demand shifter instrument has to be generally uncorrelated with unobserved changes in cost determinants for a particular leg controlling for its sea-distance $(\text{corr}(\varepsilon_{kl}, \ln z_{kl}) = 0)$. Locations that are close in sea distance are also close in land distance and may have easier access to other modes of transportation like road or rail. As a robustness check, we recalculate our instrument in equation (12) in a simplified setting by omitting the shortest 10 percentile distances for each origin $i$ and destination $j$ respectively and find similar results.

As previously noted, the observed scale economy in our setting can be generated by a number of mechanism, including but not limited to internal or external scale economies and market power. These mechanisms may generate different out of sample results, and

[31] 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.

further work should be done to isolate and test for these. In order to accommodate this multitude of mechanisms simultaneously, we implement a model-consistent and agnostic approach in our estimation of scale. Formally, we construct moments $m_1(\alpha) = Z\varepsilon$ based on Equation (11). First, however, we need to recover leg-level trade costs $t_{kl}$.

### 5.3 Recovering Trade Costs

We require two observable objects in order to recover trade costs: trade values and traffic volumes (Equation 10).[32] Our traffic data comes from our global port of call AIS shipping data.[33] We use aggregate trade data from Centre d'études Prospectives et d'Informations Internationales (CEPII) and their BACI international database for 2014, segregating containerized and non-containerized commodities.[34]

In an ideal world, estimation would recover the trade costs that directly rationalize observed bilateral containerized traffic flows—a just identified case. While we directly observe ocean containerized traffic, our data omits movement of containers overland, across and within borders. We overcome this limitation by assuming a functional form that allows for estimation without requiring the direct observation of overland links. We consider the mapping:[35]

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

where the matrix $Y$ is a vector defined as

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

where $\beta_0$ is an intercept, $\beta_1$ considers the sea distance between the nearest principal ports,[36] and $\beta_2$ considers port-to-port traffic. $\beta_3$ and $\beta_4$ consider the total incoming

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

[33]Units for traffic is in TEU. Recall we estimate ship-by-leg TEUs by combining reported ship draught and maximum TEU. This process does not rely on the merged US Customs data.

[34]We use 2014 US Customs data on containerized and non-containerized shipments to construct the share of each HS 4-digit commodity code that is transported by container. All commodities with a containerized share above 80% are labeled as containerized. This procedure shuts down the substitution between containerized and non-containerized transport. In practice we find a bimodal distribution, with some commodities being never containerized (e.g. oil and iron ore) and others always containerized (e.g. washing machines and children's toys). This process is documented in Appendix A.3.

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

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

and outgoing traffic at ports $i$ and $j$ respectively. $\beta_5$ considers the role of the backhaul problem from Wong (2020), where ship capacity is fixed by the shipping direction with the higher demand. Finally, $\beta_6$ is an indicator for a shared land border.[37]

It is crucial to note two things. First, while the equations above posit relationships between observables, our objective at this stage is not the vector $\beta$ of coefficients—which may reflect endogenous variables—but the resulting predictions for $t_{ij}$. Instead, we seek to fully saturate the variation in the data in order to generate the closest empirical prediction for the matrix of trade costs relative to the just-identified case, which yields the model-perfect estimates of trade costs for each link. This allows us to recover the trade costs while remaining agnostic to their underlying determinants, including potential economies of scale as well as possible geographic indicators. Secondly, while the parameters for $\beta$ yield estimates of every trade cost $t_{ij}$, we need not discipline $\beta$ by comparing traffic on every link. This allows us to still recover estimates of $t_{ij}$ although we do not observe within-country traffic as well as between countries traffic that share overland routes.

We create a moment $m_2$ that finds the vector $\beta$ that minimizes the difference between expected traffic, $\hat{\Xi}\left(t(\beta);Y\right)$, and observed traffic $\Xi^{data}$ for countries that do not share a land border:

$$m_2\left(\beta\right) = \left(\hat{\Xi}_{kl}\left(\beta|\mathbf{X},\mathbf{Y},\theta\right)\right) - \left(\Xi_{kl}^{data}\right)$$

where expected traffic is a function of $\beta$, trade elasticity $\theta$, as well as trade data $X$.

As noted, we do not fully observe the traffic flows of containerized goods on geographically contiguous legs, and we do not perform our estimation procedure using traffic data from these legs. Instead, our trade cost estimates, even for overland links, are disciplined by the observed traffic flows of sea-only legs that do not share a land border.

## 5.4 Joint Estimation

We combine our scale estimation and recovery of trade costs into a single stage:

$$m_1\left(\alpha,\beta\right) = Z\varepsilon\left(\alpha,\mathbf{t}(\beta)\right)$$
$$m_2\left(\beta\right) = \left(\hat{\Xi}_{kl}\left(\beta\right)\right) - \left(\Xi_{kl}^{data}\right)$$

We conduct a two-stage GMM procedure, using optimal instrumental variable weights estimation for the first set of moments $m_1$, which accounts for our casual estimates of scale, and the inverse of trade volumes on the second set of moments $m_2$, which rationalizes

[37]We do not estimate within-country trade costs directly due to data constraints and assume that they do not change in the counterfactual.

leg-level trade costs $t_{kl}$ conditional on observable world trade $X$ and traffic flows $\Xi$. We reiterate that we only conduct inference on the parameters $\alpha$. We treat $\beta$ as a set of incidental parameters, important for estimation, but not for inference. The second stage computes an optimal weighting matrix W using the first stage results.

# 6 Results

**Scale Economy** Table 1 reports our instrumented scale elasticity from our scale moments (Equation (11)). For the widely used trade elasicity value of $\theta = 4$ (Simonovska and Waugh, 2014), the interpretation of our causal estimate is that increasing traffic volume on a leg by 1% would reduce costs by 0.06%. As the median journey in our microdata has 3 legs, this translates into a 0.17% decrease in overall origin-destination trade costs.[38] These results lend support to our initial hypothesis that a major role of entrepôts is their facilitation of scale through concentration of shipments.

**Table 1:** GMM Estimation Results

| | (1) |
|---|---|
| | $\ln(c_{kl})$ |
| $\ln(\Xi_{kl})$ | -0.29 |
| | (0.13) |
| $\ln(\mathrm{d}_{kl})$ | 0.57 |
| | (0.03) |
| Constant | 4.24 |
| | (1.45) |

Notes: We conduct a two-stage GMM procedure, first using optimal instrumental variable weights estimation the first set of moments and the inverse of trade volumes on the second set of moments. The second stage computes an optimal weighting matrix W using the first stage results. $ln(c_{kl})$ is the natural log of transportation trade cost on leg $kl$. $ln\Xi_{kl}$ is the natural log of traffic volume on leg $kl$. $ln(d_{kl})$ is the natural log of sea distance between $k$ and $l$ computed using Dijkstra's algorithm. ■

**Link and average bilateral trade costs** Appendix Figure A.11 graphs our resulting matrix of pairwise trade costs. We present the vector $\beta$ estimates in the Appendix Table A.4 as purely predictive parameters, not fundamentals that we can alter in the counterfactuals. Instead, we simply need to know if our $\beta$ estimates can predict containerized traffic that reflects the actual observed traffic volumes. With a full link-level

