

Trade From Space: Shipping Networks and The Global Implications of Local Shocks*

Inga Heiland[†] Andreas Moxnes[‡] Karen Helene Ulltveit-Moe[§] and Yuan Zi[¶]

November 2022

Abstract

This paper analyzes international externalities of a local shock to the global shipping network. The 2016 Panama Canal expansion removed a bottleneck in seaborne transportation, allowing much larger ships to pass. Using both reduced-form and structural methods in combination with novel satellite data on the movements of container ships, we find that trade increased significantly among country-pairs using the canal. We find that the global real income gains from the canal expansion were over three times greater than the income gains for Panama itself. A link removal analysis reveals that most shipping links are associated with positive and quantitatively important positive international externalities.

*We thank participants at various seminars and conferences for valuable comments and discussion. We thank Clarksons Platou for providing us with access to the Clarksons World Fleet Register, and Bjørn Bodding at Clarksons Platou for valuable discussions. This project has received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation program (grant agreement 715147).

[†]Kiel University & IfW Kiel & University of Oslo & CEPR & CESifo, inga.heiland@econ.uio.no

[‡]University of Oslo & CEPR, andreas.moxnes@econ.uio.no

[§]University of Oslo & CEPR, k.h.ulltveit-moe@econ.uio.no

[¶]The Graduate Institute Geneva & University of Oslo & CEPR, yuan.zi@econ.uio.no

1 Introduction

Container ships are the engines of global trade.¹ By now, nearly all countries have container ports, constituting the nodes of the global container shipping network (Rua, 2014). The networked environment implies that a shock to a port, or a link, in the network, such as improvements in shipping infrastructure, may affect shipping costs, trade flows and real incomes for many more locations than those that are directly affected. Put differently, the network structure of transportation implies that international externalities are likely to exist. There is, however, scarce empirical evidence on the existence, and magnitude of these international externalities.

This paper quantifies the sign and magnitude of international externalities by using novel satellite data on the movements on container ships. We analyze a large shock to the transportation network: the 2016 Panama Canal expansion. The expansion removed a major bottleneck in seaborne transportation, allowing much larger ships to pass. Using both reduced-form and structural methods, this paper analyzes how the expansion led to ripple effects in terms of trade and transportation costs globally, and provides evidence of the international externalities associated with the expansion. Positive international externalities naturally arise in a networked environment because a trip from location a to c might pass through location b . Lower transport costs between b and c will not only benefit b and c , but also a . This stands in sharp contrast to canonical models of international trade, where a will typically lose when b and c integrate, i.e., a *negative* international externality.²

Our empirical analysis of global container ship movements has become possible due to the rapid advent of the global Automated Identification System (AIS) over the last years. Using an exhaustive AIS data set of all port calls made by container ships in 2016, we document novel facts about the container shipping network. First, container ships typically operate on fixed routes, i.e. they serve a stable set of ports, akin to buses serving a fixed number of stops in a city. Second, shipping activity is highly concentrated across ports, with some nodes (ports) in the network handling almost two orders of magnitude

¹Levinson (2006) and Bernhofen et al. (2016) detail the seismic changes that the worldwide adoption of container shipping technology has brought about in international trade.

²See Riezman (1979), Baldwin et al. (2003), Behrens et al. (2007), Mossay and Tabuchi (2015) and Section 7.

more ships than the median port. Third, the network is very sparse in the sense that only few countries have direct shipping routes to their trade partners.

While the AIS data provides unprecedented detail about the movement of ships, one cannot observe the movement of the cargo itself, i.e. the actual route of a shipment from country i to country j . To make progress, we propose several methods to infer the route(s) a container might take from i to j . The simplest method is to use the observed shipping network along with actual travel times between all direct port-pair links and calculate the fastest route between any potential port pair. Consider, for example, a shipping network with direct links between New York-London, New York-Hamburg, London-Oslo and Hamburg-Oslo. The fastest route between New York and Oslo might then be New York-London-Oslo if this route minimizes the sum of travel times of each leg of the journey, including waiting time at intermediate ports. The fastest path calculations reveal that 50% of all country-to-country connections involve stops in more than two other countries in between.

After computing the optimal shipping routes, we are able to assess which country pairs were exposed to the 2016 Panama Canal expansion. After 10 years of construction, the extended Panama Canal opened on June 26th of 2016. The \$5.25 billion massive construction project was a modern engineering marvel: it nearly doubled the capacity of the canal by adding a wider and deeper third lane, allowing much larger ships to pass.³ We analyze the global trade effects of the expansion, and how countries were differentially affected by the shock. Using a difference-in-difference approach, we find that country pairs whose fastest connection passed through the Panama Canal prior to the expansion traded 11% more after the expansion compared to other country pairs.

Finally, we investigate and quantify the international externalities associated with the expansion. We develop a spatial model of trade, where we – in contrast to standard trade models – take into account that goods traveling from one location to another will typically be traveling indirect rather than directly. We build on, and extend, the work of Allen and Arkolakis (2022), but while their application is on urban economics, our focus is on international trade.

Using the structure of the model, we estimate the impact of the canal expansion on

³The project required 5 million cubic meters of high-strength concrete - enough to build a highway from New York to St. Louis (Business Insider, 2016).

transportation costs using a difference-in-differences approach, comparing container traffic before and after the canal expansion. After estimating the impact on transportation costs, we use the model to quantify the effects of the Panama Canal expansion on global trade and real income. We find that the global real income gains from the canal expansion were over three times greater than the income gains for Panama itself. We therefore conclude that the expansion was associated with positive international externalities.

Finally, we show that our quantitative model can be used to address the more general question of whether the shipping network generates positive international externalities. We analyze this question by removing every possible port-pair-link from the network, one link at a time. This methodology is similar to so-called link removal analyses that have been used extensively to study real-world networks in different fields of science (Bellingeri et al., 2020). After running several thousand counterfactuals, we find that the large majority of port-pair-links are associated with positive international externalities. We document the distribution of externalities across port-pair-links and show which links are the most valuable ones for the global economy. For the average link, the global gains are 12 percent higher than the local gains, i.e., the gains enjoyed by the two countries directly linked.

To highlight the role of the global shipping network, we contrast our results to a model without a shipping network - where all trading partners are directly linked. In this case, international externalities are on average negative, i.e. reduced trade barriers between two trading partners have a negative impact on third countries. Under the special case of symmetric trade costs, international externalities are *always* negative. Our quantitative results point to the interdependences created by the global shipping network, and suggest that as long as the infrastructure costs are not shared across countries, the presence of positive externalities implies that the world invests too little in global transportation networks. We leave it to future research to quantify the socially optimal level of investment, given our findings.

The paper makes three main contributions to the literature. First, we document salient features of the world container shipping network based on unique novel data covering the worldwide movements of all container ships. Second, we develop a new methodology to calculate optimal shipping routes from our data, delivering unprecedented detail about the shipping times to and from all port-pairs worldwide. Third, this paper is the first

to analyze international externalities of the global shipping network, using the Panama Canal expansion as a case study.

Our results from contrasting international externalities of transportation cost changes and classical bilateral trade cost changes generalize the finding of Behrens et al. (2007), who show that transportation cost reductions in a network of routes can be Pareto improving at the global level, while bilateral non-transportation cost reductions necessarily harm third countries. We also relate to the literature that studies the impact of canal openings or closings. Maurer and Rauch (2019) analyze how the Panama Canal changed U.S. population patterns, whereas Feyrer (2021) studies the relationship between trade and the closing and opening of the Suez Canal.

This paper is also linked to the growing number of studies using satellite data for economic analysis. Donaldson and Storeygard (2016) provide an overview of applications which so far has focused on environmental, development and spatial issues. This paper explores how shipping satellite data can be used within the field of international trade. There are only a few recent papers that have used shipping satellite data to explore issues related to trade. Brancaccio et al. (2020) study the role of the transportation sector in world trade focusing on search frictions and the endogeneity of trade costs. They use AIS data for dry bulk ships, which typically carry commodities such as iron ore, coal, grain and sugar. Our focus is instead on container ships, which typically carry manufactured goods and account for around two-thirds of world trade based on values. In recent, and parallel, work, Ganapati et al. (2021) study the role of shipping hubs for global trade and welfare. Our paper instead focuses on international externalities of the Panama Canal expansion, using the expansion as a quasi-natural experiment.

Our paper is also tangentially related to the literature on the effects of containerization. Besides having spurred global trade as documented by Bernhofen et al. (2016), new port technology has been shown to have significantly altered countries' economic geography (Brooks et al., 2021 and Ducruet et al., 2022).

The rest of the paper is structured as follow. Section 2 documents the satellite data and the construction of the global shipping network data set and presents salient features of the network. Section 3 describes the Panama Canal expansion and explains how to compute the measure of exposure to the canal expansion. Section 4 analyzes the global impact of the Panama Canal expansion on trade. Section 5 presents a spatial model of

trade with a transportation network, while Sections 6 and 7 use the model to quantify network externalities. Section 8 concludes.

2 Data and Descriptives

2.1 Data

AIS data. Our point of departure is containerized trade. Containerized seaborne trade captures the majority of merchandise world trade (see UNCTAD, 2016), and is responsible for approximately 60 percent of the value of all seaborne trade in 2016 (Rajkovic et al., 2014). We build a comprehensive data set for the global container shipping network based on satellite data for ships. The satellite data comes from AIS (Automatic identification System) data and is provided by Marine Traffic. AIS is an automatic tracking system used on ships and by vessel traffic services (VTS). Vessels send out AIS signals identifying themselves to other vessels or coastal authorities, and the International Maritime Organization (IMO) requires all international voyaging vessels with above 300 Gross Tonnage and all passenger vessels to be equipped with an AIS transmitter. This requirement ensures a nearly universal coverage of container ships in our data, as over 99% of container-shipments around the world are made by containerships that are above 500 tonnage.

Our data set is based on a ship's port calls, i.e. the signal sent by a ship when it enters and leaves the geo boundary of a port. For each observation, we observe the ship's ID, its time stamp, transit status⁴, and current draught (i.e., maximum depth of any part of the vessel under water), as well as port information (name, country, and geographic coordinates). We use data on all port calls tracked by the AIS satellite system during the calendar year 2016. We merge the data with container ships' technical information provided by Clarkson World Fleet Register. After adjusting for some reporting errors, we were able to match 93% of global container ships sailing in 2016.

The Clarkson data provides information on each ships' scantling draught (i.e., draught when fully loaded) and deadweight tonnage (i.e. the maximum tonnes of goods that a ship can carry). Combining the two with a ship's actual draught at any time, we can back out how much cargo a ship was carrying using formulas from the marine traffic literature

⁴A ship is called 'in transit' at a port if it is not lading or unlading cargos.

(see Appendix Section B for details). Appendix Section A reports in detail our variables and how we have cleaned the data. Our final dataset includes 4,941 container ships and 514 ports for the year 2016.

Other data sets. The analysis in Section 4 requires data on trade flows, which we obtain from COMTRADE for the years 2013-2019. We aggregate monthly bilateral trade data to the quarterly level to reduce volatility that is due to seasonal effects or to lagged reporting. The analysis also requires variables such as distance and contiguity, which we obtain from the gravity database of CEPII. Data on free trade agreements come from the WTO’s RTA databases. The analysis in Section 6 requires additional information about expenditure along with a few other variables, which we obtain mainly from the Eora Global Supply Chain Database and supplement with data from the Worldbank’s World Development Indicators and from INSEE; we gather data for 149 countries for the 2015 cross-section. Appendix G provides additional details.

2.2 Stylized Facts on the Global Shipping Network

We start by documenting four salient features of the global shipping network that will guide the subsequent analysis.

Fact 1: Container ships typically operate on fixed routes. Table 1 provides descriptive statistics on the number of ports passed per ship as well the number of ships that arrive and depart per port. A key feature of container ships is that they typically visit the same port many times. The table shows that the average number of distinct ports passed per ship is roughly one sixth of the total number of ports passed per ship (12 versus 68).

Fact 2: Shipping activity is highly concentrated in space. A few ports act as major hubs in the shipping network. While the median port only serves around 200 ships per year, the top ports serve close to 14,500 ships per year. The same pattern is observed at the port-pair level, i.e. there are a few links in the network that account for a large share of total shipping activity.

Fact 3: Only 6 percent of all country pairs have a direct shipping connection. We calculate the in-degree as the number of ports to which a port is directly connected based on incoming ships, and the out-degree as the number of ports to which a port is directly connected based on outgoing ships. Table 2 shows that most ports are connected to rather few other ports. However, there is great variation between ports in how well connected

Table 1: Ships and Ports

Variable:	Obs	Median	Mean	Sd	Min	Max
Ships:						
# ports passed	4,941	64	67.81	40.16	1	312
# distinct ports passed	4,941	12	12.48	6.94	2	48
Ports:						
# incoming ships	514	206	651.82	1,457.89	5	14,486
# outgoing ships	514	201	651.82	1,454.97	5	14,421
Port pairs:						
# ships	4,158	38	80.58	168.86	5	2,779
deadweight tonnes (in millions)	4,158	0.70	2.08	4.98	9.66×10^{-3}	95.95

Note: Summary statistics are based on the port calls made by container ships in 2016.

they are. Nevertheless, even the best connected ports are only directly connected to around one sixth of the total number of ports. The 514 ports in our data are allocated across 154 countries. Only 6 percent of all country pairs have a direct shipping connection. Trade between these countries accounts for only 54 percent of world trade. This implies that a large share of global trade does not travel on direct routes, but on routes with multiple hops.

2.3 Calculation of Fastest Routes

This paper investigates the impact of the Panama canal expansion on trade by exploiting information on the underlying shipping routes. While the AIS data provides unprecedented detail about the movement of ships, one cannot observe the movement of the cargo itself, i.e. the actual route of a shipment from country i to country j . Therefore, we need a methodology to infer the shipping route between departure country i and arrival country j . This information enables us to determine to what extent trade between two countries is exposed to the Panama canal expansion.

