Entrepôt: Hubs, Scale, and Trade Costs*

Sharat Ganapati Woan Foong Wong Oren Ziv

May 2023

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 effects propagate local trade cost changes globally. Counterfactual infrastructure improvements at entrepôts generate ten times the global welfare impact relative to non-entrepôts.

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

JEL Classification: F12, F14, F62, R40

*Sharat Ganapati is Assistant Professor of International Economics, Georgetown University, Washington, DC, and Faculty Research Fellow, National Bureau of Economic Research, Cambridge, Massachusetts. Woan Foong Wong is Assistant Professor of Economics, University of Oregon, Eugene, Oregon. Oren Ziv is Assistant Professor of Economics, Michigan state University, East Lansing, Michigan. Their email addresses are sg1390@georgetown.edu, wfwong@uoregon.edu, and orenziv@msu.edu. We thank Treb Allen, Costas Arkolakis, Panle Jia Barwick, Bruce Blonigen, Johannes Boehm, Mark Colas, Kerem Coşar, Anca Cristea, Meredith Crowley, Carsten Eckel, Stefania Garetto, Matthew Grant, Keith Head, Reka Juhasz, Myrto Kalouptsidi, Steve Matusz, Ezra Oberfield, 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, 20$^{\text{th}}$ 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.

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—significantly increasing shipping times and distances.

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 endoge-

nous 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 typical journey in our microdata has 2.5 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 Singapore and Egypt (the Suez Canal) are the top two most pivotal locations in the trade network, as reflected by the strain in global supply chains when Egypt 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, 2022; 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 the role of trade networks (Allen and Arkolakis, 2019; Fajgelbaum and

[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. We find the presence of scale economies in shipping. 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 (2022) focuses on the round trip effect from container shipping: a bilateral trade cost endogeneity.

[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).

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 top two 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 within the international 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 trade and shipping traffic.

[4]The emergence of entrepôts as hubs in geographically advantageous locations is consistent with the findings of Barjamovic et al. (2019). This is related to studies of airlines hub and spoke networks. This literature takes large parts of the network as fixed (Berry, 1992) or is restricted to simple entry games (Ciliberto and Tamer, 2009)

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. We measure volume in twenty-foot equivalent container units (TEUs).

Our sample covers 4,986 unique container ships with a combined capacity of 30.6 million TEUs—over 90% of the global container shipping fleet—making 397,625 calls at 1,230 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—aggregated into containerized and noncontainerized industries according to the procedure outlined in Appendix A.3—to estimate our model in Section 5.

Figure 1: Global Network of Ships, Ports of Call Data

Notes: Dots represent 1,230 ports). Lines represent journeys between port pairs undertaken by a containership (total of 4,986 ships). We show direct distances here. Analysis uses sea-route distance.

[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.

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 and loaded onto US-bound containerships (laded) in 144 countries. This accounts for about three quarters of the 2014 TEU and tonnage imports, 77 percent and 74 percent respectively (Maritime Administration, US Department of Transportation, 2014).

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 (Appendix Figure A.1 visualizes this merge). 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.

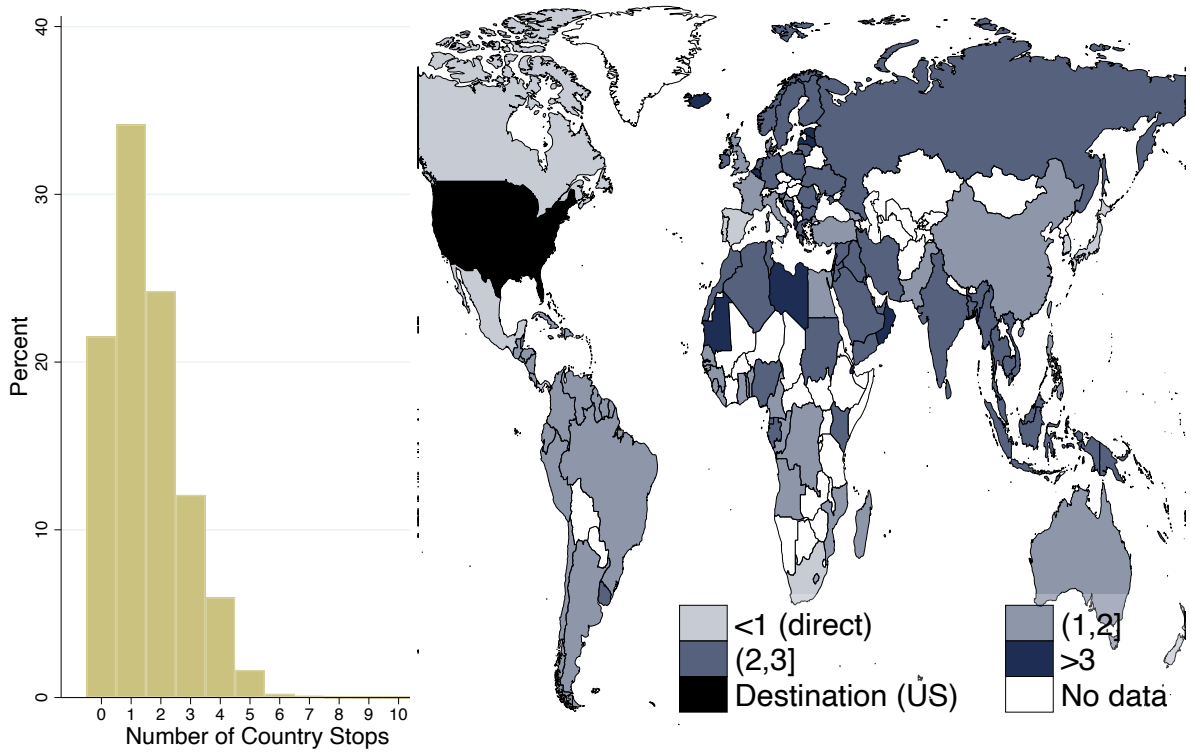
3.1 The Majority of Trade is Indirect

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 nei-

ther the origin nor destination.[7] 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.[8] Similarly, the average shipment from a majority of US trading partners is transshipped in a third-party country—60% of US trading partners transship more than 90% of their US-bound goods.[9] Figure A.5 reports the percent of goods transshipped at third-party countries.

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 number of unique third-party countries 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. 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).

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

[7]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.

[8]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.

[9]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.

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 final destination. Here the remaining gap is still substantial at 23%. 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 (Columns (2) and (5) in Table A.1 respectively). 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.*

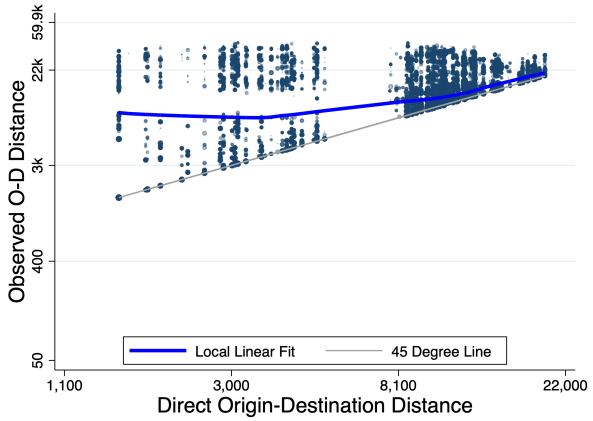
3.2 Indirect Trade Is Routed Through Entrepôts

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.

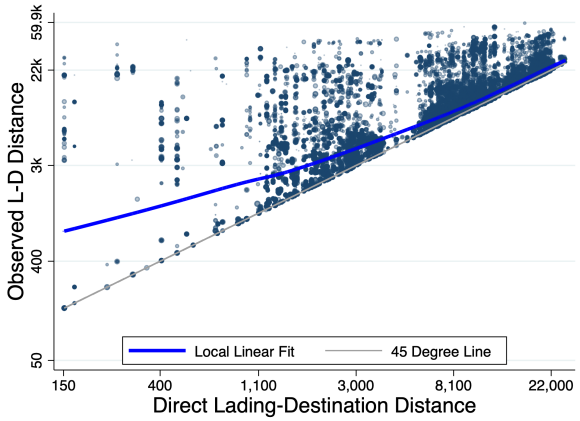
Panel (A) of Figure 4 plots each country's share of total third-party-country stops against its share of total US trade. Some locations are both popular stopping points and major countries of origin for goods like China, Germany, and Japan. Key countries like Korea, Singapore, Panama, and Egypt disproportionately participate as third-party-

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.

countries in US-bound shipments.[10] This leads to our measure of entrepôt activity:

$$Entrepôt_{l,j} \equiv \pi^l_j - \pi_{l,j} \tag{1}$$

where country j's usage of entrepôt l for its imports is the difference between π^l_j, the share of j's imports flowing through l, and $\pi_{l,j}$ the share of j's imports originating at l. This captures the use of location l above and beyond its role as an exporter to j.[11]

Panel (B) of Figure 4 repeats the exercise in (A) using *global* traffic minus trade shares.[12] 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).[13] 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

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

[11] $Entrepôt_{l,j}$ 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 robust to other functional forms—for example log differences.

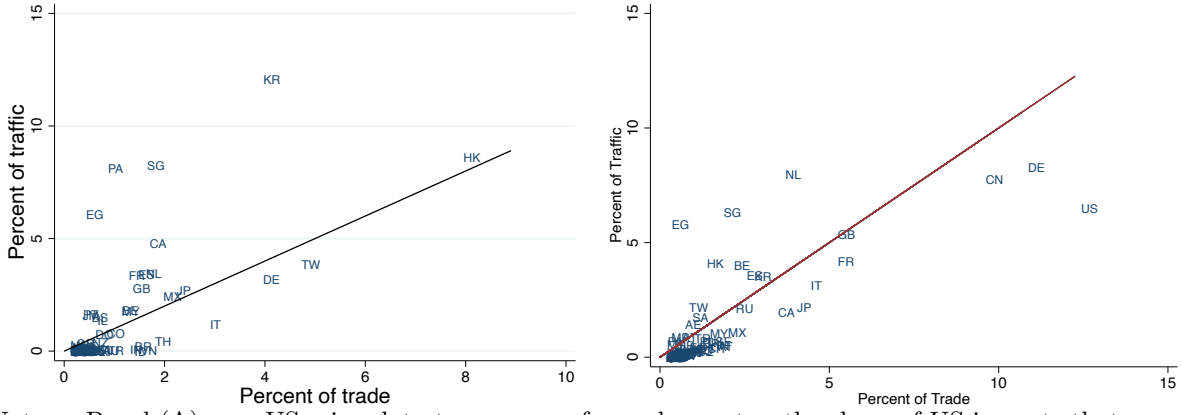
[12] We subtract country l's share of observed global containerized trade π_l from its observed share of global container traffic π^l, with an adjustment for unobserved overland traffic as described in Section 5. Appendix C clarifies how this is a consistent aggregation of the country-level measure in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

[13] Table A.3 reports the concentration ratios for trade, transshipment, and third-party-country stops.

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 US imports 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.

value of the country with the lowest measure, the US.

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.[14] 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 73% of all shipments pass through at least one entrepôt. Of indirect shipments, 92% 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.[15] 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

[14] 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.

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

spokes of the network which disproportionately choose to send goods indirectly through entrepôts.[16] The goal of our empirical estimation is to measure this trade-off within the context of the full global trading network by finding a set of node-to-node costs which describe the shipping network and is consistent with the indirect trade we observe.

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

4 Theoretical Framework

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

Entrepôts emerge as locations where goods pass through, but are neither the goods' origin nor their destination. 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.

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. Within each industry and product category, goods are homogeneous and normal.[17]

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

[16] While some entrepôts lie along lowest-cost routes, routes stopping at entrepôts are 3-9% 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.

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

factor inputs L, as well as inputs of other goods.[18] Production functions can vary across industries and countries. Cost minimization results in identical production costs among competitive firms within an industry in each country. The marginal cost of a good 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 stop to stop involves iceberg transport costs as well as product- and route-level idiosyncratic cost shocks $\epsilon_{ijnr}(\omega)$.[19] We place minimal structure on these direct leg-level costs $t_{r_{k-1},r_k}(\cdot)$ between locations r_{k-1} and r_k on route r, allowing them to be a function of exogenous and endogenous variables:

$$t_{r_{k-1},r_k} = f(\Xi, \varepsilon_{r_{k-1},r_k}) \tag{2}$$

where Ξ is a matrix of endogenous containerized traffic over the entire network and $\varepsilon_{r_{k-1},r_k}$ reflects exogenous transportation cost elements such as distance.

Route-specific idiosyncratic shocks are drawn from the Fréchet distribution such that

[18] 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.

[19] 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.

$F_{ijn}(\epsilon)$, the cumulative distribution function of the idiosyncratic draws is as follows:[20]

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

where shape parameter $\theta > 0$ captures the randomness or dispersion in the choice of routes from i to j.[21] 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:

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

where $\tilde{\tau}_{ijr}$ is the product of all leg-specific costs $t_{r_{k-1},r_k}(\Xi, \varepsilon_{r_{k-1},r_k})$ 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.[22] 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).[23]

This structure is consistent with a host of mechanisms, including but not limited to port-level effects and leg-level scale economies.[24] 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.[25] 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 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 ($p_{ijnr}(\omega)$),

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

[21] 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).

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

[23] 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)).

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

[25] 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), we do not find diseconomies of scale using non-linear least squares. See section 6 for further discussion.

is lower than all other prices of good ω from all other origin country-route combinations. 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:

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

By the law of large numbers, this is also the share of goods sold in j in industry n coming from i and taking route r. 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

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

where $\circ$ is the element-by-element Hadamard power.[26] 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:

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

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

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

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

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:

$$\pi^{kl} = \sum_{n}\sum_{j}\sum_{i} \pi_{ijn}^{kl} = \sum_{n} t_{kln}\left(\Xi, \varepsilon\right)^{-\theta} \cdot \sum_{j} \Theta_{jn}\tau_{ljn}\left(\Xi, \varepsilon\right)^{-\theta} \cdot \frac{\Phi_{kn}}{\Phi_{jn}}, \quad (8)$$

where Θ_{jn} is j's global consumption share of industry n. Because optimal route selection

[26] 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 $\left((1+\theta)/\theta\right)^{-\theta}$.

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.

Mapping our model into the data requires that 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.[27] That is, to interpret results from Equation (11) as elasticities, they are deflated by θ. The functional form in Equation (11) presumes scale economies exist at the leg level. In Section D.1, we discuss alternative specifications.

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.

[27] In our model, θ serves as both the route dispersion parameter and trade elasticity. As an alternative, we can model a nested elasticity and decompose the total trade elasticity into a transportation route elasticity of substitution and non-transportation component, estimating the former using the observed dispersion of routes in the US microdata.

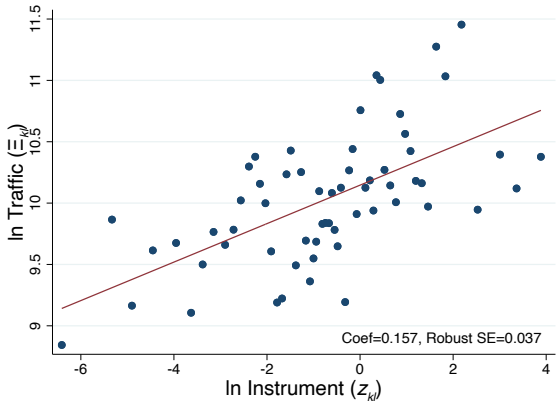
For example, consider routes from origin South Korea to destination the Netherlands. Routes that include a China-Singapore link are closer to the direct Korea-Netherlands route compared to routes that include the China-Australia link. As such, more Korea-Netherlands trade should flow through the China-Singapore 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:

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

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 $Pop_{j,1960}$.[28] Figure 5 shows the robust first-stage relationship between our instrument and traffic.

Figure 5: Residualized Plot of Correlation Between Instrument and Traffic



Notes: The figure shows a binned scatter plot of 1,946 observations of link kl with the logarithm of sea distance between k and l as a control. The x-axis is the logarithm of the instrument z_{kl}. The y-axis is the natural log of traffic on leg kl. Standard errors are clustered two ways by nodes k and l.

For plausible identification, our demand shifter instrument has to be generally uncorrelated with unobserved 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

[28] 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. While this squared deviation functional form has an intuitive interpretation, our analysis is robust to other functional forms.

land distance and may have easier access to other modes of transportation like road or rail. As a robustness check, we recalculate our instrument in equation (12) in a simplified setting by omitting the shortest 10 percentile distances for each origin i and destination j respectively and find similar results.

As previously noted, the observed scale economy in our setting can be generated by a number of mechanism, including but not limited to internal or external scale economies and market power. These mechanisms may generate different out of sample results, and 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).[29] Our traffic data comes from our global port of call AIS shipping data.[30] 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.[31] 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

[29] 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.

[30] 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.

[31] We use 2014 US Customs data on containerized and non-containerized shipments to construct the share of each HS 4-digit commodity code that is transported by container. All commodities with a containerized share above 80% are labeled as containerized. 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.

consider the mapping:[32]

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

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

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

where β_0 is an intercept, β_1 considers the sea distance between the nearest principal ports,[33] 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 trade flows from ports in i to j. Finally, β_6 is an indicator for a shared land border.[34]

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 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.

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

[33] For each country pair, we calculate the volume-weighted mean sea distance across all port pairs.

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

5.4 Joint Estimation

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

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

We conduct a two-stage GMM procedure, using optimal instrumental variable weights estimation for the first set of moments m_1, which accounts for our causal 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 inference can only be conducted on α. β contains incidental parameters, important for estimation, but not inference.[35]

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.

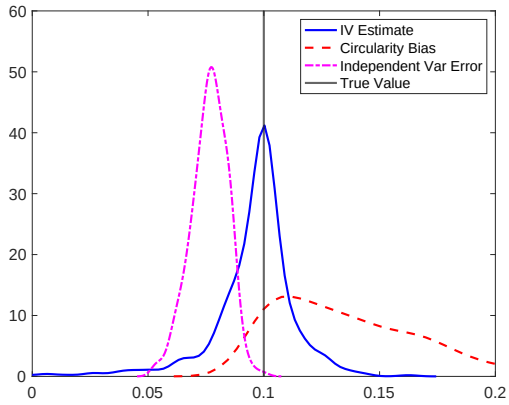
However, this approach opens the door for a mechanically-driven result. Specifically, we are concerned with estimating the causal scale impact of traffic volumes on trade cost (Equation 11) while, at the same time, our cost estimates are themselves recovered from our model prediction which is a function 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

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

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.

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 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 dependent 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.

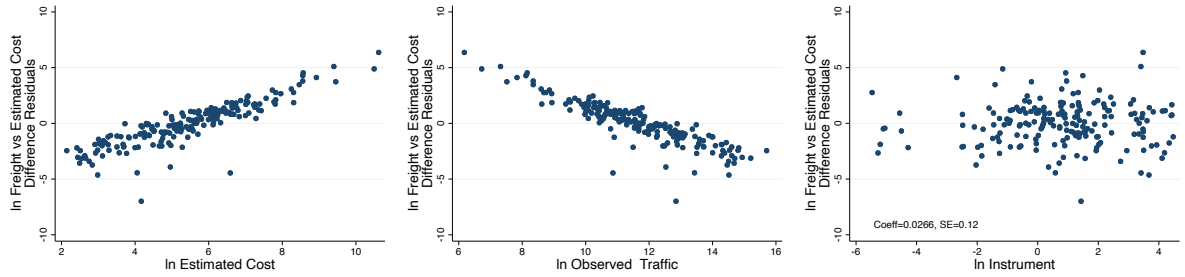
We show the lack of correlation between our instrument and an approximation of the error, estimated as the difference between our measured costs and external measures of freight costs from Wong (2022). Details for this exercise are found in Appendix D.2. Panels (A) and (B) of Figure 7 show a positive 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.