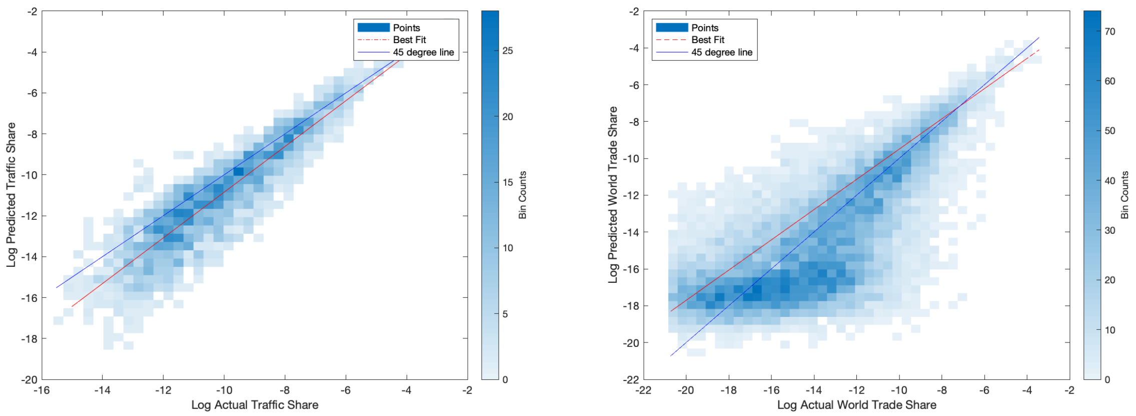
[38] This leg-level elasticity is more modest, but broadly consistent with the strong scale economies from ship size in Cullinane and Khanna (2000), which measure origin-destination elasticites that would compound, on average, three leg level elasticites. Asturias (2020) reports an origin-destination country trade-volume trade-cost elasticity of 0.23 while Skiba (2017) reports an elasticity of 0.26 using product-level import data from Latin America. We search for but do not find evidence of a declining scale elasticity at higher volumes.

trade cost matrix $[t_{kl}]$, we also can generate an average bilateral transport cost between locations $[\tau_{ij}]$. We provide our network-consistent trade-link and origin-destination cost estimates to researchers, and they are available for download on our websites. Appendix Table A.8 compares these network-consistent bilateral trade costs to more commonly used distance measures. Our cost measures have more predictive power than distance alone and both are significant in a combined specification, implying that both measures have distinct predictive power for trade.

**Model Fit** Figure 6 compares our model-predicted traffic and trade values against their observed counterparts in the data. In Panel (A), we compare actual observed global container traffic shares with the our model-predicted shares using our estimated trade costs. We include both a best fit line and a 45 degree line. We fit the data extremely well, with a correlation between the observed and predicted shares (in logs) of 0.97. Panel (B) compares our estimated trade shares to actual observed trade shares, which we do not target.[39] We fit the data well here as well with a correlation (in logs) of 0.73.

**Figure 6:** Model Fit Comparisons

(A) Traffic Volumes (Targeted) (B) Trade Value (Untargeted)




Notes: Panel (A) compares our targeted moment: predicted container traffic volumes from any two ports (y-axis) to the actual container traffic volumes (x-axis, normalized as a share to total world container traffic). Panel (B) compares untargeted aggregate trade shares (x-axis) versus predicted trade shares for containerized traffic (y-axis), where predicted trade shares are computed using the full model described in Section 8. ■

**Alternative Data Definitions** Our estimates of trade costs $t_{ij}$ are at the country-level. Estimation of a port-level cost matrix is possible. However, that requires a global set of sub-national trade data $X$, which is not broadly available. Using port traffic and national trade data, we can impute bilateral port-to-port trade data and run a

[39]To generate trade flows, we close the model using the full setup in Section 8.

version of the estimation above. Results from the port-level estimation are broadly in line with results of our main estimation, with a correlation between weighted port-pair costs and country-pair costs of 0.6. However, due to the speculative assumptions required to generate sub-national trade flows, we view the country-level estimates as more accurate.

# 7 Comparison of Model-Predicted Estimates to Data

We compare our model's results with three separate sets of data external to our estimation. First, we link our results to ship size estimates to highlight a possible scale-economy mechanism. Second, we compare our trade cost estimates with freight rates. Third, we compare our model-predicted traffic flows for US-bound shipments to our US microdata. In each, we find high correlations between our model estimates and these external data.

## 7.1 Symptoms of Scale Economies: Ship Size

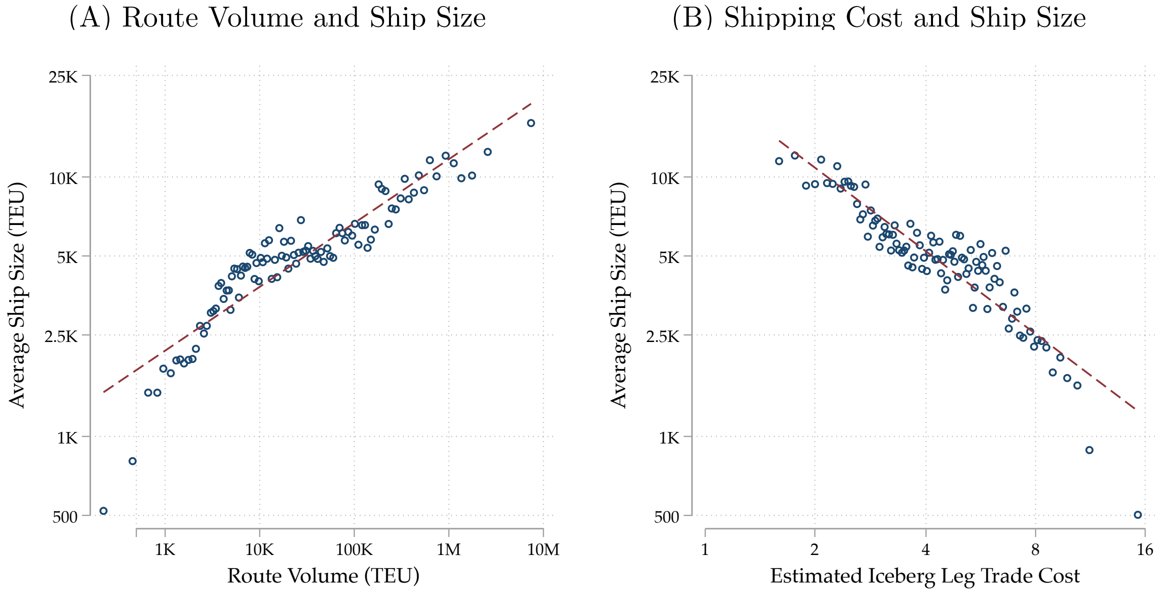
Using our model-guided estimation, we recovered leg-level shipping scale economies. A number of mechanisms can generate the cost reductions that coincide with the concentration of shipments through entrepôts. Internal or external scale economies in shipping and competition among shippers could all generate a negative relationship between volume and costs, as could factors such as port infrastructure.[40] Lacking data to directly test these mechanisms, we turn to one symptom of a scale economy observable in our US microdata which lends further credibility to our results: ship size. Relying on the idea that larger ships enable lower shipping costs (Cullinane and Khanna, 2000), we consider the correlations between ship sizes, trade volumes, and our recovered leg-level trade costs and then investigate the relationship between indirect shipping and ship size.

**Ship Sizes, Traffic Volumes, and Recovered Trade Costs** In Panel (A) of Figure 7, we show a strong positive relationship between the average containership size on a route and the traffic volume on that route, controlling for the distance between origin and destination. Using the route-level containership size measure, we show a strong positive link between ship size and our corresponding recovered trade costs (Panel (B), Figure 7). Routes with more container traffic use larger ships; a 10% increase in route volumes correspond to a 2% increase in ship size. Routes with lower trade costs use larger ships. A 10% decrease in our estimated iceberg trade costs corresponds to 6%

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

increase in ship sizes.[41]

**Figure 7:** Link Between Recovered Trade Costs and Ship Size




Notes: Figures are bin-scatter plots over all observed containership routes, with 100 bins. We control for the log(sea distance) between origin and destination ports, but add variable means back for the plots. Panel (A) plots the relationship between the total containers on a route and the average containership's size on that route (weighted by utilized capacity). Panel (B) plots the relationship between the estimated trade cost $t_{kl}$ with $\theta = 4$ and the average containership's size on that route. Containership size reflects the size of the ship for the average container on that route. ■

**Ship Size and Indirect Trade** Figure 8 further investigates the relationship between entrepôt usage and ship size, ploting ship size (x-axis) against US-bound traffic volume (y-axis) by country of origin, separately for traffic that is routed through an entrepôt and traffic that is not, such that each origin country is associated with two data points. Larger origins transport goods to the US on larger ships. However, shipments from smaller origins routed through entrepôts also arrive on large ships, such that indirect shipping through entrepôts appears to close the ship-size gap for smaller origins.[42]