Table 2: Port Networks

	Obs	Mean	Sd	Min	p50	p90	p95	p99	Max
Indegree	514	8.09	10.26	1	4.5	18	31	50	84
Outdegree	514	8.09	9.85	1	5	19	27	46	82

Note: Summary statistics are based on the port calls made by container ships in 2016.

We develop two complementary methods for inferring trade routes. The first method is to build a model-independent, brute-force route planner and find the route that minimizes travel time between two ports. The second method also relies on minimizing travel time, but infers routes by using the structure of a general equilibrium model. We present the brute-force method here, while the model-based approach is described in Section 5.

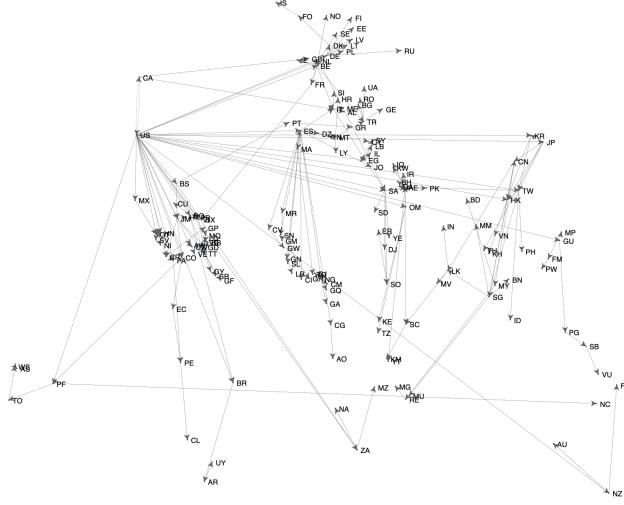
Based on the schedule of departure and arrival times, we compute the fastest route from any port i to any port j at any start time h during the year 2016 using a simple algorithm described in Appendix C. Among the set of optimal routes connecting two ports at different points in time, we select the route – the sequence of intermediate ports – that is used most frequently.⁵

Figure 1 visualizes the fastest routes for U.S. exports to all other countries based on our calculation.⁶ The figure shows that the routes typically go through hubs, e.g., U.S. shipping to Europe tends to pass through Germany and the Netherlands, whereas U.S. shipping to Africa goes through a hub in Spain. Figure 2 plots the fastest travel times between all port pairs against geodetic distance. Distance is strongly correlated with direct travel time, represented by the light blue dots in the figure. However, we observe that for indirect routes, represented by the dark blue dots, geodetic distance is much less informative for travel times. To understand further the role of shipping hubs and indirect routes in the global shipping network, we examine the number of hops on the fastest shipping routes between all ports in the network. Figure 3 shows the frequency of hops

⁵In those cases where more than one route occurs with the same highest frequency, we average over the characteristics of these routes when producing summary statistics.

⁶The figure displays only one route per country pair, namely the fastest one among the routes connecting all U.S. ports to the port(s) in the partner country.

Figure 1: Fastest Travel Times.

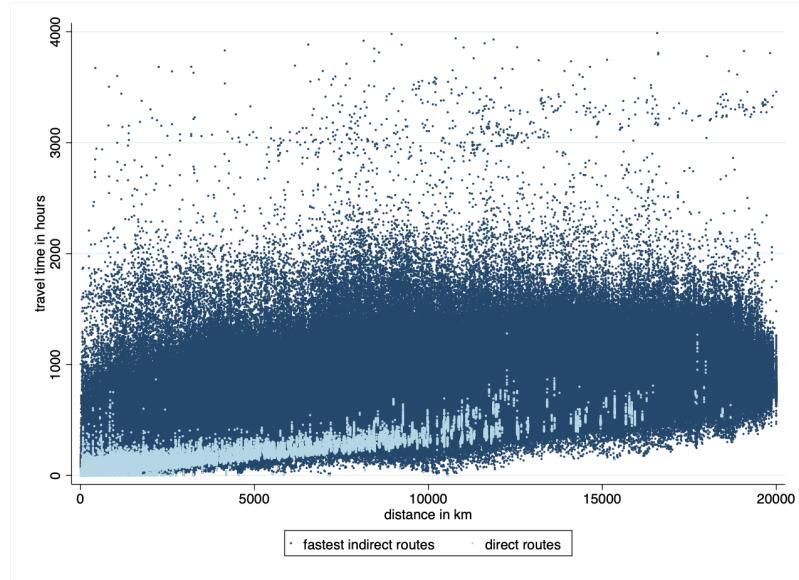


Note: The figure plots the fastest routes from the U.S. to other countries. The plotted route is the fastest one among the routes connecting any U.S. port and any port in the destination country.

after aggregating ports by country. Most country pairs are connected by routes involving at least one to four hops.

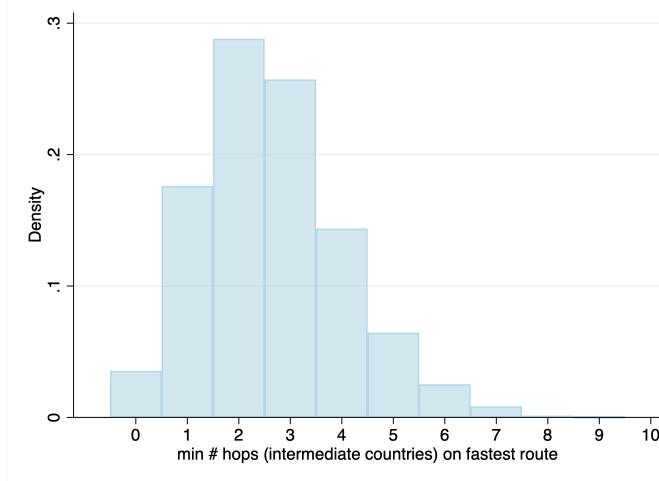
Our calculation of routes relies on the assumption that the fastest route will be the cost minimizing route, while the actual route chosen might be determined by other factors than speed, such as port costs. However, it is widely recognized that the overall cost efficiency of a ship depends on the total time it takes the ship to complete a voyage, see e.g. Cullinane and Khanna (2000). Therefore, the calculated fastest route is an approximation to the actual unobserved route. Appendix Section D.1 provides empirical evidence on the correlation between freight costs and travel time that supports this assumption. We also provide evidence that our calculated fastest routes captures the actual routes taken. Appendix Section D.2 compares our computed routes with the actual routes for Chinese trade, based on detailed Chinese customs data. The comparison shows that there is a high degree of overlap between the fastest-time routes and the actual routes in the Chinese data.

Figure 2: Travel time and distance across port-pairs.



Note: The figure plots travel times on the fastest route between two ports against their geodetic distance.

Figure 3: The distribution of the number of hops across country-pairs.



Note: The figure shows the distribution of the number of hops (intermediate countries) along the fastest route between all country pairs in the sample. The average (median) is 2.6. (3). For countries with multiple ports, the number of hops refers to the route with the lowest number of hops.

3 The Panama Canal Expansion

As background for our analysis, we first describe the Panama Canal and the 2016 expansion project. We then present the Panama Canal exposure measure, which is key to the empirical analysis.

3.1 The Panama Canal Expansion: Background

The Panama Canal, with its unique location at the narrowest point between the Atlantic and Pacific oceans, is one of the most important links of worldwide maritime trade. It reduces the time for ships to travel between the Atlantic and Pacific oceans, enabling them to avoid the lengthy, hazardous Cape Horn route around the southernmost tip of South America (Figure 4). The motivation for the Panama Canal expansion project was twofold. First, because of the rapid increase in global trade, the Panama Canal started to reach its capacity constraint. In 1914, its builders estimated that the maximum capacity of the canal would be around 80 million tons per year (Gerstle, 1944); however, the canal traffic had already reached 278.5 million tons in 2005.⁷ The Panama Canal Authority (ACP) estimated that the canal would reach its maximum sustainable capacity between 2009 and 2012. Second, container ships were getting larger thanks to the advancement of marine engineering, and many of them were too big to use the canal. The size of ships that could transit the original canal, called Panamax, was constrained by the size of the locks. By the turn of the millennium, only 41 percent of container ships and 52 percent of dry bulk ships were able to pass through the original canal (Wilson and Ho, 2018). We provide more details about the container shipping market in Appendix E.

Therefore, the ACP decided in 2006 to expand the canal by adding a new, deeper and wider lane of traffic. The expansion project was approved in April 2006, and the construction began in 2007 with an estimated total cost of US\$5.25 billion (Panama Canal Authority, 2006). The ACP initially announced that the Canal expansion would be completed by August 2014. But various setbacks, including strikes and disputes with the construction consortium over cost overruns, pushed the completion date back several times. There was, therefore, substantial uncertainty about exactly when the expanded

⁷The 2005 statistic stems from the US bureau of transportation and statistics, Panama Canal traffic by type 2005-2007.

canal would open. The expanded canal began commercial operation on 26 June 2016.

The enlarged canal was a formidable feat of modern engineering: it doubled the shipping capacity of the canal, allowing for around 90 percent of the world’s containerships to pass. In particular, the expanded canal allowed for the passage of so-called Neo-Panamax ships, which carry almost three times as much cargo compared to Panamax ships.⁸ From June to December 2016, the share of Neo-Panamax ships passing through the canal increased from 0 to 15 percent. In 2017, the canal container tonnage increased by 22% (Bowen, 2017).

Overall, the Panama Canal expansion is a good case for our study for several reasons. First, the canal is one of the most important links in the global shipping network. Therefore, the expansion was likely to have large aggregate and distributional impacts. Second, the old canal continued to operate both during and after the construction period, facilitating clean identification of the effects of the canal expansion on global trade. Third, uncertainty about the exact opening date of the expanded canal suggests that anticipation effects around the opening time were limited. Fourth, because of its unique location, traveling through the Panama Canal is substantially more time- and cost-efficient than alternative routes. This makes it possible to calculate the set of port-pairs that rely on shipping through the canal using our fastest route algorithms, and thereby to compute a suitable measure of exposure to the expansion.

3.2 Measuring Exposure to the Panama Canal Expansion

To measure the impact of the canal expansion on bilateral trade, we construct three different exposure variables, which reflect the extent to which the trade between two countries relies on using a shipping route that passes through the Panama Canal. Our first exposure variable, $PanExp_{ij}^1$, relies on the model-independent route planner described in Section 2.3 above: if a route passes through the Panama Canal in the pre-expansion period according to our fastest travel-time algorithm, we assign its exposure value equal to one,

⁸Panamax and Neo-Panamax ships refer to the size limits for ships traveling though the original and expanded canal, respectively. Panamax ships can carry roughly 5,000 twenty-foot containers (TEUs), while Neo-Panamax ships can carry roughly 13,000 TEUs. As the new lane opened, a new toll structure was introduced that differentiated across ship size. It implied higher rates for bigger ships on a per-ship basis, but lower rates for bigger ships on a per-container basis (see Wilson and Ho, 2018).

and zero otherwise. For country pairs with multiple port-to-port connections, we average the exposure of all port pairs using the source and destination port size as weights, where port size is measured in terms of total incoming (outgoing) tonnes in 2016.

The relevance of our exposure measure depends on whether the fastest time algorithm correctly predicts port-to-port connections that actually use the canal. Section 2.3 presents evidence supporting the assumption that ships travel the fastest route. We note that this assumption is more likely to hold if travel time on alternative routes is much higher. We test this for the Panama Canal by calculating, for port-pairs that use the Panama Canal, how many additional days it would take if they were to travel without passing through the canal. The results are presented in Figure 15 in the Appendix. We find that on average, travel time would increase by 14 days, or by 67%, for these port pairs. For 97.6% of the affected routes, travel time would increase by 3 days or more. As travel time is closely related to the major drivers of shipping costs, such as cost of fuel, labor compensation, and cost of running capital, it is likely to dominate the decision on whether or not to use the canal.

The relevance of our measures also depends on routes being stable over time. To address this, we compute the Panama Canal exposure measure for the post-expansion period using data for the second half of 2016, and find that the correlation between the pre- and post-period exposure measure is as high as 0.95. Specifically, 95.3% of port-pairs experience zero change, while 2.5% (2.2%) increase (decrease) their exposure, suggesting the shipping network is stable over time.

Table 3 presents summary statistics for the Panama Canal exposure measure in 2016. At the country-pair level, there are 3,623 country pairs (14% of 25,025 pairs with positive trade flows) which are to some extent connected via the Panama Canal according to our fastest route calculation. The value shipped between these countries accounts for 12% of global trade. At the country level, the majority of countries are in some way exposed to the Panama Canal: 66% of all importers have at least one fastest connection to a trade partner that passes through the canal. Across all importers, the average share of imports exposed to the Panama Canal is 7%. Figure 5 shows the share of imports passing through the Panama Canal by country. In line with our expectations, the canal is particularly important for Asian and American trade.

The second measure of exposure, $PanExp_{ij}^2$, is based on the general equilibrium model

Figure 4: International Shipping Routes with vs. without the Panama Canal



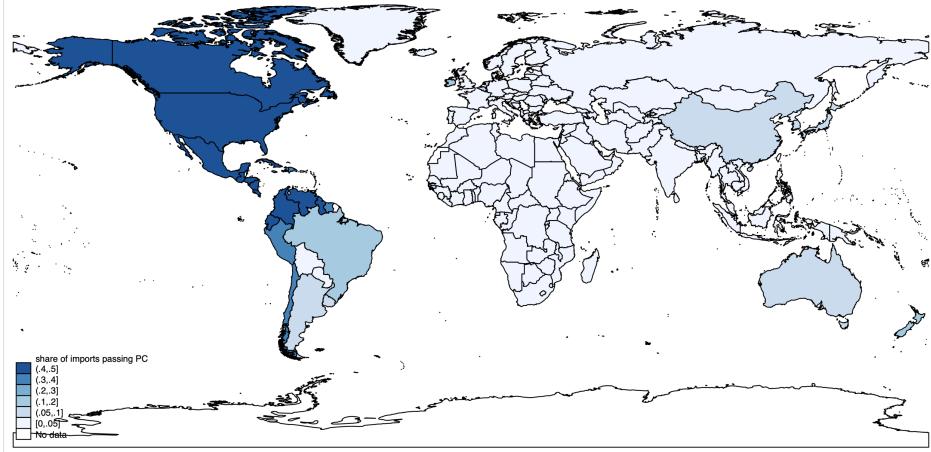
Source: <https://www.ajot.com/news/panama-canal-inauguration-a-strategic-route-for-world-trade-and-cma-cgm-gro>.

