NBER WORKING PAPER SERIES

ENTREPÔT: 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, Revised March 2022

We thank Treb Allen, Costas Arkolakis, Panle Jia Barwick, Bruce Blonigen, Johannes Boehm, Mark Colas, Kerem Co ar, Anca Cristea, Meredith Crowley, Carsten Eckel, Stefania Garetto, Matthew Grant, Keith Head, Reka Juhasz, Myrto Kalouptsidi, Steve Matusz, Ezra Ober eld, Nina Pavcnik, Andrés Rodríguez-Clare, Robert Staiger, Meredith Startz, and seminar participants at Stanford University, Princeton University, University of Michigan, University of Virginia, Boston College, the Federal Reserve Board, Ludwig-Maximilians-Universität Munich, Indiana University, University of Michigan, University of Notre Dame, 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, 2021 NBER Summer Institute 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.

Entrepôt: Hubs, Scale, and Trade Costs
Sharat Ganapati, Woan Foong Wong, and Oren Ziv
NBER Working Paper No. 29015
July 2021, Revised March 2022
JEL No. F10,F12,F14

ABSTRACT

We study the global trade network and quantify its trade and welfare impact. We document that the trade network is a hub-and-spoke system where 80% of trade is shipped indirectly, nearly all via entrepôts—major hubs that facilitate trade between many origins and destinations. 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 scale economies in shipping. Network and scale e ects in the trade network propagate local trade cost changes globally. Even when initial trade cost changes are not transportation-related, these endogenous channels alter the magnitude and distribution of welfare impacts, particularly for entrepôts. Counterfactual infrastructure improvements at entrepôts generate ten times the global welfare impact relative to 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 has been the impetus behind many policies aimed at attaining or maintaining entrepôt status (Financial Times, 2015; Reuters, 2016; Wall Street Journal, 2021).

This paper studies entrepôts, the trade network they form, and their impact on international trade. Using novel data on the trade network and developing a quantitative general equilibrium spatial trade model, we 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 and scale economies 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?

We start by constructing a new dataset mapping the journeys containerized shipments take through the global trading network. This microdata allows us to observe 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 beyond the shipment's origin and destination.

Our first contribution is to establish two stylized facts about the global trade network. Our first stylized fact is that the majority of trade—80%—is shipped indirectly. The median shipment stops at two additional countries before reaching its destination. The majority of trade is also transshipped via an additional country before its destination. This indirectness is not incidental—increasing shipping times and distances by 30%.

Our second stylized fact is that indirectness is incredibly concentrated, with over 90% of indirect trade channelled through a small number of entrepôts, establishing a hub-and-spoke network. These facts highlight a trade-off and trace the existence of a potential scale-cost relationship: indirect trade concentrated through entrepôts increases the observable distance and time costs of trade, but by revealed preference it implies lower trade costs, especially for the spokes of the network which disproportionately choose to ship via entrepôts.

In order to rationalize the documented direct and indirect trade through the global trading network, 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 foreign consumers in a generalized Ricardian setting. Low-cost routes can involve indirect shipping through additional countries, and entrepôts endogenously arise where trade costs are lowest. We allow for both scale economies and dis-economies to govern shipping costs on these network links.

Our second contribution is to use our model to estimate a global set of indirect-shipping consistent trade costs and the economies of scale in shipping. Expanding from our microdata to global seaborne container shipping and trade data, our estimation yields trade costs for each link of the global shipping network and a global set of model-consistent origin-destination trade costs that are distinct from typical distance-based costs. 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 model-predicted network flows and microdata on shipment journeys. Our trade cost estimates are publicly available online.

We use a geography-based instrument to identify 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 links have inherently higher traffic because of their geographic position in the network. For example, links that include Singapore are close to the lowest-distance route between many European and Asian countries due to Singapore's location in the Straits of Malacca. For each link, we compute the distance to and from the link relative to the shortest distance between each origin and destination, recovering a weighted average of each link's proximity to global trade. Increasing traffic volume on a link by 1% reduces costs by 0.06%. As the median journey in our microdata has 3 links, a 10% increase in overall origin-destination trade translates into a 0.17% decrease in trade costs.

Our third contribution uses our estimates and model to quantify the impact of the trade network on global trade and welfare, highlighting how trade cost changes at node countries—entrepôts and non-entrepôts—as well as links can have widespread impacts through the network that are subsequently magnified due to scale economies. Our main counterfactual quantifies the trade and welfare benefits of transport infrastructure improvements for each country in our sample. Entrepôts are pivotal to the global trade network: welfare impacts of infrastructure investment are on average 10 times higher at entrepôts than non-entrepôts. Conflating transport and non-transport trade costs impact estimated welfare effects by an order of magnitude.

This is especially true at entrepôts, which differentially concentrate infrastructure improvement benefits locally relative to non-entrepôts. Scale economies in transportation further concentrate these gains locally at and around entrepôts—highlighting that scale economies in transportation act as 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 second counterfactual investigates how non-transportation cost changes at an entrepôt can have widespread impacts beyond the countries that are directly impacted through endogenous adjustments in trade network. We illustrate this by studying the ramifications of worsening trade relations between one hub, the United Kingdom (UK), and its trading partners—Brexit. When only considering the direct impact of increased non-transportation trade costs, Brexit's consequences are largely proportional to a country's direct trade exposure with the UK. When our analysis accounts for the impact of scale economies on the trade network, 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). This illustrates how trade network and scale interactions can lead to distinct distributional outcomes in welfare even when the initial changes are unrelated to transport.

Our last counterfactual evaluates the importance of endogenous trade costs by demonstrating the welfare and trade impacts from the two endogenous mechanisms in our model: (1) network effects—allowing countries to ship indirectly and (2) scale effects—allowing countries to ship indirectly and take advantage of scale economies. To illustrate this, we study the effects of opening up the Arctic Ocean to regular year-round shipping, connecting countries in East Asia and Europe. Allowing for network effects double the welfare relative to a naïve exogenous trade cost case with no network effects and allowing for scale economies triples the welfare relative to the network effects case.

This paper ties two broad literatures 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 within a class of gravity models (Head and Mayer, 2014; Allen and Arkolakis, 2019).

With regards to the technologies underpinning trade, we make two contributions. First, 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).[1] We demonstrate the global welfare impacts of the container shipping network, which accounts for two-thirds of annual trade moved by sea (World Shipping Council). 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. Allowing for network effects double the welfare relative to a baseline case with no network effects and allowing for the effect of scale economies further triples welfare impacts.[2]

Second, we explore the general equilibrium effects of scale economies 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.[3] The role of localized scale economies in production is well known in general (Allen and Arkolakis, 2014; Allen and Donaldson, 2018), and in the context of trade in particular (Lashkaripour and Lugovskyy, 2019; Bartelme et al., 2019; Kucheryavyy, Lyn and Rodríguez-Clare, 2019). In these settings, scale economies typically generate agglomerations by acting on local productivity. By contrast, in our setting, scale economies generate agglomerations by affecting trade costs. Our counterfactuals find that, by acting on endogenous transport costs over the network, scale economies further concentrate transportation as well as trade and welfare gains at entrepôts.

With respect to the geography and structure of the trade network, we make two contributions. First, we provide empirical evidence for a growing quantitative literature investigating

[1]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.

[2]Allen and Arkolakis (2019) studies the endogeneity of trade costs to traffic congestion on highways. In our ocean shipping context, we find the presence of scale economies instead. Brancaccio, Kalouptsidi and Papageorgiou (2020) studies two aspects of trade cost endogeneity for the network of dry bulk ships carrying homogeneous commodities where all trade is direct: the loading opportunities of dry bulk ships after delivering their cargo relative to the country's trade balance (the equilibrium bargaining position of these ships), and the trade balance of neighboring countries (the network effects). Wong (2020) focuses on the round trip effect from container shipping: a bilateral trade cost endogeneity between a country's imports and exports with a specific trading partner due to containerships serving round trip routes.

[3]Our estimate is about three-quarters of the estimates in Asturias (2020) and Skiba (2017). 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).

the role of trade networks (Allen and Arkolakis, 2019; Fajgelbaum and Schaal, 2020; Redding and Turner, 2015). We provide the first and systematic documentation of indirect trade through the containerized shipping network and the pivotal role that entrepôts play within this network.[4] Our microdata on the movement of shipments through the trade network documents widespread nature of indirect trade and its concentration. In contemporaneous work, Heiland et al. (2019) study the impact of the Panama Canal expansion on global ship movements and use model-based imputations to estimate the physical movement of goods. We further estimate a set of network-consistent trade costs, distinct from and more predictive of trade than distance. Finally, our counterfactuals demonstrate how transport costs behave differently from non-transport costs, particularly at entrepôts. For example, Egypt ranks first in terms of global welfare impacts from infrastructure improvements, while it is not among the top 20 in terms of the welfare impacts from non-transportation trade cost reductions.