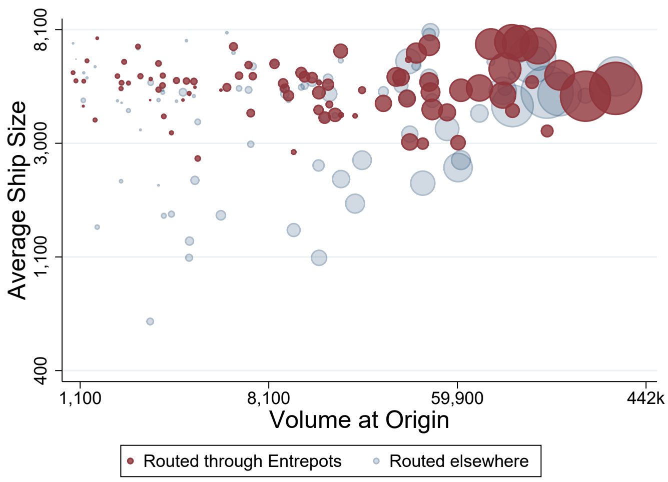
## 7.2 Cost Estimates with Freight Rates Data

Next, we compare our expected trade cost estimates $\tau_{ij}$ at the origin-destination level with container freight rates from Wong (2020). These rates are the costs paid by firms to

[41] Appendix Section D.2 reports shipment-level regressions controlling for origins, destinations, and without route distance controls. Results are similar.

[42] For shipments with the same origin, US destination, and controlling for the total number of stops, shipments stopping at entrepôts arrive on ships that are on average 15% larger. For shipments with the same origin and US destination, shipments sent directly arrive on ships that are on average 8% smaller ships. Further shipment level analysis in Appendix Section D.3 confirms the positive relationships between shipment volume and ship size and robustness to different notions of origin, lading, and transshipment.

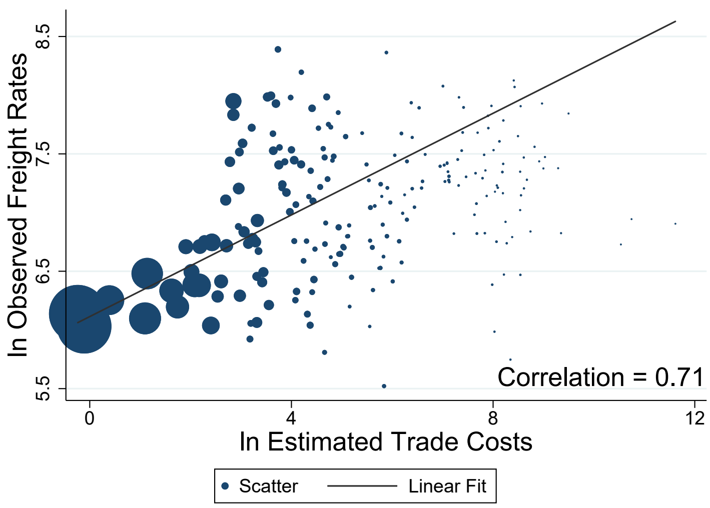
**Figure 8:** Link Between Indirect Trade and Ship Size




Notes: The x-axis shows the total exports from an origin country to the United States. The y-axis shows the average ship size which arrives from an origin country to the United States. Each country is represented by two data points, a blue and a red circle. The red circle indicates the corresponding information for trade from an origin that is routed through an entrepôt while the blue circle is for trade that is not. Circle size denotes shipping volume. Note that trade that is not routed through an entrepôt (blue circle) could either be shipped directly to the United States or shipped via a non-entrepôt. ■

transport a standard full container load between port pairs and include the base ocean rate, fuel surcharge, as well as terminal handling charges at both origin and destination. They are for the largest ports globally which handle more than 1 million containers annually and account for about 73 percent of global container volumes during this time period (World Bank). While we are only comparing a subset of the cost estimates from our entire sample with these freight rates, we find a correlation of 0.71 (Figure 9).

**Figure 9:** Correlation Between Cost Estimates With Actual Freight Rates




Notes: Data points compare origin-destination predicted costs $\tau_{ij}$ to average freight rates from Wong (2020) (210 observations). Circle size are weights for container volumes (TEU). The slope of line is the weighted regression coefficient. ■

## 7.3 Traffic Estimates with US Microdata

In order to assess our model's ability to capture actual shipment journeys and trade indirectness, we compare our model predictions for the paths of US-bound shipment traffic to the actual observed paths in our US microdata. Our estimation, which uses global traffic data rather than the US microdata, delivers predictions for how US-bound shipments travel through the shipping network. Equations (6) and (7) imply

$$\widehat{\pi_{iUS}^{kl}} = [\tau_{ik} t_{kl} \tau_{lj} \tau_{ij}^{-1}]^{-\theta} \tag{13}$$

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

We compare our model-predicted value of Equation (13) to the proportion of goods coming into the US from any origin $i$ on leg $kl$, which we call $\pi_{iUS,Data}^{kl}$, by aggregating shipments using link $kl$ in our microdata. Note that while our microdata is described in Section 2 and used to generate our stylized facts in Section 3, it is not used to estimate our trade costs in Section 5. Column (1) of Table 2 reports the univariate regression outcome between these two measures, weighted by total origin TEU. We find that a significantly positive relationship, with a coefficient of 1 in the confidence interval. Over half of the variation in the observed distribution can be explained using the predicted probabilities.

**Table 2:** Correlation Between Traffic Estimates With Microdata

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | $\widehat{\pi_{iUS}^{kl}}$ | $\widehat{\Xi^{kl}}$ | $\widehat{\pi}_{US}^{l} - \widehat{\pi}_{l,US}$ | $\widehat{\pi_{iUS}^{kl}}$ | $\widehat{\Xi^{kl}}$ | $\widehat{\pi}_{US}^{l} - \widehat{\pi}_{l,US}$ |
| $\pi_{iUS,Data}^{kl}$ | 0.846 | | | 0.872 | | |
| | (0.119) | | | (0.121) | | |
| $\Xi_{Data}^{kl}$ | | 1.225 | | | 1.241 | |
| | | (0.128) | | | (0.126) | |
| $\pi_{US,Data}^{l} - \pi_{l,US,Data}$ | | | 0.945 | | | 0.967 |
| | | | (0.111) | | | (0.115) |
| Observations | 13813 | 652 | 95 | 366010 | 2153 | 186 |
| Data | | | | All | All | All |
| $R^2$ | 0.513 | 0.659 | 0.410 | 0.513 | 0.669 | 0.415 |
| F | 50.54 | 91.60 | 22.91 | 51.75 | 96.88 | 22.53 |

Notes: $\widehat{\pi_{iUS}^{kl}}$ is the model-predicted share of goods from origin $i$ to US destination flowing through leg $k, l$, $\widehat{\Xi_{kl}}$ is the model-predicted total US-bound traffic on a given leg $k, l$, and $\widehat{\pi}_{US}^{l} - \widehat{\pi}_{l,US}$ is the model-predicted total excess US-bound traffic through node $l$. Their corresponding variables observed in the compiled microdata are indicated with subscript "*Data*": $\pi_{iUS,Data}^{kl}$, $\Xi_{kl,Data}$, and $\pi_{US,Data}^{l} - \pi_{l,US,Data}$. Columns (1) to (3) are restricted to nonzero traffic volumes in the US microdata while Columns (4) to (6) include journeys with zero traffic volumes in the US microdata (All Data). Columns (1) and (4) results are robust to tobit specifications which allow for lower and upper censoring limits. Standard errors clustered by origin and destination countries. ∎

Next, summing the predicted probabilities in Equation (13) across all origins $i$, the model delivers a prediction for the total amount of US-bound traffic on a given leg $kl$:

$$\widehat{\Xi^{kl}} = \sum_i X_{iUS} \cdot \widehat{\pi^{kl}_{iUS}}$$

where $X_{iUS}$ is the total trade flow from origin $i$ to the US. Column (2) compares this to the total volume of shipments moving between a given leg in the microdata, which we call $\Xi^{kl}_{Data}$, again finding a positive, significant coefficient with 1 in the confidence interval.