Table 3: Panama Canal Exposure: Summary Statistics for 2016.

Country pairs with exposure		Global trade exposed		Importers with exposure	
(1) # pairs	(2) % of total	(3) value in trn \$	(4) % of total	(5) # importers	(6) % of total
3,623	14 %	1.8	12 %	144	66 %

Note: The table shows in column (1) ((2)) the number (share) of country pairs with a fastest and most frequent connection passing the Panama Canal; in column (3) ((4)) the value of (share of global) trade between country pairs whose fastest and most frequent connection passes the Panama Canal; in columns (5) and (6), respectively, the number of importers with at least one fastest connection passing the Panama Canal and their share in the total number of importers.

Figure 5: Panama Canal Exposure by Country.



Note: The figure shows the share of imports passing through the Panama Canal in total imports by country.

presented in Section 5. $PanExp_{ij}^2$ is the likelihood that port pair ij uses the canal. We provide more details about the alternative exposure measure in Appendix I.

4 The Impact of the Panama Canal Expansion on Global Trade

In this section, we estimate the impact of the Panama Canal expansion on global trade. Section 4.1 specifies our empirical strategy and Section 4.2 presents the main estimation results. In Section 4.3, we perform a set of robustness checks.

4.1 Empirical Strategy

To analyze the effect of the canal expansion on global trade, we use the quarterly Comtrade data from the time period 2013Q1 to 2019Q4. Table 12 in the Appendix Section G summarizes the estimation sample. We employ a simple difference-in-differences specification:

$$y_{ijt} = \beta Post_t \times PanExp_{ij}^k + \delta \cdot Z_{ijt} + \delta_{ij} + \delta_{it} + \delta_{jt} + \varepsilon_{ikt}, \quad (1)$$

where y_{ijt} is log of imports from country i to country j in quarter t .⁹ The dummy variable $Post_t$ equals 1 for quarters after June 2016 and 0 otherwise. Panama Canal exposure is captured by the variable $PanExp_{ij}^k$, with $k = 1, 2$. Our hypothesis is that the canal expansion has a benign impact on trade. We would therefore expect the sign of $PanExp_{ij}^1$ and $PanExp_{ij}^2$ to be positive and of similar magnitude.

To account for the potential importance of geography and pre-expansion trading conditions, Z_{ijt} includes a set of bilateral controls: a dummy for joint membership in a free trade agreement (FTA), bilateral geographical variables (distance, contiguity and common language) and the share of deadweight tonnes traveling on Neopanamax ships on the route connecting i and j prior to the expansion, all of which are interacted with the $Post_t$ dummy. We also include a large set of fixed effects: origin-time δ_{it} and destination-time fixed effects δ_{jt} , which control for country-specific exporting and importing trends, and source-destination fixed effects δ_{ij} , which control for all time-invariant country-pair characteristics.

4.2 Empirical Results

We present the estimation results in Table 4. Columns (1) and (2) use the baseline $PanExp_{ij}^1$ measure, with the full set of fixed effects. Column (2) further includes the control variables Z_{ijt} . The point estimates are similar in both cases: for country pairs whose fastest routes are fully exposed to the Panama Canal, trade increased by around 10 percent after the expansion. On average, a one standard deviation increase in $PanExp_{ij}^1$ (0.21) implies about a 2.27 percent increase in trade. Columns (3) and (4) use the $PanExp_{ij}^2$ measure. We note that the model-derived exposure measure produces results that are in line with our baseline results.

Pre-trends. Figure 6 shows the estimated coefficient β by quarter, when using the $PanExp_{ij}^1$ treatment. We find no significant difference in the pre-trends for the country pairs with high and low Panama Canal exposure, indicating that the identifying assumption holds. The treatment effects in the post-expansion period are positive and centered around 0.10, although not individually significant, but the sum of them, which corresponds to the regression in Column (2) in Table 4, is strongly significant.

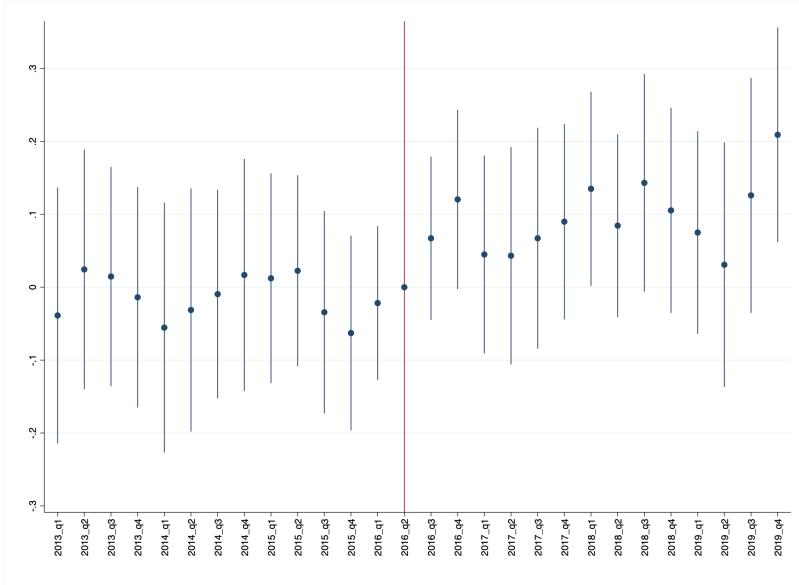
⁹We use imports by country of consignment, rather than country of origin. Country of consignment is the country where the last ownership change occurred before goods arrive in the importing country.

Table 4: The Impact of the Panama Canal Expansion on Trade

	(1)	(2)	(3)	(4)
$Post_t \times PanExp_{ij}^1$.105*** (.038)	.108*** (.040)		
$Post_t \times PanExp_{ij}^2$.156*** (.055)	.151** (.060)
Controls	No	Yes	No	Yes
FEs	ij, it, jt	ij, it, jt	ij, it, jt	ij, it, jt
Observations	199,177	199,177	199,177	199,177
Exporters/Importers	140/105	140/105	140/105	140/105
adj. R^2	.937	.937	.937	.937

Note: Dependent variable is the log of imports from country i to country j in quarter t over the period 2013Q1 – 2019Q4. The control variables are: an FTA indicator and geographical variables (distance, contiguity and common language), all interacted with $Post_t$, and the share of deadweight tonnes traveling on Neopanamax ships on the route connecting i and j in the pre period interacted with $Post_t$. Standard errors are clustered by i, j . Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure 6: Panama Canal Exposure Coefficient by Quarter.



Note: Graph illustrates regression equation: $y_{ikt} = \beta \sum_{q=2013:q1}^{2019:q4} I[t = q] \times PanExp_{ij}^1 + \delta \cdot Z_{ijt} + \delta_{ij} + \delta_{it} + \delta_{jt} + \varepsilon_{ijt}$ where Z_{ijt} further includes $\ln Dist$ interacted with quarter dummies. Solid lines indicate 90% confidence intervals.

4.3 Robustness

To check the robustness of our results, we re-estimate the specification from column (2) of Table 4 for a shorter time span, for a balanced panel, for monthly data, and for a narrower measure of trade which excludes raw materials. In addition we present results using a set of alternative Panama exposure measures, based on the same brute-force algorithm used for $PanExp_{ij}^1$. The results are reported in Table 5. Overall, we find that the results are relatively insensitive to various perturbations of the data and the exposure measure, underscoring the robustness of the baseline results.

Column (1) of Table 5 shows that the estimated magnitude is not driven by the disproportionately large effect in Q4 of 2019 visible in Figure 6. Column (2) complements our quarterly baseline analysis with monthly-level data and finds a similar result: a one standard deviation increase in $PanExp$ (0.21) implies about a 1.72 percent increase in monthly trade. The slightly smaller coefficient estimate in the monthly regression (.082 vs. .108 in the baseline) is consistent with the fact that there is more measurement error in monthly flows, which is smoothed out by aggregating to quarters.

In columns (3) and (4) of Table 5, we use all the optimal routes found by our brute-force algorithm, instead of relying on the most frequent routes, to construct alternative Panama exposure measures. In column (4), the exposure measure at the port-to-port level is the simple average across the binary exposure measure of all the optimal routes that appeared at least once in the first half of 2016. In column (4), the routes are weighted by the length of the time periods during which they were optimal. Aggregation from the port-to-port level to the country-pair level is done the same way as the baseline. The estimation results remain quantitatively similar in both cases, suggesting our results are not sensitive to the particular way we constructed the Panama Canal exposure.

An hypothesis is that country pairs for which avoiding the Panama Canal takes more time would be relatively more positively affected by the canal expansion. To test this hypothesis, we propose an additional measure, $\log(HoursSaved_{ij})$, based on the relative attractiveness of the first-best versus the second-best route. Specifically, for port pairs that use the Panama Canal in their fastest route, we set $\log(HoursSaved_{ij})$ equal to the log travel-hour difference between the fastest route and the fastest alternative which does not use the canal. For port pairs that do not use the Panama Canal in their fastest

Table 5: Robustness

	(1)	(2)	(3)	(4)	(5)	(5)
	2019Q4 dropped	Monthly data	Simple avg.	Weighted avg.	$\log(HoursSaved_{ij})$	No raw materials
$Post_t \times PanExp_{ij}^1$.102** (.040)	.082** (.040)	.098** (.048)	.103** (.048)	.023*** (.008)	.082*** (.027)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
FEs	ij, it, jt	ij, it, jt	ij, it, jt	ij, it, jt	ij, it, jt	ij, it, jt
Observations	193,450	600,884	199,177	199,177	199,177	198,696
Exporters/Importers	140/105	140/107	140/105	140/105	140/105	140/105
adj. R^2	.937	.900	.937	.937	.937	.939

Note: In column (1) the last quarter of 2019 is dropped. Column (2) is based on monthly data for the full sample period. Columns (3) and (4) are based on a $PanExp_{ij}^1$ measure computed as a simple (column (3)) and weighted (column (4)) average across the exposure of all paths between two ports in i and j , rather than the exposure of the most frequent route. Weights in column (4) are given by the amount of time for which a certain path was optimal, that is, the number of hours between the start date of the path and the start date of the previous optimal path relative to the length of the pre period. Column (5) is based on the estimated time saved by using the Panama Canal. Column (6) excludes raw materials. Standard errors are clustered by i and j . Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

route, $\log(HoursSaved_{ij})$ is set equal to zero.¹⁰ We then aggregate the variable to the country-pair level in the same way as for the baseline exposure measure. The results are presented in column (5): country pairs that would save more time by passing through the Panama Canal experienced a relative increase in trade after the expansion, consistent with our baseline findings.

Finally, we aim to limit the analysis to trade that is typically carried by container ships. We do this by excluding raw materials from the measure of bilateral trade.¹¹ As raw materials are rarely transported in containers, we do not expect that the fastest route algorithm (and therefore the Panama Canal exposure measure) to correctly identify their dependence on the Panama Canal. Column (6) presents the results. As expected, we find

¹⁰The time difference variable $HoursSaved_{ij}$ is computed using the same data as in Figure 15.

¹¹Raw materials comprise all products classified under the HS two-digit levels 25-27 and 72-81.

a significantly positive treatment effect for trade excluding raw materials. Overall, these results strengthen the baseline findings that the results are indeed driven by the role of the Panama Canal expansion in facilitating more efficient container shipping.

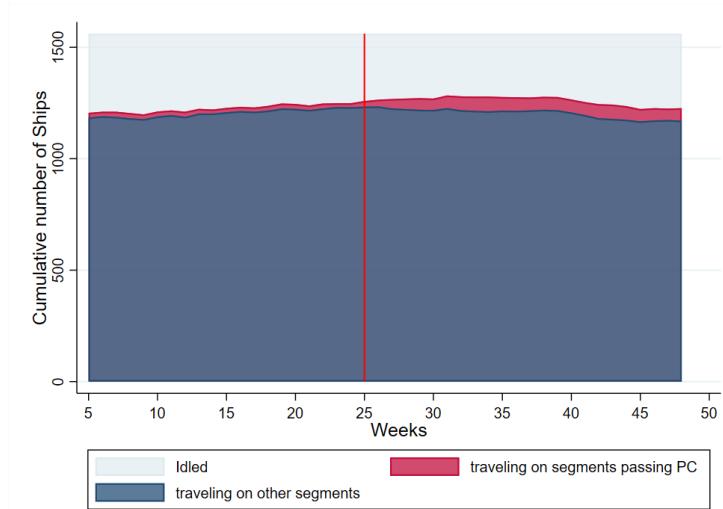
4.4 Re-routing of Ships

We end this section by discussing the interpretation of the results above. As is well known in differences-in-differences analyses, the coefficient estimate can only identify the impact on the treatment group relative to the control group. In our context, this means that we can only identify the increase in trade for the treatment country-pairs relative to the control country-pairs. For example, if larger container ships are re-routed from control to treatment pairs, one might observe more trade between treatment pairs, but also less trade between control pairs.

We investigate this question by analyzing the movements of Post-Panamax ships, i.e. ships that are beyond the size limits for traveling through the old canal, before and after the canal expansion. Specifically, we analyze Post-Panamax ships' weekly activities by classifying them into three mutually exclusive categories: (i) ships that are traveling on segments of the shipping network passing through the canal, (ii) on other segments, (iii) or staying idle. The results are presented in Figure 7. First, the figure shows that, even though the number of large ships serving Panama Canal segments increased after the canal expansion, they account for a small fraction of the fleet of Post-Panamax ships.¹² This finding suggests that large containerships had already been widely adopted globally. Second, the figure shows that there was a significant number of idle containerships in every given week, suggesting that the shipping industry had free capacity during this period. All in all, this suggests that re-routing was relatively limited, and as such that the control group was not much affected by the canal expansion. Appendix E provides more details about the container shipping industry and demonstrates that our findings are consistent with the industry's general development since 2006.