Figure 7: Balancing Test

(A) Approximated Error vs Estimated Costs

(B) Approximated Error vs Observed Traffic

(C) 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 (2022) 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.

6 Results

Scale Economy Table 1 reports our instrumented scale elasticity from our scale moments (Equation (11)). For the widely used trade elasticity value of $\theta = 4.4$ (Simonovska and Waugh, 2014), the interpretation of our causal estimate is that increasing traffic volume on a link by 1% would reduce costs by 0.06%. As the typical journey observed in our microdata has 2.5 links, this translates into a 0.17% decrease in overall origin-destination trade costs. Our estimate is within one standard error of Hummels and Skiba (2004) who estimates an elasticity of freight to quantity of 0.18 using an IV and trade data from six importers and Asturias (2020) who reports an elasticity of 0.23 using US port data.[36] Additionally, Skiba (2017) reports an elasticity of 0.26 using product-level import data from Latin America. Our estimate is also broadly consistent with the literature on scale in production. Bartelme et al. (2019) estimates a sector-level scale elasticity of 0.13 while Lashkaripour and Lugovskyy (2019) finds an elasticity of 0.19 after jointly estimating both scale and trade elasticities.

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

[36]The six importers in Hummels and Skiba (2004) are Argentina, Brazil, Chile, Paraguay, Uruguay, and the US. We compare our estimates to theirs for all countries since our scale estimate is based on global data.

Table 1: GMM Estimation Results

	(1)
	$\ln(c_{kl})$
$\ln\left(\Xi_{kl}^{data}\right)$	-0.28
	(0.02)
$\ln(\mathrm{d}_{kl})$	0.59
	(0.03)
Constant	4.06
	(0.35)

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 link kl. $ln\Xi_{kl}^{data}$ is the natural log of traffic volume on link kl. $\ln(d_{kl})$ is the natural log of sea distance between k and l computed using Dijkstra's algorithm.

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.

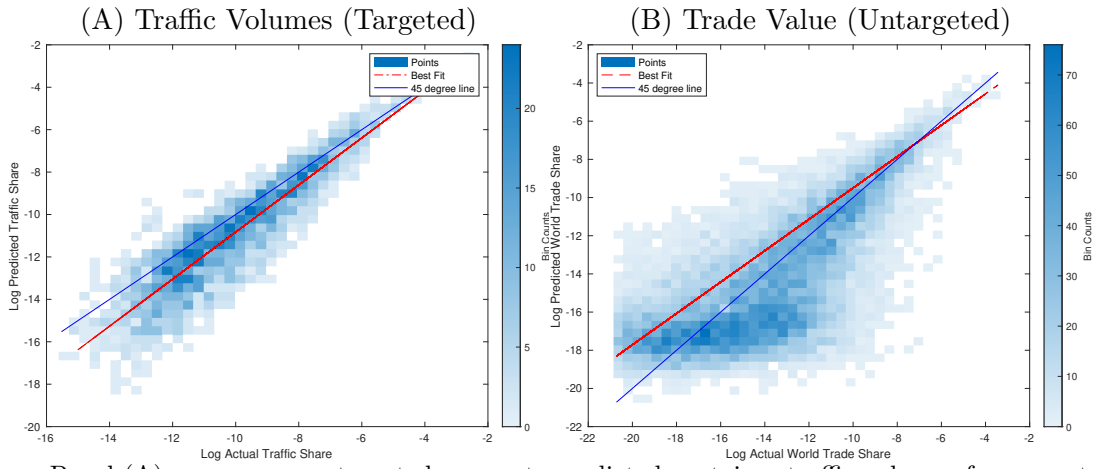
Robustness and Alternative Specifications First, to mitigate the risk of model misspecification (in Equation (11))—e.g. port-level economies or diseconomies of scale, we explore alternative specifications. Adding origin or destination fixed effects increases the magnitude of leg-level scale economy. We choose our current specification which yields a more conservative scale measure. We also search for, but do not find, nonlinearities in our estimated scale economy indicative of port congestion or scale economies that would result in altered counterfactual costs. Additionally, as an alternative estimation approach for Equation (11), we use external cost measures—freight rates from 140 bilateral pairs from Wong (2022)—and find similar, larger, but noisier point estimates (Section D.2.4). These pecuniary freight rates are available for just a subset of routes compared to our setting and do not include all possible elements of link trade costs that are consistent with our model.

Finally, locations that are strategically close to each other in sea distance are close in land distance and potentially have easier access to alternative modes of transportation like road or rail. We recalculate our instrument omitting the shortest 10 percentile distances

for each origin-destination-pair and find that our results retain the same signs and stay within a standard error of our baseline estimates.

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 0.9. Panel (B) compares our estimated trade shares to actual observed trade shares, which we do not target.[37] We fit the data well here as well with a correlation (in logs) of 0.7.

Figure 8: Model Fit Comparisons



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 external data. 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.

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

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

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.[38] Lacking data to directly test these mechanisms, we turn to one symptom of a scale economy observable in our US microdata which lends further credibility to our results: ship size. Relying on the idea that larger ships enable lower shipping costs (Cullinane and Khanna, 2000), we consider the correlations between ship sizes, trade volumes, and our recovered leg-level trade costs and then investigate the relationship between indirect shipping and ship size.

Ship Sizes, Traffic Volumes, and Recovered Trade Costs In Panel (A) of Figure 9, we show the 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. In Panel (B), using the route-level containership size measure, we show the positive link between ship size and our corresponding recovered trade costs. Routes with more container traffic use larger ships; a 10% increase in route volumes correspond to a 2% increase in ship size (Column (1), Table A.8). Routes with lower trade costs use larger ships. A 10% decrease in our estimated iceberg trade costs corresponds to 6% increase in ship sizes (Column (1), Table A.9).[39]

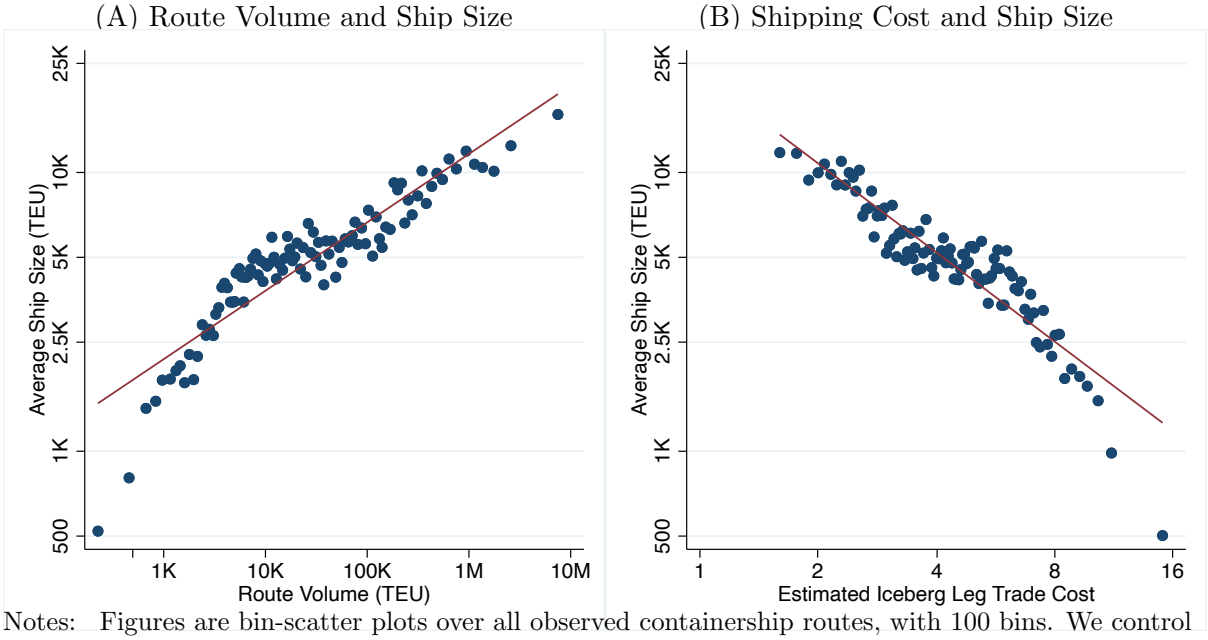
Ship Size and Indirect Trade Figure 10 further investigates the relationship between entrepôt usage and ship size, plotting ship size (x-axis) against US-bound traffic volume (y-axis) by country of origin, separately for traffic that is routed through an entrepôt and traffic that is not, such that each 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.[40]

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

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

[40] 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 9: Link Between Recovered Trade Costs and Ship Size



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

7.2 Cost Estimates with Freight Rates Data

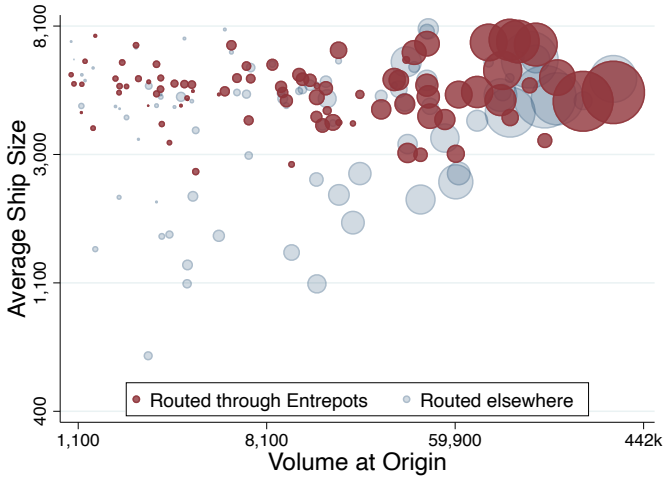
Next, we compare our expected trade cost estimates τ_{ij} at the origin-destination level with container freight rates from Wong (2022). These rates are the costs paid by firms to transport a standard full container load between port pairs and include the base ocean rate, fuel surcharge, as well as terminal 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 0.7 (Figure 11).

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

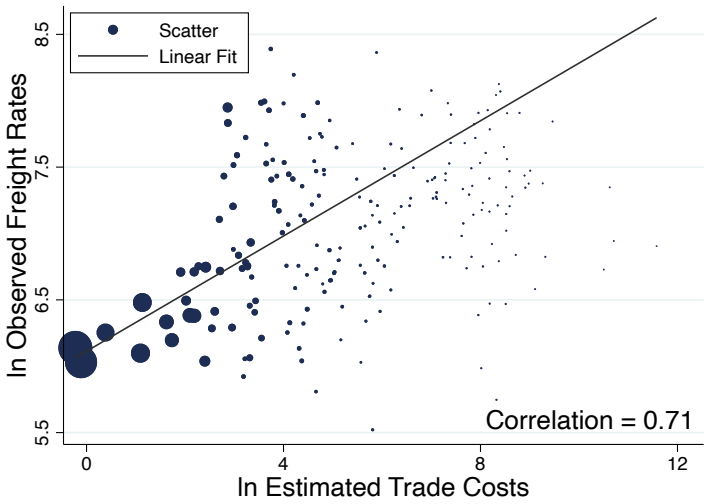
$$\widehat{\pi_{iUS}^{kl}} = [\tau_{ik} t_{kl} \tau_{lj} \tau_{ij}^{-1}]^{-\theta} \tag{13}$$

Figure 10: Link Between Indirect Trade and Ship Size



Notes: The x-axis shows the total export volume in TEUs 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 specific to the route (either through an entrepôt or not). 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.

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 (2022); Drewry Maritime Research (2014). Circle size are weights for container volumes (TEU).

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 $\pi^{kl}_{iUS,Data}$, by aggregating shipments using link kl in our microdata. Note that while our microdata is described in Section 2 and used to generate our stylized facts in Section 3, it is not used to estimate our trade costs in Section 5. Column (1) of Table 2 reports the univariate regression outcome between these two measures, weighted by total origin TEU. We find that a significantly positive relationship, with a coefficient of 1 in the confidence interval. Over half of the

variation in the observed distribution can be explained using the predicted probabilities.

Table 2: Correlation Between Traffic Estimates With Microdata

	(1) $\widehat{\pi^{kl}_{iUS}}$	(2) $\widehat{\Xi^{kl}}$	(3) $\widehat{\pi}^l_{US} - \widehat{\pi}_{l,US}$	(4) $\widehat{\pi^{kl}_{iUS}}$	(5) $\widehat{\Xi^{kl}}$	(6) $\widehat{\pi}^l_{US} - \widehat{\pi}_{l,US}$
$\pi^{kl}_{iUS,Data}$	0.844			0.870		
	(0.117)			(0.119)		
Ξ^{kl}_{Data}		1.217			1.233	
		(0.123)			(0.121)	
$\pi^l_{US,Data} - \pi_{l,US,Data}$			0.942			0.963
			(0.224)			(0.220)
Observations	13763	650	95	365330	2149	186
Data				All	All	All
R^2	0.518	0.666	0.420	0.517	0.675	0.425
F	51.88	97.98	17.68	53.15	103.9	19.20

Standard errors clustered by origin and destination countries.

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

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

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

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

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

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

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

In the microdata, a number of legs have zero traffic volumes. However, our model predicts some small amount of traffic on every leg. In Columns (4) through (6), we re-run the regressions for each corresponding predicted traffic estimate including legs with zero observed volumes (increasing our observation count). Our results do not significantly change 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.[41] 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

We quantify the welfare importance of the trade network and the specific role entrepôts play within that network in three counterfactual exercises. 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.

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.

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

8.1 Counterfactual Methodology

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.

Closing the model We adopt the Caliendo and Parro (2015) framework and 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 $\dot{t}_{kl} = t'_{kl}/t_{kl}$, which change expected trade costs $\dot{\tau}_{ijn} = \tau'_{ijn}/\tau_{ijn}$, and (2) changes in non-transportation trade costs $\dot{\kappa}_{kl} = \kappa'_{kl}/\kappa_{kl}$. Both alter the endogenous costs of production, price indices, wage levels, trade flows, and welfare. We solve for how wages and prices change $\left\{\dot{w}_i, \dot{P}_i\right\}$ as a function of changes to model primitives, $\left\{\dot{\tau}_{ijn}, \dot{z}_{in}, \dot{\kappa}_{ijn}\right\}$, and compute changes in marginal costs $\dot{c}_{in}$ and trade volumes $\dot{X}_{ij}$.

Additional Data We combine our trade volume data with country-level input-output data from the EORA database, aggregating to three sectors: non-traded, containerized traded, and noncontainerized 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 conservatively set $\theta = 4$ (Simonovska and Waugh, 2014).[42]

Procedure Changes to transport costs are implemented as changes to link costs $\dot{t}_{kl}$, which, translated through the model, generate changes in the expected trade cost between every bilateral trading pairs in our data—even those that are not directly connected with each other. Once calculated, these bilateral changes enter isometrically to changes in bilateral non-transportation costs. For analysis which includes the impact of scale, we model a new equilibrium in the short-to-medium run, by following an iterated procedure in Algorithm 1 in Appendix F.1. In this procedure, we start at today's equilibrium and allow all shippers to optimize their transportation patterns. We then recalculate trade costs at new volumes according to Equation (11). We iterate, allowing re-optimization until a new stable equilibrium is reached. Our model theoretically admit multiple equil-

[42] An alternative approach decomposes the total trade elasticity into a transportation route elasticity of substitution and non-transportation component, estimating the former using the observed dispersion of routes in the US microdata.

libria (as in (Brancaccio, Kalouptsidi and Papageorgiou, 2020)), however we focus on the unique equilibrium from our current starting point—the world today.[43]

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 in the network excluding countries' own welfare change, for both cases with and without scale responses. Our 15 entrepôts dominate this list. Singapore and Egypt are top two, evocative of the strain in global supply chains when the Suez Canal was blocked in March 2021 (WSJ, FT, AP).[44] Scale economies' impact on the transportation network (overlaid grey bars) further augment the differential impact of entrepôts.[45] Infrastructure investments at entrepôts generate on average 10

[43] 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.

[44] Within drybulk shipping, Brancaccio, Kalouptsidi and Papageorgiou (2020) finds that removing the Suez Canal has a higher welfare impact than the Panama Canal and Strait of Gibraltar.

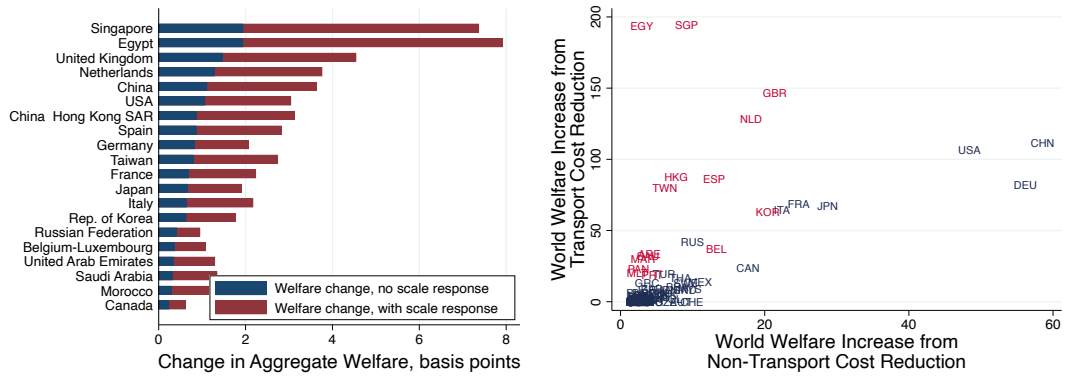
[45] Panel (A) of Appendix Figure A.14 repeats the exercise for cases with non-transportation cost reductions, finding that the top 20 list is dominated by the largest economies instead.

times the global welfare impact relative to investment elsewhere (Columns (3) and (4), Appendix Table A.13).[46]

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 aggregate net change in global welfare after infrastructure investment in the targeted country, excluding the country's own welfare change, for the 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, by 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). Entrepôts are labelled 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: the effects of one type of counterfactuals will be a poor predictor of the other for *any given country*. For entrepôts, (red in Panel (B)), the 1-to-1 relationship is violated. For example, Egypt ranks top two 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.[47] 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

[46] Appendix Tables A.13 and A.14 examines the differential impact of targeting entrepôts.

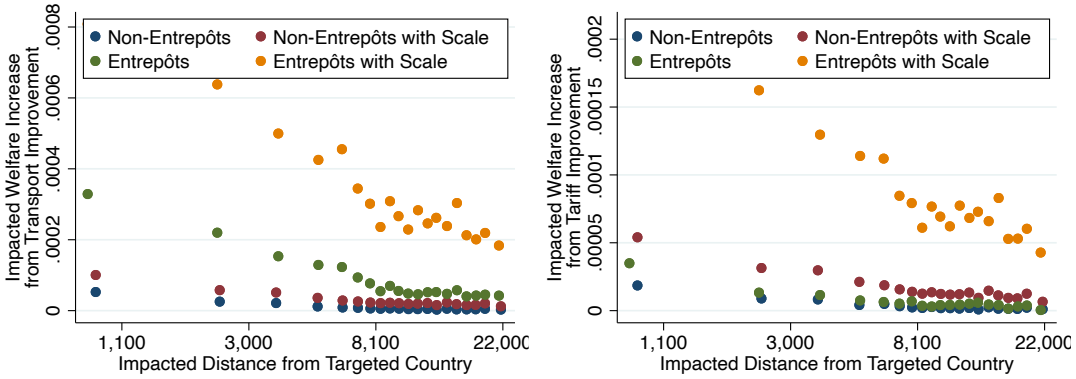
[47] Panel (B) Appendix Figure A.14 finds similar results comparing non-transportation cost reductions with and without an endogenous scale response. Country-pair bilateral trade results are similar.