Second, our model embeds transportation networks within a class of gravity models (Head and Mayer, 2014). 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 with non-transport barriers to trade and in the presence of unobserved traffic flows. Methodologically, we adopt an approach from the literature on marginal cost estimation (Ackerberg et al., 2007), combining market level data and exogenous instruments with equilibrium assumptions—the indirect routing of trade in our case, or market conduct in the Industrial Organization literature's case—to recover unobserved costs. We establish that our estimates reflect actual costs and indirect flows by comparing our model predictions to external cost estimates, ship sizes, and observed trade routes in our microdata. These results serve as a check to the validity of the Allen and Arkolakis (2019) framework in the trade setting.

2 Data

Our paper uses two distinct sets of data. To establish the stylized facts of the international trade network (Section 3), we use a microdata on the detailed journey of *US-bound* shipments. To estimate *global* trade costs that are network-consistent (Section 5), we use global data on

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

trade and shipping traffic.

To construct the microdata on US shipments, we merge two proprietary data sets: 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.[5]

Our ports of call data captures vessel movements using Automatic Identification System (AIS) transponders.[6] 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.

Our sample covers 4,986 unique container ships with a combined capacity of 18.1 million twenty-foot equivalent container 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. We do not observe containers being loaded or unloaded. 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

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

[6]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: Global Network of Ships, Ports of Call Data

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). ■

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).[7] While the shipments' exact journey between origin and the first stop (the port 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.

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

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 between. The average container entering the US stops at around two *third-party-countries* who are neither the origin nor destination.[8] 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.[9] 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.[10] 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)).

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? 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. As an example, goods transiting the Straits of Malacca can perhaps stop at Singapore since it is "on-the-way." However, the significant additional distance and time incurred by indirect travel relative to the direct path, 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

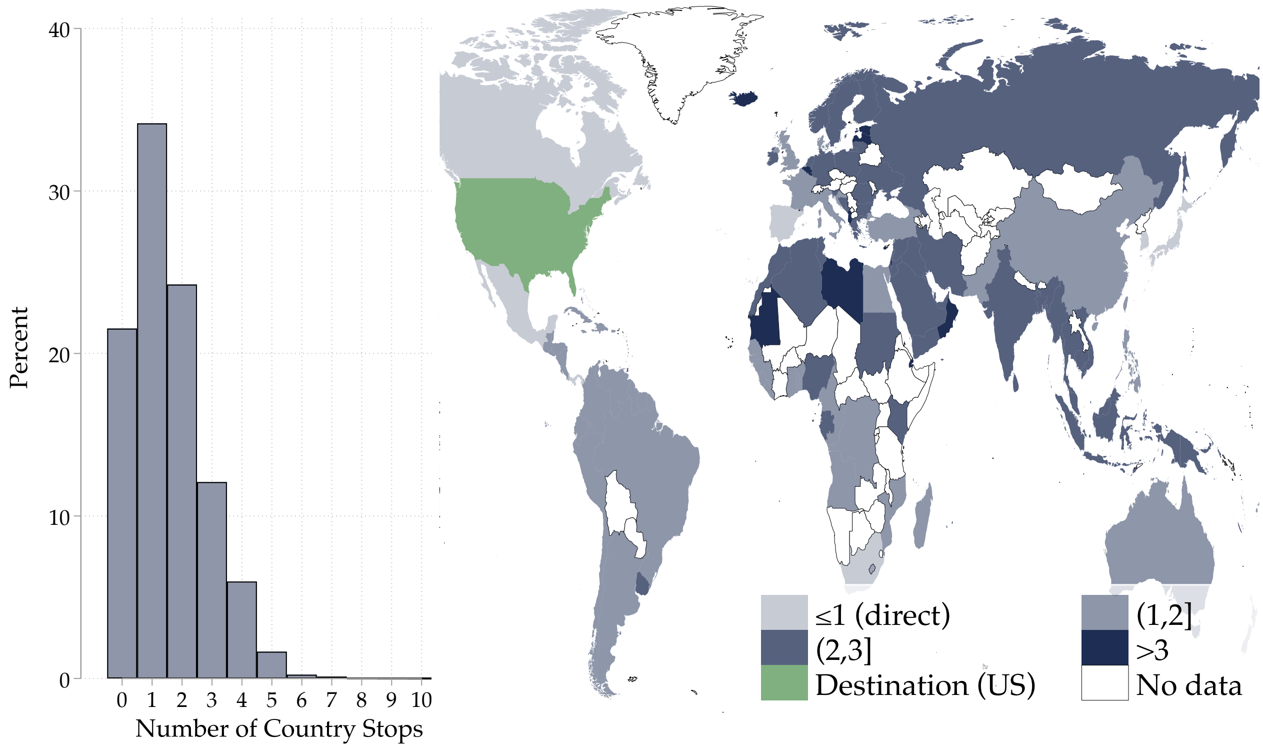
[8]Mean of 1.5 and s.d. of 1.3. Landlocked countries are excluded. The average number of port stops is higher (Figure A.3, mean of 4.6 and standard deviation of 3.5). This result is robust for shipment weight and value (Figure A.4). Multiple stops at the same third-party country are not counted.

[9]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.

[10]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



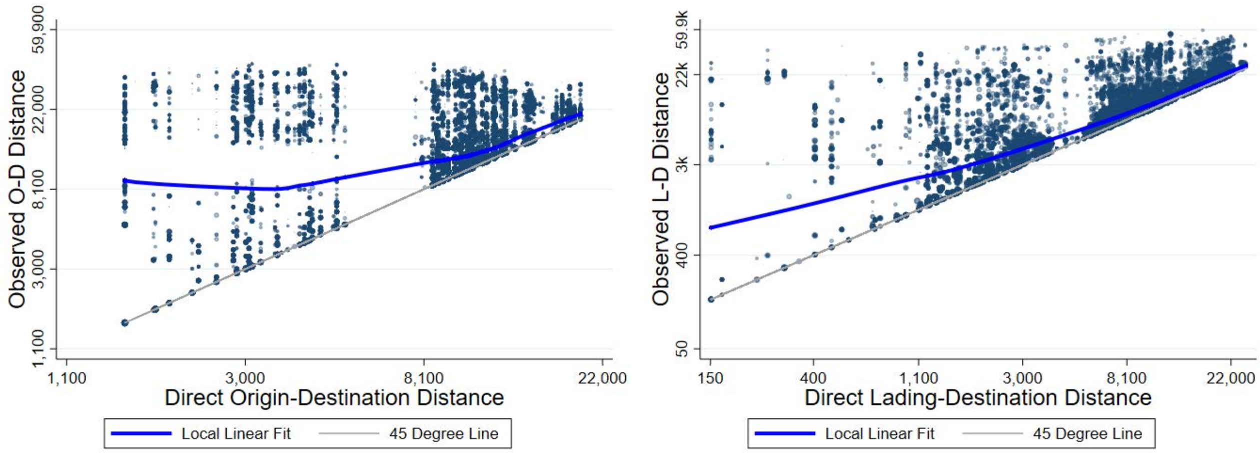
Notes: Panel (A) shows the distribution of containers by the number of unique third-party countries the containers visited. In Panel (B), for each origin country, we calculate the average number of third-party country. The destination country (US) is excluded (in white). Plots are at the shipment level 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). ■

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

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



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.[11] 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 π_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