¹²Vessels that are slightly above the Panamax size limit could still pass if they were not fully loaded, therefore the number of Post-Panamax ships passing through the canal before expansion is not zero.

Figure 7: Number of Post-Panamax Ships by activity type (2016).



Note: The total number of Post-Panamax ships refers to the number of Post-Panamax ships built before 2016 as reported by Clarkson World Fleet Register. Post-Panamax ships are defined as ships with a maximum carrying capacity of more than 5,500 TEU (or 52,500 deadweight tonnage (DWT) if TEU is missing).

5 Theoretical Framework

Our empirical analysis provides evidence of the impact of the Panama Canal expansion on global trade. In order to assess the international externalities associated with the shipping network, we introduce a parsimonious quantitative model of world container traffic and trade. The model differs from standard trade models in how the transportation of goods is modeled: goods are not only shipped between two trading partners, but pass through a shipping network when departing from an origin and arriving in a destination port. Agents choose the optimal route endogenously in order to minimize transport costs.

We build on and extend the work of Allen and Arkolakis (2022): while their application is on urban economics, where individuals choose where to live and commute, our focus is on international trade, where goods move across borders subject to transport costs. After presenting the model in this section, we apply the model to quantify the network externalities of the Panama Canal expansion in Section 6. Finally, we quantify the network externalities of all links in the transportation network, and contrast our finding with a canonical trade model without a transportation network.

5.1 Model Setup

Consider a world with N locations indexed by i and j , each endowed with L_i units of labor. There is a continuum of varieties indexed by $\nu \in [0, 1]$. Individuals have constant elasticity of substitution (CES) preferences over varieties with elasticity of substitution $\sigma \geq 0$. Labor is the only input, and is inelastically supplied for producing and shipping goods. Shipping from an origin i to a destination j entails taking a route r through the network, which is subject to multiplicative iceberg transport costs $\prod_{k=1}^K t_{r_{k-1}, r_k}$. Here, K is the number of links on route r and t_{r_{k-1}, r_k} denotes the transport costs of traveling through the k^{th} link of r . We let R_{ij} denote the set of all possible shipping routes from i to j . The efficiency of producing and shipping each variety ν from i to j via route r is characterized by $\varepsilon_{ij,r}(\nu)$. We assume that $\varepsilon_{ij,r}(\nu)$ is independently and identically Frechet distributed with level parameter A_i and dispersion parameter θ .

Individuals purchase each variety from the cheapest location-route source. The idiosyncratic shocks $\varepsilon_{ij,r}(\nu)$ imply that not all varieties are traveling though the same route even if the source and destination ports are the same, e.g. variety ν_1 going from Lisbon

to Oakland may pass through Rotterdam while variety ν_2 may pass through Houston. A possible micro-foundation for the shocks $\varepsilon_{ij,r}(\nu)$ is heterogeneity in the preferred time of shipment, e.g. route planners report multiple routes between Lisbon and Oakland, and those routes are available on different dates.¹³ Furthermore, it buys us tractability in terms of producing analytical expressions for many key objects of interest.

5.2 General Equilibrium

We now turn to solving the general equilibrium and characterizing global trade and container traffic. We impose two market clearing conditions, total income Y_i equals total sales, and total expenditure E_i equals total purchases:

$$Y_i = \sum_j X_{ij} \quad E_i = \sum_j X_{ji}, \quad (2)$$

where X_{ij} is the total value of goods shipped from i to j . Using the market clearing conditions and the properties of the Frechet distribution, it can be shown that X_{ij} equals

$$X_{ij} = \tau_{ij}^{-\theta} \frac{Y_i}{\Pi_i^{-\theta}} \frac{E_j}{P_j^{-\theta}}, \quad (3)$$

where

$$\tau_{ij} = \left(\sum_{r \in R_{ij}} \prod_{l=1}^K t_{r_{l-1}, r_l}^{-\theta} \right)^{-1/\theta} \quad (4)$$

is the shipping cost from i and j . The variable Π_i is the standard multilateral resistance term known from gravity models:

$$\Pi_i^{-\theta} = (A_i L_i)^{-\theta} Y_i^{\theta+1}, \quad (5)$$

and P_j is the consumer price index:

$$P_j^{-\theta} = \sum_i \tau_{ij}^{-\theta} Y_i \Pi_i^\theta. \quad (6)$$

With balanced trade, $E_i = Y_i$, we now formally define the equilibrium of the model:

¹³See e.g. <https://www.cma-cgm.com/ebusiness/schedules/routing-finder>.

Definition 1. Given $\{L_i\}$, $\{A_i\}$ and $\{\tau_{ij}\}$, an equilibrium is a output vector $\{Y_i\}$, expenditures $\{E_i\}$, bilateral trade flows $\{X_{ij}\}$, $\{\Pi_i\}$ and $\{P_i\}$ that satisfies equilibrium conditions (2), (3), (5), (6), as well as the balanced trade condition, for all i, j .

At the bilateral level, the model aggregates to a standard Ricardian trade model. However, the shipping costs τ_{ij} are no longer “bilateral”; instead, they are an endogenous outcome of consumers’ optimal routing problem. Its value depends on the number of routes available linking locations i and j , and the transport cost of each route, which depends on the shipping costs t_{kl} of all segments on that route. Since wages are the only source of income, we can solve for nominal wages w_i in location i using $w_i = Y_i/L_i$. Welfare of individuals is then simply w_i/P_i .

Solving for the Equilibrium

Defining $\mathbf{A} \equiv [t_{ij}^{-\theta}]$, Allen and Arkolakis (2022) show that transport costs τ_{ij} can be rewritten as

$$\tau_{ij} = b_{ij}^{-1/\theta}, \quad (7)$$

where b_{ij} is the ij^{th} elements of the matrix \mathbf{B} and $\mathbf{B} = (I - \mathbf{A})^{-1}$. Using equation (7) along with the gravity equation and the market clearing conditions (3) and (2), we can write the equilibrium conditions as

$$\Pi_i^{-\theta} = \frac{E_i}{P_i^{-\theta}} + \sum_j t_{ij}^{-\theta} \Pi_j^{-\theta} \quad (8)$$

$$P_i^{-\theta} = \frac{Y_i}{\Pi_i^{-\theta}} + \sum_j t_{ji}^{-\theta} P_j^{-\theta}, \quad (9)$$

When trade is balanced, $E_i = Y_i = \Pi_i^{-\theta/(\theta+1)} (A_i L_i)^{\theta/(\theta+1)}$, and given values of t_{kl} , A_i and L_i , the $2N$ equations (8) and (9) can be solved for the $2N$ equilibrium outcomes Π_i and P_i . In the quantitative application below, we will write equations (8) and (9) in changes following the “exact hat algebra” approach by Dekle et al. (2007), to solve for counterfactual equilibrium changes.

Trade and Traffic

We end this subsection by characterizing traffic flows according to the model. We define traffic as the total value of all cargo passing through a segment (k, l) in the

network. It can be shown that, in equilibrium, the value of traffic between k and l is

$$\Xi_{kl} = t_{kl}^{-\theta} P_k^{-\theta} \Pi_l^{-\theta}. \quad (10)$$

Furthermore, there is a simple mapping between trade and traffic: One can express equilibrium trade flows as:

$$X_{ij} = c_{ij}^X Y_i E_j \quad (11)$$

where c_{ij}^X is the $(i, j)^{th}$ element of the matrix $\mathbf{C}^X \equiv (\mathbf{D}^X - \boldsymbol{\Xi})^{-1}$, where \mathbf{D}^X is a diagonal matrix with i^{th} element $d_i \equiv \frac{1}{2}(Y_i + E_i) + \frac{1}{2} \sum_j (\Xi_{ji} + \Xi_{ij})$ and $\boldsymbol{\Xi} = [\Xi_{ij}]$ (Allen and Arkolakis, 2022).

6 Network Externalities of the Canal Expansion

We now turn to applying the model to assess whether the Panama Canal expansion gave rise to network externalities, and if so, to quantify the sign, and magnitude, of the externalities. First, we estimate the change in transportation costs associated with the expansion. Second, we calculate a counterfactual equilibrium with lower transportation costs and quantify the local and global impact on real incomes. Finally, we evaluate the fit and performance of the quantified model.

6.1 Estimating the impact on Transportation Costs

We use the structure of the model to estimate the impact of the canal expansion on transport costs. We start by assuming that transport costs t_{kl} are a function of the canal expansion:

$$t_{kl} = e^{-\delta PanExpand_{kl}} \nu_{kl}, \quad (12)$$

where ν_{kl} captures unobserved factors that determine transportation costs and $PanExpand_{kl}$ is one if link kl is using the canal and zero otherwise. Recall that the model gives us the following equilibrium expression for traffic flows: $\Xi_{kl} = t_{kl}^{-\theta} P_k^{-\theta} \Pi_l^{-\theta}$. Taking logs, differencing and inserting equation (12) yields

$$\Delta \ln \Xi_{kl} = \delta \theta PanExpand_{kl} - \theta \Delta \ln \Pi_l - \theta \Delta \ln P_k - \varepsilon_{kl}, \quad (13)$$

where $\varepsilon_{kl} \equiv -\theta\Delta \ln \nu_{kl}$ and Δ denotes the change from the 1st to the 2nd half of 2016. The multilateral resistance variables $\Delta \ln \Pi_l$ and $\Delta \ln P_k$ will be controlled for using origin and destination fixed effects, respectively.

While the model gives us an expression for the value of traffic, Ξ_{kl} , our data has information about the volume of traffic, Ξ_{kl}^V . We proceed by assuming that the volume and value of traffic are proportional, i.e. $\Xi_{kl} = \alpha \Xi_{kl}^V$, so that $\Delta \ln \Xi_{kl}^V = \Delta \ln \Xi_{kl}$. We acknowledge that this assumption may be overly restrictive in the cross-section, e.g. traffic between some port-pairs (k, l) may have higher unit values than between other port pairs (k', l') . For the purposes of inferring $\theta\delta$, however, the key (and less restrictive) requirement is simply that the unit value of traffic through the canal does not change pre/post the canal expansion, relative to the control group.

The estimate of $\delta\theta$ then allows us to infer the change in transport costs from equation (12), given knowledge about the trade elasticity θ . Specifically, the change in transport costs is:

$$\Delta \ln t_{kl} = -\frac{\hat{\delta}\theta}{\theta} PanExpand_{kl}. \quad (14)$$

We estimate equation (13) using the AIS data, for every port-pair kl , from the 1st to the 2nd half of 2016 (recall, the expanded canal opened 26 June 2016). By construction, the volume of traffic on a segment kl , Ξ_{kl}^V , is the product of three margins:

$$\Xi_{kl}^V = ShipSize_{kl} \times Frequency_{kl} \times Utilization_{kl},$$

where $ShipSize_{kl}$ is the average size of ships on link kl , $Frequency_{kl}$ is the number of ships and $Utilization_{kl}$ is the percentage of used ship capacity. All these margins are observed directly in the AIS data. The construction of the dataset is described in detail in Appendix Section B. We estimate equation (13) using each of these margins as outcome variables.

Results. The estimation results are shown in Table 6. Columns (1), (4) and (7) show results without any controls, whereas columns (2), (5) and (8) are results with the following controls: source- and destination country fixed effects, source and destination port latitude and longitude, source and destination port capacity (total traffic), and average travel time between k and l . Columns (3), (6) and (9) are estimation results when we instead include source- and destination port fixed effects. We find that average ship size

Table 6: The Margins of Shipping: Results

Dep. var.	$\Delta \ln ShipSize_{kl}$			$\Delta \ln Frequency_{kl}$			$\Delta \ln Utilization_{kl}$			$\Delta \ln Time_{kl}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$PanExpand_{kl}$.22*** (.05)	.14** (.06)	.12* (.06)	.01 (.12)	.02 (.14)	.02 (.15)	.05 (.04)	.02 (.05)	.02 (.05)	.07 (.06)	.01 (.04)	.01 (.04)
Controls	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes	No
Source/dest. FE	No	No	Yes	No	No	Yes	No	No	Yes	No	No	Yes
Obs	3,595	3,566	3,403	3,595	3,566	3,403	3,595	3,566	3,403	3,595	3,566	3,403

Notes: The difference Δ refers to the change from the 1st to 2nd half of 2016. $ShipSize_{kl}$ is calculated as the average across all trips on a given segment. $Frequency_{kl}$ is the number of ships using the segment. $Utilization_{kl}$ is traffic Ξ_i relative to capacity ($ShipSize_{kl} \times Frequency_{kl}$). $Time_{kl}$ is the average travel time from k to l . $PanExpand_{kl}$ is an indicator taking the value one if the segment uses the Panama Canal. Regressions are weighted by the initial level of traffic Ξ_{kl} . Controls are: source- and destination country fixed effects, source and destination port latitude and longitude, source and destination port capacity (total traffic), and average travel time between k and l . Source/destination FE refers to source- and destination port fixed effects. Robust standard errors in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

increased by $.12 - .22$ log points, depending on the specification, for Panama Canal segments relative to other segments. This is as expected, because the expansion facilitated much larger ships passing through the canal, see Section 3.1. The other two margins, ship utilization and frequency, are estimated to be around zero and are statistically insignificant. This suggests that the impact of the expansion on trade runs mainly through allowing bigger ships to pass through the canal. We also investigate whether travel time changed due to the canal expansion. The last three columns in Table 6 report the results when (the change in log) average travel time from k to l is the outcome variable. This margin is also close to zero and statistically insignificant, reinforcing the conclusion that ship size was the main margin of adjustment.

Economic magnitudes. The estimate of $\delta\theta$ equals to the sum of the estimates for the three margins: ship size, frequency and utilization from Table 6. Across specifications, this yields a value of $\theta\delta$ between 0.15 and 0.25. In our baseline specification we use the value $\theta = 8$, implying that the canal expansion lowered transport costs by around 2-3

percent ($\hat{\theta}\delta/\theta \approx .2/8$).