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.[48]

Figure 13: Spatial Decay of Benefits By Entrepôt Status

(A) Decrease in Transportation Costs (B) Decrease in Non-Transportation Costs



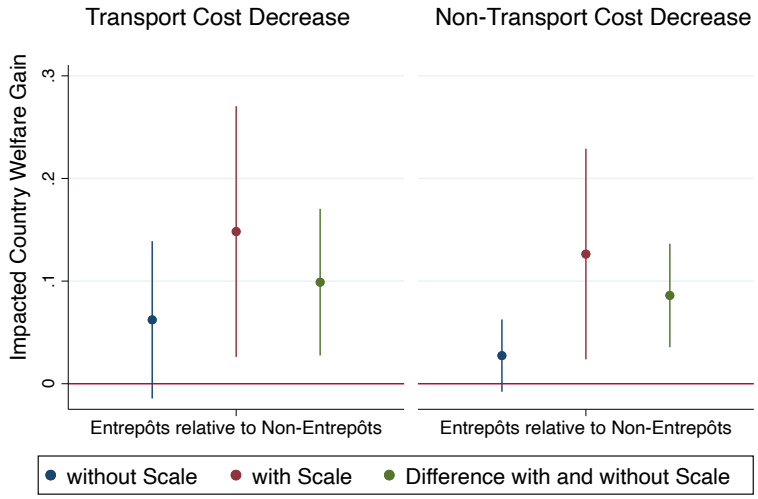
Notes: Panel (A) is a binned scatter of welfare effects of transportation infrastructure on impacted countries vs distance between targeted and impacted countries. Targeted countries receive the cost reduction and impacted countries trade with them. Blue and red dots are the no-scale and scale cases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

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,

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

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

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, controlling for impacted country GDP, bilateral distance, and targeted country fixed effects. Targeted countries receive the cost reduction and impacted countries trade with them. Standard errors are clustered by targeted country.

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.

We model two cases: first without, then with the impact of scale on the trade network. In our first case, as in a traditional model, outcomes are only affected through changes in trade with the UK or multilateral resistance. However, with scale economies, the decrease in UK trade will raise trade costs of neighboring countries through the 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 3 reports aggregate effects. The direct effect decreases global welfare by 2.3 basis points (Column (1)). The introduction of scale economies leads to a decrease of 9 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 that rely on UK feeder routes to get their goods to large vessels that ply transoceanic trade (Table A.15).

Table 3: Welfare and Trade Impact of Brexit and Artic Passage Opening, Basis Points

	Direct Effect	Network Effect	Total Effect (Network & Scale)
	(1)	(2)	(3)
Panel (A) Brexit: Impact of Non-Transport Trade Costs			
Δ Average Global Welfare	-2.3		-9.2
Δ Container Trade Volumes	-19.5		-100.9
Panel (B) Arctic Passage: Impact of Endogenous Trade Costs			
Δ Average Global Welfare	1.4	2.9	6.4
Δ Container Trade Volumes	23.2	44.0	94.8

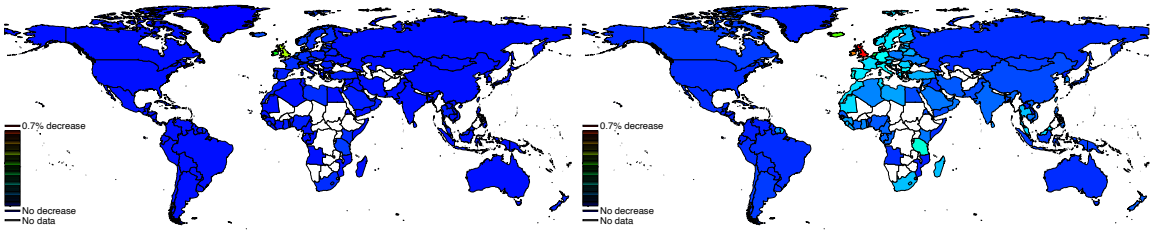
Notes: Panel (A) presents results for Brexit, a 5% increase in non-transportation trade costs κ_{ij} between UK and its trading partners. The direct effect of Brexit only accounts for changes in direct trade with the UK or multilateral resistance. The total effect allows for the change in direct UK trade 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 scale impact.

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

Figure 15: Welfare Changes - Brexit

(A) Direct Effect from Tariff Change (B) Total Effect: Network and Scale



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. 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.

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 existing shipping routes today and shortest ocean-going distance of these routes after the Arctic sea passage is viable.

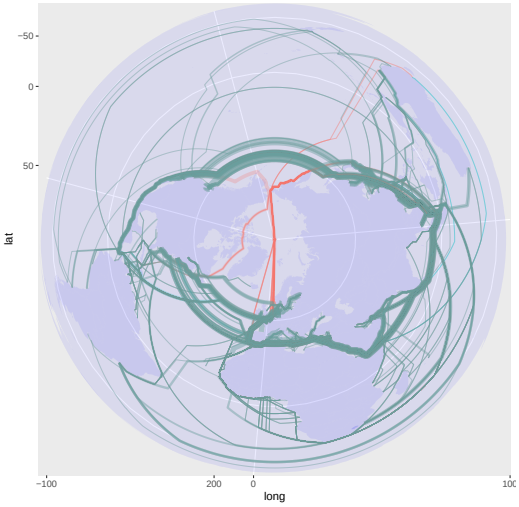
We compare three different cases. First, we consider a network-naive exogenous trade cost case where we only allow for changes in origin-destination trade costs between country pairs for which the direct bilateral distance decreased. Second, for all observed links with positive traffic, we recalculate t_{kl} using new distances with the option of traveling through the Arctic Passage and α_2 in Equation (11). Here, countries without direct connections through the passage—e.g. China and Ukraine—experience trade cost changes due to the trade network effects. Third, we repeat the second case accounting for the impact of scale: as trade costs change, trade volumes change, reducing trade costs further.

Assuming exogenous trade costs with our input-output structure, Column (1) of Table 3 Panel (B) shows that the network-naive and direct effects of the Arctic Passage are positive, with aggregate welfare increasing 1.4 basis points, and container trade volumes increasing 23 basis points. Endogenizing trade costs to allow for the trade network impact of the passage—including indirect shipping—doubles the aggregate welfare effect to 2.9 basis points and increases worldwide container volumes by 44 basis points (Column (2), Table 3 Panel (B)). Allowing for both scale and network effects triples and doubles the welfare and trade impact relative to the network results.

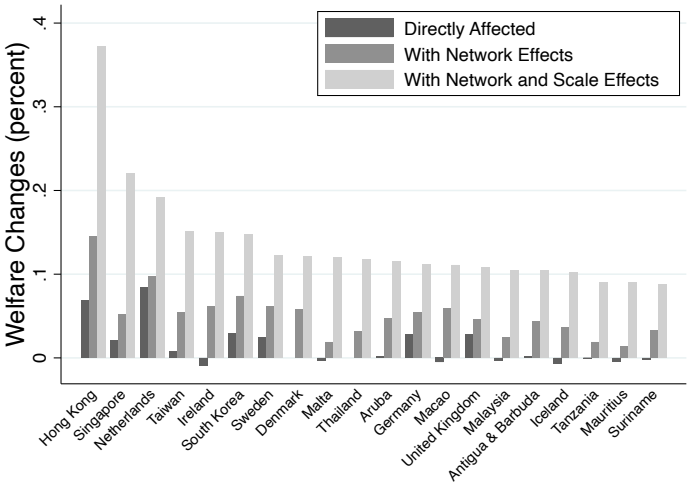
Panel (B) in Figure 16 plots the top 20 most impacted countries, showing gains are particularly pronounced in East Asian entrepôts like Hong Kong and Singapore which disproportionately benefit from the scale economy. Scandinavian countries also gain due

Figure 16: The Opening of the Arctic 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. Most global shipping utilize similar routes which results in many overlapping brown lines. Route width reflects the number of containers (TEU). Panel (B) shows the percent change in welfare of the simulated opening of the Arctic 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.

to their geography. Denmark and Finland, which in the baseline first case have zero or a small trade diversion impact, gain due to the trade network and scale response. Appendix Figure A.17 shows changes in the relative wage-adjusted price index.

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 stylized facts, we develop a general equilibrium model of world trade with endogenous trade costs and entrepôts, estimating both the underlying trade costs on all routes, and scale economies. We quantify the impact of the trade network on global trade and welfare, highlighting how changes at nodes operate through the network, entrepôts, and scale economies to create widespread impacts. We find that infrastructure investments at entrepôts generate on average 10 times the global welfare impact relative to investment elsewhere.

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. And while our estimates of scale economies are agnostic to underlying mechanisms, future work should consider the roles of fixed costs in enabling the scale economies in containerized shipping, especially the costs incurred by potential oligopolies in setting shipping networks and the endogenous creation of firm-specific hub-and-spoke networks. In particular, we can account for leg-level monopolies and variable markups but not within-firm spillovers in sea route selection. While sector-specific research has been done on these networks, future work should consider a tractable general equilibrium framework 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*.

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.

Berry, Steven T. 1992. "Estimation of a Model of Entry in the Airline Industry." *Econometrica: Journal of the Econometric Society*, 889–917.

Bertoli, Simone, Michaël Goujon, and Olivier Santoni. 2016. "The CERDI-seadistance database." HAL Working Papers halshs-01288748.

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."

Ciliberto, Federico, and Elie Tamer. 2009. "Market structure and multiple equilibria in airline markets." *Econometrica*, 77(6): 1791–1828.

Conte, Maddalena, Pierre Cotterlaz, Thierry Mayer, et al. 2021. "The CEPII gravity database." *CEPII: Paris, France.*

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.

Drewry Maritime Research. 2014. "Freight Rate Data."

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.

Encyclopedia Britannica. 2022. "Landlocked Country."

Fajgelbaum, Pablo D, and Edouard Schaal. 2020. "Optimal transport networks in spatial equilibrium." *Econometrica*, 88(4): 1411–1452.

Feenstra, Robert C, Robert Inklaar, and Marcel P Timmer. 2019. "Penn World Table Version 9.1." *Am. Econ. Rev*, 105: 3150–3182.

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, and Alexandre Skiba. 2004. "Shipping the good apples out? An empirical confirmation of the Alchian-Allen conjecture." *Journal of Political Economy*, 112(6): 1384–1402.

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."

Kelso, Nathaniel Vaughn, and Tom Patterson. 2010. "Introducing natural earth data - naturalearthdata. com." *Geographia Technica*, 5(82-89): 25.

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.

Maritime Administration, US Department of Transportation. 2014. "U.S. Waterborne Foreign Container Trade by U.S. Customs Ports 2000 – 2017." last accessed April 2023.

OECD. 2018. "Annex Table 55: Import Penetration."

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. 2022. "The round trip effect: Endogenous transport costs and international trade." *American Economic Journal: Applied Economics*, 14(4): 127–66.

World Bank. 2018. "Development Indicators."

Online Appendix

Entrepôt: Hubs, Scale, and Trade Costs

Sharat Ganapati, Woan Foong Wong, and Oren Ziv

Appendix A Data Construction

A.1 Shipment Microdata

We compile and combine two proprietary microdata sets in this project: global ports of call data for all containerships, which allows us to reconstruct the routes taken by specific ships, and United States bill of lading data for containerized imports, which gives us shipment-level data on imports into the United States. Independently, each of these datasets allows us to partially describe the global shipping network. By merging them, we are able to reconstruct nearly the entire journey most shipments entering the United States take, from their initial origin point or place of receipt to the port of entry into the United States. To our knowledge, we provide the most comprehensive reconstruction of the global shipping network and routes undertaken by individual shipments into the United States (Panjiva, 2014; Astra Paging, 2014; CEPII, 2017; KGM Associates, 2014).

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

Our sample covers a six-month period, from April to October 2014. Over this period, we have information on 4,986 unique container ships with a combined capacity of 30.6 million TEU. This represents over 90% of the global container shipping fleet. Ports with no AIS receivers or where information is not shared do not show up in our data. In addition, if transponders are turned off or transmissions are not recorded, ports of call can be missed. However, transponders are required to be operational by the International

Maritime Organization on ships engaging in international voyages 300 gross tons, applying to all containerships in our sample (International Maritime Authority, 2003).

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

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

Over six months of US imports from April to October 2014, we see a total of 14.8 million TEUs weighing 106 million tons were imported into the US from 221 shipment origin countries and 144 countries with ports of lading. This accounts for about three quarters of the 2014 TEU and tonnage imports, 77 percent and 74 percent respectively (Maritime Administration, US Department of Transportation, 2014).[2] The countries in our data are categorized using the three-digit alphabetical codes assigned by the International Organization for Standardization (ISO) by the Statistics Division of the United Nations Secretariat. Accordingly, Hong Kong, Macau, Taiwan, as well as dependencies and areas of special sovereignty like Guam have their own designated codes. Non-containerized goods, including goods on roll-ons (vehicle carriers), bulk cargo liners (for commodities), and non-containerized cargo ships are not observed in our data.

Our data captures the following location information for each shipment into the US: the foreign location where the shipment originated from (*shipment origin*), the foreign port where it was loaded on the containership which brings it into the US (*port of lading*), and the US port where it was unloaded from the containership (*port of unlading*). In addition, we know the name and identification number of the containership (IMOs) which transported the shipment as well as the shipment's weight, number of containers (TEUs), and product information.

This data set allows us to start tracing the journey of a shipment from its origin to its

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

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

destination US port. In particular, we can determine whether this shipment was loaded at its origin location onto the vessel that brings it directly to its final US destination, or if it went through at least one other location during its journey. When matched with the port of call data, we can reconstruct most of its remaining journey after the port where it was loaded onto a US-bound vessel (from its port of lading).

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

First, we identify containerships using Vessel IMOs. Vessel IMOs are identifiers unique to containership vessels and stay constant for the lifetime of their operation. By IMO, we identify about 4000 ships in the Bills of lading data. An additional (roughly) 2,000 ships are matched to IMOs using a fuzzy string match, after which matches are made by hand with the help of undergraduate research assistants.

Second, we match the port calls that the containerships make with the ports of arrival of shipments. Ports of arrival are recorded using UNLOCODEs in the AIS port of call data and US Census Schedule D codes in the Bills of Lading data. We construct a crosswalk to match these ports with the help of undergraduate research assistants.

Third, we match the port calls that the containerships make with the ports of lading of shipments. Ports of lading are recorded using UNLOCODES in the AIS port of call data and the US Customs and Border Protection's listing of foreign ports (Schedule K) in the Bills of Lading data. We construct a crosswalk to match these ports with the help of undergraduate research assistants.

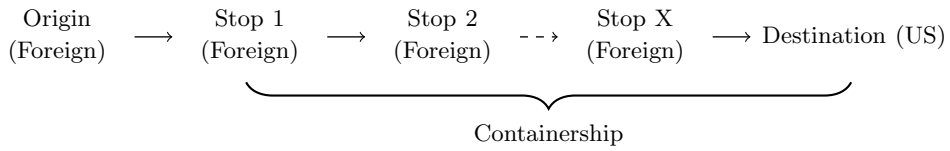
What remains unobserved is the shipment's journey between its Origin and its first stop (port of lading location). This portion of the shipment's journey takes place in a container and could be transported overland (by trucks or rail) or by sea on another containership. While this information is not recorded by both our datasets and therefore unobserved, the amount of indirectness that we establish in our stylized facts is a lower bound since we assume that this portion of the journey is direct. The amount of transshipment that we establish in our stylized facts is also a lower bound since at most we observe one transshipment port. To the best of our knowledge, we capture the most detailed information on shipments' journeys by merging these two datasets.

For each bill of lading, we match ship, date of unlading, and port of unlading to

the AIS data on ships' port of call. Once we match shipments to ships, we record each port of call in the AIS data before the port of unlading as a stop the shipment makes, then remove all stops observed before the ship stopped at the port of lading. If the port of lading is not observed, the route is discarded and the shipment remains unmatched. Furthermore, any routes that include the port of unlading before the date of unlading are discarded, as they represent loops where the port of call for the port of lading is missing.

Over 90% of containerized TEUs entering the US in the bills of lading data can be matched to routes using this method.[3] Appendix Figure A.1 visualizes this merge.

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



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

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

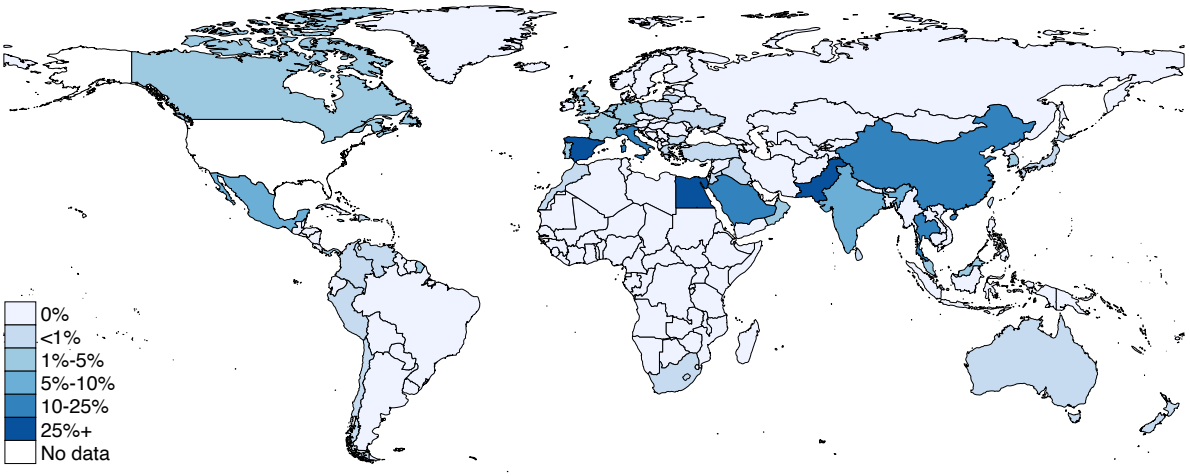
A.2 Geographic Distance Data

Geographic distance data is computed using two rasterized (with pixels) world maps. One map consists of all the navigable oceans and large seas, with a polar ice cap, as well as the Suez and Panama canals. The second map assumes that the Arctic ice sheet melts away due to anthropogenic climate change. In both maps, we compute the sea distance between ports of call, and aggregate to the national level using port-to-port container flows. We do this computation in R using Dijkstra's algorithm on a world map with and without Arctic ice caps.[4] We argument this with distance data from Bertoli, Goujon

[3] Unmatched shipments may have missing and unrecoverable ship information, or ports of call that do not match lading and unlading records on bills of lading. In addition, a small number of reconstructed routes have implicit voyage speeds above 50KPH, and are discarded.

[4] For more information, see the 'gdistance' package and mapping files from Kelso and Patterson (2010).

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



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

and Santoni (2016) and Conte et al. (2021), as well a data on landlocked countries from Encyclopedia Britannica (2022).

A.3 Aggregate Economic and Trade Statistics

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

To aggregate industry trade to industries that use container shipments versus trade that does not, we use aggregate data from 2014 from the United State Customs, as disseminated by Schott (2008).[5] This data reports the share of shipments by HS Codes that arrive by containerships. We consider 4-digit HS Codes as a consistent level of aggregation. The distribution of containership share by HS code is bi-modal, with one peak around 0% and another around 100%. We use a cutoff of 80%. So HS codes that are shipped by containership to the US over 80% of the time are classified as "containerizable" trade.