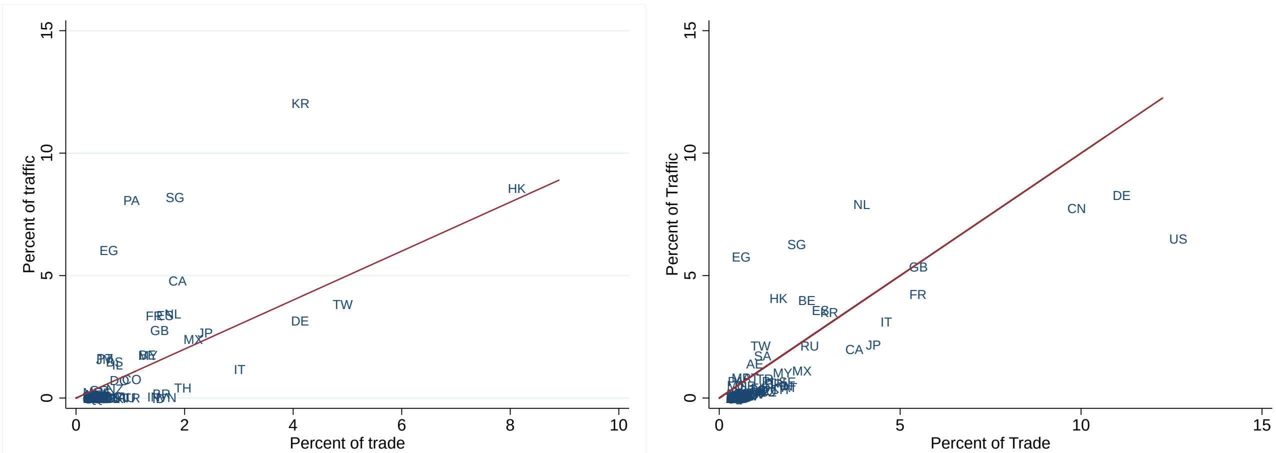
[11]Figure A.6 tabulates the percent of all goods entering the US stopping in that country, broken into goods originated there and elsewhere.

the use of location l above and beyond its role as an exporter to j.[12]

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. ■

Panel (B) of Figure 4 repeats the exercise in (A) using *global* traffic minus trade shares.[13] While the results are broadly consistent with the microdata in Panel (A), 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. In both panels, third-party-country stops (the Y-axes) are significantly more concentrated than trade (the X-axis).[14] Our measure of entrepôt activity in Equation (1) is the distance to the 45-degree line. Appendix Table A.2 lists our measure for all the countries and territories in our data, normalized by the value of the country with the lowest measure, the US.

[12] $Entrepôt_{l,j}$ 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 $Entrepôt_{l,j}$. Our results here and throughout are be robust to other functional forms—for example log differences.

[13] We subtract country l's share of observed global containerized trade π_l from its observed share of global container traffic π^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.

[14] Table A.3 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).

Definition of Entrepôts We define the top 15 countries using this metric as our set of global entrepôts, a natural break after which the measure rapidly flattens (Appendix Table A.2). This list of 15 includes several well-known global hubs, but our results are robust to changes in this threshold as well as to using a continuous measure.[15] This threshold and definition will be used again in counterfactual analyses, where we explore the impacts of cost changes at these hubs. For US shipments, we see 72% of all shipments pass through at least one entrepôt. Of indirect shipments, 90% pass through an entrepôt.

Additionally, we find that smaller origin countries disproportionately use entrepôts. They are simultaneously more likely to ship their goods indirectly and more likely to use entrepôts (see Appendix B.3 and Figure A.7 for further details). Jointly, this confirms that smaller countries are spokes which disproportionately use entrepôts for their trade.[16] 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 observable distance and time costs of trade, but by revealed preference implies lower costs, especially for the spokes of the network which disproportionately choose to send goods indirectly through entrepôts.[17] 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.

[15]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.

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

[17]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.

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.

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 link-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 ω_n in industry $n \in N$ in country j.[18] Within each industry and product category, goods are homogeneous and normal.[19]

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 ω is common for all goods in the same industry n, and includes a vector of factor inputs L, as well as inputs of other goods.[20] Production functions can vary across industries and countries. Cost minimization results in identical production costs among competitive firms within an industry

[18] 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.

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

[20] 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 ω from all industries. All inputs are treated as homogeneous.

in each country. The marginal cost of a good ω is

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

where P_i is the matrix of prices of all goods ω in industries n in 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 ω in industry n must pay non-transport trade costs κ_{ijn} and iceberg transport costs $\tau_{ijnr}(\omega)$ 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 ω 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 $\epsilon_{ijnr}(\omega)$.[21] 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 Ξ 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 $F_{ijn}(\epsilon)$,

[21]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 θ.

the cumulative distribution function of the idiosyncratic draws is as follows:[22]

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

where shape parameter $\theta > 0$ captures the randomness or dispersion in the choice of routes from i to j.[23] Higher $\epsilon_{ijnr}(\omega)$ draws mean industry n has lower costs for route r.

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

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

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 ω 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.[24] We treat t_{k_{r-1},k_r} in Equation (3) as ad valorem, corresponding to the iceberg costs typically considered in the literature (Allen and Arkolakis, 2019; Fajgelbaum and Schaal, 2020). To test the validity of this modeling approach, we consider the fit between our cost estimates with two sets of external data and find significant correlations (Section 7).[25]

This structure is consistent with a host of mechanisms, including but not limited to port-level effects and leg-level scale economies.[26] 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.[27] 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 ω being

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

[23]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).

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

[25]Using an additive cost assumption through the network, Allen and Arkolakis (2019) derives a similar expression for the iceberg cost structure (Appendix D.1, Allen and Arkolakis (2019)).

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

[27]We omit discussion of the optimal shipping network from the perspective of a firm with market power, and focus on leg-level scale instead. In our time period (2014), most ports do not appear to be at capacity and we do not exhibit port-level congestion effects.

shipped on route r from i to j only if the final price of ω, which includes both the marginal cost of production and shipping cost on route r from i to j ($p_{ijnr}(\omega)$), is lower than all other prices of good ω from all other origin country-route combinations.

We then consider the probability that a given country and route r' will be selected as the lowest cost route-supplier combination for good ω conditional on price p:

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

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

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

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.[28] 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 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.[29] 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

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

[29]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 γ is the function $\Gamma(t) = \int_0^{\infty} x^{t-1}\exp^{-x} dx$ evaluated at $\left(\left(1+\theta\right)/\theta\right)^{-\theta}$.

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 Θ_{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 ($t_{kl}(\Xi, \varepsilon_{kl})$) 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 trade network, and estimate a scale elasticity in shipping.

5.1 Taking the Model to Data

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

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

In our setting, expensive trade routes suffer from Ricardian selection at destination markets—the route's impact on prices make them less competitive relative to other routes. Yet, this does not impact the trade cost estimation as seen in Equation (9), which is identical to Allen and Arkolakis (2019), despite differences in framework. 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 links kl. Accordingly, conditioning on the total observed origin-destination trade values X_{ijn}, the trade between i and j contributing 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 origin-destination trade $\left(X_{\bar{N}} \equiv \sum_{n\in\bar{N}} X_n\right)$ and link-level traffic $\left(\Xi_{\bar{N}} \equiv \sum_{n\in\bar{N}} \Xi_{kln}\right)$ are observable. Summing Equation (9) over industries $n \in \bar{N}$ yields:

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

Equation (9) tells us that to accurately measure transport costs, we only need data on origin-destination trade and link-level 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 link 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:

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

where α_0 is a constant, Ξ_{kl}^{data} is the traffic volume between link kl which we observe in the ports call data, α_1 is the relationship between price and quantity (traffic volumes), $\alpha_2 \cdot \ln d_{kl}$ is the coefficient and measure of log sea-distance from k to l respectively. $(\hat{t}_{kl}^{-\theta} - 1)$ allows us to interpret α_1 as the elasticity between cost and traffic volumes to a trade elasticity θ. That is, to interpret results from Equation (11) as elasticities, they are deflated by trade elasticity θ.

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 a China-Singapore link are closer to the direct Korea-Netherlands route compared to routes that include the China-Australia link. 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 Australia on routes between East Asia and Europe compared to Singapore.

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:

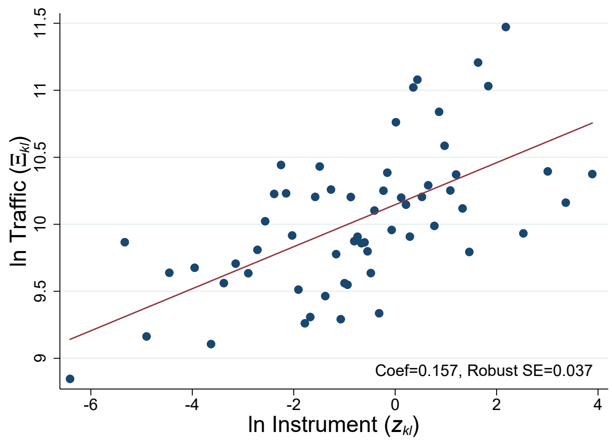
$$z_{kl} = \sum_{i \backslash k,l} Pop_{i,1960} \sum_{j \backslash \{k,l\}} Pop_{j,1960} \frac{d_{ij}^2}{(d_{ik} + d_{lj})^2}, \tag{12}$$

where d_{ij} is the sea distance between origin i and destination j, and the square of the relative excess distance between links 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 $Pop_{j,1960}$.[30] Figure 5 shows the robust first-stage relationship between our instrument and traffic.