6.2 Network Externalities

Exact hat algebra. The general equilibrium of the model presented above can be written in changes, using the “exact hat algebra” approach from Dekle et al. (2007). We focus on a change in transportation costs (estimated in the section above), holding everything else constant. Section H.1 in the Appendix shows that the system of equations can be simplified to:

$$\hat{\Pi}_i^{-\theta} = \frac{Y_i}{Y_i + \sum_j \Xi_{ij}} \frac{\hat{\Pi}_i^{-\theta/(\theta+1)}}{\hat{P}_i^{-\theta}} + \sum_j \left(\frac{\Xi_{ij}}{Y_i + \sum_j \Xi_{ij}} \right) \hat{t}_{ij}^{-\theta} \hat{\Pi}_j^{-\theta} \quad (15)$$

$$\hat{P}_i^{-\theta} = \frac{Y_i}{Y_i + \sum_j \Xi_{ji}} \frac{\hat{\Pi}_i^{-\theta/(\theta+1)}}{\hat{\Pi}_i^{-\theta}} + \sum_j \left(\frac{\Xi_{ji}}{Y_i + \sum_j \Xi_{ji}} \right) \hat{t}_{ji}^{-\theta} \hat{P}_j^{-\theta}, \quad (16)$$

where the hat notation is the relative change from the initial to the counterfactual equilibrium, e.g. $\hat{t}_{kl} = t'_{kl}/t_{kl}$ where t'_{kl} is transportation costs in the counterfactual equilibrium. After solving this system of equations, the relative change in real income is simply $\hat{w}_i/\hat{P}_i = \hat{\Pi}_i^{-\theta/(\theta+1)}/\hat{P}_i$.

The data and parameters required to solve this system are modest: We need the estimate of the change in Panama Canal transportation costs from Section 6.1, \hat{t}_{kl} , data on initial (i.e. 1st half of 2016) traffic Ξ_{kl} , initial expenditure Y_i and the trade elasticity θ . We set $\theta = 8$, which is consistent with previous estimates in the literature (see e.g. Eaton and Kortum (2002)).¹⁴ The procedure to calibrate the value of α (the relationship between the volume and value of traffic) is described in Appendix H.3. To calculate total expenditure by port, Y_i , we use data for total country expenditure and allocate expenditure to ports based on the relative port size, see details in Appendix Section G.3. We summarize the data and parameters used in Table 7.

Results. A convenient measure of network externalities is the change in global real income relative to the change in real income for Panama. If this ratio is greater than one, then the canal expansion was associated with positive international externalities. According to our results, this ratio is 3.14, i.e. the gains for the world was more than

¹⁴The quantification exercise is based on a slightly smaller set of ports (492) and countries (149), due to the fact that it requires a balanced dataset of port-to-port flows for the 1st half of 2016.

Table 7: Data and Parameters.

Variable	Description	Value	Source
θ	Trade elasticity	8	Previous literature
α	The value per ton of traffic	USD 1,734	Calibrated, Appendix H.3
Ξ_{kl}^V	Initial traffic flows (volume)		AIS, Marine Traffic, 1st half 2016
Y_i	Initial Expenditure		Eora Global Supply Chain Database, 2015; World Development Indicators; INSEE

three times greater than the gains for Panama alone. Our results therefore suggest that the expansion was associated with large positive international externalities. Figure 8 shows the change in real income, w_i/P_i , for the top 20 locations with the largest gains. Not surprisingly, the ports closest to the canal are gaining the most. However, we also observe ports further away, such as in Colombia, Ecuador, Peru and the Caribbean, that obtain large welfare gains from the canal expansion.

6.3 Mechanisms and Model Fit

This section evaluates the out-of-sample fit of the model. According to the model, the canal expansion caused trade to grow by \hat{X}_{ij} . In order to evaluate the performance of the model, we estimate a version of the reduced-form equation (1) from Section 4.1, but we replace real data on trade flows with the simulated trade data from the model (i.e. we replace the left hand side variable with simulated data).

The results are shown in Table 8. Column (1) performs the analysis at the port-pair level, while column (2) aggregates the data to the country level, similar to Table 4 in Section 4.2. Interestingly, the treatment effect using simulated data is very close to the treatment effect using real data. Furthermore, the high adjusted R^2 values also suggest that the model fits the data quite well. Recall that the counterfactual model only required data on initial traffic Ξ_{kl} and expenditure Y_i , in addition to an estimate of the change in transport costs \hat{t}_{kl} and the trade elasticity θ . Therefore, the model-generated growth in trade is an out-of-sample prediction, i.e. we did not use trade flows, neither in levels nor in changes, when parameterizing the model.

Figure 8: Real income, % change. Top 20 ports.

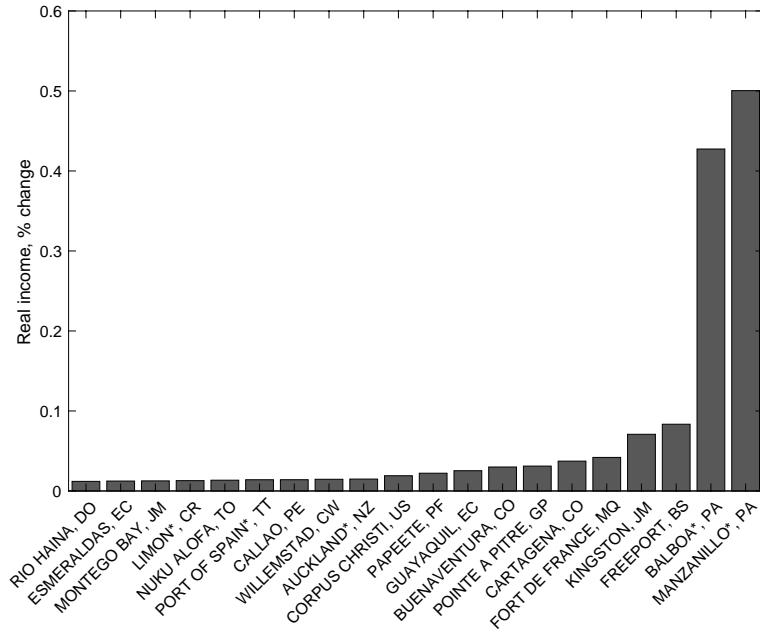


Table 8: Reduced-form regression: Simulated data.

Dependent variable: \hat{X} (simulated)	Port-pair	Country-pair
$PanExp_{ij}^1$.083*** (.002)	.085*** (.006)
Source/destination FE	Yes	Yes
Observations	240,064	10,253
Dep. ports/Arr. ports	490/492	138/102
adj. R^2	.726	.736

Notes: The dependent variable is the relative change in exports from port i to port j implied by the model, \hat{X} . All columns include i and j fixed effects. Standard errors clustered by i and j . Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

7 The Distribution of Network Externalities

Throughout the paper, we have emphasized the role of the shipping network for international externalities. In this section we show how our framework also can be used to quantify the magnitude of international externalities associated with all links in the global shipping network.

We proceed as follows: For each link in the shipping network with observed positive traffic $\Xi_{kl} > 0$ (in total 3,821 links), we remove that link from the network, i.e. set $t_{kl} = \infty$, and calculate the counterfactual equilibrium for the world economy, as described in Section 6.2 above. This methodology is similar to so-called link removal analyses that have been used extensively to study real-world networks in different fields of science (Bellingeri et al., 2020).

For each dropped link kl , we calculate the global change in real income:

$$\Delta Global^{kl} = \sum_i Y_i \left(\frac{\hat{w}_i^{kl}}{\hat{P}_i^{kl}} - 1 \right),$$

and the local change in real income (for locations k and l):

$$\Delta Local^{kl} = \sum_{i=k,l} Y_i \left(\frac{\hat{w}_i^{kl}}{\hat{P}_i^{kl}} - 1 \right).$$

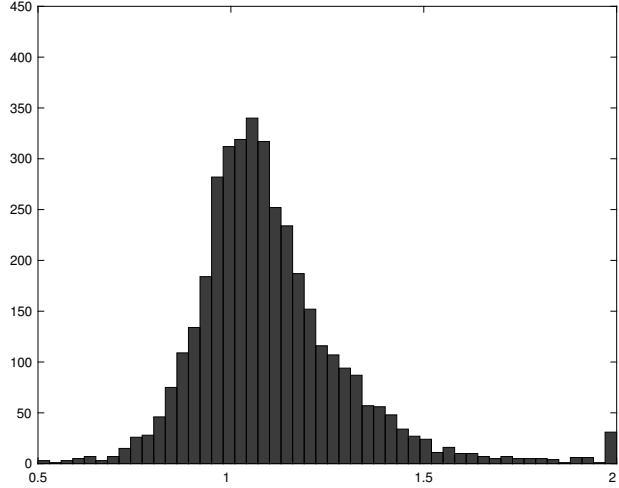
In these expressions, superscript kl refers to the counterfactual where link kl is dropped, whereas subscripts refer to ports. Our preferred measure of international externalities is

$$Ext^{kl} = \frac{\Delta Global^{kl}}{\Delta Local^{kl}}. \quad (17)$$

Similar to the analysis in Section 6.2, if $Ext^{kl} > 1$, then the drop in global income is larger than the drop in local income, implying that the link produces a positive externality for countries other than k and l . If $Ext^{kl} < 1$, then the opposite is true, so that the link produces a negative externality for countries other than k and l .

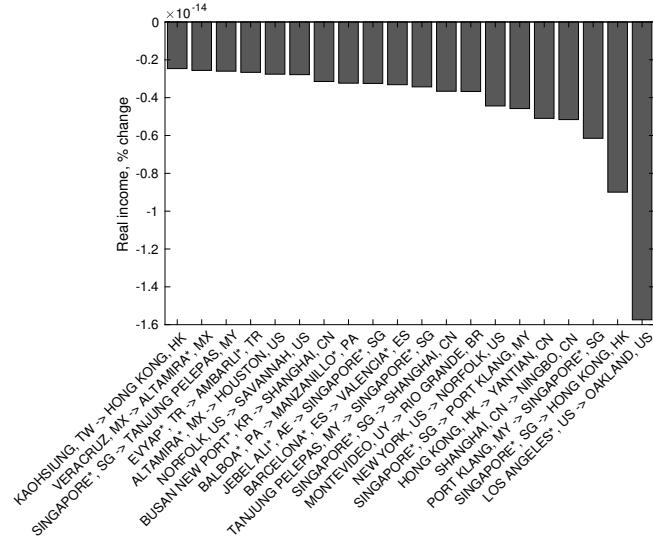
Figure 9 shows the distribution of Ext^{kl} across all dropped links kl . The mean of Ext^{kl} is 1.12, i.e. the global losses are 12 percent greater than the local losses. Therefore, our results show that the average link in the global shipping network produces a positive, and quantitatively important, international externality. For 71 percent of the links, $Ext^{kl} > 1$, showing that positive network externalities are widespread.

Figure 9: International Externalities.



Note: The figure shows the distribution of Ext^{kl} across dropped links kl . To aid visibility, the histogram is censored at 0.5 and 2.

Figure 10: International Externalities : Top 20 Links.



Note: The figure shows $(\hat{RealInc}^{-kl} - 1) \cdot 100$ for the 20 links with the largest negative relative change.

Next, we explore which links produce the most positive externality in an absolute sense. Specifically, we calculate the relative change in global income, but exclude locations k and l from the calculation:

$$Real \hat{I}nc^{-kl} = \sum_{i \neq k,l} \frac{Y_i}{\sum_{j \neq k,l} Y_j} \frac{\hat{w}_i^{kl}}{\hat{P}_i^{kl}}.$$

Figure 10 shows the links associated with the largest drop in global income (excluding locations k and l). Many of the world's busiest links are associated with the greatest positive externalities: According to our results, Los Angeles to Oakland, Singapore to Hong Kong and Port Klang to Singapore are among the top externality links.

A No-network Benchmark We end this section by contrasting the results above with counterfactual results coming from a canonical model of international trade without a transportation network. Our hypothesis is that the positive externalities identified above will disappear, and even turn negative. The reason is that a better link between location b and c is likely to hurt a rather than help a in a non-networked environment: Lower trade costs will divert trade from a to b and c , which may lower real incomes in a .¹⁵

We proceed as follows. In the absence of a shipping network, trade costs τ_{ij} are exogenous. Using the same definition of the equilibrium as described in Section 5, we solve the model in changes. Appendix Section J shows that the system of equations is

$$\hat{\Pi}_i^{-\theta} = \sum_j \hat{\tau}_{ij}^{-\theta} \frac{\hat{E}_j}{\hat{P}_j^{-\theta}} \frac{X_{ij}}{Y_i} \quad (18)$$

$$\hat{P}_i^{-\theta} = \sum_j \hat{\tau}_{ji}^{-\theta} \frac{\hat{Y}_j}{\hat{\Pi}_j^{-\theta}} \frac{X_{ji}}{E_i} \quad (19)$$

After imposing trade balance, $\hat{E}_j = \hat{Y}_j$, and using the fact that $\hat{Y}_i = \hat{\Pi}_i^{-\theta/(\theta+1)}$, we can solve this system of equations given data on initial trade flows X_{ij} , expenditure E_i , output

¹⁵This is a common finding from the literature on preferential trade agreements. While Viner (1950) raised the (in his eyes, unlikely) possibility of positive welfare effect in third countries arising from the “general diffusion of the increased prosperity of the customs union area” (p.105), a wide range of theoretical models predict that negative externalities dominate (Riezman, 1979, Baldwin et al., 2003, Behrens et al., 2007 and Mossay and Tabuchi, 2015). While empirical evidence on welfare effects is hard to come by, several studies document negative effects on trade between third countries and members of a preferential trade agreement (see, e.g. Freund and Ornelas, 2010 for a summary of findings).

Y_i , as well as the change in trade costs, $\hat{\tau}_{ij}$, and the elasticity, θ . As in the counterfactuals above, we set $\hat{\tau}_{ij} = \infty$ for each possible link with positive traffic. Initial trade flows X_{ij} , expenditure E_i and output Y_i are calculated as follows. In the network model, we only used data for initial traffic Ξ_{ij} , and no data for trade X_{ij} . To make the two counterfactuals comparable, the initial values need to be consistent across models. We do this by first converting the traffic data Ξ_{kl} to trade data X_{ij} , using equation 11. Total income and expenditure is then simply the sum across rows and columns in the trade matrix, $Y_i = \sum_j X_{ij}$ and $E_i = \sum_j X_{ji}$. By doing so, the two models are calibrated to the same initial steady state.