Finally, summing probabilities in Equation (13) across origins $i$ and nodes $k$, we obtain the total traffic through node $l$. Subtracting volume of exports from $l$, we obtain the entrepôt usage of $l$ for US-bound shipments:

$$\widehat{\pi}^l_{US} - \widehat{\pi}_{l,US} \propto \sum_k \widehat{\Xi^{kl}} - X_{l,US} = \sum_k \sum_i X_{iUS} \cdot \widehat{\pi^{kl}_{iUS}} - X_{l,US}$$

Column (3) compares this to its counterpart in the microdata, which we call $\pi^l_{US,Data} - \pi_{l,US,Data}$, finding a positive and significant result with 1 within the confidence interval as well.

In the microdata, a number of legs have zero traffic volumes. However, our model predicts that there should be some small amount of traffic on every leg. Columns (4) through (6), re-run the regressions for each corresponding predicted traffic estimate including legs with zero observed volumes. Accordingly, there is a big jump in the number of observations. Including these links do not significantly change our results because our model predicts extremely low volumes on these legs.

Our paper provides a new set of global trade costs which accounts for the trade network. The tight fit between our estimates—trade costs and traffic—and separate sets of observed data external to our estimation demonstrates that our estimates reflect actual costs and indirect traffic flows in the trade network. Additionally, these results serve as a check to the validity of our modeling approach and the Allen and Arkolakis (2019) framework. Allen and Arkolakis (2019) impute traffic and trade flows within the US highway system for their estimation.[43] Despite the strong structural assumptions made and the limited data requirements, our checks curtail the risk that our estimates are wildly off the mark. In addition to our leg and origin-destination cost estimates, we provide model-implied indirectness measures for ocean shipping as well as resulting market access measures to researchers on our websites.

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

# 8 Counterfactuals

We explore three sets of counterfactuals with the goal of quantifying the global trade and welfare impact of the trade network. Our primary counterfactual highlights the importance of modeling network-consistent trade costs by comparing a reduction in transport costs to and from a country—equivalent to an improvement to its local transportation infrastructure—versus a reduction in traditional non-transportation costs for the same country—like a unilateral tariff liberalization or reduction in information frictions. We then consider two illustrative scenarios. The first considers the role of a negative trade shock, the United Kingdom leaving the European Union. The second studies the effects of global warming decreasing trade costs, with the Arctic opening up to trade between the Pacific and Atlantic Oceans, bypassing the Suez and Panama canals.

To estimate these counterfactuals, we first introduce structural assumptions into our general framework as well as factor and goods market clearing and balanced trade conditions in order to deliver a quantifiable general equilibrium model.

## 8.1 Counterfactual Methodology

**Closing the model** We adopt the Caliendo and Parro (2015) framework. We assume there are three sectors ($N = 3$): containerized tradables $c$, non-containerized tradables $nc$, and nontradables $nt$ ($n \in [c, nc, nt]$), all three of which are used as final goods and intermediates in roundabout production. See Appendix E for full details.

**Equilibrium in changes** Defining the general equilibrium using hat algebra, we consider two sets of changes: (1) link-level transport costs $\hat{t}_{kl} = t'_{kl}/t_{kl}$, which change expected trade costs $\hat{\tau}_{ijn} = \tau'_{ijn}/\tau_{ijn}$, and (2) changes in non-transportation trade costs $\hat{\kappa}_{kl} = \kappa'_{kl}/\kappa_{kl}$. Both alter the endogenous costs of production, price indices, wage levels, trade flows, and welfare.[44] We solve for how wages and prices change $\left\{\hat{w}_i, \hat{P}_i\right\}$ as a function of changes to model primitives, $\left\{\hat{\tau}_{ijn}, \hat{z}_{in}, \hat{\kappa}_{ijn}\right\}$, and compute changes in marginal costs $\hat{c}_{in}$ and trade volumes $\hat{X}_{ij}$.

**Additional Data** 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. We use country-level consumption and production data to compute Cobb-Douglas shares $\eta$ and $\gamma$. This gives

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

us a sample size of 136 countries. We follow the literature and conservatively set $\theta = 4$ (Simonovska and Waugh, 2014).

**Procedure** Changes to transport costs are implemented as changes to link costs $t_{kl}$, which, translated through the model, generate changes in the expected trade cost between every bilateral trading pairs in our data—even those that are not directly connected with each other. Once calculated, these bilateral changes enter isometrically to changes in bilateral non-transportation costs. For analysis which includes the impact of scale, we model a new equilibrium in the short-to-medium run, by following an iterated procedure in Algorithm 1 in Appendix F.1. In this procedure, we start at today's equilibrium and allow all shippers to optimize their transportation patterns. We then recalculate trade costs at new volumes according to Equation (11). We iterate, allowing re-optimization until a new stable equilibrium is reached. There may be alternative equilibria, however we focus on the unique equilibrium from our current starting point—the world today.[45]

## 8.2 The Global Impact of Local Infrastructure Improvements

**Overview** We consider the role of the shipping network in international trade and the specific importance of entrepôts in that network. We run two types of counterfactuals. For all countries, we consider the impact of transportation infrastructure investment in the form of a 1% reduction in transportation costs ($t_{kl}$) to and from a *targeted* country. We contrast this with a 1% reduction in non-transportation trade costs ($\kappa_{ij}$) to and from the targeted country, such as a unilateral tariff reduction or reduction in information frictions. For each type of counterfactual, we evaluate two cases—equilibrium changes with and without accounting for the endogenous impact of scale economies on transport costs throughout the shipping network. Reductions in $\kappa_{ij}$ without scale effects consider changes in a manner which ignores the shipping network, while the other three cases involve exogenous and/or endogenous changes to the shipping network. In each of these 4 cases, we consider welfare and bilateral trade changes to the targeted country as well as to all other *impacted* countries, and focus specifically on differences between entrepôts and non-entrepôts. With 136 targeted countries and 4 cases, we have 544 counterfactuals.

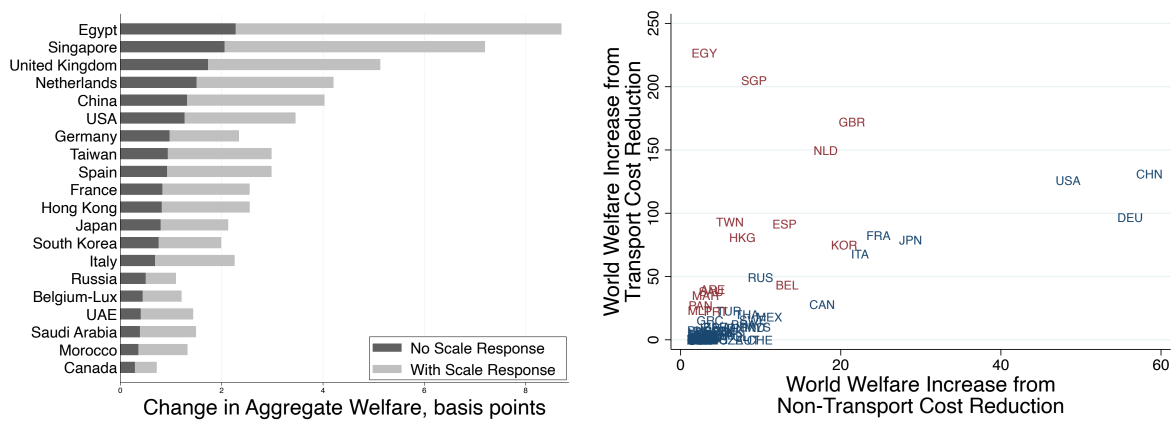
[45] Kucheryavyy, Lyn and Rodríguez-Clare (2019) establishes a common mathematical structure that characterizes the unique equilibrium in multi-industry gravity trade models with industry-level external economies of scale. Their structure requires that the product of the trade and scale elasticities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

**Which Countries are Pivotal to the Trade Network?** Our general equilibrium model yields a convenient metric for how pivotal a country or node is within the trade network: the impact of changes at the country on global welfare excluding a country's own. Pivotal locations are those which generate the largest adjustments throughout the network. Panel (A) in Figure 10 lists the global welfare impact of infrastructure improvements at the 20 most pivotal nodes in the network excluding countries' own welfare change, for both cases with and without scale responses. Our 15 entrepôts dominate this list. Egypt tops it, evocative of the strain in global supply chains when the Suez Canal was blocked in March 2021 (WSJ, FT, AP). Scale economies' impact on the transportation network (overlaid grey bars) further augment the differential impact of entrepôts.[46] Infrastructure investments at entrepôts generate on average 10 times the global welfare impact relative to investment elsewhere.[47]