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

[30] 1960 Population here stands in place of GDP, which may be endogenous to the trade costs in our model. The year is chosen because immigration and populations prior to 1960 could not plausibly be impacted by 2014 containerized shipping costs.

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. ■

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 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 $m_1(\alpha, \beta) = Z\varepsilon\left(\alpha, \hat{\mathbf{t}}\right)$ based on Equation (11) with vector α and matrix of trade costs $\hat{\mathbf{t}}$. First, however, we need to recover leg-level trade costs $\hat{t}_{kl}$.

5.3 Recovering Trade Costs

We require two observable objects in order to recover a global set of trade costs: origin-destination trade values and link-level traffic volumes (Equation 10).[31] Our traffic data comes

[31]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.

from our global port of call AIS shipping data.[32] We use aggregate origin-destination 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.[33] Note that we do not rely on the merged US microdata in our estimation.

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:[34]

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

where the matrix $\mathbf{Y}$ is a vector defined as

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

where β_0 is an intercept, β_1 considers the sea distance between the nearest principal ports,[35] and β_2 considers port-to-port traffic. β_3 and β_4 consider the total incoming and outgoing traffic at ports i and j respectively. β_5 considers the role of the backhaul problem from Wong (2020), where ship capacity is fixed by the shipping direction with the higher demand. Finally, β_6 is an indicator for a shared land border.[36]

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 β of coefficients—which may reflect endogenous variables—but the resulting predictions for $\hat{t}_{ij}$. Instead, we seek to fully saturate the variation in the data in order to generate the closest empirical prediction for the matrix of

[32] 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.

[33] 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.

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

[35] 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.

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

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 β yield estimates of every trade cost $\hat{t}_{ij}$, we need not discipline β by comparing traffic on every link. This allows us to still recover estimates of $\hat{t}_{ij}$ 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 β that minimizes the difference between the matrix of expected traffic, $\hat{\Xi}\left(\beta|\mathbf{X},\mathbf{Y},\theta\right)$, and observed traffic Ξ^{data} for countries that do not share a land border:

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

where expected traffic is a function of β, trade elasticity θ, as well as observed trade values $\mathbf{X}$.

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:

$$\begin{aligned} m_1\left(\alpha,\beta\right) &= Z\varepsilon\left(\alpha,\hat{\mathbf{t}}(\beta)\right) \\ m_2\left(\beta\right) &= \left(\hat{\Xi}\left(\beta|\mathbf{X},\mathbf{Y},\theta\right)\right) - \left(\Xi^{data}\right) \end{aligned}$$

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 trade volumes on the second set of moments m_2, which rationalizes a global set of link-level trade costs t_{kl} conditional on observable origin-destination trade values $\mathbf{X}$ and link-level traffic flows Ξ^{data}. We reiterate that we only conduct inference on the parameters α. We treat β as a set of incidental parameters, important for estimation, but not for inference.[37]

[37] The second stage computes an optimal weighting matrix using the first stage results.

5.5 Simultaneous Identification of Scale and Trade Costs

Our approach parallels the Industrial Organization literature, which seeks to recover unobserved cost structures, and identification depends both on instrumental variables and behavioral assumption. For example, Ackerberg et al. (2007) take market level data and instruments to recover demand and then use equilibrium assumptions on behavior to recover marginal costs, which are then projected on product attributes. Similarly, we rely jointly on the structure of equilibrium shipping flows embedded in the (Allen and Arkolakis, 2019) framework and our demand-shifting instrument.

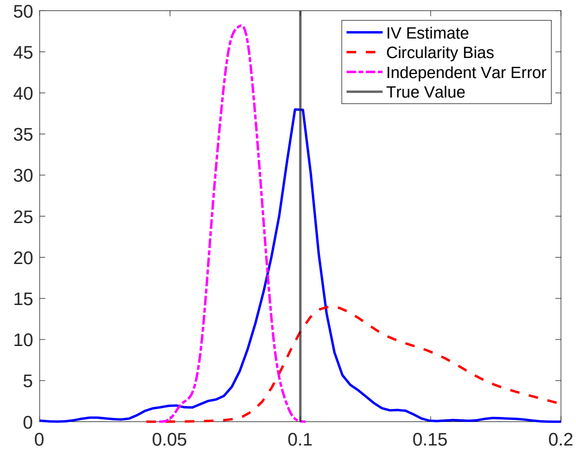
However, this approach opens the door for a mechanically-driven result. Specifically, we are concerned with estimating the causal scale impact of traffic volumes on trade cost (Equation 11) while, at the same time, our cost estimates are themselves recovered from our model prediction which is a function of traffic volumes (Equation 10). This circularity can introduce a mechanical correlation if, for example, measurement error in traffic feeds both into trade cost estimates and traffic.

We approach this problem through multiple methods. First, we establish how this issue can arise due to measurement error in our context. We show how this error can be considered a form of omitted variable bias, and the conditions under which an instrumental variable can correct for this bias. Second, we run Monte Carlo simulations that confirm the existence of this bias in the presence of measurement error and show how our instrument eliminates it. Third, we use external data on freight costs to estimate potential traffic-correlated errors, both to illustrate the potential bias in the OLS and show how our instrument removes this bias. Finally, we run a parallel scale estimation purely on our external freight costs and find similar results. See Appendix D.2 for full details.

Figure 6 summarizes our findings using Monte Carlo Simulations. First we show that with true trade costs, typical measurement error in traffic volumes would bias ordinary least squares (OLS) estimates downward (purple dot-dash line). If measurement error in traffic affects the trade cost estimates, the OLS estimates would bias upward (red dash line), since the dependant variable (trade costs) is partially derived from the independent variable (traffic). However, a valid instrumental variable can correct for this bias (blue solid line). Appendix D.2 further elaborates on the simulation procedure and each specification.

It is not possible to directly test the validity of our exclusion restriction and the conditions

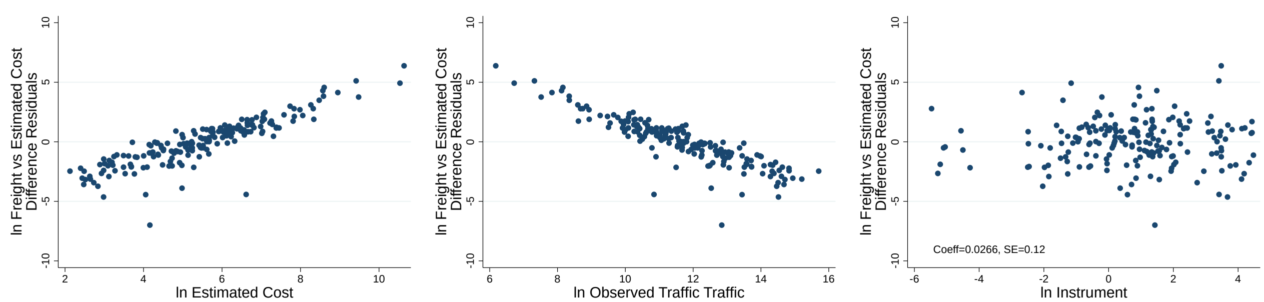
Figure 6: Monte Carlo Simulations Illustrating Estimation Biases



Notes: The figure shows 500 simulated estimates. The blue solid line is our preferred instrumental variable estimator. Our instrument is correlated with the true shipping traffic on a particular route. The purple dot-dash line illustrates classic measurement error in the independent variable (shipping traffic on a route), leading to classic attenuation bias in OLS. The red dash line illustrates our principle worry, an upward bias in OLS, due to our recovered trade costs being a function of observed shipping traffic that could be measured with error. A valid IV can correct for this bias (blue solid line). See Appendix D.2 for full details. ■

Figure 7: Balancing Test

(A) Approximated Error vs Estimated Costs | (A) Approximated Error vs Observed Traffic | (B) Approximated Error vs Instrument



Notes: Figures are scatter plots of, on the X-axis, the natural log of the estimated leg costs in Section 5, the observed traffic, and the geography-based instrument used in Section 5, in Panels (A), (B), and (C), respectively, against the difference between the natural logs of the estimated leg costs in Section 5 and from Wong (2020) on the Y-axis, residualized after controlling for sea-distance for 209 legs for which both costs exist. Standard errors are clustered two-ways by the nodes on each link. See Appendix D.2 for full details. ■

derived in Appendix D.2 under which our instrument eliminates this bias: no correlation between the error in our estimated trade costs and the instrument. However, we can show the lack of correlation between our instrument and an approximation of the error, estimated as the difference between our measured costs and external measures of freight costs from Wong (2020). Details for this exercise are found in Appendix D.2.