For aggregate trade and economic statistics for using in the counterfactual, we use the Eora global supply chain database with a multi-region input-output table (EORA-

[5]This data has been continually updated by the author following the initial publication

MRIO).[6] We collapse all world trade into three categories; those that are non-tradable, those that are typically traded over oceans by containerized vessels, and those that are not typically traded over oceans by containerized vessels.[7] We again classify industries using the methods of Schott (2008). We augment this with GDP data from Feenstra, Inklaar and Timmer (2019); World Bank (2018); OECD (2018).

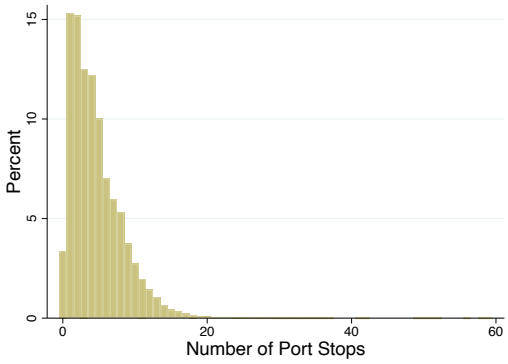
Appendix B Additional Descriptive Results

In this section we report additional results and robustness checks related to the analysis in Section 3.

B.1 Additional Indirectness Results

Figure A.3 reports the histogram of number of port stops minus the port of lading if the port of lading is in the country of origin, and the port of unlading. We exclude landlocked countries. The mean number of third-port stops is 4.6 and fewer than 5% of shipments do not stop at additional ports.

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



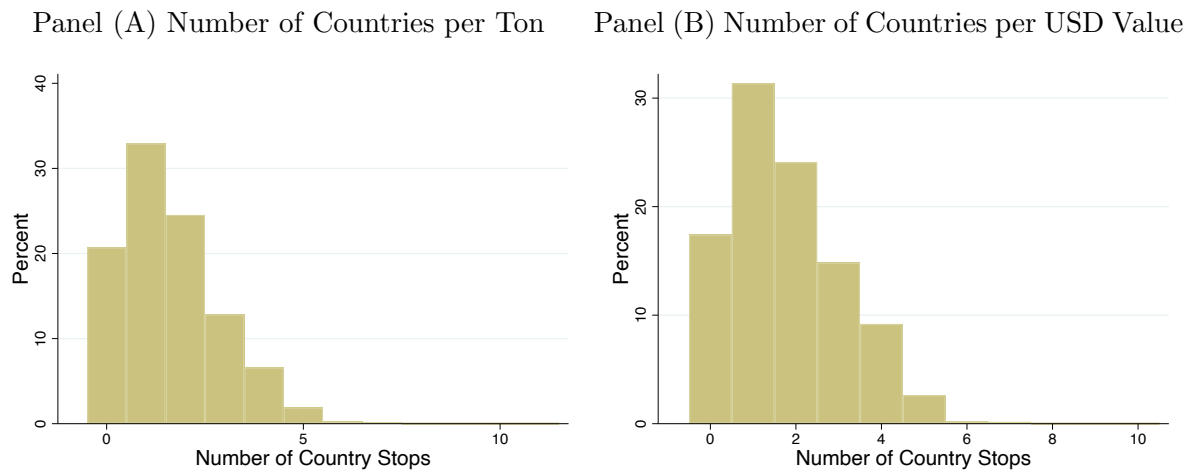
Notes: This figure reports the distribution at the shipment level of the number of unique port stops minus the port of lading if the port of lading is in the country of origin, and the port of unlading, weighted by shipment TEU. Shipments from landlocked countries are excluded.

Next, in Figure A.4, we rerun the analysis in Panel (A) of Figure 2 weighting by Tons in Panel (A) and USD in Panel (B). For the latter, a minority of shipment data report dollar values. Overall, the results are similar to our main results using TEU.

Figure A.5 reports the percent of shipments loaded onto a US-bound ship in a third-party country by country of origin. Countries that are closer and trade more with the US are less likely to transship goods at third-party countries—a fact we explore in more detail in Appendix B.4.

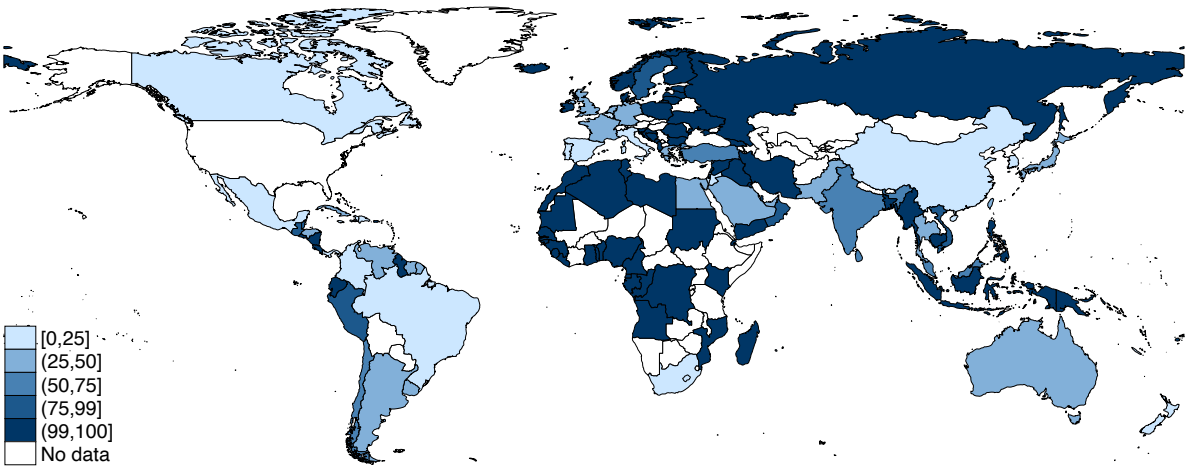
[6] Freely available for academic use from https://worldmrio.com/.
[7] This includes bulk shipping, roll-on roll-off ships, as well as air freight.

Figure A.4: Distribution of Third-Party Countries Involved in Bilateral Trade by Weight and Value



Notes: Panel (A) reports the distribution of the total number of unique third-party country stops made by shipments entering the US, weighted by shipment tons (kg). Panel (B) reports the same but weights by value for the portion of shipments for which value measures are reported.

Figure A.5: Transshipped Trade Share between Origin and US Destination



Notes: This figure plots for each country the share of its originated shipments transshipped in a third-party country, weighted by TEU. Lighter colors indicated lower levels of transshipped trade share (ie. more direct trade). The US is not included since it is the destination country. Landlocked countries are also excluded. 34 of the shipment origin countries are landlocked accounting for 1.6 percent of total TEUs. The missing remaining countries are either due to lack of overall trade with the US (e.g. Somalia) or due to the merge process (e.g. Namibia).

Finally, we further explore the result that additional stops increase the distance and time costs of trade. In Table A.1, we regress, at the shipment level, log of observed distance (Columns (1)-(4)) and time (Column (5)), on the number of country stops made by a shipment. All port distances are computed using Dijkstra's algorithm, and time is computed by the difference in AIS logs for port of lading and unlading. Results are clustered two ways by port of unlading and port of lading.

Column (1) reports the baseline relationship: an elasticity of 0.112 (SE 0.022) on

stops, controlling for the computed direct sea distance between the port of lading and port of unlading. Adding port of lading fixed effects (Column (2)) or port of unlading fixed effects (Column (3)) does not significantly change the result. In Column (4), we add port of lading-by-unlading fixed effects. Here identification comes from variation between routes where goods come on and off boats at exactly the same ports, but where different ships take different routes (the existence of this variation is explored further in Appendix B.4). The elasticity here remains stable as 0.104 (SE 0.03). Column (5) repeats our most heavily controlled-for exercise in Column (4) but with time traveled as the variable of interest. We find an elasticity of 0.333 (SE 0.0819) which implies that for shipments loaded and unloaded at the same ports, routes with double the stops along the way increase journey time by 33%.

Table A.1: The Relationship Between Indirectness, Distance, and Time

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

Notes: This table presents regression coefficients for regression of ln Observed Distance, the natural log of sea distance traveled between all reported ports of call, or ln Time Travelled, the natural log of time between port of lading and port of unlading, and ln Country Stops, the natural log of unique third-country stops, as well as ln Direct Distance, the natural log of the sea distance between the port of lading and unlading. Distances are calculated using Dijkstra's algorithm and measured in kilometers while time is measured in hours. Observations are shipment level and weighted by TEU. Shipments originating in landlocked countries are omitted. Standard errors in parentheses are clustered two ways by the port of lading and port of unlading.

B.2 Additional Concentration Results

List Countries by Entrepôts Activity. Table A.2 reports our index of entrepôt activity for all countries in our data, using data on trade and transportation that have been adjusted as in Section 5 and normalized so that the lowest value (for the US) is zero.

Countries towards the top of the list have more third-country activity, with the 15 countries at the top of this list defined as entrepôts for the purposes of our counterfactual analyses. These include Egypt (Suez Canal) and Singapore. Countries in the middle of the list neither differentially depend on nor are used as third countries. Countries that

are small and/or less open dominate this section like Papua New Guinea or North Korea. Countries towards the end of the list differentially depend on others as third countries (like Ireland and Malaysia), with the bottom of the list dominated by the largest economies, who account for large portions of global trade but not large portions of global traffic (like China and Germany).

Table A.2: Entrepôt Index by Country

Country Name	Index Value	Country Name	Index Value	Country Name	Index Value
Egypt	11.42	Congo	5.77	Ecuador	5.73
Singapore	10.39	Barbados	5.77	Bangladesh	5.72
Netherlands	10.22	Suriname	5.77	Tunisia	5.71
China Hong Kong SAR	8.65	Aruba	5.77	Angola	5.70
Belgium-Luxembourg	7.78	Guinea	5.77	Iraq	5.70
Taiwan	7.26	New Caledonia	5.77	Croatia	5.70
Spain	6.99	Lao Peoples Dem. Rep.	5.77	Qatar	5.69
Saudi Arabia	6.72	Mauritania	5.77	Peru	5.69
Rep. of Korea	6.68	Ghana	5.77	Bulgaria	5.68
United Arab Emirates	6.63	Cyprus	5.77	Viet Nam	5.65
Morocco	6.47	Nicaragua	5.77	Nigeria	5.64
Panama	6.44	Georgia	5.77	Chile	5.62
Malta	6.30	Dem. Peoples Rep. of Korea	5.77	New Zealand	5.60
Portugal	6.17	Madagascar	5.76	Kazakhstan	5.60
United Kingdom	6.09	Albania	5.76	Algeria	5.60
Greece	5.99	Honduras	5.76	Venezuela	5.56
Bahamas	5.94	Lithuania	5.76	Kuwait	5.56
Pakistan	5.90	United Rep. of Tanzania	5.76	Romania	5.55
Israel	5.88	Mauritius	5.76	Malaysia	5.54
Lebanon	5.87	Papua New Guinea	5.76	Finland	5.50
Russian Federation	5.85	Mongolia	5.76	So. African Customs Union	5.50
Jamaica	5.83	Cambodia	5.76	Ukraine	5.49
Uruguay	5.83	Slovenia	5.76	Iran	5.49
Dominican Rep.	5.82	Cameroon	5.76	Poland	5.46
Sri Lanka	5.81	Gabon	5.76	Philippines	5.45
Djibouti	5.79	Brunei Darussalam	5.75	Australia	5.45
Benin	5.78	Côte dIvoire	5.75	Argentina	5.44
Senegal	5.78	Guyana	5.75	Indonesia	5.30
Togo	5.78	Trinidad and Tobago	5.75	Brazil	5.30
Colombia	5.77	Belarus	5.75	Denmark	5.24
Gambia	5.77	Yemen	5.75	Ireland	5.23
Liberia	5.77	Iceland	5.75	Thailand	5.23
Somalia	5.77	Latvia	5.75	Norway	5.21
Eritrea	5.77	Paraguay	5.75	Czech Rep.	5.12
Antigua and Barbuda	5.77	Kenya	5.75	Mexico	5.02
Cabo Verde	5.77	Turkey	5.75	Sweden	4.99
Greenland	5.77	Cuba	5.75	Switzerland	4.93
Cayman Isds	5.77	Libya	5.74	France	4.90
Belize	5.77	Guatemala	5.74	India	4.80
Sierra Leone	5.77	Bolivia Plurinational State of	5.74	Austria	4.70
Montenegro	5.77	China Macao SAR	5.74	Italy	4.63
Mozambique	5.77	Syria	5.74	Canada	4.45
Maldives	5.77	Estonia	5.73	China	4.10
Haiti	5.77	Costa Rica	5.73	Japan	4.09
Bahrain	5.77	Oman	5.73	Germany	3.38
				USA	0.0

Notes: Table presents measure of entrepôt activity, calculated, as defined in Section 3, as the percent of global trade minus the percent of global traffic, with adjustments made for overland traffic, with the US normalized to zero.

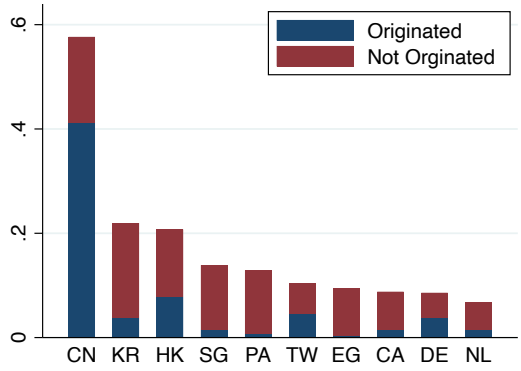
Concentration of US-Bound Shipments Panel (A) of Figure A.6 tabulates, for each of the top ten countries, the percent of all goods entering the US stopping in that

country. The share of shipments accounted for by shipment origination is in blue while shipments observed stopping in the country but not originating in the country is in red. Unsurprisingly, many recognizable entrepôts are listed, including Korea, Panama, Singapore, and Egypt. Perhaps more surprisingly, more than 50% of the containers entering into the US stop in China. While this panel sums to over 1, since each container stops in more than one country, over 80% of shipments to the US stop in at least 1 of 5 countries: China, Panama Singapore, Korea, or Egypt.[8]

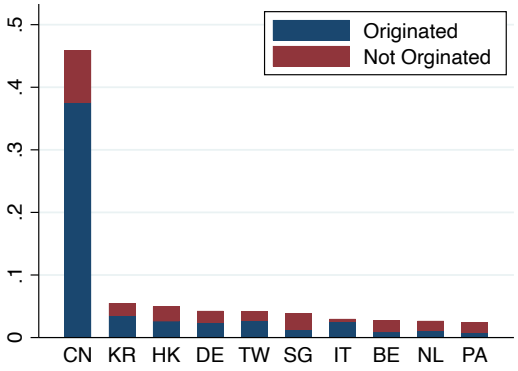
Panel (B) replicates Panel (A) but for the country of lading. Here the total of all bars (including those not graphed) sum to 1, and China again dominates as a source of lading. A few of these top countries, like Germany in (A) and Italy in (B) are majority blue, implying they are important to the US because of their role as an origination country. Other countries, like Singapore, are differentially red, and appear important as entrepôt rather than as countries of origin.

Figure A.6: Roles of Countries in Bilateral Trade: Origin vs Entrepôts

(A) Share of Shipments Stopping in Country, for Top Ten Countries



(B) Share of Shipments Laded in Country, for Top Ten Countries



Notes: The blue portion in Panel (A) highlights the share of all incoming US shipments that originate in the indicated country while the red accounts for the percent of all incoming US shipments stopping in that country (not originated), weighted by TEU.
Panel (B) replicates Panel (A) but for country of lading.

B.3 Spokes Disproportionately Use Entrepôts

Conceiving the shipping network as a hub and spoke system implies that spokes largely access their trading markets using hubs. While in Section 3 we find the network is characterized by having hubs, we clarify here that the excess concentration of shipments at entrepôts are in part due to their disproportionate use by smaller, less well-connected

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

Table A.3: Concentration Ratios

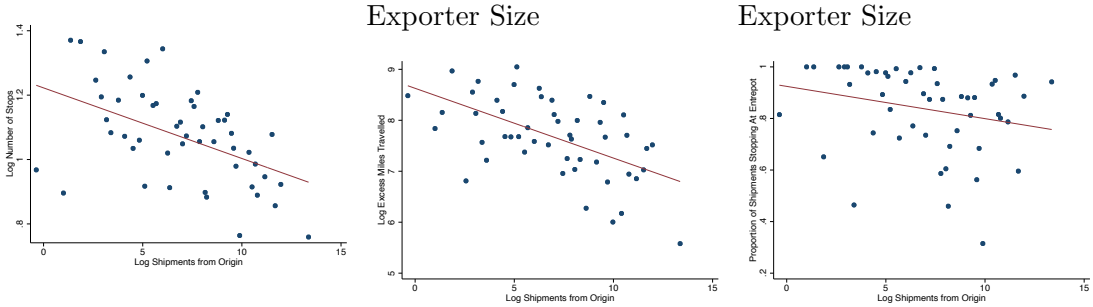
	Third-Party Stops	Transshipment	Trade
Max/50	429	476	400
99/50	390	476	76
95/50	215	135	27
90/50	120	91	15

Notes: Data present concentration ratios across countries in our data. Third-party Stops are the sum total TEU-weighted shipments that use a country as a third-party country. Transhipments are the TEU-weighted sum total of shipments transshipped at a country, and Trade is the total volume of trade from a country. Countries are ranked and percentile ratios are presented. For example, the country used the most (by TEU shipments) as a third-country stop acts as such for 429 times the number of shipments stopping at the median (50th-percentile) country.

origins, or, in other words, the spokes of the network.

Figure A.7: Smaller Exporters Are Disproportionately Indirect and More Likely to Use Entrepôts

(A) Stops vs Exporter Size (B) Excess Distance vs Exporter Size (C) Entrepôt Usage vs Exporter Size



Notes: Binned scatter plots with observation at the origin level weighted by total TEU with 50 bins. The x-axis for each of the three panels features the size of each origin's exports to the US. Panel (A) shows the relationship between the origin's size and its average number of stops before its US destination. Panel (B) shows the relationship between the origin's size and the average excess distance traveled by its exports before its US destination. Panel (C) shows the relationship between the origin's size and the share of its exports which stopped at an entrepôt before its destination.

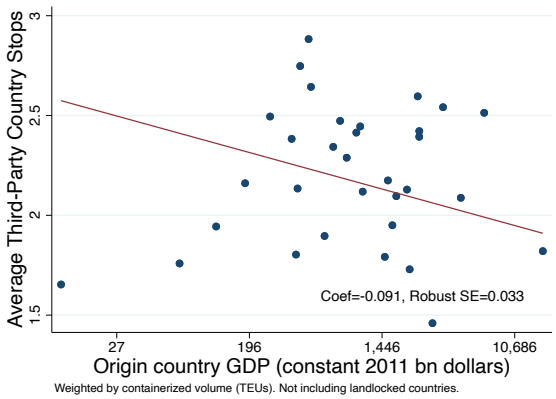
In Figure A.7, we zoom in on the set of origins that are simultaneously the most indirect and most likely to send goods through hubs. The three panels are binned scatterplot with 50 bins of origin-level measures of average TEU (A) number of stops, (B) excess distance, and (C) likelihood of passing through an entrepôt. Panel (A) of Figure A.7 confirms that smaller origins are more indirectly connected to the US, and Panel (B) confirms that shipments from smaller origins move further distances to get to their destination. Panel (C) shows that shipments from smaller origins are more likely to use entrepôts. These relationships are echoed in shipment-level regressions which add additional controls and cluster by origin. In sum, the smallest origins constitute the spokes of the hub-and-spoke network.