**Figure 10:** Most Pivotal Countries in the Network: Change in Global Welfare

(A) Transportation Cost Reductions: Highest Global Welfare Changes

(B) Transportation Costs Reductions vs Non-Transportation Cost Reductions




Notes: Panel (A) shows absolute values for aggregate net change in global welfare after infrastructure investment in the targeted country, excluding the country's own welfare change, for the 20 countries with the largest global impact calculated without scale economies. Overlaid grey bars represent welfare changes allowing for the network's endogenous response to scale economies. Panel (B) compares, for each country, the change in world welfare, excluding the country's own welfare, from a 1% decrease in non-transportation costs (X-axis) vs a 1% decrease in transportation costs (Y-axis). Markers are ISO Country codes. Entrepôts are in red. ■

**When Does Accounting for the Transportation Network Matter?** Panel (B) of Figure 10 plots the average welfare impact, excluding the targeted country's own

[46] Panel (A) of Appendix Figure A.14 repeats the exercise for cases with non-transportation cost reductions, finding that the top 20 list is dominated by the largest economies instead.

[47] Appendix Tables A.9 and A.10 examines raw and conditional mean differential impact of targeting entrepôts.

welfare change, of a transportation cost reduction (as in Panel (A)) against the same for non-transportation trade costs. While, driven by gravity, there is a strong overall relationship between the two counterfactuals, the average difference is roughly an order of magnitude (100-160 log points): the effects of one type of counterfactuals will be a poor predictor of the other for *any given country*. For entrepôts, (red in Panel (B)), the 1-to-1 relationship is violated. At the extreme, while the effect of non-transportation cost reductions in Egypt has a similar global welfare effect to that of Colombia, Egypt's impact is larger than that of the US in the transportation cost reduction exercise.[48] The pivotal nature of the entrepôts are specific to their role in the trade network.

Ignoring the trade network impacts of policy rolls the quantitatively large network impacts into the effects of non-transportation cost changes. On the one hand, the impact from any one individual trade cost change will be highly non-predictive. On the other hand, this may not qualitatively impact analysis at the spokes of the network—those origins or destinations which do not significantly participate in trade as third countries—but substantially obfuscates the role of entrepôts in trade.

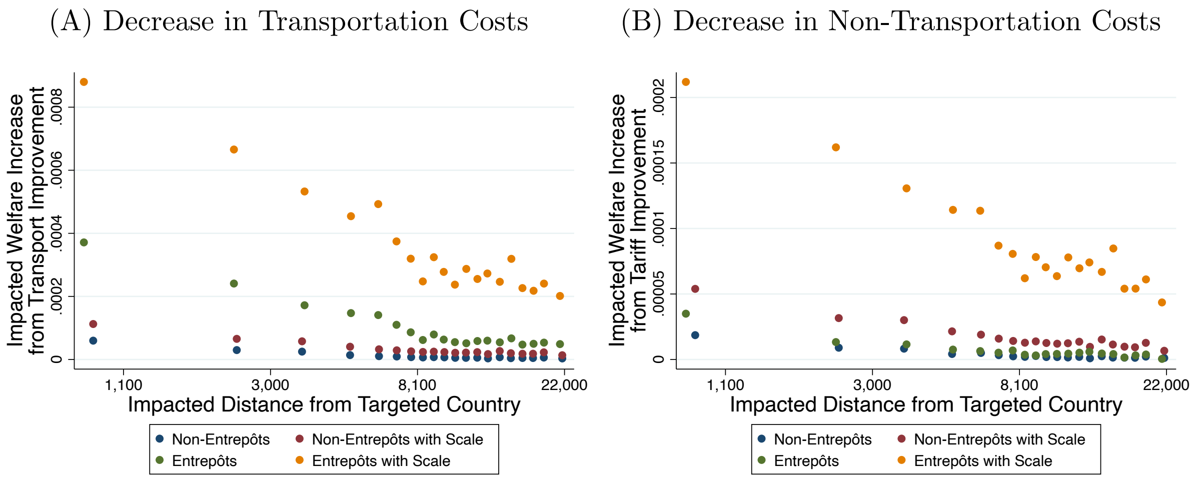
**The impact of entrepôts are localized** To account for the differential impacts of entrepôts, we drill down to one particular margin at which the impact appears most distinct: locally. Figure 11 is a binned scatter plot considering the welfare effects on the impacted country (y-axis) relative to its distance from the targeted country (x-axis), adjusting for the impacted country fixed effects. Nearly overlapping blue and green dots in Figure 11 Panel (B) show a nearly identical distance gradient for non-entrepôts and entrepôts respectively for counterfactual non-transportation cost reductions without scale economies. The blue and green dots in Figure 11 Panel (A) show the overall larger impact of infrastructure investments at entrepôts is relatively more localized—decaying at 5 times the rate. Scale economies amplify the localization, with orange dots decaying at 7-times the rate compared to red.[49]

**Scale economies concentrate gains to entrepôts** Finally, we turn our attention to how these cost reductions differentially affect the impacted countries when they are entrepôts versus non-entrepôts. Figure 12 plots the differential welfare gains to entrepôts

[48] Panel (B) Appendix Figure A.14 finds similar results comparing non-transportation cost reductions with and without an endogenous scale response. Country-pair bilateral trade results are qualitatively and quantitatively similar.

[49] The orange dots in Panel (B) which include the endogenous scale response through the transportation network, echo these results.

**Figure 11:** Spatial Decay of Benefits By Entrepôt Status



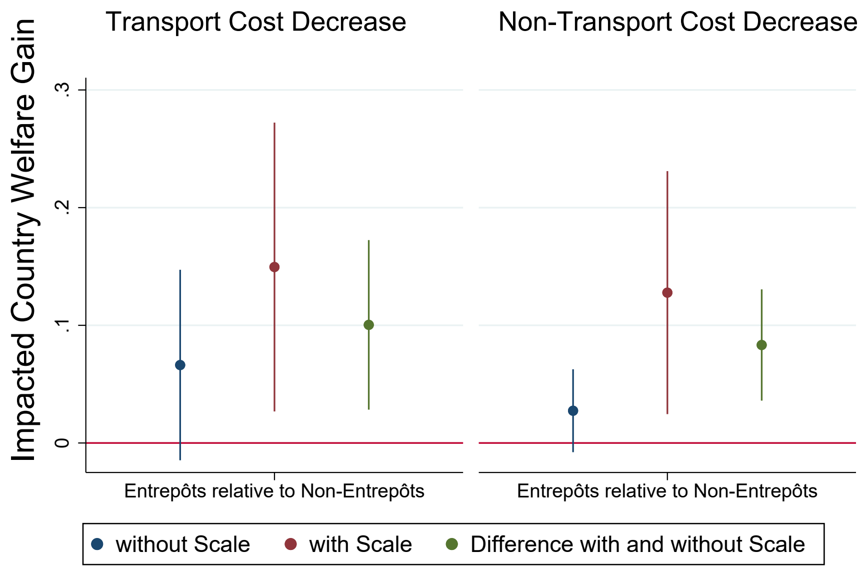

Notes: Panel (A) shows binned scatter residual for welfare effects on impacted countries of transportation infrastructure in targeted countries vs distance between the targeted and impacted countries. Blue and red dots are the no-scale and scale cases for counterfactuals where targeted countries are not entrepôts, respectively. Green and orange dots are no-scale and scale cases, respectively, for counterfactuals where the targeted countries are entrepôts. Panel (B) presents the same for reductions in non-transportation trade costs. ■

relative to non-entrepôts, as impacted countries, controlling for impacted country size, distance between targeted and impacted countries, as well as targeted country fixed effects. Without scale economies, we find that the welfare gains for both entrepôts and non-entrepôts are not significantly different (in blue). However, the differential benefits *to* entrepôts is significant and large when allowing for scale economies (in red). Scale economies disproportionately accrue gains to entrepôts as impacted countries. The coefficient on the entrepôt dummy is 0.15 (SE of 0.06) and 0.13 (SE of 0.05) for transportation and non-transportation counterfactuals, respectively. The pairwise difference between the two cases (in green) is statistically significant. These results—that scale economies in transportation concentrate gains locally at and around hubs—highlight scale economies in transportation as a source of agglomeration.