Panels (A) and (B) of Figure 7 show a positive and negative correlation between this approximation of the error and estimates link costs and link traffic, respectively, controlling for distance, consistent with the circularity bias in Appendix D.2 and the Monte Carlo. Panel (C) shows a weak and insignificant correlation between this residualized approximation of the error and our instrument, again controlling for sea-distance. The lack of correlation is consistent with an instrument which is uncorrelated with the true error. While this is insufficient to validate our instrument, it performs the same role as a balancing test, showing an absence of evidence of exclusion restriction violations.

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)
	$\ln(c_{kl})$
$\ln\left(\Xi_{kl}^{data}\right)$	-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. $ln\Xi_{kl}^{data}$ 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. ■

[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.

Link and average bilateral trade costs Appendix Figure A.11 graphs our resulting matrix of pairwise trade costs. We present the vector β estimates in the Appendix Table A.5 as purely predictive parameters, not fundamentals that we can alter in the counterfactuals (see Appendix D.1 for further details). Instead, we simply need to know if our β estimates can predict containerized traffic that reflects the actual observed traffic volumes. With a full link-level 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.11 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.

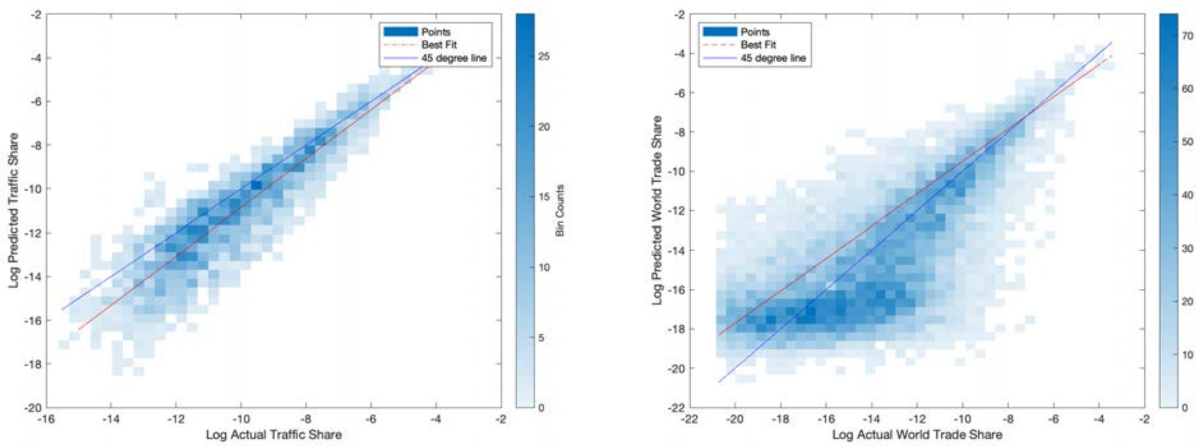
Model Fit Figure 8 compares our model-predicted traffic and trade values against their observed counterparts in the data. In Panel (A), we compare actual observed global container traffic shares with the our model-predicted shares using our estimated trade costs. We include both a best fit line and a 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.

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 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.

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

Figure 8: 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. ■

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 one 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, we estimate leg-level shipping scale economies. A number of mechanisms can generate the cost reductions that coincide with these scale economies. 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

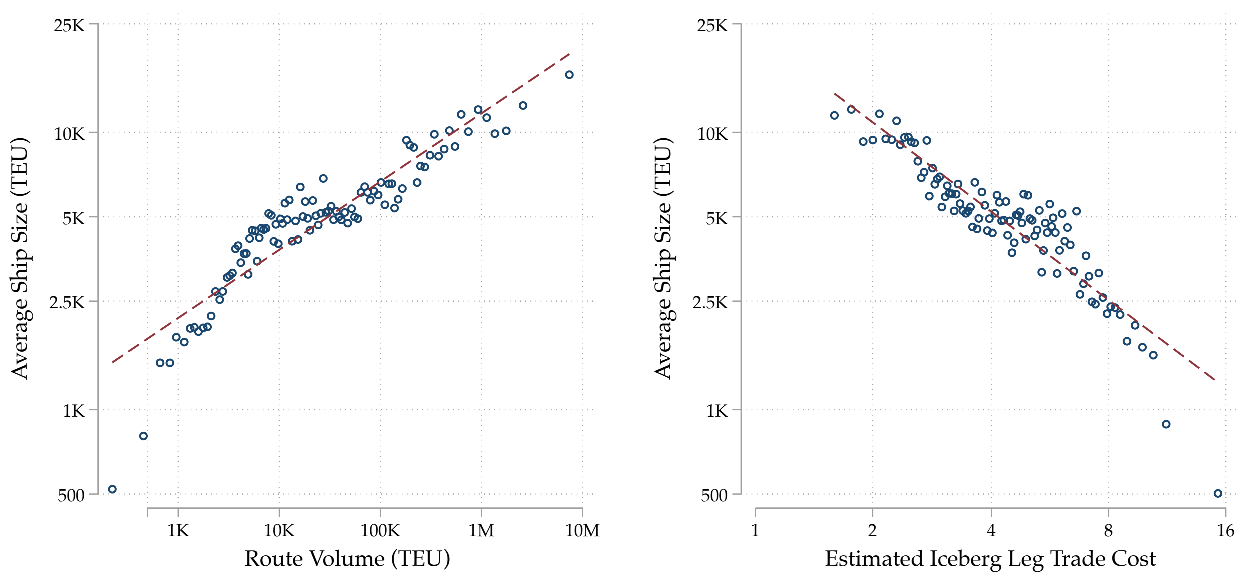
[40]High-traffic routes are served by many carriers, using ships capable of carrying 25,000 containers with automated loading and unloading.

then investigate the relationship between indirect shipping and ship size.

Ship Sizes, Traffic Volumes, and Recovered Trade Costs In Panel (A) of Figure 9, 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 9). 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% increase in ship sizes.[41]

Figure 9: Link Between Recovered Trade Costs and Ship Size

(A) Route Volume and Ship Size (B) Shipping Cost and Ship Size



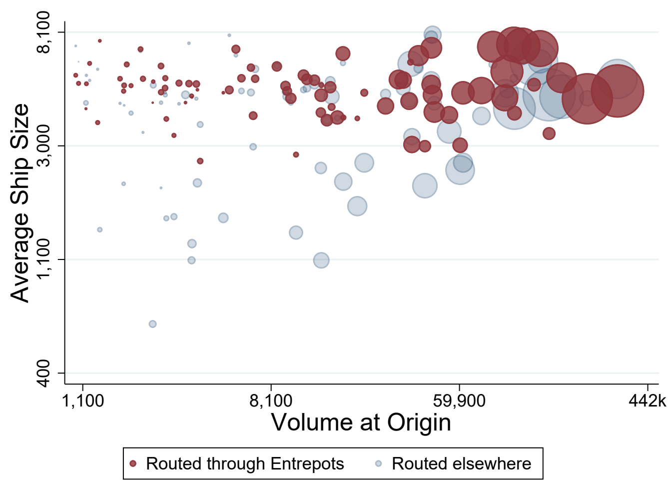
Notes: Figures are bin-scatter plots over all observed containership routes, with 100 bins. We control for the log(sea distance) between origin and destination ports, but add variable means back for the plots. Panel (A) plots the relationship between the total containers on a route and the average containership's 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 10 further investigates the relationship between entrepôt usage and ship size, plotting 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

[41] Appendix Section D.3 reports shipment-level regressions controlling for origins, destinations, and without route distance controls. Results are similar.