B.4 Variation in Connectivity

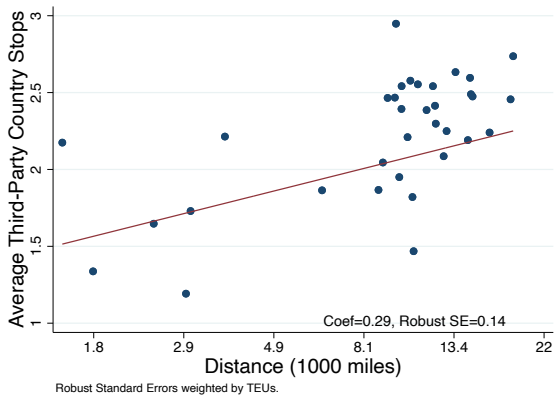
There is a high degree of variance in indirectness across countries, as shown in Figures 2 and A.5. This variation is reasonable explained by traditional gravity variables. In Panel (A) of Figure A.8, we find that countries with higher GDPs are more likely to have less stops on their journeys to the US. In Panel (B), we find that countries which are closer to the US are more likely to have less stops on their journeys (i.e. have more direct trade with the US). These results are robust to using port stops instead of country stops (Table A.4) as well as to weighting by containers, tons, and value. One natural interpretation of this would be the endogenous response of shippers to the scale of shipments from these countries. Of course, the availability of direct trade to the US could in principle reverse the causality.

Figure A.8: Larger and Closer Countries Have Lower Number of Average Stops

(A) Stops vs Country Size



(B) Stops vs Distance



Notes: Binned scatter plots with observation at the origin level weighted by total TEU. Landlocked countries are excluded.

Table A.4: Relationship Between Stops and Country Size as Well as Distance

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

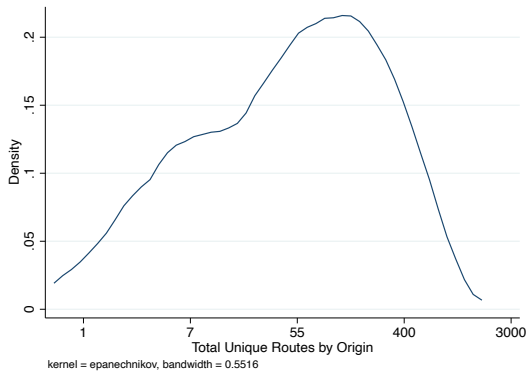
Notes: This table presents coefficients from country-level regression of ln Cty Stops, the natural log of the TEU-weighted average number of third-party country stops made for shipments from country against ln GDP, the natural log of the country's GDP, and ln Distance, the natural log of the sea distance between the countries. Observations are weighted by total TEU. Landlocked countries are excluded. Robust standard errors in parentheses.

Do shipments from a given origin follow a unique path to the US? Panel (A) in Figure A.9 shows the distribution in the number of unique routes to the US by origin country. With an average of about 397 routes with wide variation (sd 681), observed routes from a single origin are indeed varied. The countries with the highest number of unique routes are big trading partners like China, the United Kingdom, Germany, and well-established entrepôts like Hong Kong. Countries with the lowest unique routes are smaller trading partners like American Samoa, Nauru, Tonga, and Montserrat. The existence of this within-origin route variation will be a particularly important assumption in our model and external validity checks.

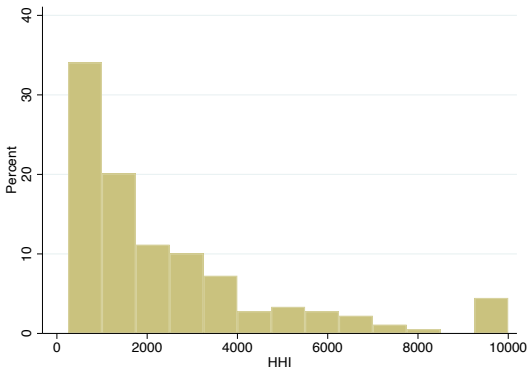
We can measure the concentration of these unique routes by constructing a Herfindahl-Hirschman Index (HHI) for each origin country using the container shares of each route. Panel (B) in Figure A.9 shows that almost 70 percent of origin countries have fairly low concentrations of routes (HHI less than 1500) The average HHI overall is 1475 (sd 1974). Examples of countries with high levels of concentration are like Vanuatu, Cuba, and Liberia while countries with low levels of concentration are Macau, Hong Kong, and Belgium-Luxembourg.

Figure A.9: Variation in Trade Indirectness

(A) Number of Unique Routes by Origin-Destination Pair



(B) Distribution of Route Concentration



Notes: Panel (A) plots the kernel density plot for the total number of unique routes from a given origin country. Panel (B) plots the distribution of the HHI index for routes from each country.

Appendix C Additional Theoretical Results

C.1 Definition of Entrepot

The share of imports from origin i to destination j in industry n which passes through leg kl is:

$$\pi_{ij}^{kl} = \left[(c_i\kappa_{ij})^{-\theta} \cdot b_{ik}a_{kl}b_{lj}\right] \cdot \Phi_j^{-1}, \tag{14}$$

where $a_{ij} = t_{ij}^{-\theta}$ and $b_{ij} = \tau_{ij}^{-\theta}$.

Summing over shipment origins, we write the share of global shipping to a destination j that goes through kl as follows:

$$\begin{aligned}\pi_j^{kl} &= \sum_i \left[(c_i\kappa_{ij})^{-\theta} \cdot b_{ik}a_{kl}b_{lj}\right] \cdot \Phi_j^{-1} \\ &= \sum_i \left[(c_i\kappa_{ij})^{-\theta} \cdot b_{ik}\right] a_{kl}b_{lj} \cdot \Phi_j^{-1} \\ &= \Phi_k a_{kl}b_{lj} \cdot \Phi_j^{-1}\end{aligned}$$

Summing over all k's generates the share of traffic to j which flows through node l:

$$\begin{aligned}\pi_j^{l} &= \sum_k \Phi_k a_{kl}b_{lj} \cdot \Phi_j^{-1} \\ &= b_{lj} \cdot \Phi_j^{-1} \cdot \sum_k \Phi_k a_{kl}\end{aligned}$$

Now, the total share of shipments originating at l and going to j is:

$$\pi_{lj} = \left(c_l\kappa_{lj}\right)^{-\theta} b_{lj}\Phi_j^{-1}.$$

We define node l's measure of Entrepôt with respect to destination j as

$$Entrepôt_{lj} \equiv \pi_j^l - \pi_{lj} = \sum_k \Phi_k a_{kl}b_{lj} - \left(c_l\kappa_{lj}\right)^{-\theta} b_{lj} \tag{15}$$

which is the difference between node l's weighted network position with respect to destination j – how close to j other locations k in the network are when moving through l, where weights are multilateral resistance at k – and the marginal cost of production and transport from l. The former predicts transit, as higher values here mean l is in a more important position in the network to move goods to j, while the latter predicts exports from l to j.

We further note that, holding constant all other leg level costs $a_{k'l'}$ for $k' \neq k \vee l' \neq l$, a reduction in leg-level trade cost to a node l –i.e. an increase in a_{kl}, increases this measure

for l. In particular,

$$\frac{d\pi_j^l - \pi_{lj}}{da_{kl}} > 0 \tag{16}$$

noting that $d\Phi_j/da_{kl} > 0$ and $db_{lj}/da_{kl} > 0$.

This functional form has the convenient property that it aggregates from micro- to macro-data – as we see below – and thus allows for consistency in measure between our micro-level approach in Figure 4 and our macro-level approach when estimating the model and counterfactuals.

Similarly we can write share of global trade moving through kl as

$$\pi^{kl} = a_{kl} \cdot \sum_j b_{lj} \Theta_j \frac{\Phi_k}{\Phi_j}. \tag{17}$$

where Θ_j is country j share of global GDP. And, summing across locations k,

$$\pi^l = \sum_k \Phi_k a_{kl} \cdot \sum_j \Theta_j \frac{b_{lj}}{\Phi_j}. \tag{18}$$

The global share of trade from l as

$$\pi_l = \sum_j \Theta_J \frac{\left(c_l \kappa_{lj} \tau_{lj}\right)^{-\theta}}{\Phi_j} \tag{19}$$

The difference between the two is

$$\begin{aligned}
\pi^l - \pi_l &= \sum_k \Phi_k a_{kl} \cdot \sum_j \Theta_j \frac{b_{lj}}{\Phi_j} - \sum_j \Theta_j \frac{\left(c_l \kappa_{lj} \tau_{lj}\right)^{-\theta}}{\Phi_j} \\
&= \sum_j \frac{\Theta_j}{\Phi_j} \left[\sum_k \Phi_k a_{kl} \cdot b_{lj} - \left(c_l \kappa_{lj} \tau_{lj}\right)^{-\theta} \right] \\
&= \sum_j \frac{\Theta_j}{\Phi_j} \left[\pi_j^l - \pi_{lj} \right]
\end{aligned}$$

which is a weighted average of our individual country j measure, where each destination country j measure is weighted by its share of total global trade and network proximity, respectively Θ_j and Φ_j. That is, we can either take (on the left hand side) measure of global shares of traffic and trade, or (on the right hand side) an average of the same difference for each country j derived from micro-data.

C.2 The Network Effect of Adjustments on Trade

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

cost t_{kl}, trade volumes between i and j will adjust according to the following equation:

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

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

What can we say about the signs on these terms? As the trade cost matrix is endogenous to trade volumes, these terms are ambiguous, as a change in t_{kl}, by changing trade volumes, changes traffic volumes at each leg, and therefore equilibrium effects on the full matrix of trade costs.

However, if we consider a change in t_{kl} which holds fixed all other leg costs $t_{k'l'}$ for $k' \neq k \vee l' \neq l$, only the final term can be negative. Intuitively, a reduction in trade costs between k and l can increase consumption at j, reduce expected trade costs between i and j, and reduce production costs at i, all of which result in an increase in trade volumes between i and j. However, a reduction in trade costs between k and l also stiffens competitions at j. If this last effect is large enough, it can overturn the sign of the first three.

In the scale-free case, the total effect is positive if and only if the elasticities of consumption at j ($\epsilon_{X_{jn},t_{kl}}$), production costs at i ($\epsilon_{c_{in},t_{kl}}$), and trade costs between i and j ($\epsilon_{\tau_{in},t_{kl}}$) with respect to t_{kl} are larger than the elasticity of multilateral resistance at j with respect to t_{kl} ($\epsilon_{\Phi_j,t_{kl}}$). Furthermore, $\frac{\partial X_{ijn}}{\partial t_{kl}} > 0$ if and only if:

$$\epsilon_{X_{jn},t_{kl}} + \left[\epsilon_{c_{in},t_{kl}} + \epsilon_{\tau_{in},t_{kl}}\right](1 - \pi_{ijn}) > \sum_{i' \neq i} (\epsilon_{c_{i'n},t_{kl}} + \epsilon_{\tau_{i'jn},\tau_{kl}})\pi_{ijn}. \tag{20}$$

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

C.3 Extension: Market Power

This section addresses the question of firm behavior and how does it fit in with the estimation of scale economies. We investigate this issue using a simple adaptation of the Cournot framework with endogenous entry (Sutton, 1991). Suppose we have origin and destination countries denoted by k and l respectively. Demand for shipping on this route is $\Xi_{kl}(t_{kl})$, where t_{kl} is the equilibrium cost of shipping on that leg, determined by the shippers on that route.

Consider a game with two stages. First, shippers with constant marginal costs c decide to enter after paying cost ϵ. Second, shippers play a nash-in-prices entry game to determine the shipping price t. A particular shipper's i market share on route kl is $s_{i,kl} = \frac{\exp\left(at_{i,kl}\right)}{\sum \exp\left(at_{i,kl}\right)}$ where $i \in 1...N_{kl}$. Total demand for shipping on the route is $\Xi_{kl} = \delta \times \left(\sum_{i=1}^{N_{kl}} \exp\left(at_{i,kl}\right)\right)^{\gamma}$, where a, δ, and γ are constants that governs consumer sensitivity to shipping prices.

Starting with backward induction and the first stage, each symmetric shipper i on route kl will charge a shipping cost

$$t_{i,kl} = \frac{\delta}{a\left[1-(1-\gamma)\, s_{i,kl}\right]} + c.$$

In the first stage, we then determine the number of shippers N_{kl} who are willing to pay entry cost ϵ. This is pinned down by the equation:

$$\frac{\Xi_{kl}}{N_{kl}} = \left[\frac{\delta \Xi_{kl}}{a\epsilon} + (1-\gamma)\right]^{-1}.$$

So in equilibrium,

$$\ln\left(t_{kl}-c\right) = -\ln\left(a\right) - \ln\left(1-(1-\gamma)\left[\frac{\delta \Xi_{kl}}{a\epsilon} + (1-\gamma)\right]^{-1}\right)$$

This relationship is sensitive to competition and market share forms. Additionally we write trade costs in iceberg form and if marginal costs are low, then $c \equiv 1$. A rough approximation is close to our main estimating relation, for some constant ϕ:

$$\ln\left(t_{kl}-1\right) \approx \ln\left(\phi\right) + \alpha \ln\left(\Xi_{kl}^{data}\right).$$

Effectively one source for scale economies comes from an increase market size that increases entry and thus drives down prices.

Appendix D Estimation

This section reports additional details, results, and robustness checks from our estimation strategy, as well as discusses the potential threats to identification.

D.1 Recovery of Predicted Trade Costs

Table A.5 shows the results of our estimation that predicts leg-level trade costs. Positive values for β indicate increases in trade costs and negative values indicate decreases in trade cost.

However, these estimates are not causal, and cannot be used for either inference or counterfactuals. They represent the power of various (including highly endogenous) variables in predicting a trade cost matrix that rationalizes leg-level containerized traffic flow. We find high correlations between observed and model-predicted shares, including for shares that we do not target (Figure 8). We find a correlation between trade shares of 0.7 (which we do not target) and traffic shares of 0.9 (which we target). If we had more possible useful predictive variables, we could use a machine learning technique to tease out the best basis of variables to predict model-consistent trade costs.

Table A.5: Predictive Trade Cost Estimates

Coefficient	Estimate
β_0 (intercept)	7.968
β_1 (log distance)	-0.006
β_2 (log route traffic)	-1.033
β_3 (log outgoing port traffic)	0.273
β_4 (log incoming port traffic)	0.275
β_5 (land borders)	-0.386
β_6 (trade volume)	-0.000

Notes: Results presented here are the moments from the GMM estimation in Section 5. These results are not causal, and cannot be used for either inference or counterfactuals. They represent the predictive power of various (possibly endogenous) variables in predicting a trade cost matrix that rationalizes leg-level containerized traffic flow.

This analysis reflects the spirit of pure prediction and cannot satisfy the "Lucas Critique" as they are purely observational and do not reflect fundamental economic parameters, forces, or relationships. In Section 5 we address endogeneity and causality, using an instrument to find the relationship between route-level volume and trade costs.

D.2 Scale Elasticity and Trade Cost Estimation

In this Appendix we explore the potential for a mechanical relationship between traffic volumes and costs in our model, which is also present in the Allen and Arkolakis (2019)

framework. We first show how such a potential correlation is a form of omitted variable bias and conditions under which an instrument corrects it. We then run Monte Carlo simulations confirming the existence of the bias in the model and showing how our instrument can remove it. We proxy for the bias using the difference between model-generated costs and a small set of external cost estimates, and show both the existence of the bias in the OLS relationship between volumes and cost and that the bias is eliminated by the instrument. Finally, we use our external cost estimates in a parallel estimation and find a similar scale economy.

D.2.1 Identification Strategy

As mentioned in the main text, we recognize that there is traffic volumes and trade costs are endogenous. As a result, we introduce a demand shifter as our instrument to recover the causal impact of traffic on trade costs. In this section, we show how a potential mechanical relationship between traffic volumes and costs in our model can be a form of omitted variable bias and conditions under which an instrument can correct for it. This issue is also present in the Allen and Arkolakis (2019) framework.

Suppose we observe traffic volumes with measurement error ($\Xi_{kl}^{data} = \Xi_{kl} + \chi_{kl}$) and $\hat{t}_{kl}$ is our estimated trade cost as part of the estimation procedure. Our OLS specification for our scale elasticity from Equation (11) would be slightly modified to the following:

$$\ln(\hat{t}_{kl}^{\theta} - 1) = \alpha_0 + \alpha_1 \cdot \ln \Xi_{kl}^{data} + \alpha_2 \cdot \ln d_{kl} + \varepsilon'_{kl}, \tag{21}$$

Suppose the error on trade costs as part of the estimation process, due to mismeasured traffic volumes, is as follows:

$$ln(\hat{t}_{kl}^{\theta} + 1) = ln(t_{kl}^{\theta} + 1) + \nu_{kl}, \tag{22}$$

where t_{kl}^{θ} is the true cost and ν_{kl} is some error in the estimation. The measurement error from ν_{kl} can create a mechanical correlation between $\hat{t}_{kl}^{\theta}$ and Ξ_{kl}^{data} if $Cov(\nu_{kl}, \Xi_{kl}^{data}) \neq 0$, i.e. when the error between the true and estimated costs are correlated with observed traffic flows.

Using Equation (22) in order to recover the true trade costs t_{kl}^{θ} from the OLS specification in Equation (21):

$$\begin{aligned} \ln(t_{kl}^{\theta} + 1) + \nu_{kl} &= \alpha_0 + \alpha_1 \cdot \ln \Xi_{kl}^{data} + \alpha_2 \cdot \ln d_{kl} + \varepsilon'_{kl} \\ \ln(t_{kl}^{\theta} - 1) &= \bar{\alpha}_0 + \bar{\alpha}_1 \cdot \ln \Xi_{kl}^{data} + \bar{\alpha}_2 \cdot \ln d_{kl} + \nu'_{kl} \end{aligned} \tag{23}$$

where $\nu'_{kl} = \varepsilon'_{kl} - \nu_{kl}$, and the mechanical correlation can be interpreted as a stan-

dard concern that the error is correlated with the regressor, in this case through ν'_{kl} if $Cov(\nu_{kl}, \Xi_{kl}^{data}) \neq 0$.

While measurement error on the dependent variable is not unique to this setting, the specific concern here is the measurement error is correlated with traffic. For example, measurement error in observed traffic flows will show up both in Ξ_{kl}^{data} as well as in estimated costs $\hat{t}_{kl}$, and using the estimated costs may recover a mechanical correlation which could bias our scale economy estimates. However, even if this is the case, the instrument can recover the true scale parameter so long as the instrument is correlated with traffic but uncorrelated with the measurement error – i.e. under a specific version of an exclusion restriction.

Using an instrument z_{kl}, the estimate of α_1 from Equation (21) is

$$\alpha_{1,IV} = \frac{Cov(z_{kl}, \hat{t}_{kl})}{Cov(z_{kl}, \Xi_{kl}^{data})}$$

whereas the IV estimate of $\bar{\alpha}_1$ from Equation (23) is

$$\bar{\alpha}_{1,IV} = \frac{Cov(z_{kl}, t_{kl})}{Cov(z_{kl}, \Xi_{kl}^{data})}$$

Crucially, these two estimates are identical if and only if

$$Cov(z_{kl}, t_{kl}) = Cov(z_{kl}, \hat{t}_{kl}) = Cov(z_{kl}, t_{kl} + \nu_{kl})$$

where all variables are residualized for distance. This condition holds if our instrument is uncorrelated with the error in our estimation of leg costs, $Cov(z_{kl}, \nu_{kl}) = 0$. As such, in order to recover the correct coefficient on scale, we have a second restriction, $Cov(z_{kl}, \nu_{kl}) = 0$ in addition to the standard exclusion restriction, $Cov(z_{kl}, \varepsilon_{kl}) = 0$.