## 8.3 Brexit

Further illustrating the trade network consequences of bilateral trade cost changes, we study the effects of Brexit—as a 5% increase in trade costs for goods that originate or are destined for the UK. We assume these increases will not be charged to goods that temporarily stop or are transshipped at British ports. For example, Irish exports destined for Britain will face higher tariff costs, while Irish exports destined for the United States will not—even if those goods stop in Felixstowe en route.

**Figure 12:** Differential Welfare Gains of Impacted Countries by Entrepôts Status




Notes: Figure plots the coefficients (dots) and confidence intervals (lines) for indicators for entrepôt status from a country-pair level regression of impacted countries' log percent welfare gains from a transportation cost reduction or an infrastructure improvement (left panel) or non-transportation trade cost reduction (right panel) at targeted countries. Regressions control for impacted country GDP, distance to targeted country, and targeted country fixed effects. Standard errors are clustered by the targeted country. The blue dots represent the welfare gains for cases without scale economies. The red dots represent the welfare gains allowing for the network's endogenous response to scale economies. The green dots use the difference in logs between the two cases on the left-hand side. ■

We model two cases: first without, then with the impact of scale on the trade network. In our first case, as in a traditional model, outcomes are only affected through changes in trade with the UK or multilateral resistance. However, with scale economies, the decrease in UK trade will raise trade costs of neighboring countries through the transportation network. Lower trade volumes lead to increased transport costs, not only for the UK, but also countries that use the UK as an entrepôt. Irish exports to the US will now be more costly, as they will either pay the increased costs of travelling through Britain, use an alternative entrepôt, or take a low-volume, more costly direct trip.

Panel (B) of Table 3 reports aggregate effects. The direct effect decreases global welfare by 2.3 basis points. The introduction of scale economies leads to a more than 4 times decrease of 10 basis points. Trade volumes follow a similar pattern. Figure 13 highlights the distributional effects in terms of welfare (see Appendix Figure A.16 for trade volumes). Scale economies amplify the Brexit impact, especially for Europe. Notably, the impact of scale is not well-predicted by the non-scale case. We document significant negative welfare impacts on Ireland, Iceland and other Nordic countries, many of which rely on UK feeder routes to get their goods to large vessels that ply transoceanic

**Table 3:** Aggregate Counterfactual Outcomes, Basis Points

| | Welfare | | | | Container Trade | | | |
|---|---|---|---|---|---|---|---|---|
| | $\Delta\kappa_{kl}$ | $\Delta\kappa_{kl}$ Scale | $\Delta t_{kl}$ | $\Delta t_{kl}$ Scale | $\Delta\kappa_{kl}$ | $\Delta\kappa_{kl}$ Scale | $\Delta t_{kl}$ | $\Delta t_{kl}$ Scale |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| | (A): Local Trade Cost Reductions | | | | | | | |
| Global Changes | | | | | | | | |
| Mean | 0.08 | 0.26 | 0.20 | 0.58 | 0.87 | 2.93 | 2.31 | 6.62 |
| Standard Deviation | (0.20) | (0.59) | (0.47) | (1.41) | (2.22) | (6.74) | (5.31) | (16.00) |
| | (B): Brexit | | | | | | | |
| Global Changes | -2.3 | -10.0 | | | -24.5 | -112.7 | | |
| | (C): Opening of the Arctic Passage | | | | | | | |
| Global Changes | 1.5 | | 3.3 | 8.9 | 17.4 | | 38.2 | 101.8 |

Notes: Columns (1)-(4) present aggregate welfare changes. Columns (5)-(8) present changes to aggregate container trade. Odd columns correspond to cases where no scale economy feedback loops are allowed. Even columns present results allowing for scale economy feedback. Panel (A) reports results for our first counterfactual, transportation and non-transportation cost declines for each of 136 countries. Columns (1), (2), (5), and (6) present results for cases where non-transportation trade costs are reduced. Columns (3), (4), (7), and (8) present results for cases where transportation costs are reduced (infrastructure improvements). Panel (B) presents results for Brexit, a 5% increase in non-transportation trade costs $\kappa_{ij}$ between the UK and its trading partners. Case 1 of Brexit corresponds to Columns (1) and (5), while Case 2 corresponds to Columns (2) and (6). Panel (C) presents results for the Arctic Passage counterfactual. Case 1, naive impacts on origin-destination pairs, corresponds to Columns (1) and (5). Case 2, allowing for network trade, corresponds to Columns (3) and (7). Case 3, adding in the impact of the scale, correspond to Columns (4) and (8). ■

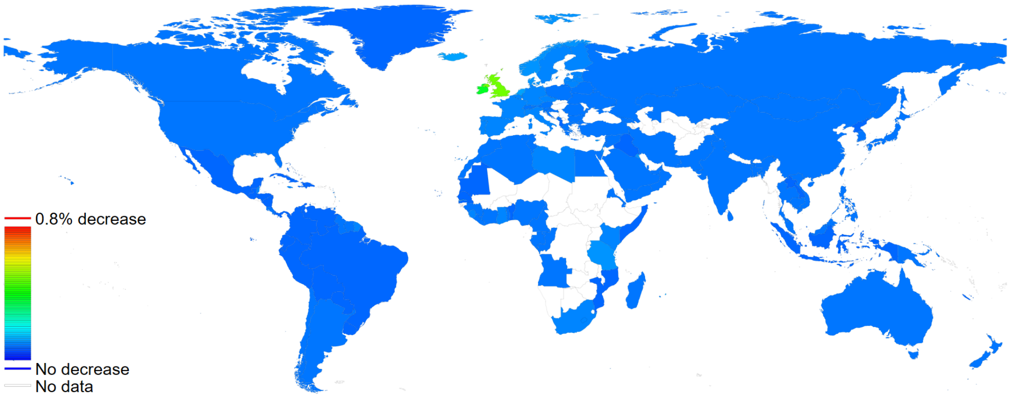
trade with Shanghai and New York (Table A.15). Its trade not easily routed through alternatives, Ireland is most affected outside of the UK (as recognized in FT 2020).

## 8.4 Opening the Arctic Passage

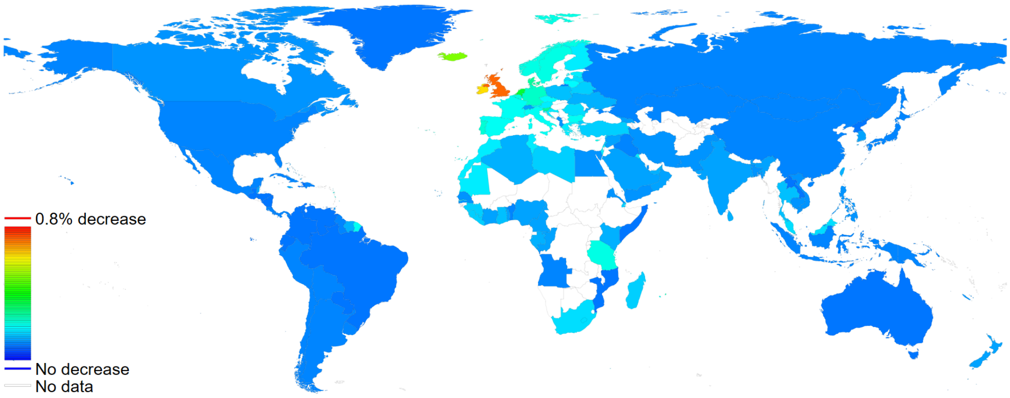
To consider the role of the trade network in reflecting physical trade route changes, we model the opening of the once-fabled Northeast and Northwest Passages through the Arctic Ocean between North America, Northern Europe and East Asia as a viable shipping route due to global warming. For example, a 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 (Economist, 2018). For every $kl$ pair, we compute the difference in sea distance using Dijkstra's algorithm between world maps with and without arctic ice caps (Appendix A.2). Panel (A) of Figure 14 compares the top 150 existing shipping routes today and shortest ocean-going distance of these routes after the Arctic sea passage is viable. New routes going through the Arctic passage are in red,

**Figure 13:** Welfare Changes - Brexit

(A) Tariff Change, No Network Scale Effects




(B) Full Trade Network Effects and Scale Economies




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

non-changing routes are in brown, and abandoned routes are in blue.