As above, we calculate our measure of externality from equation (17). Figure 11 shows the distribution of international externalities, Ext^{kl} , across all links. The mean of Ext^{kl} is now 0.96, showing that the average link produces a negative externality for other countries. For 64 percent of the links, there is a negative externality, i.e $Ext^{kl} < 1$.

For completeness, we also report the density of Ext^{kl} when imposing symmetry in trade costs, i.e. $\tau_{ij} = \tau_{ji}$.¹⁶ Figure 12 shows the distribution of Ext^{kl} across all links. The mean of Ext^{kl} is 0.93 and for 100 percent of the links, $Ext^{kl} < 1$. Therefore, the canonical model with trade cost symmetry will always produce negative externalities from changes in trade costs. We therefore conclude that in the absence of a transportation network, reduced trade barriers will typically harm rather than benefit third countries.

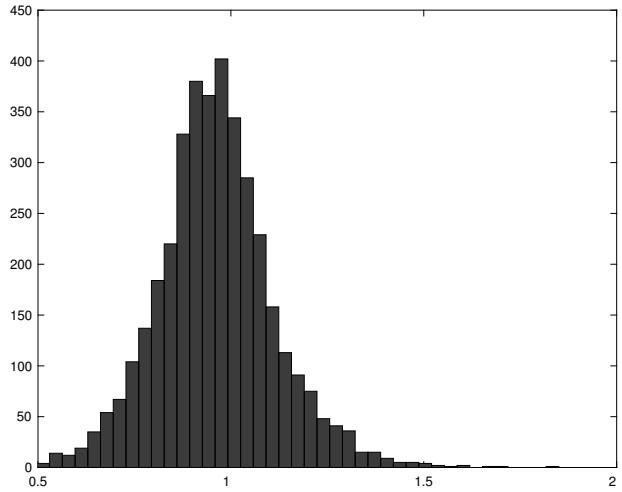
8 Concluding remarks

In this paper we exploit novel satellite data on all port calls made by container ships in 2016. This allows for the construction of a new comprehensive dataset on the global shipping network and optimal shipping routes. We apply this dataset to analyze how local shocks hitting a segment of the shipping network affect all trading partners worldwide to varying degrees based on their exposure to the shock.

Using the 2016 Panama Canal expansion as a quasi-natural experiment, we show that this expansion not only had an effect on trade flows directly exposed to the canal, but also had widespread indirect effects on world trade due to countries' indirect exposure to the canal through the global shipping network. Based on a counterfactual analysis we

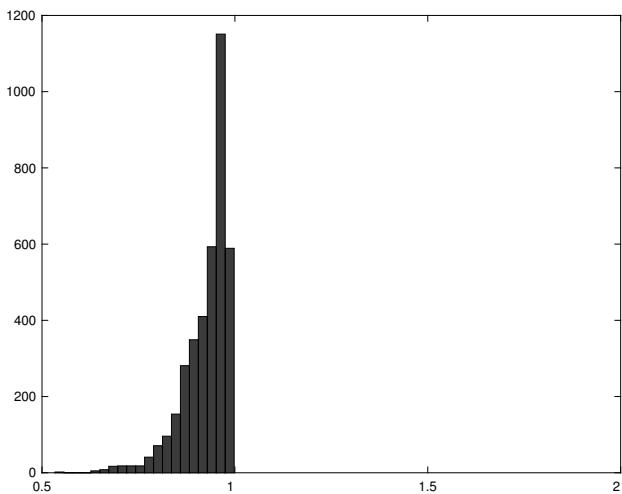
¹⁶In practice, we take the average of X_{ij} and X_{ji} before iterating on equations (18) and (19).

Figure 11: International Externalities : No Shipping Network.



Note: The figure shows the distribution of Ext^{kl} across dropped links kl .

Figure 12: International Externalities : No Shipping Network and Trade Cost Symmetry.



Note: The figure shows the distribution of Ext^{kl} across dropped links kl .

find that the Panama Canal expansion produced sizable gains in real income, and that the real income gains were shared by many countries. Therefore, our research points to positive and quantitatively important international externalities associated with the canal expansion.

Finally, we show that our framework can be used to address the more general question of international externalities in a global transportation network. Using a link removal strategy, we find that the large majority of the port-pair-links are associated with positive international externalities. Our results illustrate that in the presence of global transportation networks, reduced trade barriers between two trading partners may indirectly also benefit third countries. We compare our findings to an economic environment without a transportation network, and find that in the absence of a network, reduced barriers between trade partners typically harm rather than benefit third countries.

Our quantitative results point to the interdependences created by the global shipping network, and suggest that as long as infrastructure costs are not shared across countries, the presence of positive externalities implies that the world invests too little in global transportation networks. We leave it to future research to quantify the socially optimal level of investment, given our findings. We view the analysis as a first step in unpacking the overall impact of the network structure on global trade and income. A promising area of future research is to quantify how improvements in transportation technology may trigger upgrades in the shipping fleet and global port infrastructure with far-reaching implications for the level and distribution of world income.

References

- Adland, R., H. Jia, and S. P. Strandenes (2017). Are ais-based trade volume estimates reliable? the case of crude oil exports. *Maritime Policy and Management* 44(5), 657–665.
- Allen, T. and C. Arkolakis (2022). The welfare effects of transportation infrastructure improvements. *The Review of Economic Studies*.
- Baldwin, R., R. Forslid, P. Martin, G. Ottaviano, and F. Robert-Nicoud (2003). *Economic Geography and Public Policy*. Princeton University Press.
- Behrens, K., A. R. Lamorgese, G. I. Ottaviano, and T. Tabuchi (2007). Changes in transport and non-transport costs: Local vs global impacts in a spatial network. *Regional Science and Urban Economics* 37(6).
- Bellingeri, M., D. Bevacqua, F. Scotognella, R. Alfieri, and D. Cassi (2020). A comparative analysis of link removal strategies in real complex weighted networks. *Scientific Reports* 10(3911).
- Bernhofen, D. M., Z. El-Sahli, and R. Kneller (2016). Estimating the effects of the container revolution on world trade. *Journal of International Economics* 98, 36 – 50.
- Bowen, M. (2017). Moody's upgrades panama canal.
- Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2020). Geography, search frictions and endogenous trade costs. *Econometrica* (2), 657–691.
- Brooks, L., N. Gendron-Carrier, and G. Rua (2021). The local impact of containerization. *Journal of Urban Economics*.
- Cullinane, K. and M. Khanna (2000). Economies of scale in large containerships: optimal size and geographical implications. *Journal of Transport Geography* 8, 181–195.
- David, M. (2015). Vessels and ballast water. In M. David and S. Gollasch (Eds.), *Global Maritime Transport and Ballast Water Management*. Springer.
- Dekle, R., J. Eaton, and S. Kortum (2007). Unbalanced trade. *American Economic Review: Papers and Proceedings* 97(2), 351–355.

- Donaldson, D. and A. Storeygard (2016). The view from above: Applications of satellite data in economics. *Journal of Economic Perspectives* 30(4), 171–198.
- Ducruet, C., R. Juhasz, D. Nagy, and C. Steinwender (2022). All aboard: The aggregate effects of port development.
- Eaton, J. and S. Kortum (2002). Technology, geography and trade. *Econometrica* 70(5), 1741–1779.
- Feyrer, J. (2021). Distance, trade, and income - the 1967 to 1975 closing of the suez canal as a natural experiment. *Journal of Development Economics* 153.
- Freund, C. and E. Ornelas (2010). Regional trade agreements. *Annual Review of Economics* 2(1).
- Ganapati, S., W. F. Wong, and O. Ziv (2021). Entrepot: Hubs, scale, and trade costs.
- Gerstle, M. (1944). The land divided - a history of the panama canal and other isthmian canal projects.
- Levinson, M. (2006). *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger*. Princeton University Press.
- Maurer, S. and F. Rauch (2019). Economic geography aspects of the panama canal.
- Mossay, P. and T. Tabuchi (2015). Preferential trade agreements harm third countries. *The Economic Journal* 125(589).
- Panama Canal Authority (2006). Proposal for the expansion of the panama canal: Third set of locks project.
- Rajkovic, R., N. Zrnic, O. Cokorilo, S. Rajkovic, and D. Stakic (2014). Multi-objective container transport optimization on intermodal networks based on mathematical model.
- Rau, P. and S. Spinler (2016). Investment into container shipping capacity: A real options approach in oligopolistic competition. *Transport Research Part E* 93, 130–147.
- Riezman, R. (1979). A 3x3 model of customs unions. *Journal of International Economics* 9(3).

- Rua, G. (2014). Diffusion of containerization.
- UN (2013). *International Merchandise Trade Statistics Compilers Manual - Revision 1 (IMTS 2010-CM)*.
- UNCTAD (2016). *Review of Maritime Transport*.
- Wilson, W. W. and J. D. Ho (2018). The panama canal. In B. Blonigen and W. W. Wilson (Eds.), *Handbook of International Trade and Transportation*. Edward Elgar.
- Zhang, X., B. Podobnik, D. Y. Kenett, and H. E. Stanley (2014). Systemic risk and causality dynamics of the world international shipping market. *Physica A: Statistical Mechanics and its Applications* 415, 43–53.

Appendix

A Constructing the Container Traffic Data Set

Our point of departure are the AIS data containing all port calls made by ships in 2016 that has been provided by MarineTraffic. Based on the ship categories used by MarineTraffic, we limit the data set to the ships categorized as “container ship” and “Cargo/containership”. MarineTraffic provides each ship with a unique identifier (Ship ID). We start out with close to 5,300 ships based on this identifier. We use this to identify each ship’s travel history. A ship also has an IMO number and an MMSI number as well as a Ship Name. We use this information to merge the AIS data set with the World Fleet register data base constructed by Clarkson, which has vessel specific information on a range of time invariant ship characteristics, such as the vessels carrying capacity measured in deadweight tonnes (dwt) and cargo capacity of container ship measured in twenty-foot equivalent unit (TEU).

Ideally there should be a perfect match between ship identifiers (IMO, MMSI and Ship ID). However, for around 5% of the ships this is not the case. The mismatch could either because of misreporting, or changing of owners (containerships typically change their MMSI number when changing the owner). We correct for both misreporting and the change of identifiers by cross checking a ship’s IMO and MMSI number, as well as ship’s characteristics, like its deadweight tons (dwt). We are able to correct for most of the misreporting and end up with 5,165 distinct containerships. Finally, as we want to focus on global container traffic, we introduce a threshold of 15,800 deadweight tons. This leaves us with 4,941 ships.

We then proceed by cleaning the routes of each container ship. The AIS data are very rich with information on not just ports, but also on whether the ship is lading/unlading in a port, or is just *in transit* (e.g. due to need for additional fuels). In addition the data set has information on anchorages, i.e. stops made by ships in places that are not ports.

We sort trips for each ship by their time stamp, so that their travel records are listed as Arrival-Departure-Arrival-Departure, etc. A *trip* is defined as a direct port-to-port voyage. If a ship departs a port A, makes several *in transit* stops at other ports, or stops at anchorages, before finally arriving at port B, we define the voyage from A to B as one trip of the ship. We use the draught reported when the ship reaches the arrival port as

the draught of the trip. Moreover, we drop a small number trips for which the arrival time stamp erroneously equals the departure time stamp. We also drop trips that are taken by less than 5 ships during the year. Finally, we aggregate ports located within 30 kilometers of each other and within the same country and we drop ports that do not appear both as arrival and departure ports. We lose less than two percent of the shipped volume by imposing these restrictions on ship size, non-zero travel time, and the set of ports.

B Calculating Global Container Traffic

Based on the container traffic data set described above, we compute a set of measures to characterize the global container traffic for any port pair for a given period: (i) *frequency*, i.e. the number of ships traveling between the two ports; (ii) *ship size*, i.e. average ship size traveling in terms of deadweight tonnes (dwt); (iii) *shipments* (cargo), computed based on AIS data matched with data on ship characteristics; and (iv) *utilization*, calculated as shipments/(ship size \times frequency).

Due to the availability of AIS data, the use of draught-based estimates of ships' cargo has recently emerged in the maritime transport literature, see e.g. Adland et al. (2017). The draught of a ship refers to the vertical distance between the surface of the water and the lowest point of a vessel. We build on this approach, and as we limit the analysis to one type of ships, namely container ships, we are able to establish a relatively simple rule for the computation of the ships' container shipment. For each sailing ship we observe the draught reported by the ship en route, H_A , which will vary depending on the ship's cargo. A ship sailing without cargo is referred to as a ship sailing in ballast. In practice, a ship sails in ballast if its draught is smaller than a given threshold value, which we refer to as ballast draught (H_B). Specifically, we define $H_B = 0.55H_S$, where H_S is the ship's scantling draught. Scantling draught is the draught the ship will have when it is fully loaded, and it is also referred to as design draught, as it is this draught it is built for, and is thus a constant. We have access to technical information on ships' scantling draught as well as the vessel's carrying capacity (*dwt*) from the Clarkson World Fleet Database (see Section A above). We use 0.55 as the weight to define ballast draught based on the

maritime engineering literature.¹⁷ Letting H_A refer to the draught reported by the ship en route, we calculate the shipments carried by a ship on a specific voyage, as

$$EffectiveDWT = dwt * (H_A - H_B) / (H_S - H_B). \quad (20)$$

A ship's draught as well as estimated cargo relates to one specific *trip*, i.e. to a voyage between two ports.

Table 9 shows that, based on our draught-based estimates, on average container ships do merely 1% of their trips without cargo (in ballast). This stands in sharp contrast to other types of vessels that are typically involved in very different trades, and do not operate on “bus routes” like container ships. Brancaccio et al. (2020) focus on dry bulk ships and report that 42% of the ships travel without cargo. We also observe that there is substantial variation across trips with respect to draught, effective dwt, and across ports with respect to total incoming and outgoing cargo.