D.2.2 Monte Carlo Simulations

Here we run Monte Carlo simulations to demonstrate the potential sources of bias in the OLS estimate of the scale elasticity, and how an instrument satisfying the conditions outlined in the previous section recovers an unbiased estimate. The results are shown in both the main text in Figure 6 and here in Table A.6.

Simulation Procedure We run the following simulation procedure:

1. Generate distances between 15 countries from a unit uniform distribution.

2. Generate a graph with links between countries, where 1/3 of pairs have a bilateral transport link kl.

3. Generate the invariant part of trade costs on link between k and l. This is a linear transformation of the invariant part of trade cost in the main text (Equation (2)):

$$a_{0,kl} = \frac{1}{300} Distance_{kl}.$$

4. Generate origin-destination trade values between countries i and j:

$$X_{ij} = 3 \times Distance_{ij} + \nu_{ij},$$

where ν_{ij} is drawn from a normal distribution with mean 1 and standard deviation 1 ($N(1,1)$).

5. Generate model consistent trade costs that satisfy the following relationship in Equation (10) using matrix notation:[9]

$$\Xi = A \otimes \left(B' \left(X \oslash B\right) B'\right) \tag{24}$$

where Ξ is the true matrix of traffic volumes, lower case a_{kl} denotes the k,l element of the matrix A such that $a_{kl} = \exp\left(a_{0,kl} + \alpha_1 \ln\left(\Xi_{kl}\right)\right)$, lower case b_{kl} denotes the k,l element of the matrix B such that $b_{kl} = \tau_{ij}^{\theta}$, and X is the matrix of trade values (each element is X_{ij}). To translate this back to Equation (10), note that $a_{kl} = t_{ij}^{-\theta}$ and $b_{kl} = \tau_{ij}^{\theta}$ (Equations (2) and (3)). We use $\alpha_1 = 0.01$ for this simulation (True Value, first row, Table A.6).

6. Assume that the econometrician observes mis-measured traffic $\ln\left(\Xi^{data}\right) = \ln\left(\Xi\right) + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is drawn from $N(1,1)$.

7. Generate an instrumental variable Z such that $E\left(Z\boldsymbol{\epsilon}\right) = 0$, but $E\left(Z\Xi\right) \neq 0$.

8. Use our routine from the main text to recover $\hat{A}\left(\tilde{\Xi}, X, Distance\right)$, based on the mismeasured $\tilde{\Xi}$ from step 6. Each element in this $\hat{A}$ matrix is denoted as $\hat{a_{kl}}$.

Specifications We run the four specifications below, 500 times each, and report the median estimate of α_1 and its standard deviation in Table A.6. The true value of α_1 is 0.01 for this simulation (first row, Table A.6). The distributions of each specification are plotted in Figure 6 in the main text.

1. **No Errors Scenario** If we perfectly observe Ξ_{kl} without measurement error, we would be able to generate the true model-consistent trade costs a_{kl} (Equations (10)

[9] We use the notation of Allen and Arkolakis (2019) (See Corollary 1, Equation (22)).

and (24)) and run the following OLS specification:

$$\ln(a_{kl}) = \alpha_0 + \hat{\alpha}_{1,OLS} \ln(\Xi_{kl}) + \alpha_2 \ln d_{kl} + \psi_{kl}$$

where Ξ_{kl} denotes the k, l element of the matrix Ξ and $\ln d_{kl}$ is the log of distance. Note that even though the estimate of a_{kl} is generated from Ξ, there is no bias in our estimates and recovers the true value (second row, Table A.6).

2. **Circularity Bias Scenario** If we do not perfectly observe traffic and instead observe traffic with error $\tilde{\Xi}$ (Step 6 above), we will generate trade costs with measurement error ($\hat{a}_{kl}$) per Step 8. This will lead to the following OLS specification:

$$\ln(\hat{a}_{kl}) = \alpha_0 + \hat{\alpha}_{1,OLS,noise} \ln\left(\Xi_{kl}^{data}\right) + \alpha_2 \ln d_{kl} + \psi_{kl}.$$

The measurement error here biases our OLS estimates upwards, due to the mechanical relationship of our traffic to implied trade costs (third row, Table A.6).

3. **Independent Variable Error Scenario** Here we consider classic measurement error in observed traffic volumes: $\ln\left(\check{\Xi}_{kl}\right) = \ln(\Xi_{kl}) + N(0,1)$. Assuming that our trade costs are estimated correctly, this will result in the following OLS specification:

$$\ln(a_{kl}) = \alpha_0 + \hat{\alpha}_{1,OLS,measurement} \ln\left(\check{\Xi}_{kl}\right) + \alpha_2 \ln d_{kl} + \psi_{kl}$$

This classic measurement error will lead to a classic attenuation bias in the results (third row, Table A.6).

4. **IV with Circularity Bias Scenario** With mismeasured traffic volumes that generate mismeasured trade costs, we run two-stage least squares using the simulated instrument Z from Step 7. The first and second stages of our specification are as follows:

$$\ln\left(\Xi_{kl}^{data}\right) = \beta_0 + \hat{\beta}_{1,IV,noise} \ln(Z_{kl}) + \beta_2 \ln d_{kl} + \psi'_{kl}$$
$$\ln(\hat{a}_{kl}) = \alpha_0 + \hat{\alpha}_{1,IV,noise} \ln\left(\Xi_{kl}^{data}\right) + \alpha_2 \ln d_{kl} + \psi_{kl}$$

The instrumental variable approach restores the upward bias of the measurement error (last row, Table A.6).

D.2.3 Circularity Bias and Geographic Instrument

As is generally the case with an exclusion restriction, we cannot directly test the condition $Cov(z_{kl}, \nu_{kl}) = 0$ (Section D.2.1). However, we can proxy for ν_{kl} by comparing our model's

Table A.6: Monte Carlo Estimates - With Scale

Estimate		Median Estimate	Standard Deviation
α_1	True Value	0.10	
$\hat{\alpha}_{1,OLS}$	No Errors	0.10	0.00
$\hat{\alpha}_{1,OLS,noise}$	Circularity Bias	0.13	0.04
$\hat{\alpha}_{1,OLS,meausrement}$	Independent Var Error	0.08	0.01
$\hat{\alpha}_{1,IV,noise}$	IV with Circularity Bias	0.10	0.02
	N	500	

Notes: The distribution of these results is shown in both the main text in Figure 6. The estimate for $\hat{\alpha}_{1,OLS}$ in the No Errors specification (second row) shows no bias and recovers the true value. The estimate for $\hat{\alpha}_{1,OLS,noise}$ in the third row illustrates our upward circularity biases and shows a larger scale economy that the true scale economy. The estimate for $\hat{\alpha}_{1,OLS,meausrement}$ shows how different our upward bias is from classic attenuation bias (the fourth row). Lastly, our estimate for $\hat{\alpha}_{1,IV,noise}$ shows how our instrument corrects for this upward attenuation bias and recovers the true value for our scale economy a_1 (last row).

estimates of leg costs t_{kl} with external estimates of pecuniary shipping costs from Wong (2022).

We calculate an estimate of our model's mismeasurement $\hat{\nu}_{kl}$ as follows:

$$\hat{\nu}_{kl} = t_{kl} - t_{Wong,kl}.$$

for the 209 links for which we have external freight cost estimates from Wong (2022), residualizing both for distance.

In Panels (A) and (B) of Figure 7, we plot a scatter plot of this estimated mismeasurement against traffic and our estimated costs, controlling for sea distance. The existence of a correlation between the three is consistent with a potential circularity bias. In the context of our Monte Carlo Simulations, this correlation opens the door to an upward bias in an OLS estimate of the scale elasticity.

In Panel (C), we plot the same estimated mismeasurement against our geography-based instrument z_{kl}. Here the correlation in Panel (A) vanishes. While we caution that this lack of correlation is not evidence that the exclusion restriction is met, as that condition is inherently unknowable, this exercise can be thought of as a balancing test, where the observed correlation between proxies for the unobserved error and the endogenous variable is not present with the instrument.

D.2.4 Estimating Scale Elasticities without Imputed Costs

In order to test the robustness of our identification strategy, we find a similar scale elasticity using observed freight rates for only a subset of routes from Wong (2022). First, we find suggestive evidence for potential scale economies for this subset of routes

where we directly observe freight rates for. These findings further support our findings in Section 7.1 on the presence of scale economies in our context. This result holds even when we include origin and destination-level port fees, which would be correlated with port-level congestion. Second, we apply our instrument to this subset of routes, but due to a small number of observations and a relatively weak first stage, cannot reject the null hypothesis.

We note that estimating our scale elasticity using this approach has two main drawbacks. First, external pecuniary freight rates such as those in Wong (2022) do not include all possible elements of network leg costs that are consistent with our model. Second, our goal is to estimate a global set of leg-level trade costs and a global dataset on observed freight rates does not exist. Estimating a scale elasticity within this context creates both a power issue (as we show below) and an external validity issue. Nevertheless, these estimates provide a measure of scale elasticity that is free of any potential bias from our trade cost estimation and has the potential to indirectly confirm the results from our instrumented estimation.

First, we find a statistically significant and negative correlation between freight rates and traffic in Table A.7. In Column (1), we find this negative correlation using all-in freight rates which is the sum of base freight rates of the route, origin port fees, destination port fees, and bunker fuel. Distance between routes is included as a control and is positively correlated with freight rates. Using an even smaller set of routes for which we observe base freight rates directly, we find that the coefficient between freight rates and traffic in Column (2) retains the same sign and is within one confidence interval of the results in Column (1). Given that we observe origin and destination port fees for this even smaller subset of routes, we can include these fees in Column (3). These fees are potentially correlated with congestion at the origin and destination ports. Including these proxies for origin and destination port congestion, the coefficient between base freight rates and traffic retains the same sign and is within one standard error of the results in Column (2). We conclude that this is further suggestive evidence for the presence of scale economies in this context.

Second, we apply our instrument to this subset of routes and find a scale elasticity that is within a standard error of our results in Table 1. Due to the small number of observations, however, this estimate is noisy and our first stage is weak.

Table A.7: Correlation between Freight Rates and Traffic for Subset of Routes

	(1) All-in FR	(2) Base FR	(3) Base FR
Traffic	-0.0478	-0.116	-0.108
	(0.0241)	(0.0337)	(0.0256)
Distance	0.404	0.555	0.516
	(0.0848)	(0.137)	(0.104)
Origin Fees			-0.164
			(0.241)
Dest Fees			0.557
			(0.308)
Specification	OLS	OLS	OLS
Observations	142	142	142
R^2	.38	.24	.29

Notes: All variables are in logs. Robust standard errors clustered by origin and destination ports in parentheses. weighted by route trade values. The all-in freight rates used in Column (1) are the costs paid by firms, in dollar terms, to transport a standard full container load between port pairs. These all-in rates include the base ocean rate, fuel surcharge, as well as port handling fees at both origin and destination. For a much smaller subset of routes, we observe the direct origin and destination fees breakdown of these freight rates. The base rate is used in Column (2) while the base rate and port fees are used in Column (3). In order to make this comparison directly, the observations are restricted to routes where the direct breakdown is observed.

D.3 Ship Sizes, Trade Volumes, and Recovered Trade Costs Robustness

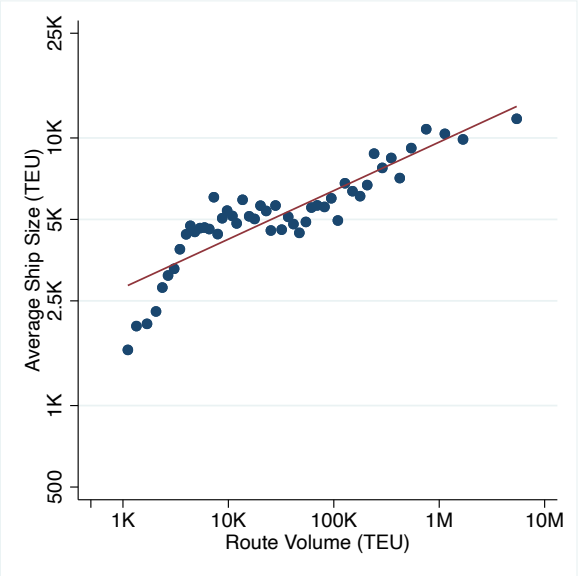
Figure A.10 replicates main text Figure 9 without distance controls. Results are broadly similar. Tables A.8 and A.9 replicate Panels (A) and (B) in Figure 9 in regression form respectively. Columns (2)-(4) sequentially add controls for route distance, origin fixed effects, and destination fixed effects. Results are broadly consistent with baseline results. Table A.8 shows a 10% increase in trade volumes is correlated with a 2.2-2.5% in average ship sizes. Table A.9 shows a 10% decrease in estimated trade costs corresponds to 6.3-10.6% increase in average ship sizes.

D.4 Shipment-Level Data: Ship Size

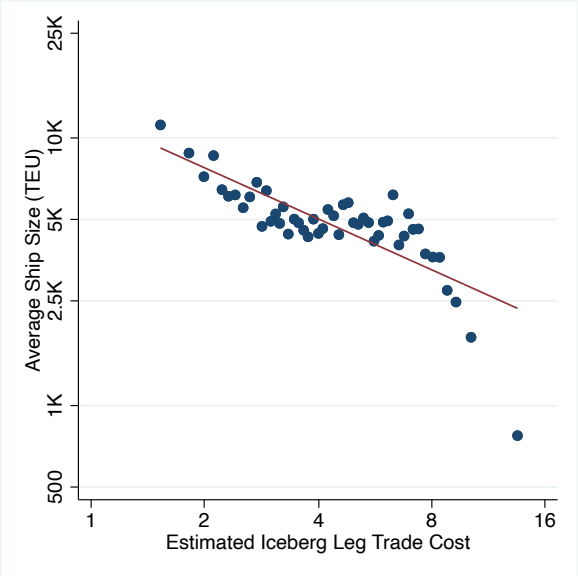
We pair the visual analysis in Figure 10 with Table A.10, which displays shipment-level regressions. Column (1) regresses, for our sample of shipments, the log of ship size against the log of total origin country volumes shipped (TEUs), confirming a positive relationship. Column 2 adds the log of quantity laded at each shipment's port of lading– the port where the shipments are loaded onto a US-bound ship (Stop 1 in Figure A.1). Both coefficients are positive but the coefficient on origin volume is almost halved (0.084 in Column (1) compared to 0.043 in Column (2)), indicating that much of the correlation

Figure A.10: Link Between Recovered Trade Costs and Ship Size - No Distance Controls

(A) Route Volume and Ship Size



(B) Shipping Cost and Ship Size



Notes: These figures are bin-scatter plots over all observed containership routes, with 100 bins. (A) plots the relationship between the total containers on a route and the average containership's size on that route. (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.

Table A.8: Correlation Between Route Volume and Ship Size

	(1) log(Ship Size)	(2) log(Ship Size)	(3) log(Ship Size)	(4) log(Ship Size)
log(Trade Volume)	0.223	0.245	0.238	0.216
	(0.00671)	(0.00609)	(0.00607)	(0.00727)
FE Origin	0	1	1	1
FE Good	0	0	1	1
FE Origin-Good	0	0	1	1
R^2	0.315	0.515	0.627	0.703
N	2304	2304	2304	2304

Notes: We consider the relationship between the total containers on a route and the average containership's size on that route. Containership size reflects the size of the ship for the average container on that route. We use robust standard errors. Column (2), controls for logarithm of shipping distance. Column (3), adds controls for the origin port. Column (4) adds fixed effects for the destination port.

between origin volume and ship size acts through the size of the lading port. Column (3) fully interacts the variables in Column (2) with an indicator variable for shipments that are laded in their origin countries. As suggested by the figure, for shipments whose origin country differs from lading country—an indicator value of 0—the correlation between ship size and lading volume is considerably higher (0.130), and shipments' ship sizes are not strongly correlated with origin country volumes when they lade in third countries (0.009).

Finally, stopping at larger ports matters, even when goods remain on board: goods

Table A.9: Correlation Between Shipping Costs and Ship Size

	(1) log(Ship Size)	(2) log(Ship Size)	(3) log(Ship Size)	(4) log(Ship Size)
log(Trade Cost)	-0.632	-1.060	-0.986	-0.842
	(0.0353)	(0.0310)	(0.0279)	(0.0282)
FE Origin	0	1	1	1
FE Good	0	0	1	1
FE Origin-Good	0	0	1	1
R^2	0.136	0.457	0.596	0.703
N	2304	2304	2304	2304

Notes: We consider 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. We use robust standard errors. Column (2), controls for logarithm of shipping distance. Column (3), adds controls for the origin port. Column (4) adds fixed effects for the destination port.

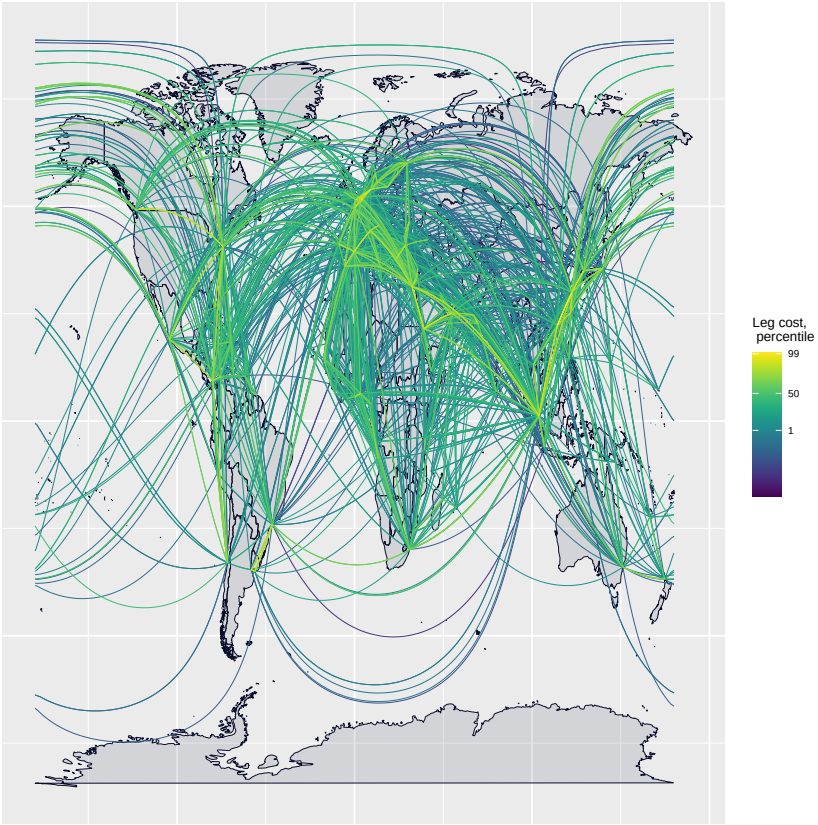
Table A.10: Determinants of Ship Size

	(1) ln Ship Size	(2) ln Ship Size	(3) ln Ship Size	(4) ln Ship Size
ln Volume at Origin	0.0843	0.0432	0.00924	
	(0.0163)	(0.0179)	(0.0121)	
ln Volume at Lading		0.0804	0.127	0.0283
		(0.0202)	(0.0230)	(0.0182)
$\mathbb{1}$(Lading is Origin)=1			-0.0219	
			(0.300)	
$\mathbb{1}$(Lading is Origin)=1 × ln Volume at Lading			-0.0937	
			(0.0295)	
$\mathbb{1}$(Lading is Origin)=1 × ln Volume at Origin			0.0861	
			(0.0220)	
ln Largest Port Stop				0.121
				(0.0250)
Observations	215,642	215,642	215,642	215,642
R^2	.124	.174	.199	.21
F-stat	26.82	14.66	13.51	26.73

Notes: Observations are at the shipment level, weighted by TEU, representing all matched imported containers to the United States. ln Ship Size is the natural log of maximum ship capacity in TEU. Ln Volume at Origin is the natural log of the sum of all shipments' TEU by shipment origin country. Ln Volume at Lading is the sum of all shipments' TEU by shipment lading country. The indicator takes a value of 1 if the shipment is laded at the country of origin. Ln Largest Port Stop is the maximum of the natural log of the volume of lading at all ports visited between the port of lading and unlading. Standard errors are clustered two ways by lading and destination ports.

lading at smaller transshipment points that travel along major routes are also on larger ships. Column (4) of Table A.10 regresses shipments' log ship size against the log volume laded at their port of lading and the log volume laded at the largest port at which we observe the shipment making a port call. The effect of the max-port-size variable is large, positive, and overall stronger than the effect of lading port volumes alone. Additional

Figure A.11: Trade Cost Estimates, All Legs



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

stops that move through entrepôts allow shipments laded in smaller ports to travel on larger ships. Indirectness facilitates larger ship sizes beyond transshipment alone.