We compare three different cases. First, we consider a network-naive case where we only allow for changes in origin-destination trade costs between country pairs for which the direct bilateral distance decreased.[50] Second, we lower $t_{kl}$ for all observed links with positive traffic according to $\alpha_2$ in Equation (11) calculating new distances with the option of traveling through the Arctic Passage. Here, even countries that do not ship directly to each other—e.g. China and Ukraine—experience changes in expected transport trade costs.[51] Third, we repeat the second case accounting for the impact of scale: as trade costs change, trade volumes change, reducing trade costs further.

Column (1) of Table 3 Panel (C) shows that the direct effects of the Arctic Passage are positive, with aggregate welfare increasing 1.5 basis points, and container trade volumes

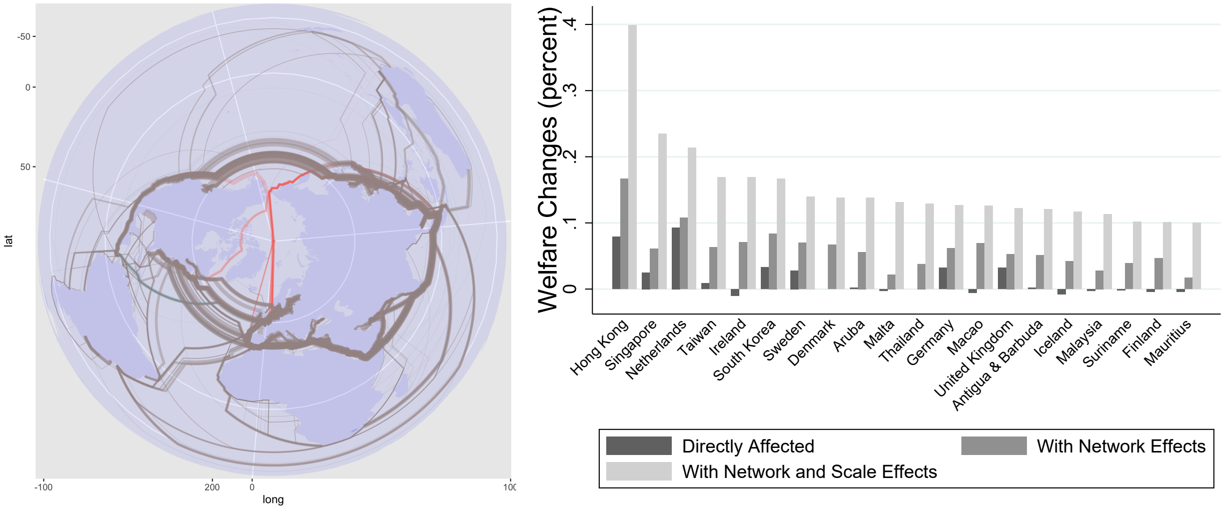
[50]We do not allow third-party-countries to take advantage of these reductions.

[51]For countries affected in cases one and two, the magnitude of changes are mechanically identical.

**Figure 14:** The Opening of the Artic Passage

(A) Shipping Routes: Before and After

(B) Welfare Changes for Top 20 Countries




Notes: The red lines in Panel (A) indicate counterfactual shipping. Blue lines in indicate existing shipping. Their overlap is brown. Route width reflects the number of containers (TEU). Panel (B) shows the percent change in welfare of the simulated opening of the Nordic Passage for the 20 countries with the largest welfare changes. The first bar reflects only the trade cost changes on routes that are directly affected from the opening. The second bar allows for the trade costs to affect indirect trade with network effects while the third bar allows for the endogenous response to scale economies. ■

increasing 17 basis points. Case two, accounting for the full trade network impact of the passage including indirect shipping, doubles the aggregate welfare effect to 3.3 basis points and increases worldwide container volumes by 38 basis points. Allowing for scale economies triples and doubles the welfare and trade impact relative to the network results (9 basis points welfare gain and 102 basis points increase in trade).

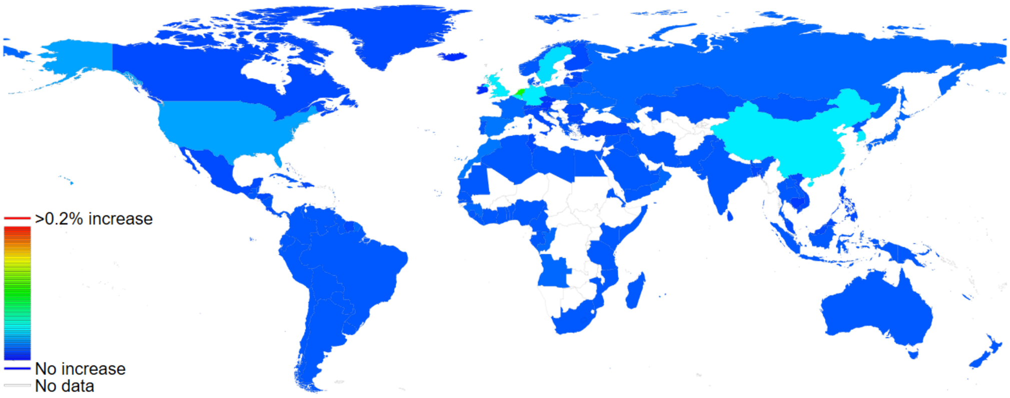
Panel (B) in Figure 14 plots the top 20 most impacted countries, showing gains are particularly pronounced in East Asian entrepôts like Hong Kong and Singapore which disproportionately benefit from the scale economy. Scandinavian countries also gain due to their geography. Denmark and Finland, which in the baseline first case have zero or a small trade diversion impact, gain due to the ability to leverage the trade network and scale response from the opening.

Figure 15 show changes in the relative wage-adjusted price index (interpreted as national welfare, if we omit the costs of climate change) across the three cases.[52] In the baseline scenario in Panel (A), we see increases from trade between countries that are along the Northeast passage, and spillover impacts at countries not directly impacted—reflecting classic multilateral resistance and cascading effects from value chains. Some countries see small trade diversion effects (Panel (B), Figure 14). Figure 15 Panel (B)

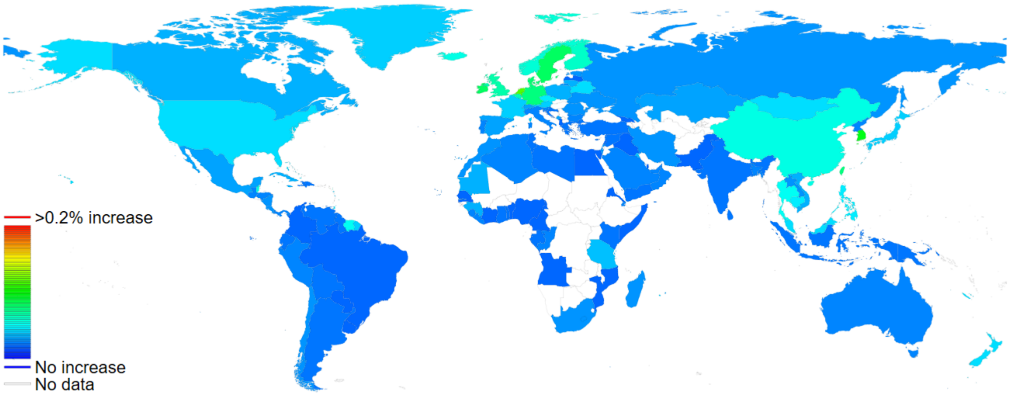
[52] Appendix Figure A.18 shows related changes in country-by-country containerized exports.