Table 9: Ships, Trips and Port

Variable:	Obs	Mean	Sd	Min	Max
Ships:					
Share of trips in ballast (<55%)	4,937	0.01	0.05	0	1
Trips:					
Actual draught (% of scantling draught)	331,249	0.94	0.07	0.55	1
Effective dwt on loaded trips	331,249	26,113.93	24,559.94	1.23	199,744
Ports:					
Total incoming effective dwt (in millions)	514	16.83	44.36	0.01	498.70
Total outgoing effective dwt (in millions)	514	16.83	44.34	0.01	499.98

Note: Summary statistics are based on the port calls made by container ships in 2016. Effective dwt is calculated based on dwt and draught and is used as a measure for cargo.

¹⁷The threshold for ballast water is chosen based on information from MarineTraffic supported e.g. David (2015).

C Fastest Route Calculation

Using the schedule of actual departure times and arrival times of all container ships in our dataset, we compute the fastest path from port i to port j at time h , where h measures hours since Jan 01 2016 00:00. Our algorithm solves the same problem as a public transport routing algorithm that delivers the fastest bus route from i to j at time h , allowing passengers to switch to feasible connecting buses at any stop. The only difference is that we compute these paths ex post, that is, we use the observed departure and arrival times in 2016, whereas the most common application of public transport algorithm is the calculation of optimal routes for specific times in the future.

Our algorithm works as follows. Every time a ship departs from i to anywhere, we compute all feasible paths a container can take to get to j through the network of connections available at that point in time, allowing the container to switch to any ship that leaves from the current port in the future. To limit the computational burden, we make two simplifying assumptions. First, we consider only paths involving up to 15 intermediate ports. Second, we convert time stamps from seconds to hours. That is, we convert the arrival and departure time stamps from the original format $yy:mm:dd\ hh:mm:ss$ to $yy:mm:dd:hh$, attributing arrivals or departures occurring within 30 minutes before and after hour hh to hour hh . Without loss of generality, we then convert the time stamp $yy:mm:dd:hh$ to an integer h that counts hours since Jan 01 2016 00:00. We treat all ships leaving port k at hour $h+1$ or later as a feasible connection for a container arriving at port k at hour h .

From the set of feasible paths, we drop all those that are dominated by other paths that start at the same time or later, but arrive earlier. We also drop paths that are identical to others in terms of travel time and arrival time, but involve more stops in intermediate ports. Note that any feasible path from i must start with one of the ships that actually departed from i . Then, suppose we have identified the fastest path from i to j starting with a ship departing i at time h_0 (denoted (i,j,h_0)) and another fastest path from i to j starting with a ship departing at time h_1 , which we denote by (i,j,h_1) . Then, for any hour $h' \in (h_0, h_1]$, the fastest way to get to j is to wait at i until hour h_1 and then take the path (i,j,h_1) . The total travel time of (i,j,h') is then the travel time along (i,j,h_1) plus the waiting time $h_1 - h'$. The result of the algorithm is then a set of paths, (at least) one for each port pair i,j at every starting time h in 2016. For every optimal

path (i, j, h) , we retain the information on the total travel time as well as the sequence of intermediate ports.

The algorithm is programmed in Stata. We ran the algorithm in parallel on 514 cores (one for each departure port) endowed with an Intel Xeon-Gold 6138 2.0 GHz processor on the Saga supercomputer (<https://www.sigma2.no/node/537>). The average (maximum) required CPU time per core denoted as $hh : mm : ss$ is 01:55:55 (05:50:37), the total CPU time is 993 hours. The maximum RAM required per core is 78920K. The result is 13,915,115 unique paths described by departure port i , arrival port j , departure time h , arrival time h_a and up to 15 intermediate ports.

D Empirical Evidence on Shipping Costs and Actual Shipping Routes

D.1 Freight costs and Fastest Routes

Our empirical analysis relies on the assumption that cargoes from a country i to a country j are shipped on the fastest route between the two countries. To justify this assumption we use trade data for the US by customs district and country of origin that allows us to back out freight costs and examine the correlation between freight costs and travel time on direct routes observed in the AIS data. The results are reported in Table 10. The dependent variable is freight costs computed as cif/fob margin relative to import value. The unit of observation is the freight cost of containerized imports by US customs district, country of origin, and product (10-digit HTS code). U.S. customs districts are matched to U.S. container ports based on names. Independent variables are travel time ($\ln Hours_{ij}$), geodetic distance ($\ln Dist_{ij}$), total dwt of ships traveling to US port j from country i in 2016 ($\ln DWT$), total number of ships traveling to US port i from country j in 2016 ($\ln Ships_{ij}$) and average ship size based on the latter two variables ($\ln AvgDWT_{ij}$). The analysis shows that there is a positive correlation between travel time and freight costs. This positive correlation remains also when we control for other potential determinants of freight costs such as distance and characteristics of the cargo flow. We note that there is also a negative correlation between freight costs and average ship size, indicating economies of scale in transport at the ship level.

Table 10: Correlations with Freight costs

	(1)	(2)	(3)	(4)	(5)	(6)
$\ln Hours_{ij}$.004*** (.000)		.005*** (.001)	.005*** (.001)	.004*** (.001)	.004*** (.001)
$\ln Dist_{ij}$.005* (.001)	-.001 (.001)	-.001 (.001)	.002 (.001)	-.001 (.001)
$\ln DWT$				-.000 (.000)		
$\ln Ships_{ij}$					-.000 (.001)	.000 (.001)
$\ln AvgDWT_{ij}$					-.005*** (.000)	-.005*** (.000)
FEs	j,p	j,p	j,p	j,p	j,p	j,p
Observations	167,227	167,227	167,227	167,227	167,227	187,011
Exporter/US ports	61/20	61/20	61/20	61/20	61/20	61/20
Products	13,086	13,086	13,086	13,086	13,086	13,296
adj. R^2	0.152	0.152	0.152	0.152	0.152	0.156

Note: Dependent variable is freight costs computed as cif/fob margin as share of the import value. Unit of observation is the freight cost of containerized imports by US port, country of origin, and product (10-digit HTS code). The sample is based on US trade in 2016 and include only transactions where the US port of entry is also the port of unlading. Columns (1)-(5) include only those port-country pairs where a US port is connected to only one port in the partner country. Column (6) includes all port-country pairs and the values of all independent variables are computed as averages across multiple ports in the exporting country. All regressions include fixed effects for US ports and products. Standard errors are clustered at the product level. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D.2 Evidence on Actual Routes of Chinese Trade

We have access to Chinese customs data for 2006, where we observe both transportation method and one transit country. We can therefore check whether the transit ports according to our fastest route algorithm overlap with the transit country in the Chinese data.

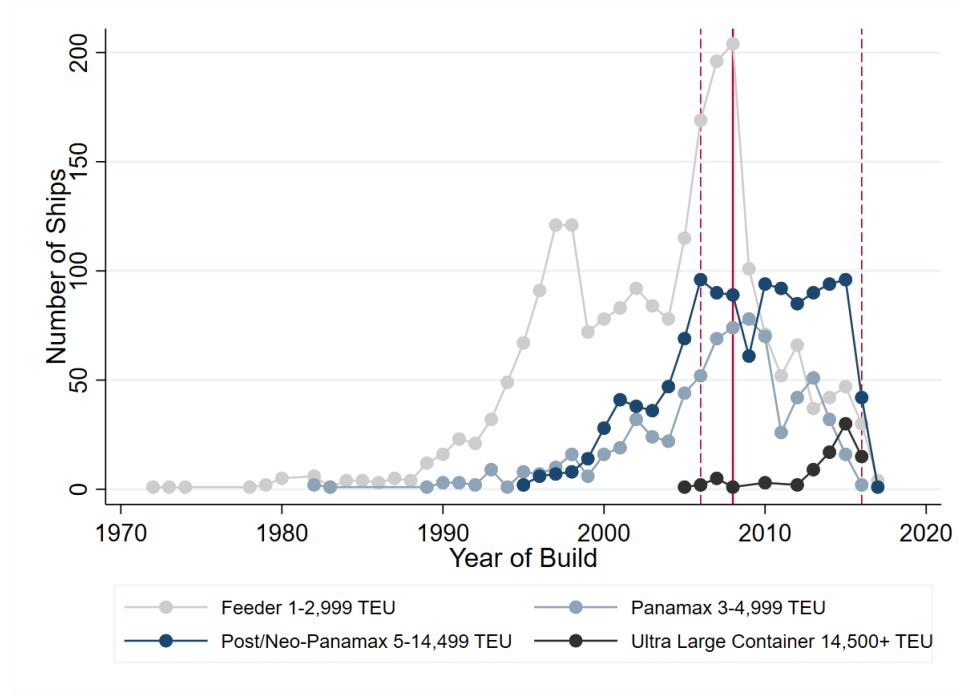
We perform the following analysis. First, we aggregate the Chinese data to the origin-transit country-destination level (imports or exports), and only keep observations where transportation method is by sea. Second, for each origin-transit-destination triplet in the Chinese data, we check whether we find a similar triplet according to the fastest route algorithm. We find that 87% of the fastest-time routes we identified for Chinese imports and exports include transit countries that are matched with origin-transit-destination triplets in the Chinese data. At the same time 30% of the origin-transit-destination triplets in the Chinese data are matched with triplets in our constructed fastest-time data set. However, the trade values of the matched triplets are on average 10 times higher than the unmatched ones, in total making up about 81% of the Chinese marine trade in 2016. Our finding suggest that fastest-time routes correctly capture the main routes Chinese trade takes.

E The Container Shipping Market

We have compiled data on the size distribution of all container ships over time. Specifically, using data from Clarksons, we can look the ship size distribution *by year of construction*. Figure 13 plots the number of ships by different size bins and by construction year. Interestingly, the number of Neopanamax ships (i.e., ships that cannot pass through the old canal, but can pass through the new canal), has been relatively stable between the announcement year and completion year (marked by dashed lines in the figure). Except for ultra large container ships, which cannot pass the Canal even after expansion, all three categories experienced a decline in newly build ships after the financial crisis (solid-line). The numbers support our view that large ships (Post/Neopanamax ships) were already widely adopted globally before the expansion project started, while the Panama Canal was a bottleneck of global container shipping. The numbers strongly indicate that the canal expansion was not sufficient to incentivize owners to invest in new ships. In addition to this, the container shipping industry has been characterized by over-investment and idle ship capacity for many years (see Figure 14), in the wake of the 2007 financial crisis and trade collapse, see e.g. Zhang et al. (2014). Data from the consulting industry shows

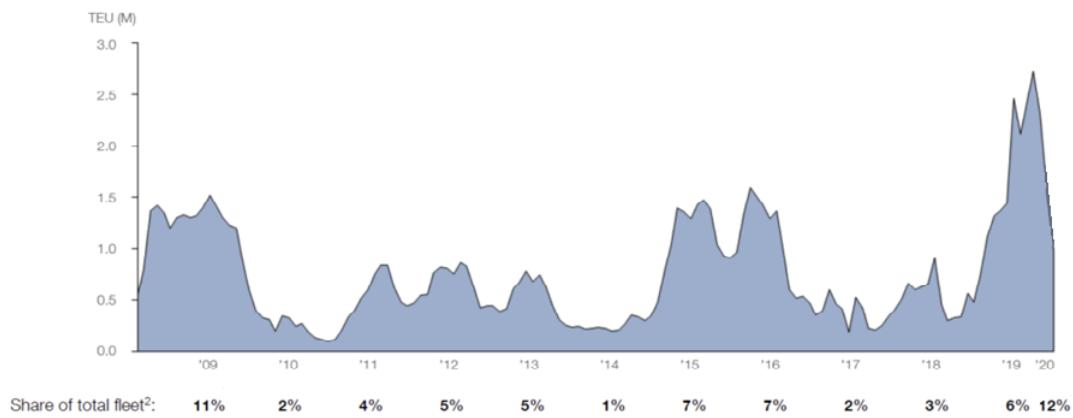
that around 5% of container ships were idle over the period 2009 to 2016, see Figure 14.¹⁸. Moreover, container freight rates have also been relatively low over the 2006-2016 time period, consistent with the finding that there was ample capacity in the market, see e.g. Rau and Spinler (2016).

Figure 13: Ship Building



¹⁸<http://www.globaltrademonitor.com/2020/09/21/flexport-idle-container-ship-capacity-is-returning-to-normal-levels-after-increases-in-q2/>

Figure 14: Idle Containership Capacity 2009-2020



F Panama Canal Exposure

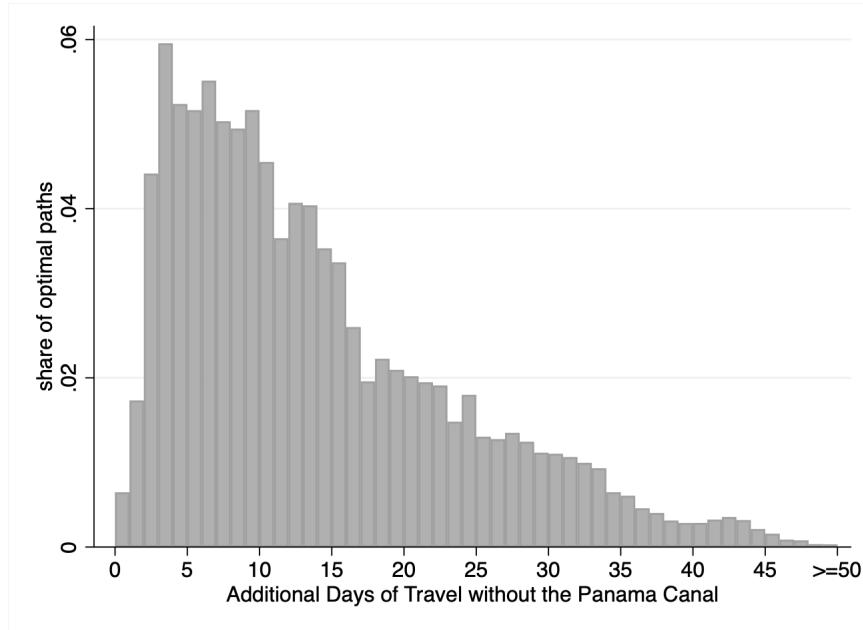
F.1 Summary Statistics

Table 11: Summary statistics on Panama Canal exposure

Rank	Importer	Share of total		Exporter	Share of total	
		imports passing PC	imports		exports passing PC	exports
1	USA	50.8	14.0	USA	30.5	9.0
2	MEX	10.2	2.5	CHN	16.0	14.9
3	CAN	9.6	2.7	MEX	12.1	2.6
4	CHN	4.2	7.7	CAN	10.1	2.5
5	JPN	2.8	3.7	JPN	5.8	4.3
6	KOR	1.7	2.6	DEU	3.4	8.5
7	DEU	1.6	6.5	KOR	3.4	3.4
8	GBR	1.6	4.2	GBR	1.3	2.6
9	CHL	1.2	0.4	FRA	1.3	3.3
10	COL	1.2	0.3	ITA	1.2	3.1
11	BRA	1.1	0.9	CHL	1.1	0.4
12	BLX	1.1	2.6	BRA	1.1	1.3
13	NLD	1.0	3.0	IRL	0.9	1.1
14	AUS	0.9	1.2	PER	0.8	0.2
15	FRA	0.9	3.8	COL	0.8	0.2

F.2 Sailing without the Panama Canal

Figure 15: Travel Time without use of the Panama Canal



Note: The figure shows the distribution of port-to-port travel time differences with vs. without the Panama Canal for the pairs that are using the Panama canal according to our algorithm.