D.5 Additional Estimation Results

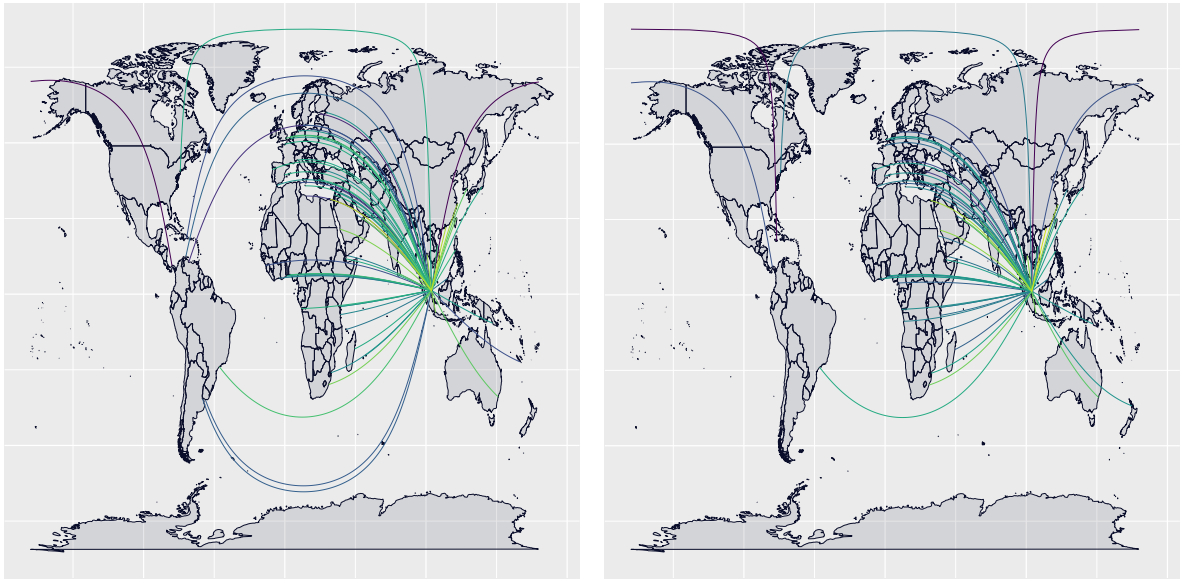
We plot our estimated route costs in Figure A.11. Thicker and lighter colors indicate lower-cost routes. Shorter and more heavily trafficked routes are the cheapest. The effect of scale is observable here: Syria to France is one of the highest cost legs, significantly higher than Singapore to Gibraltar, a much longer distance. Even among the subset of bilateral pairs for which we observe traffic, the triangle inequality is violated 280 times.

Figure A.12 plots bilateral incoming and outgoing trade costs for Singapore and Lebanon separately. Singapore is not only well-connected both as an origin and destination, but also has some of the cheapest legs. Lebanon, on the other hand, has both fewer and shorter connections.

Figure A.12: Trade Costs by Country

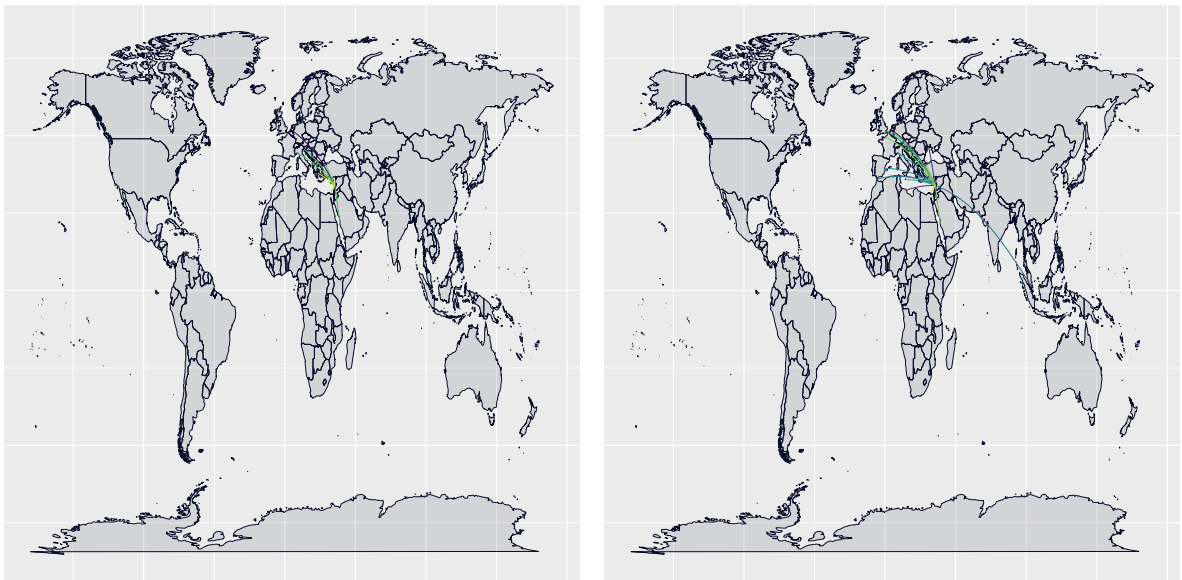
(A): Singapore, Origin

(B): Singapore, Destination



(C): Lebanon, Origin

(D): Lebanon, Destination



Notes: This map plots estimated link costs from Singapore in Panel (A) to Singapore in Panel (B), from Lebanon in Panel (C), and to Lebanon in Panel (D). Lighter colors indicate lower trade costs.

D.6 Analysis of Trade costs

Figure A.13 plots country-level market access for producers and consumers, which are averages of the expected trade cost (from the B-matrix) weighted by the GDP of origins and destinations, respectively. Entrepôts such as Egpyt, Panama, and (not visible) Singapore and Gibraltar have generally cheaper trade costs, as does China, due to the scale of shipping as well as access to nearby low-cost entrepôt (Korea, Singapore, and Japan).

Table A.11 reflects the log-linear relationship between our estimated trade cost τ, aggregate bilateral trade values, and distance. These results highlight the reduced form relationships between these three variables, as well as the predictive power of our computed trade costs. Without origin or destination fixed effects, our trade costs alone can explain 29% variation of global trade. The logarithm of distance can account for less than 3%. We do not take this as a horse race, but rather indication that these two measures are distinct: Our cost estimates τ measure network proximity and real shipping network relationships. Distance is a proxy for other orthogonal variables which impact trade volumes as well.

Table A.11: The Relationship between Trade Volumes and Network-Consistent Trade Costs and Distance

	(1)	(2)	(3)	(4)	(5)	(6)
	Log trade values					
Log $\tau_{ij}^{-\theta}$	0.462***		0.444***	0.756***		0.516***
	(0.0297)		(0.0306)	(0.0310)		(0.0278)
Log Dist		-0.755***	-0.393***		-1.372***	-0.673***
		(0.0993)	(0.0937)		(0.0626)	(0.0597)
Constant	12.71***	14.82***	16.04***	15.67***	20.36***	19.29***
	(0.354)	(0.909)	(0.772)	(0.311)	(0.561)	(0.439)
Orig, Dest FEs	No	No	No	Yes	Yes	Yes
Observations	23,344	22,985	22,985	23,344	22,985	22,985
R-squared	0.290	0.028	0.292	0.762	0.753	0.771

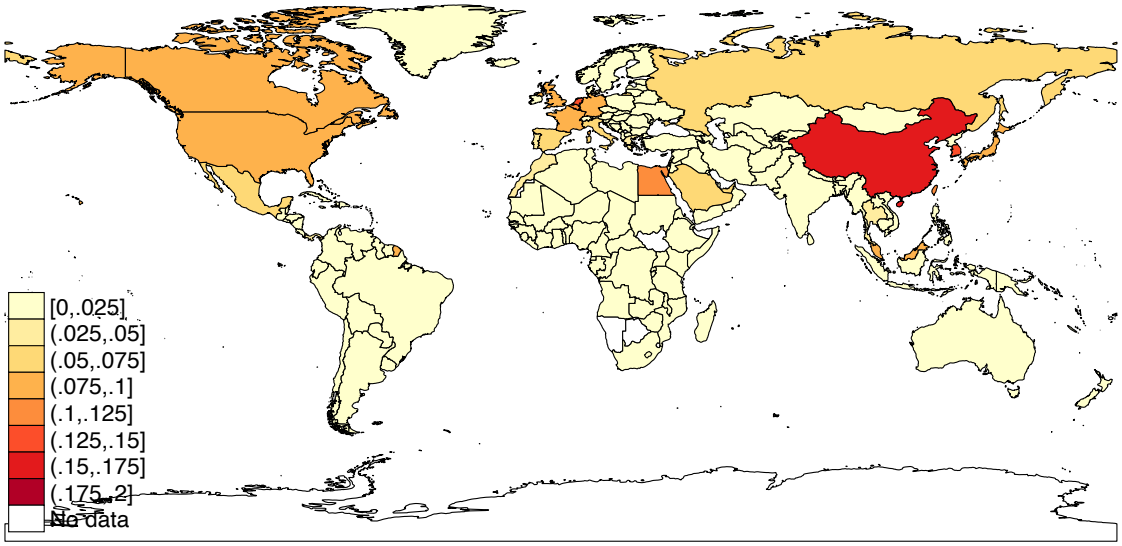
Notes: This table presents regression coefficients from the regression of the natural log of trade volumes on the natural log of $\tau_{ij}^{-\theta}$, the natural log of model-estimated origin-destination trade costs raised to the trade elasticity, and the natural log of distance, and the sea distance between the origin and destination measured in kilometers. Column (1)-(3) report results for cost and distance independently, then combined. Columns (4)-(6) rerun regressions in (1)-(3), respectively, adding origin and destination fixed effects.

Appendix E General Equilibrium Model in Changes

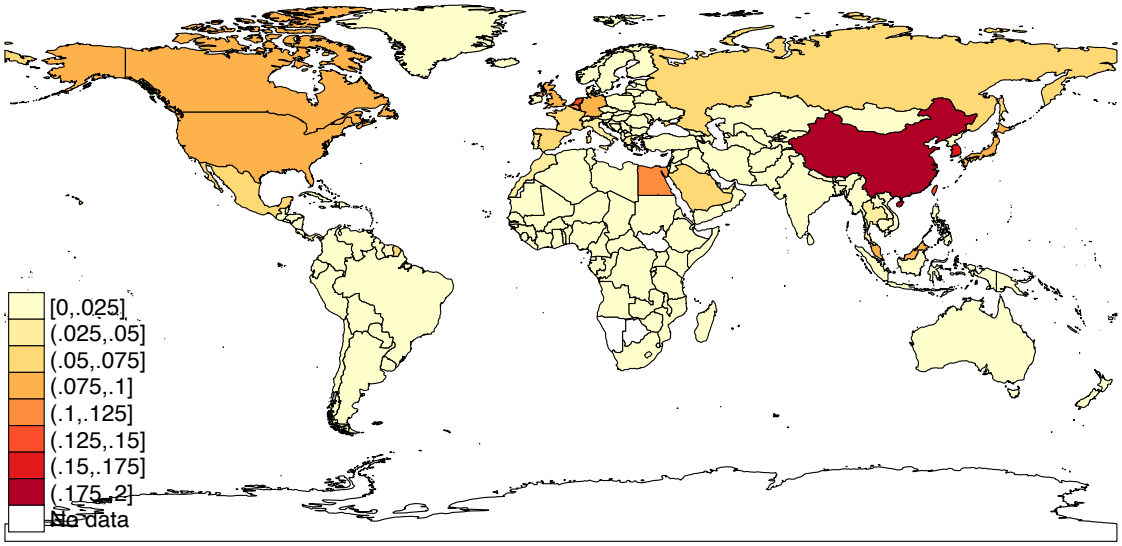
To close our model, we adopt the Caliendo and Parro (2015) framework. A continuum of intermediate goods ω_n are used in the production of composite goods that are in turn used domestically both as final goods and as materials for intermediate production by firms in

Figure A.13: Market Access

(A) Consumer Market Access



(B) Producer Market Access



Notes: Figure plots producer and consumer market access for each country according to the transportation costs estimated in Section 5. Countries with higher market access are in darker reds, while countries with lower market access are in lighter yellows. Countries with missing data are in white.

each industry n. We assume there are three sectors ($N = 3$): containerized tradables c, non-containerized tradables nc, and nontradables nt ($n \in [c, nc, nt]$). Intermediates in the nt sector are only sourced domestically while ω_{nc} and ω_c goods are sourced internationally. Trade routes are modeled for all three sectors but we only consider transportation cost changes for the containerized sector ω_c.

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

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

where η_n is the Cobb-Douglas industry share, $\sum_n \eta_n = 1$.

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

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

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

$$c_{in} \equiv \frac{\Upsilon_{in} w_i^{\gamma_{in}} \prod_{n'}^{N} P_{in'}^{\gamma_{in}^{n'}}}{z_{in}}, \tag{25}$$

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

Composite goods production In each country i, composite goods in industry n are produced using a CES aggregate of intermediates Ω_n, purchased and sold domestically at marginal cost. In traded industries, intermediates are sourced internationally from lowest-cost suppliers. Using the standard aggregation, the resulting price at j of the

composite in industry n is expected to be the following (where A_n is a constant):

$$P_{jn} = A_n \left[\sum_{i=1}^{I} c_i^{-\theta_n} \kappa_{ijn}^{-\theta_n} \tilde{\tau}_{ijn}^{-\theta_n} \right]. \tag{26}$$

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

$$\dot{c}_{in} = \dot{w}_i^{\gamma_{in}} \prod_{k=1}^{N} \dot{P}_{ik}^{\gamma_{ink}}. \tag{27}$$

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

$$\dot{P}_{in} = \left[\sum_{i=1}^{J} \pi_{ijn} [\dot{\tau}_{ijn} \dot{c}_{in}]^{-\theta_n} \right]^{-1/\theta_n}. \tag{28}$$

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

$$\dot{\pi}_{ijn} = \left[\frac{\dot{c}_{in} \dot{\tau}_{ijn}}{\dot{P}_{in}} \right]^{-\theta_n}. \tag{29}$$

Trade volumes similarly adjust:

$$X'_{in} = \sum_{k=1}^{N} \gamma_{ink} \sum_{j=1}^{I} \frac{\pi'_{ijn}}{1 + \kappa_{ijn}} X'_{jk} + \alpha_{in} I'_i. \tag{30}$$

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

$$\sum_{n=1}^{N} \sum_{i=1}^{I} \frac{\pi'_{ijn}}{1 + \kappa_{ijn}} X_{in} - D_i = \sum_{n=1}^{N} \sum_{i=1}^{I} \frac{\pi'_{jin}}{1 + \kappa_{jin}} X_{jn}, \tag{31}$$

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

Appendix F Counterfactual Results

F.1 Counterfactual Procedure

Algorithm 1 describes the algorithm for finding the new equilibrium after an adjustment that induces an endogenous scale response in the network. We limit the counterfactual trade cost on any route to be no lower that the minimum observed initial trade cost.

F.2 Counterfactuals: Additional Figures and Tables

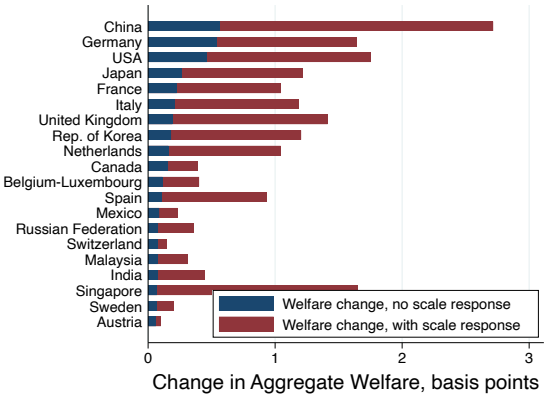
The Global Impact of Local Infrastructure Improvements Table A.13 reports, for each case (denoted by column), the mean global welfare impact (first row) and

Algorithm 1 Scale Counterfactual Algorithm

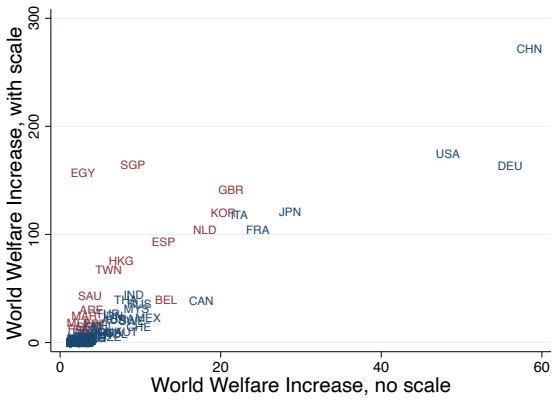
procedure WELFARE CHANGE($X_0, \Xi_0, \dot{t}$) ▷ Find a new equilibrium
  Initialize current trade flows X_0 and traffic Ξ_0
  Initialize changes in cost fundamentals $\dot{\tau}$ ▷ Example: shipping distances changes
  Compute $A_0 = A(\Xi_0; \dot{\tau})$ ▷ Following equation 11
  Compute $B_0 = (I - A_0)^{-1}$
  Initialize difference $= \infty$, tolerance $= \epsilon$
  while $difference < tolerance$ **do**
    Update trade flows $X_1 = X(B_0)$ ▷ Solving 8.1
    Update traffic $\Xi_1 = \Xi(X_1, A_0, B_0)$ ▷ Following equation 10
    Update leg costs $A_1 = A(\Xi_1)$
    Update trade costs $B_1 = (I - A_1)^{-1}$
    Compute $difference = \Sigma_{ij}(B_1 - B_0)^2$
    Update $A_0 = A_1$ and $B_0 = B_1$
  Return final trade flows X_1
  Compare welfare and price index changes between X_1 and X_0 ▷ Solving 8.1

Figure A.14: Most Pivotal Nodes: Change in Welfare Excluding Own

(A) Non-Transportation Cost Reductions: Highest Global Welfare Changes



(B) Non-Transportation Costs Reductions with vs without scale



Notes: Panel (A) shows absolute values for aggregate net change in global welfare after non-infrastructure cost reductions in the listed country, excluding the country's own, 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 excluding the endogenous scale response (X-axis) vs a the same including the scale response (Y-axis). Markers are ISO Country codes. Entrepôts are in red.

standard deviation (second row) across all 136 targeted countries. Rows three through six consider results separately for counterfactuals where targeted countries are entrepôts and non-entrepôts. Column (1) reports welfare changes from non-transportation cost reductions without scale responses. Raw effects from counterfactuals targeting entrepôts are roughly twice as large, reflecting entrepôts' greater global integration—a difference eliminated below in Table A.14. In Column (2), the scale response, which incorporates the

Table A.12: Welfare and Trade Outcomes from Improvements in Transportation and Non-Transportation Costs, Basis Points

	Non-Transportation Improvement		Transportation Improvement	
	Baseline Effect	Total Effect (Network & Scale)	Network Effect	Total Effect (Network & Scale)
	(1)	(2)	(3)	(4)
Δ Average Global Welfare				
Mean	0.08%	0.26%	0.18%	0.54%
Standard Deviation	(0.20)	(0.59)	(0.41)	(1.33)
Δ Container Trade Volumes				
Mean	0.87%	2.89%	2.02%	6.11%
Standard Deviation	(2.22)	(6.65)	(4.67)	(14.98)

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.