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]

Figure 10: Link Between Indirect Trade and Ship Size



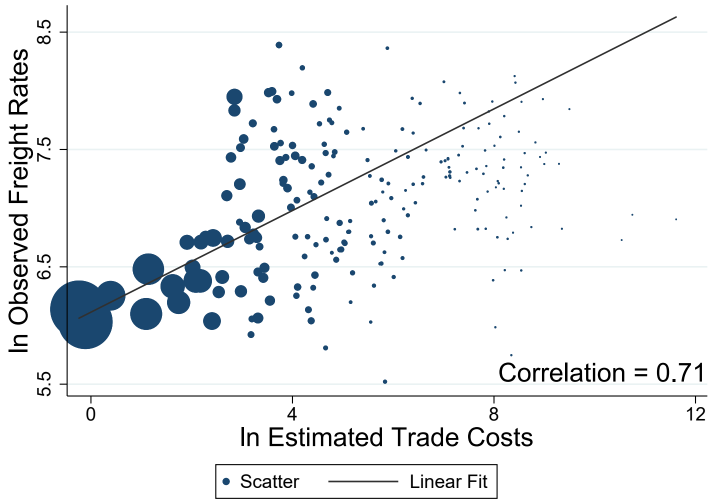
Notes: The x-axis shows the total exports from an origin country to the United States. The y-axis shows the average ship size which arrives from an origin country to the United States. Each country is represented by two data points, a blue and a red circle. The red circle indicates the corresponding 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. ■

7.2 Cost Estimates with Freight Rates Data

Next, we compare our expected trade cost estimates τ_{ij} at the origin-destination level with container freight rates from Wong (2020). These rates are the costs paid by firms to 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 11).

[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.4 confirms the positive relationships between shipment volume and ship size and robustness to different notions of origin, lading, and transshipment.

Figure 11: Correlation Between Cost Estimates With Actual Freight Rates



Notes: Data points compare origin-destination predicted costs τ_{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.

Next, summing the predicted probabilities in Equation (13) across all origins i, the model

Table 2: Correlation Between Traffic Estimates With Microdata

	(1) $\widehat{\pi_{iUS}^{kl}}$	(2) $\widehat{\Xi^{kl}}$	(3) $\widehat{\pi}_{US}^{l} - \widehat{\pi}_{l,US}$	(4) $\widehat{\pi_{iUS}^{kl}}$	(5) $\widehat{\Xi^{kl}}$	(6) $\widehat{\pi}_{US}^{l} - \widehat{\pi}_{l,US}$
$\pi_{iUS,Data}^{kl}$	0.846			0.872		
	(0.119)			(0.121)		
Ξ_{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. ■

delivers a prediction for the total amount of US-bound traffic on a given leg kl:

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

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 Ξ_{Data}^{kl}, 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:

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

Column (3) compares this to its counterpart in the microdata, which we call $\pi_{US,Data}^{l} - \pi_{l,US,Data}$, 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 matches 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.

8 Counterfactuals

In three counterfactual exercises, we quantify the welfare importance of the trade network and the specific role entrepôts play within that network.

In our first counterfactual, we demonstrate that (1) transportation improvements at entrepôts have significant global welfare impacts (not including their own gains), as well as localized benefits for nearby neighboring countries as a result of the trade network, (2) the global impact of transportation improvements differs meaningfully from non-transportation improvements for all countries—not just, but especially for especially entrepôts—due to the network structure of trade, and (3) scale economies in transportation further magnifies these impacts.

In our second counterfactual, we illustrate how non-transportation cost changes at an entrepôt generate widespread impacts through the trade network—beyond directly impacted countries—by considering the impact of a negative trade shock on an entrepôt node country in the form of the United Kingdom leaving the European Union. Changes to the trade network due to scale economies generate different consequences for Brexit, both in effects' magnitudes and their distributions.

[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.

Our third counterfactual evaluates the welfare and trade impacts of the two endogenous mechanisms in our model: (1) network effects—allowing countries to ship indirectly and (2) scale effects—allowing countries to ship indirectly and take advantage of scale economies. To illustrate this, we study the effects of the Arctic opening up to trade between the Pacific and Atlantic Oceans, bypassing the Suez and Panama canals and decreases in trade cost on specific links.

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 $\hat{X}_{ij}$.[45]

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 η and γ. This gives 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 $\hat{t}_{kl}$, which, translated through the model, generate changes in the expected trade cost between every bilat-

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

[45] We use hat notation here in a different context than previous sections, where it denoted estimated and model-derived results.

eral 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. Our model theoretically admit multiple equillibria (as in (Brancaccio, Kalouptsidi and Papageorgiou, 2020)), however we focus on the unique equilibrium from our current starting point—the world today.[46]

8.2 Importance of Entrepôts in the Trade Network

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 (κ_{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 κ_{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.

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 12 lists the global welfare impact of infrastructure improvements at the 20 most pivotal nodes

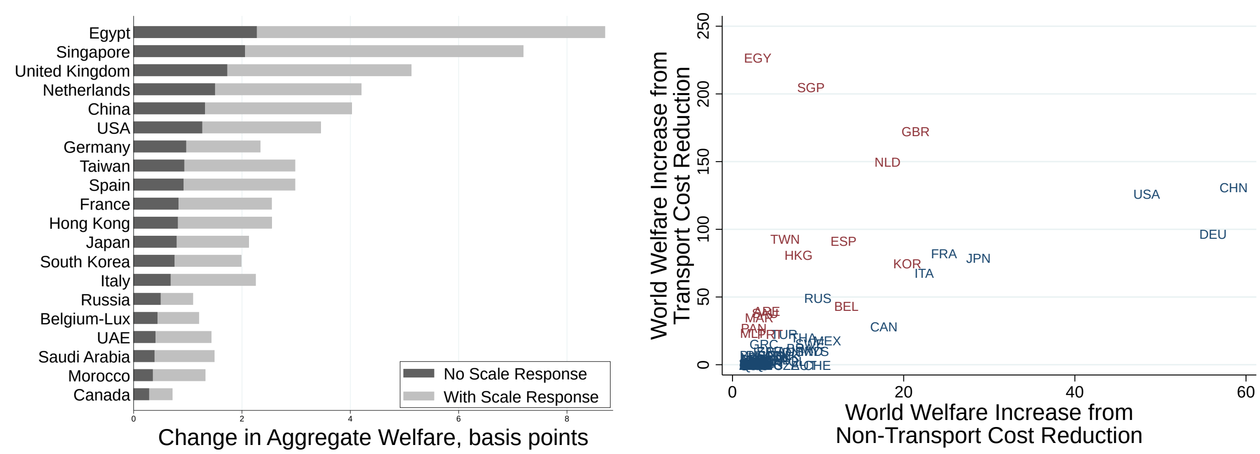
[46] Kucheryavyy, Lyn and Rodríguez-Clare (2019) establishes a common mathematical structure that characterizes the unique equilibrium in multi-industry gravity trade models with industry-level external economies of scale. Their structure requires that the product of the trade and scale elasticities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

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).[47] Scale economies' impact on the transportation network (overlaid grey bars) further augment the differential impact of entrepôts.[48] Infrastructure investments at entrepôts generate on average 10 times the global welfare impact relative to investment elsewhere.[49]

Figure 12: Most Pivotal Countries in the Network: Change in Global Welfare

(A) Transportation Improvements: Highest Global Welfare Changes

(B) Transportation vs Non-Transportation Improvements



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 Trade Network Matter? Panel (B) of Figure 12 plots the average welfare impact, excluding the targeted country's own 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*.

[47]Within drybulk shipping, Brancaccio, Kalouptsidi and Papageorgiou (2020) finds that removing the Suez Canal has the highest impact on welfare decline relative to the Panama Canal and Strait of Gibraltar.

[48]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.

[49]Appendix Tables A.12 and A.13 examines raw and conditional mean differential impact of targeting entrepôts.

For entrepôts, (red in Panel (B)), the 1-to-1 relationship is violated. For example, Egypt ranks first in terms of global impact from infrastructure improvements, while it is not among the top 20 in terms of non-transportation trade cost reductions. 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.[50] 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.