The figure is created as follows. First, we identify all the paths that pass the Panama Canal according to our algorithm. Second, we remove the canal from the shipping network and recalculate new, second-fastest, paths using our algorithm. Third, we compare travel time before and after removing the canal.

G Additional Data Sources

G.1 COMTRADE Trade Flows

The monthly COMTRADE data was downloaded via the API call "<https://comtrade.un.org/api/get/plus?max=250000&type=C&freq=M&px=HS&ps=inserttimeperiod&r=insertreportercode&p=all&rg=all&cc=TOTAL&fmt=csv&token=inserttoken>" between Jan 8-10, 2021.

We aggregate monthly observations to quarters and keep only quarters where trade flows were reported in every month. We use the total value of imports by destination and country of consignment (i.e., the country from which goods were dispatched to the final destination; see UN (2013, p 185)). For imports of raw materials used in Table ??, we downloaded monthly trade data from the same source using the API call “<https://comtrade.un.org/api/get/plus?max=250000&type=C&freq=M&px=HS&ps=inserttimeperiod&r=insertreportercode&p=all&rg=all&cc=25,26,27,72,73,74,75,76,77,78,79,80,81&fmt=csv&token=inserttoken>”.

G.2 Estimation Data: Summary Statistics

The estimation sample covers about 82% of global imports reported to Comtrade. The missing 18% are due to countries not reporting trade data to Comtrade on a monthly basis (which are aggregated to the quarterly level).

Table 12: Summary Statistics of the Estimation Sample

Variable	N	Mean	Std. Dev	Min	Max	Source
ln Value (in \$, by quarter)	199,177	16.2	3.27	1.39	25.74	monthly COMTRADE
FTA	199,177	.30	.46	0	1	WTO RTA database
ln Distance	199,177	8.66	.82	4.55	9.89	CEPII
Contiguity	199,177	.02	.15	0	1	CEPII
Common Language	199,177	.14	.35	0	1	CEPII
Pan Exposure	199,177	.14	.35	0	1	AIS data

Note: Export data in rows 1 is aggregated from monthly to quarterly frequency and covers the period 2013Q1- 2019Q4.

G.3 Data for the Model-based Quantification

We distribute total expenditure by country across ports according to the relative size of ports measured by the total incoming tonnes in the first half of 2016, according to the AIS data. In our sample, 46 percent of countries have only one port. Expenditure by country is taken from the Eora Global Supply Chain Database (MRIO) (<https://worldmrio.com/>).

For 19 out of the 149 countries (small islands and overseas territories) expenditure data is not available. We construct the missing expenditure level using GDP data for these countries obtained from the Worldbank's World Development Indicators and from INSEE together with the average expenditure/GDP ratio of small islands for which we do observe both expenditure and GDP.

H The Model

H.1 Solving the model in changes

This section shows how to solve the general equilibrium of the model in changes, using the “exact hat” notation developed in Dekle et al. (2007).

The first equilibrium condition is

$$\begin{aligned} Y_i &= \sum_j X_{ij} \\ Y_i &= \frac{Y_i}{\Pi_i^{-\theta}} \sum_j \tau_{ij}^{-\theta} \frac{E_j}{P_j^{-\theta}} \\ \Pi_i^{-\theta} &= \sum_j \tau_{ij}^{-\theta} \frac{E_j}{P_j^{-\theta}}, \end{aligned}$$

where we substituted in for the gravity equation and solved for Π_i . In matrix notation, this can be rewritten as:

$$\begin{aligned} [\Pi_i^{-\theta}] &= [1 - A]^{-1} \left[\frac{E_i}{P_i^{-\theta}} \right] \\ [1 - A] [\Pi_i^{-\theta}] &= \left[\frac{E_i}{P_i^{-\theta}} \right] \\ [\Pi_i^{-\theta}] - A [\Pi_i^{-\theta}] &= \left[\frac{E_i}{P_i^{-\theta}} \right] \\ \Pi_i^{-\theta} - \sum_j t_{ij}^{-\theta} \Pi_i^{-\theta} &= \frac{E_i}{P_i^{-\theta}} \\ \Pi_i^{-\theta} &= \frac{E_i}{P_i^{-\theta}} + \sum_j t_{ij}^{-\theta} \Pi_i^{-\theta}. \end{aligned}$$

In a similar fashion, the second equilibrium condition can be rewritten as:

$$\begin{aligned}
E_i &= \sum_j X_{ji} \\
E_i &= \frac{E_i}{P_i^{-\theta}} \sum_j \tau_{ji}^{-\theta} \frac{Y_j}{\Pi_j^{-\theta}} \\
P_i^{-\theta} &= \sum_j \tau_{ji}^{-\theta} \frac{Y_j}{\Pi_j^{-\theta}} \\
[P_i^{-\theta}] &= [1 - A']^{-1} \left[\frac{Y_i}{\Pi_i^{-\theta}} \right] \\
[1 - A'] [P_i^{-\theta}] &= \left[\frac{Y_i}{\Pi_i^{-\theta}} \right] \\
P_i^{-\theta} - \sum_j t_{ji}^{-\theta} P_j^{-\theta} &= \frac{Y_i}{\Pi_i^{-\theta}} \\
P_i^{-\theta} &= \frac{Y_i}{\Pi_i^{-\theta}} + \sum_j t_{ji}^{-\theta} P_j^{-\theta}.
\end{aligned}$$

Expressed in changes, the two equilibrium conditions become

$$\begin{aligned}
\hat{\Pi}_i^{-\theta} &= \frac{E_i}{E_i + \sum_j \Xi_{ij}} \frac{\hat{E}_i}{\hat{P}_i^{-\theta}} + \sum_j \left(\frac{\Xi_{ij}}{E_i + \sum_j \Xi_{ij}} \right) \hat{t}_{ij}^{-\theta} \hat{\Pi}_j^{-\theta} \\
\hat{P}_i^{-\theta} &= \frac{Y_i}{Y_i + \sum_j \Xi_{ji}} \frac{\hat{Y}_i}{\hat{\Pi}_i^{-\theta}} + \sum_j \left(\frac{\Xi_{ji}}{Y_i + \sum_j \Xi_{ji}} \right) \hat{t}_{ji}^{-\theta} \hat{P}_j^{-\theta}.
\end{aligned}$$

Since trade is balanced, $E_i = Y_i$. Furthermore, by using the fact that $\hat{\Pi}_i = \hat{Y}_i^{-(\theta+1)/\theta}$, we can write the system as

$$\hat{\Pi}_i^{-\theta} = \frac{Y_i}{Y_i + \sum_j \Xi_{ij}} \frac{\hat{\Pi}_i^{-\theta/(\theta+1)}}{\hat{P}_i^{-\theta}} + \sum_j \left(\frac{\Xi_{ij}}{Y_i + \sum_j \Xi_{ij}} \right) \hat{t}_{ij}^{-\theta} \hat{\Pi}_j^{-\theta} \quad (21)$$

$$\hat{P}_i^{-\theta} = \frac{Y_i}{Y_i + \sum_j \Xi_{ji}} \frac{\hat{\Pi}_i^{-\theta/(\theta+1)}}{\hat{\Pi}_i^{-\theta}} + \sum_j \left(\frac{\Xi_{ji}}{Y_i + \sum_j \Xi_{ji}} \right) \hat{t}_{ji}^{-\theta} \hat{P}_j^{-\theta}. \quad (22)$$

H.2 Algorithm for solving the equilibrium in changes

The system of equations (21) and (22) can be solved with a simple fixed point procedure. We start with a guess of $\hat{\Pi}_i^{-\theta}$ and $\hat{P}_i^{-\theta}$. We then update equation (21) to get a new value of $\hat{\Pi}_i^{-\theta}$. We then update equation (22) to get a new value of $\hat{P}_i^{-\theta}$. We iterate on the two fixed points until the system converges.

World output is the numeraire, $\sum_i Y_i = Y^W = 1$. Specifically, when iterating on the fixed points above, for each iteration, we rescale $\hat{\Pi}_i$ and \hat{P}_i so that $\hat{Y}^W = 1$ holds. We have

$$\hat{Y}^W = \sum_i \frac{Y_i}{Y^W} \hat{Y}_i = \sum_i \frac{Y_i}{Y^W} \hat{\Pi}_i^{-\theta/(\theta+1)}.$$

After each iteration of equation (21), we calculate \hat{Y}^W and then rescale $\hat{\Pi}_i^{-\theta}$ by dividing by $\hat{Y}_i^{\theta+1}$.

H.3 Converting Ξ_{kl}^V to Ξ_{kl}

This section describes how to calibrate the value α in the expression $\Xi_{kl} = \alpha \Xi_{kl}^V$. The methodology is as follows: First, start with a guess of the value of α , α^0 , and obtain values of Ξ_{kl} . According to the model, there is a mapping between traffic Ξ_{kl} and trade X_{ij} according to equation (11). After converting traffic to trade, we calculate the value of world container trade flows, i.e. $\tilde{X}^W = \sum_{ij, i \neq j} X_{ij}$, according to the model. If \tilde{X}^W is different than the true value of world container trade, X^W , i.e. $\tilde{X}^W - X^W \neq 0$, we update the guess of α , and continue to do so until $\tilde{X}^W - X^W = 0$. The value of α that delivers $\tilde{X}^W - X^W = 0$ is USD 1734 per deadweight tonnage of traffic.

The world value of container trade, X^W , is calculated as follows. According to Rajkovic et al. (2014), the global value of container trade was 5.6 trillion USD in 2010. According to the WTO, world merchandise trade increased by 4.6 percent from 2010 to 2016. Under the assumption that the share of container trade in total merchandise trade is constant, world container trade in 2016 is 5.9 trillion USD (5.6 trillion USD \times 1.046)

I Alternative Exposure Measures

This section discusses the alternative Panama Canal exposure measure. We parameterize transport costs t_{kl} as

$$t_{kl} = TravelTime_{kl}^\delta, \quad (23)$$

where $TravelTime_{kl}$ refer to average travel time across all trips on a link kl . We set $\theta = 8$ (as in the main text) and $\delta = 0.15$, consistent with a gravity distance elasticity of roughly 1 ($\delta \times \theta \approx 1$). We calculate trade costs τ_{ij} by invoking equation (7). We can then

calculate the likelihood of using a link kl for trade between i and j . The likelihood is (see Allen and Arkolakis, 2022):

$$\pi_{ij}^{kl} = \left(\frac{\tau_{ij}}{\tau_{ik} t_{kl} \tau_{lj}} \right)^\theta.$$

Define \mathbb{P} as the set of links that use the Panama canal, according to the container traffic data. The model-derived likelihood of using the canal is calculated as

$$\pi_{ij}^{PA} = \max_{kl \in \mathbb{P}} \pi_{ij}^{kl}.$$

The alternative exposure measure is $PanExp_{ij}^2 = \pi_{ij}^{PA}$.

J Solving the no-network model in changes

This section shows how to solve the general equilibrium of the no-network model in changes, using the “exact hat” notation developed in Dekle et al. (2007).

The first equilibrium condition is

$$\begin{aligned} Y_i &= \sum_j X_{ij} \\ Y_i &= \frac{Y_i}{\Pi_i^{-\theta}} \sum_j \tau_{ij}^{-\theta} \frac{E_j}{P_j^{-\theta}} \\ \Pi_i^{-\theta} &= \sum_j \tau_{ij}^{-\theta} \frac{E_j}{P_j^{-\theta}}, \end{aligned}$$

where we substituted in for the gravity equation and solved for Π_i .

The second equilibrium condition can be rewritten as:

$$\begin{aligned} E_i &= \sum_j X_{ji} \\ E_i &= \frac{E_i}{P_i^{-\theta}} \sum_j \tau_{ji}^{-\theta} \frac{Y_j}{\Pi_j^{-\theta}} \\ P_i^{-\theta} &= \sum_j \tau_{ji}^{-\theta} \frac{Y_j}{\Pi_j^{-\theta}}. \end{aligned}$$

Expressed in changes, the two equilibrium conditions become

$$\begin{aligned}\hat{\Pi}_i^{-\theta} &= \sum_j \hat{\tau}_{ij}^{-\theta} \frac{\hat{E}_j}{\hat{P}_j^{-\theta}} \frac{X_{ij}}{Y_i} \\ \hat{P}_i^{-\theta} &= \sum_j \hat{\tau}_{ji}^{-\theta} \frac{\hat{Y}_j}{\hat{\Pi}_j^{-\theta}} \frac{X_{ji}}{E_i}.\end{aligned}$$