Table A.13: Counterfactual Reductions in Local Trade Costs, by Targeted Country Entrepôt Status

	Welfare Change from Cost Reduction				Trade Change from Cost Reduction			
	Non-Transportation		Transportation		Non-Transportation		Transportation	
	$\Delta\kappa_{kl}$	$\Delta\kappa_{kl}$ with Scale	Δt_{kl}	Δt_{kl} with Scale	$\Delta\kappa_{kl}$	$\Delta\kappa_{kl}$ with Scale	Δt_{kl}	Δt_{kl} with Scale
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Global Changes								
Mean	0.08%	0.26%	0.18%	0.54%	0.87%	2.89%	2.02%	6.11%
Standard Deviation	(0.20)	(0.59)	(0.41)	(1.33)	(2.22)	(6.65)	(4.67)	(14.98)
Reductions at Entrepôts								
Mean	0.16%	0.88%	0.87%	2.91%	1.56%	9.64%	9.73%	32.77%
Standard Deviation	(0.16)	(0.64)	(0.67)	(2.41)	(1.64)	(7.03)	(7.43)	(27.03)
Reductions at Non-Entrepôts								
Mean	0.07%	0.18%	0.09%	0.24%	0.78%	2.05%	1.07%	2.81%
Standard Deviation	(0.20)	(0.54)	(0.27)	(0.71)	(2.28)	(6.13)	(3.13)	(8.21)

Notes: This Table replicates the results for our first counterfactual, transportation and non-transportation cost declines for each of 136 countries (Table A.12), breaking out the mean results from 136 targeted countries (rows one and two) into those 15 targeting entrepôts (rows three and four) and all others (rows five and six). Columns (1)-(4) present aggregate welfare changes. Columns (5)-(8) present changes to aggregate container trade. Columns (1), (2), (5), and (6) present results for cases where non-transportation trade costs are reduced. Columns (3), (4), (7), and (8) present results for cases where transportation costs are reduced (infrastructure improvements). Odd columns correspond to cases where no scale economy feedback loops are allowed. Even columns present results allowing for scale economy feedback. In each case, we report the mean impact and its standard deviation in parentheses.

effects of each shock on the transportation network, augments this to 5-fold. In Columns (3) and (4), infrastructure investments at entrepôts generate on average 10 times the global welfare impact relative to investment elsewhere.

Table A.14 compares the welfare impact of 136 counterfactuals in each of the four cases, the relationship between a welfare increase from a given reduction in trade costs and the entrepôt status of the targeted location, controlling for GDP at the targeted location, the distance between targeted and impacted countries, and a fixed effect for impacted country. The latter controls for whether a specific impacted country is particularly sensitive to trade cost reductions. There are 18,340 bilateral pairs of targeted and impacted countries. Regressions are weighted by impacted country's GDP.

The strong controls in these regressions reduce the differential impact of entrepôts: Column (1) controls for gravity variables and impacted fixed effects, fully accounting for entrepôt countries' raw positive impact (double non-entrepôts' in Table A.12). However, once scale economies' impact on the transportation network are accounted for (in Column (2)), the impact from counterfactuals targeting entrepôt countries are an order of magnitude larger. In Column (3), when the transportation network is directly impacted by infrastructure investment, entrepôts are at baseline more than two-thirds more impactful (52 log points) and over 200% more impactful when scale economies are allowed.

Note that because some welfare effects are negative, we add a constant to all results before taking logs. This makes the indicator variable not directly comparable to the raw numbers in Table A.12. However, the relative size and direction of results are robust to using raw percent changes on the left-hand side.

Figure A.14 repeats the exercise in Figure 12 for non-transportation cost reductions with and without the endogenous response of costs throughout the network when accounting for scale economies. The black bars in Panel (A) underscore that without transportation network impacts, smaller entrepôts are generally not pivotal. The grey bars show results accounting for the endogenous response of the transportation network. The dramatic difference for Singapore in particular underscores that conflating network changes with non-network adjustments such as tariff changes can bias results.

In Panel (B), we plot the results for each country with and without scale. Here the average relationship as well as the average error is nearly identical as in Figure 12, as is the bias at entrepôts.

Brexit Figure A.15 shows the impact of our two counterfactual cases on the UK's 20 largest trading partners in welfare percent changes. Black bars show the impact of

Table A.14: Bilateral Welfare Impacts, by Entrepôt status

	ln %Δ Welfare from Non-Transport Cost Reduction		ln %Δ Welfare from Transport Cost Reduction	
	$\Delta\kappa$	Δt with Scale	Δt	Δt with Scale
	(1)	(2)	(3)	(4)
$\mathbb{1}_{entrep\hat{o}t} = 1$	-0.0428	0.655	0.517	1.145
	(0.0393)	(0.117)	(0.0870)	(0.181)
ln GDP, targeted 0.220	0.364	0.495	0.568	
	(0.0130)	(0.0219)	(0.0335)	(0.0255)
ln Distance	-0.352	-0.551	-0.353	-0.435
	(0.0540)	(0.0531)	(0.0455)	(0.0491)
Obs	18340	18340	18340	18340
R^2	0.575	0.729	0.817	0.809

Notes: Results weighted by impacted country GDP. Outcome values are shifted by a constant in order to include negative values. Standard errors in parentheses are clustered two ways by targeted and impacted countries. 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.

increased non-transportation trade friction with the UK. Grey bars show the impact with scale effects changing transportation costs through the UK. All partners experience outsized losses due to scale economies. Most of these losses come through increased trade costs in the Netherlands and Belgium, which far from benefiting from our counterfactual, lose because of decreased volumes as well. Ireland in particular, which our microdata tells us sends 50% of goods to the US through the UK, experiences large additional losses.

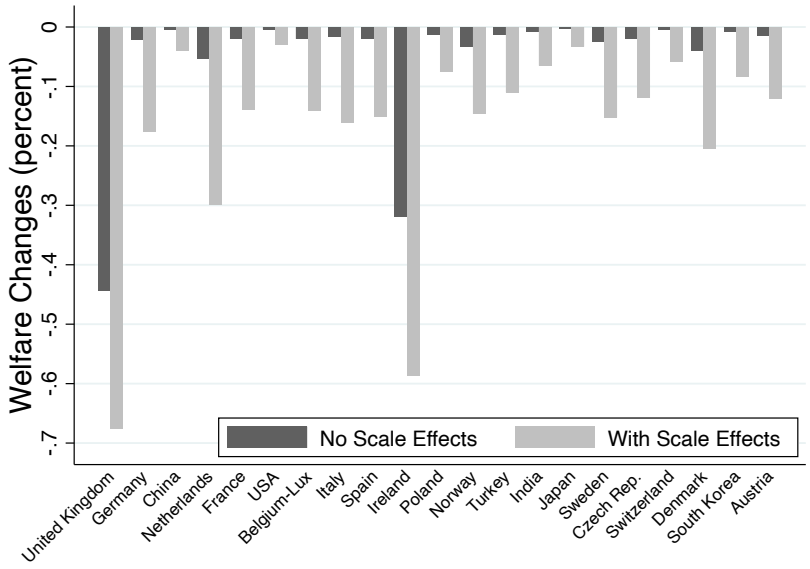
Global trade volume changes under these two cases are reported in Figure A.16. These results largely mirror our welfare results in the main text.

The Opening of the Arctic Passage Figure A.17 shows changes in the relative wage-adjusted price index (interpreted as national welfare, if we omit the costs of climate change) across the three cases.[10] In the baseline scenario in Panel (A), we see increases in 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 A.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.

Since some of the Asian entrepôts are smaller and harder to see on a global map,

[10] Appendix Figure A.19 shows related changes in country-by-country containerized exports.

Figure A.15: Welfare Changes - Brexit - Largest Trading Partners



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

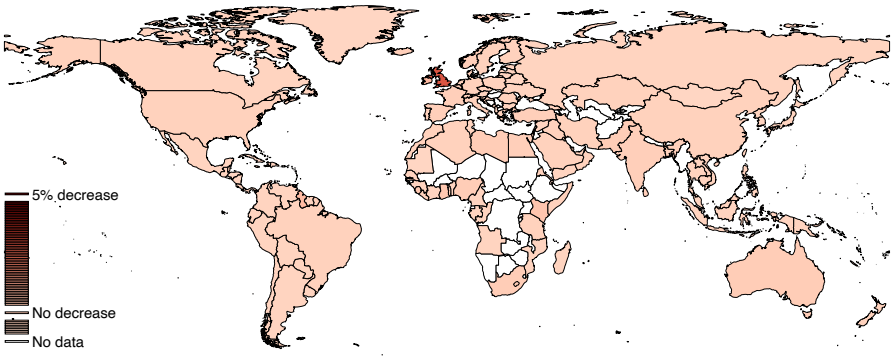
Figure A.18 zooms in on the welfare changes of Singapore, Hong Kong, and Taiwan as well as their surrounding countries as a result of the opening of the Arctic Passage. In the baseline scenario in Panel (A), we see that these entrepôts have a direct welfare increase from the passage opening since they have direct routes to Northern European countries and North America. When allowing for indirect trade in Panel (B), the neighboring countries of these entrepôts see an increase in welfare because they are now able to benefit from using these entrepôts to trade with the Northern European countries and North America. When allowing for scale economies to amplify effects in Panel (C), the entrepôts and their neighboring countries are going to benefit even further as a result of this indirect trade.

The concentration of welfare gains in entrepôts from this counterfactual highlights a novel source of agglomeration—scale economies in transportation and transport networks can help contribute to and shape entrepôts. This is further explored in our first counterfactual in Subsection 8.2.

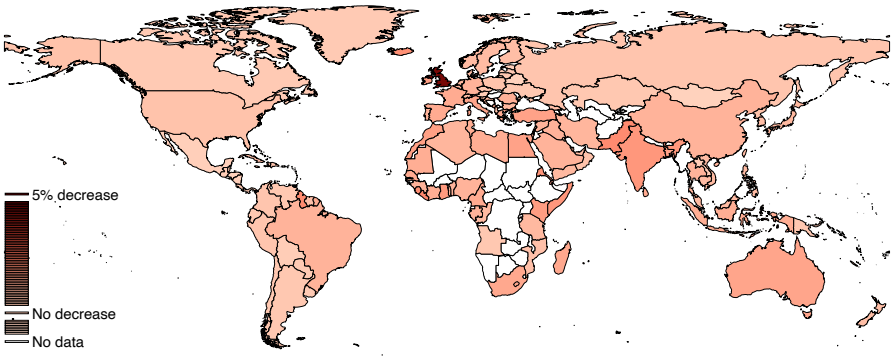
Figure A.19 reports global trade volume changes under the three cases. These results highlight the significant heterogeneity in trade changes across countries and largely mirror our welfare results in the main text.

Figure A.16: Export Volume Changes - Brexit

(A) Trade Cost Change, No Network Scale Effects



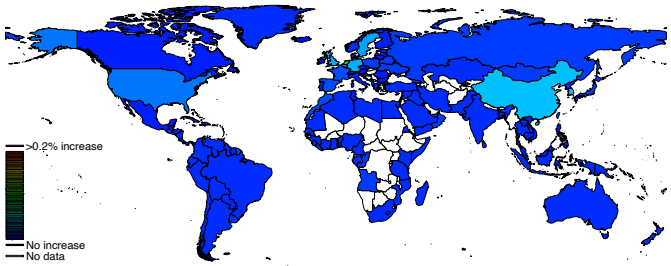
(B) Full Trade Network Effects and Scale Economies



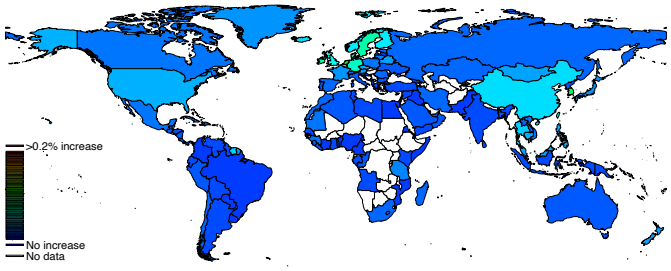
Notes: These two plots show the percent change in exports of a simulated 5% increase in trading costs with the United Kingdom for all countries in our dataset. Darker reds reflect a greater increase. White represents omitted countries. Panel (A) reflects changes if shipping costs remain constant, reflecting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

Figure A.17: Welfare Changes - Arctic Passage

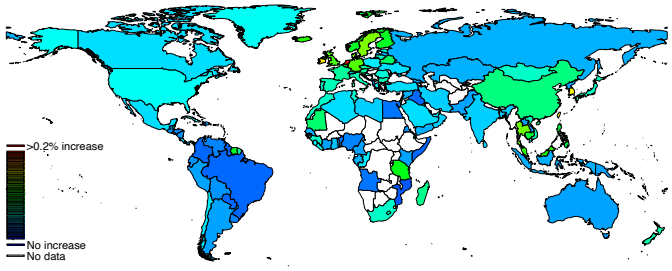
(A) Only Directly Affected Routes (Exogenous Trade Costs)



(B) Full Trade Network Effects



(C) Full Trade Network Effects and Scale Economies



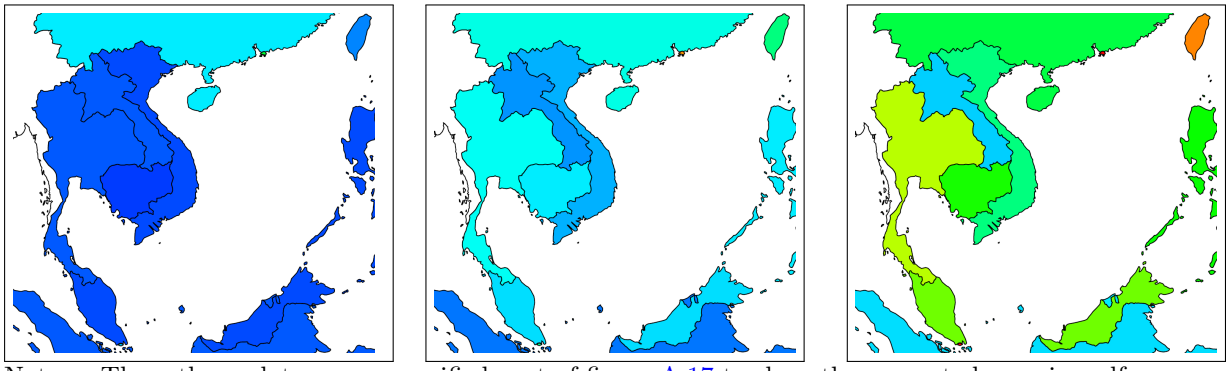
Notes: Plots show the percent change in welfare (the relative price index). Darker reds reflects a greater increase and blue represents no change. Omitted countries are white. Panel (A) reflects changes only allowing trade costs to decrease on routes whose distance is directly reduced to the Arctic Passage. 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.

Figure A.18: Welfare Changes on Asian Entrepôts - Arctic Passage

(A) Only Directly Affected Routes

(B) Full Trade Network Effects

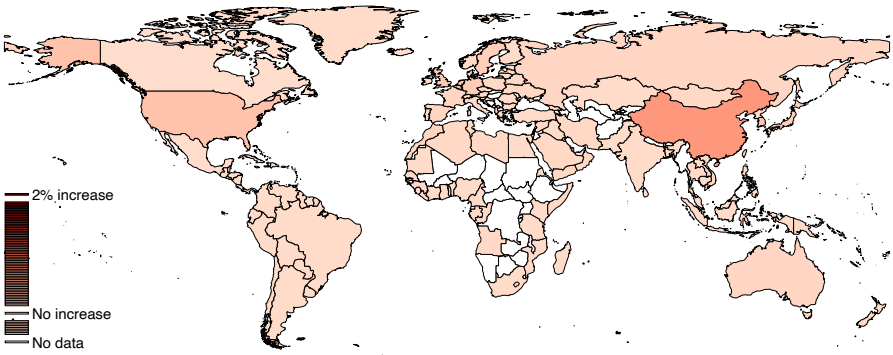
(C) Full Trade Network Effects & Scale Economies



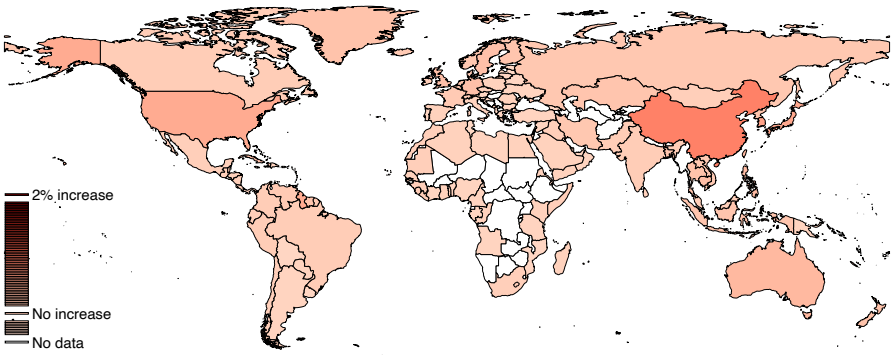
Notes: These three plots are a magnified part of figure A.17 to show the percent change in welfare (the relative price index) for a subset of Asian Entrepôts in our dataset. Darker reds reflects a greater increase and blue represents no change. White represents omitted countries. Panel (A) reflects changes if we only allow trade costs to decrease on routes whose distance is directly reduced to the Arctic Passage. Panel (B) reflects changes if we allow all countries to indirectly access the Arctic Passage through the trade network. Panel (C) allows for the endogenous network response to scale economies.

Figure A.19: Export Volume Changes - Arctic Passage

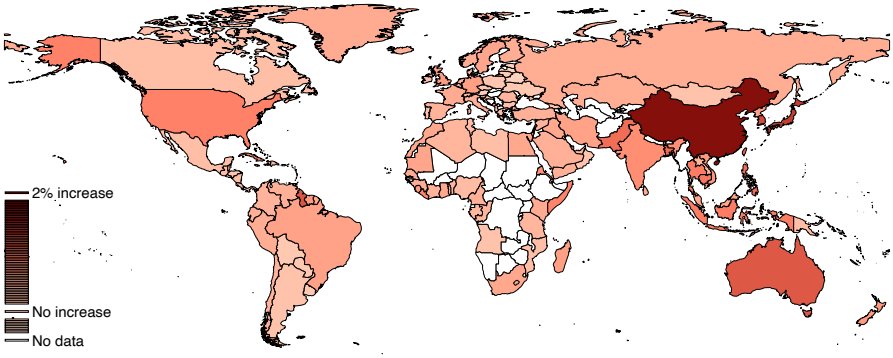
(A) Only Directly Affected Routes



(B) Full Trade Network Effects



(C) Full Trade Network Effects and Scale Economies



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