**Figure 15:** Welfare Changes - Arctic Passage

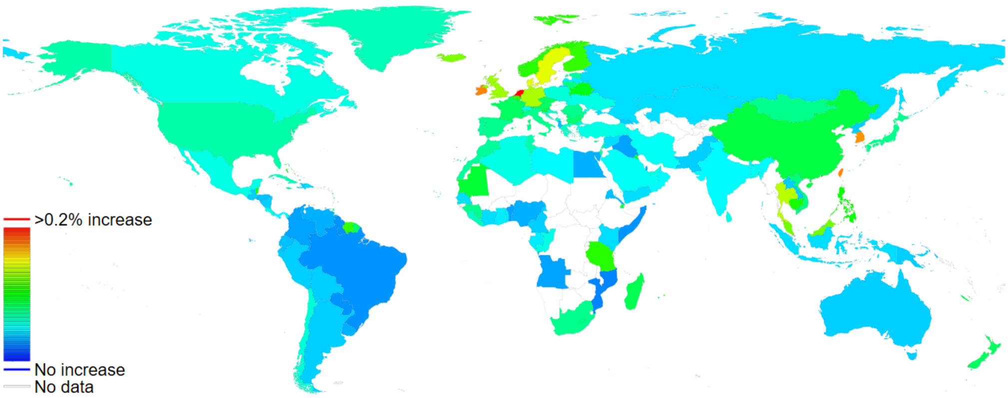
(A) Only Directly Affected Routes




(B) Full Trade Network Effects




(C) Full Trade Network Effects and Scale Economies




Notes: Plots show the percent change in welfare (the relative price index). Darker reds reflects a greater increase and blue represents no change. Omitted countries are white. Panel (A) reflects changes only allowing trade costs to decrease on routes whose distance is directly reduced to the Arctic Passage. Panel (B) reflects changes allowing all countries to indirectly access the Arctic Passage through the trade network. Panel (C) allows for the network's endogenous response to scale economies. ■

shows how, through indirect trade, the benefits of the passage pass on to nearby countries not directly impacted. In Panel (C), scale economies amplify these effects.

# 9 Conclusion

This paper studies entrepôts, the trade network they form, and their impact on international trade. We characterize the global container shipping network as a hub-and-spoke system by documenting that the majority of trade is indirect and flows from origins to destinations through entrepôts (hubs). To rationalize these novel and salient facts, we develop a general equilibrium model of world trade with endogenous trade costs and entrepôts, estimating both the underlying trade costs on all routes, and scale economies. We quantify the impact of the trade network on global trade and welfare, highlighting how changes at nodes operate through the network, entrepôts, and scale economies to create widespread impacts.

While the focus of this paper is on the general equilibrium effects of the trade network, there are two aspects that lend themselves to further study. First, while we are singularly focused on containerized shipping because containerized trade accounts for the majority of global seaborne trade, the hub and spoke network is not specific to just containerized trade (Rodrigue, Comtois and Slack, 2013). Such networks are also prevalent in freight services like UPS or DHL in addition to air transport. Second, our estimates of scale economies are agnostic to underlying mechanisms. Future work should especially be done to consider mechanisms the roles of fixed costs in enabling the scale economies in containerized shipping, such as the costs incurred by potential oligopolies in setting shipping networks and the endogenous creation of firm-specific hub-and-spoke networks.[53] While sector-specific research has been done on these networks, fruitful research should take into consideration a tractable general equilibrium framework to be able to quantify welfare effects.

# References

**Alder, Simon.** 2015. "Chinese roads in India: The effect of transport infrastructure on economic development." Working Paper.

**Allen, Treb, and Costas Arkolakis.** 2014. "Trade and the Topography of the Spatial Economy." *The Quarterly Journal of Economics*, 129(3): 1085–1140.

**Allen, Treb, and Costas Arkolakis.** 2019. "The welfare effects of transportation infrastructure improvements." National Bureau of Economic Research.

**Allen, Treb, and Dave Donaldson.** 2018. "The geography of path dependence." *Working Paper.*

[53] In particular, we can account for leg-level monopolies and variable markups, but we cannot account for within-firm spillovers in sea route selection.

**Anderson, James E, Mykyta Vesselovsky, and Yoto V Yotov.** 2016. "Gravity with scale effects." *Journal of International Economics*, 100: 174–193.

**Asturias, Jose.** 2020. "Endogenous Transportation Costs." *European Economic Review.*

**Barjamovic, Gojko, Thomas Chaney, Kerem Coşar, and Ali Hortaçsu.** 2019. "Trade, merchants, and the lost cities of the bronze age." *The Quarterly Journal of Economics*, 134(3): 1455–1503.

**Bartelme, Dominick, Arnaud Costinot, Dave Donaldson, and Andres Rodriguez-Clare.** 2019. "The Textbook Case for Industrial Policy: Theory Meets Data." Mimeo.

**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.

**Brancaccio, Giulia, Myrto Kalouptsidi, and Theodore Papageorgiou.** 2017. "Geography, Search Frictions and Endogenous Trade Costs." National Bureau of Economic Research.

**Caliendo, Lorenzo, and Fernando Parro.** 2015. "Estimates of the Trade and Welfare Effects of NAFTA." *The Review of Economic Studies*, 82(1): 1–44.

**Coşar, A Kerem, and Banu Demir.** 2018. "Shipping inside the box: Containerization and trade." *Journal of International Economics*, 114: 331–345.

**Cullinane, Kevin, and Mahim Khanna.** 2000. "Economies of scale in large containerships: optimal size and geographical implications." *Journal of transport geography*, 8(3): 181–195.

**Ducruet, César, Réka Juhász, Dávid Krisztián Nagy, and Claudia Steinwender.** 2019. "All aboard: The aggregate effects of port development." Working paper.

**Eaton, Jonathan, and Samuel Kortum.** 2002. "Technology, geography, and trade." *Econometrica*, 70(5): 1741–1779.

**Fajgelbaum, Pablo D, and Edouard Schaal.** 2017. "Optimal transport networks in spatial equilibrium." National Bureau of Economic Research.

**Grant, Matthew, and Meredith Startz.** 2020. "Cutting Out the Middleman: The Structure of Chains of Intermediation." Working Paper.

**Head, Keith, and Thierry Mayer.** 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." *Handbook of International Economics*, 4: 131.

**Heiland, Inga, Andreas Moxnes, Karen Helene Ulltveit-Moe, and Yuan Zi.** 2019. "Trade From Space: Shipping Networks and The Global Implications of Local Shocks." *CEPR.*

**Holmes, Thomas J, and Ethan Singer.** 2018. "Indivisibilities in Distribution." National Bureau of Economic Research.

**Hummels, David, Volodymyr Lugovskyy, and Alexandre Skiba.** 2009. "The trade reducing effects of market power in international shipping." *Journal of Development Economics*, 89(1): 84–97.

**International Maritime Authority.** 2003. "AIS Guidelines."

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

**Kucheryavyy, Konstantin, Gary Lyn, and Andrés Rodríguez-Clare.** 2019. "Grounded by gravity: A well-behaved trade model with external economies." Working Paper.

**Lashkaripour, Ahmad, and Vova Lugovskyy.** 2019. "Scale economies and the structure of trade and industrial policy." mimeo Indiana University.

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

**Moretti, Enrico.** 2019. "The effect of high-tech clusters on the productivity of top inventors." National Bureau of Economic Research.

**Redding, Stephen J, and Matthew A Turner.** 2015. "Transportation costs and the spatial organization of economic activity." *Handbook of regional and urban economics*, 5: 1339–1398.

**Rodrigue, Jean-Paul, Claude Comtois, and Brian Slack.** 2013. *The geography of transport systems.* Routledge.

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

**Schott, Peter K.** 2008. "The relative sophistication of Chinese exports." *Economic policy*, 23(53): 6–49.

**Simonovska, Ina, and Michael E Waugh.** 2014. "The elasticity of trade: Estimates and evidence." *Journal of international Economics*, 92(1): 34–50.

**Skiba, A.** 2017. "Regional economies of scale and regional welfare." *Review of International Economics, Forthcoming.*

**Sutton, John.** 1991. *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration.* MIT press.

**Wang, Chengjin, and Jiaoe Wang.** 2011. "Spatial pattern of the global shipping network and its hub-and-spoke system." *Research in Transportation Economics*, 32(1): 54–63.

**Wong, Woan Foong.** 2020. "The Round Trip Effect: Endogenous Transport Costs and International Trade." Working Paper.