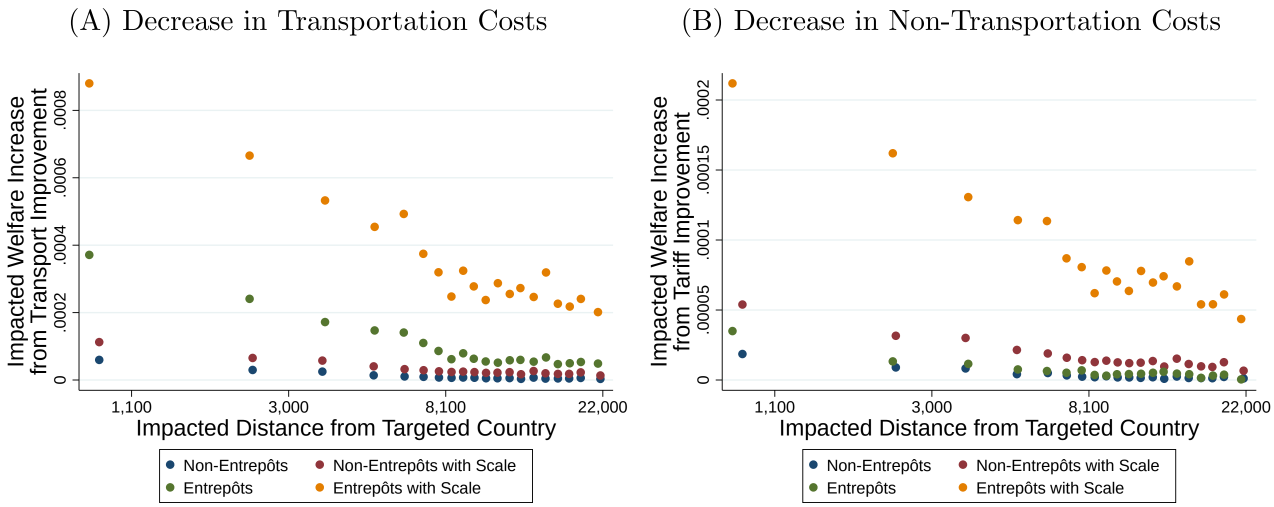
The impact of entrepôts are localized To account for the differential impacts of entrepôts, we drill down to one particular margin at which the impact appears most distinct: locally. Figure 13 is a binned scatter plot considering the welfare effects on the impacted country (y-axis) relative to its distance from the targeted country (x-axis), adjusting for the impacted country fixed effects. Nearly overlapping blue and green dots in Figure 13 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 13 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.[51]

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 14 plots the differential welfare gains to entrepôts 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,

[50]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.

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

Figure 13: 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. ■

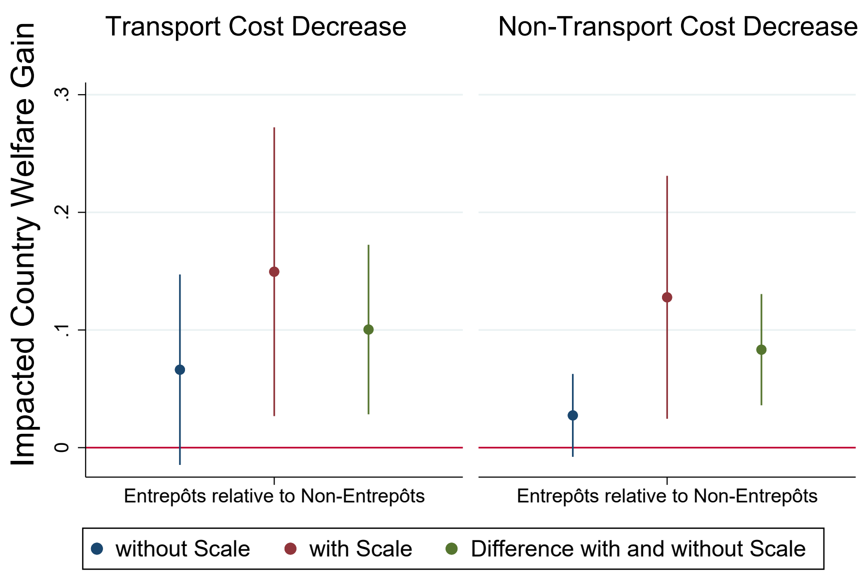
we find that the welfare gains for both entrepôts and non-entrepôts are not significantly different (in blue). However, the differential benefits *to* entrepôts is significant and large when allowing for scale economies (in red). Scale economies disproportionately accrue gains to entrepôts as impacted 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 Impact of Non-Transport Trade Costs on the Network

In order to illustrate the trade network consequences of non-transportation trade cost changes on a node, we study the effects of Brexit—a 5% increase in non-transportation 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.

We model two cases: first without, then with the impact of scale on the trade network. In

Figure 14: 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. ■

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 trade 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 (A) of Table 4 reports aggregate effects. The direct effect decreases global welfare by 2.3 basis points (Column (1)). The introduction of scale economies leads to a more than 4 times decrease of 10 basis points. Trade volumes follow a similar pattern. Figure 15 highlights the distributional effects in terms of welfare (see Appendix Figure A.16 for trade volumes). Scale economies amplify the Brexit impact, especially for European countries. Notably, the impact of scale is not well-predicted by the non-scale case (Panel (B), Figure 15). We document significant negative welfare impacts on Ireland, Iceland and other Nordic countries, many of which rely on

Table 3: Welfare and Trade Outcomes from Improvements in Transportation and Non-Transportation Costs, Basis Points

	Non-Transportation Improvement		Transportation Improvement	
	Baseline Effect (1)	Total Effect (Network & Scale) (2)	Network Effect (3)	Total Effect (Network & Scale) (4)
Δ Average Global Welfare				
Mean	0.08	0.26	0.20	0.58
Standard Deviation	(0.20)	(0.59)	(0.47)	(1.41)
Δ Container Trade Volumes				
Mean	0.87	2.93	2.31	6.62
Standard Deviation	(2.22)	(6.74)	(5.31)	(16.00)

Notes: This table reports results for our first counterfactual, transportation and non-transportation cost declines for each of 136 countries. Columns (1) and (2) present results for cases where non-transportation trade costs are reduced. Columns (3) and (4) present results for cases where transportation costs are reduced (infrastructure improvements). The top panel presents aggregate welfare changes. The bottom panel presents changes to aggregate container trade. Columns (1) and (3) correspond to cases where no scale economy feedback loops are allowed. Columns (2) and (4) present results allowing for scale economy feedback. ■

UK feeder routes to get their goods to large vessels that ply transoceanic 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).

Table 4: Welfare and Trade Impact of Brexit and Artic Passage Opening, Basis Points

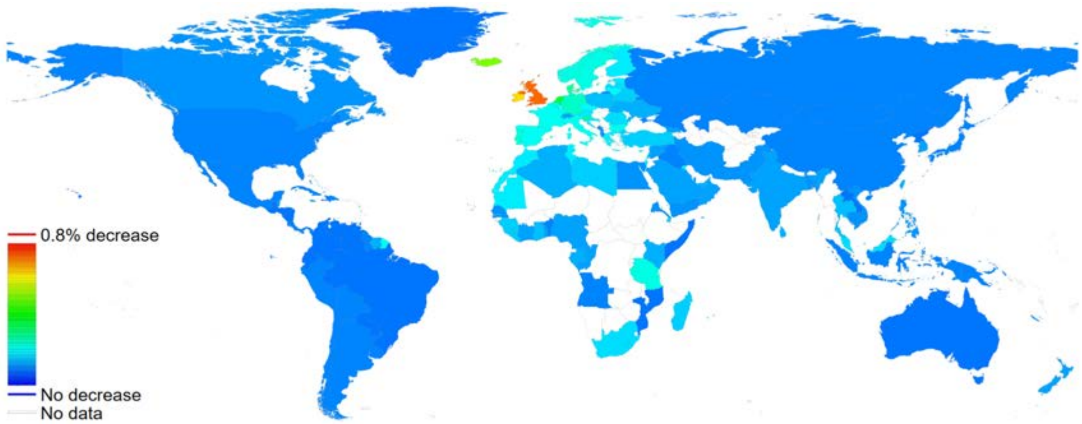
	Direct Effect (1)	Network Effect (2)	Total Effect (Network & Scale) (3)
Panel (A) Brexit: Impact of Non-Transport Trade Costs			
Δ Average Global Welfare	-2.3		-10.0
Δ Container Trade Volumes	-24.5		-112.7
Panel (B) Arctic Passage: Impact of Endogenous Trade Costs			
Δ Average Global Welfare	1.5	3.3	8.9
Δ Container Trade Volumes	17.4	38.2	101.8

Notes: Panel (A) presents results for Brexit, a 5% increase in non-transportation trade costs κ_{ij} between the UK and its trading partners. The direct effect of Brexit only accounts for the changes in direct trade with the UK or multilateral resistance. The total effect allows for the change in direct UK trade to impact trade costs with neighboring countries through the trade network. Panel (B) presents results for the Arctic Passage counterfactual. The direct effect of the passage opening only accounts for direct changes in physical distance between countries. The network effect results allow for indirect shipping through the trade network as a result of the passage opening. The total effect adds in the impact of the scale. ■

Figure 15: Welfare Changes - Brexit

(A) Direct Effect from Tariff Change

(B) Total Effect: Full Trade Network and Scale Effects



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. ■

8.4 Impact of Endogenous Trade Costs on the Network

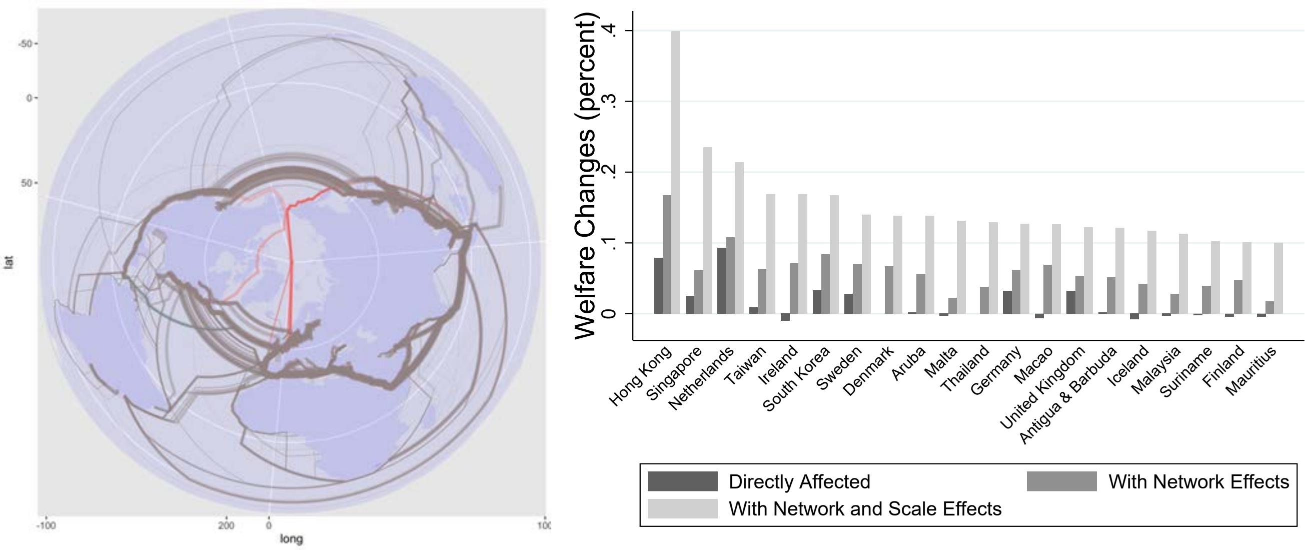
We evaluate the importance of endogenous trade costs by demonstrating the welfare and trade impacts from the two endogenous mechanisms in our model: (1) network effects—allowing countries to ship indirectly and (2) scale effects—allowing countries to ship indirectly and take advantage of scale economies. We achieve this by studying the physical trade route changes due to 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 link within the network, 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 16 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, non-changing routes are in brown, and abandoned routes are in blue.

Figure 16: 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. ■

We compare three different cases. First, we consider a network-naive exogenous trade cost case where we only allow for changes in origin-destination trade costs between country pairs for which the direct bilateral distance decreased. Note that we do not allow third-party-countries to take advantage of these reductions. Second, we lower t_{kl} for all observed links with positive traffic according to α_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 due to the trade network effects.[52] Third, we repeat the second case accounting for the impact of scale: as trade

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

costs change, trade volumes change, reducing trade costs further.

Assuming exogenous trade costs with our input-output structure, Column (1) of Table 4 Panel (B) shows that the network-naive and direct effects of the Arctic Passage are positive, with aggregate welfare increasing 1.5 basis points, and container trade volumes increasing 17 basis points. Endogenizing trade costs to allow for the 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 (Column (2), Table 4 Panel (B)). Allowing for both scale and network effects 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, Column (3)).

Panel (B) in Figure 16 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 17 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.[53] In the baseline scenario in Panel (A), we see increases from trade between countries that are along the Northeast passage, and small spillover impacts at countries not directly impacted—reflecting classic multilateral resistance and cascading effects from value chains. Figure 17 Panel (B) 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

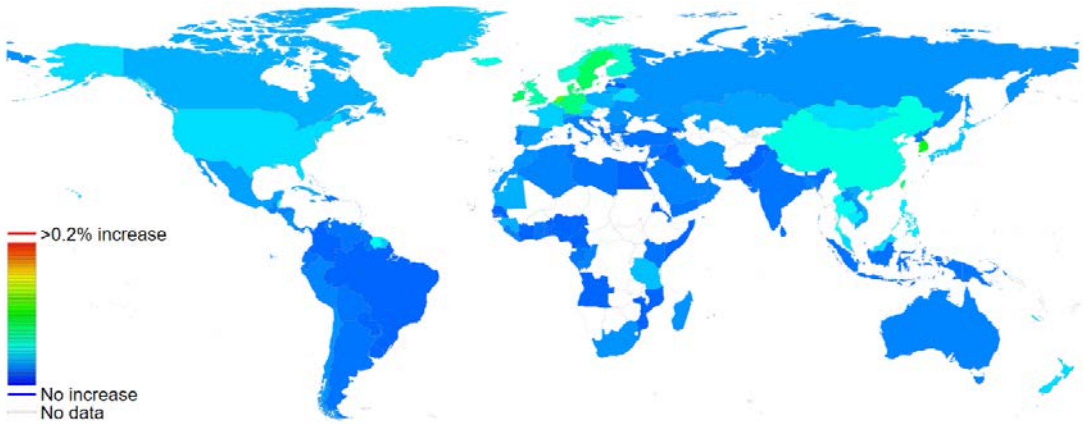
This paper studies entrepôts, the trade network they form, and their impact on international trade. We characterize the global container shipping network as a hub-and-spoke system by documenting that the majority of trade is indirect and flows from origins to destinations through entrepôts (hubs). To 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

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

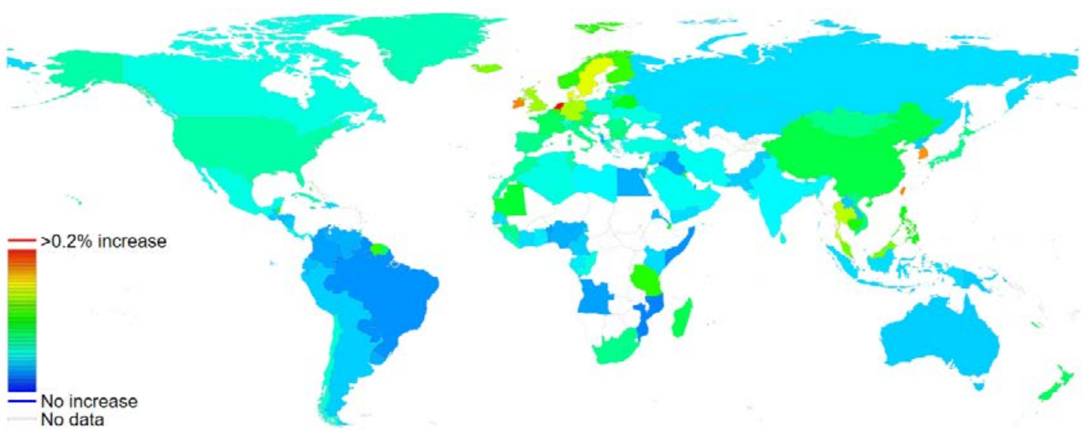
Figure 17: Welfare Changes - Arctic Passage

(A) Only Directly Affected Routes (Exogenous Trade Costs)

(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. ■

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. We find that infrastructure investments at entrepôts generate on average 10 times the global welfare impact relative to investment elsewhere.

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.[54] 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

Ackerberg, Daniel, C Lanier Benkard, Steven Berry, and Ariel Pakes. 2007. "Econometric tools for analyzing market outcomes." *Handbook of Econometrics*, 6: 4171–4276.

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

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

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

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

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

Astra Paging. 2014. "Automatic Identification System Data."

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

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

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 G, Arnaud Costinot, Dave Donaldson, and Andres Rodriguez-Clare. 2019. "The textbook case for industrial policy: Theory meets data." National Bureau of Economic Research.

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. 2020. "Geography, transportation, and endogenous trade costs." *Econometrica*, 88(2): 657–691.

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

CEPII. 2017. "BACI: International Trade Database at the Product-Level."

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. 2020. "Optimal transport networks in spatial equilibrium." *Econometrica*, 88(4): 1411–1452.

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."

KGM Associates. 2014. "Eora Global Supply Chain Database."

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.

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

Panjiva. 2014. "U.S. Bills of Lading."

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.*

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

Wong, Woan Foong. 2020. "The Round Trip Effect: Endogenous Transport Costs and International Trade." *American Economic Journal: Applied Economics, Forthcoming.*