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

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

March 2021

Abstract

This paper examines the structure of the shipping network and its implications on global trade and welfare. Using novel data on the movements of container ships, we calculate optimal travel routes. We then estimate the impact of a shock to the network on global trade by analyzing the effect of the 2016 Panama Canal Expansion. Trade between country pairs using the canal increased by 10% after the expansion. While the building costs were borne by Panama alone, a model-based quantification analysis shows that the welfare gains were shared by many countries, due to the network structure of shipping.

*We would like to 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).

†University of Oslo & Statistics Norway & 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

¶University of Oslo & CEPR, yuan.zi@econ.uio.no

1 Introduction

Container ships are the engines of global trade. Levinson (2006) and Bernhofen et al. (2016) detail the seismic changes that the worldwide adoption of container shipping technology has brought about in international trade. As documented by Rua (2014), by now, nearly all countries have container ports, constituting the nodes of the global container shipping network. However, there is scarce empirical evidence on the structure of the shipping network, e.g. which route a container might travel from the dock of port i to port j. At the same time, the structure of the shipping network is an essential determinant of the costs of trade, and there is increasing evidence suggesting that connectivity is at least as important as geographical distance in determining freight costs.[1] The networked environment also implies that a shock to a port, or a link, in the network, such as improvements in shipping infrastructure, may affect shipping costs and trade flows for many more countries than those that are directly affected. In this paper, we use satellite data on the movement of container ships to establish novel evidence on the routes that form the global shipping network. This in turn allows us to analyze the global reach of local shocks to the shipping network with respect to trade costs, trade flows and real incomes.

Our contribution is threefold. First, we document salient features of the world container shipping network based on unique novel data covering the worldwide movements of all container ships in 2016. Second, we calculate optimal shipping routes, inferred from the satellite data, and estimate the impact of a local shock on global trade. We do so by using the Panama Canal expansion in 2016 as natural experiment: From the optimal routes, we infer which port-pairs use the canal. This enables us to estimate the reduced-form impact of the expansion on global trade. Third, we quantify the trade and welfare effects of the shock using a quantitative model of trade. In contrast to standard trade models, our model features a shipping network, as observed in the data, and routes form endogenously based on the cost of shipping between ports. Comparing the welfare effects arising based on our network model with those that would arise from a model without a shipping network, our analysis shows that the existence of the shipping network translates into heterogeneous and widespread changes in trade costs, trade flows and real income.

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. AIS reporting of vessel positions offers a degree of automation in data processing and aggregation that was not previously possible.

Using an exhaustive data set based on AIS 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 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. Less than 6% of all 22,562 pairs of countries with container ports are directly connected.

[1] See Limao and Venables (2001) on the weak relationship between geographical distance and shipping costs, Wilmsmeier and Hoffmann (2008) on the importance of connectivity, and UNCTAD (2015) for a review of the role of distance and connectivity.

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 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 Of course, 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). As such, the calculated fastest route is an approximation to the actual unobserved route. The fastest path calculations reveal that 50% of all country-to-country connections involve stops in more than two other countries in between.

Besides adding to the distance traveled by a container, indirect routes expose bilateral flows to the shipping infrastructure of other countries. To demonstrate the importance of exposure to third-country infrastructure, we analyze the global trade effects of a large improvement in local shipping infrastructure in 2016: the expansion of the Panama Canal. 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.[2] Our information on shipping routes allows us to explore how exporters and importers worldwide were differentially affected by this local change in the shipping infrastructure. Using a difference-in-difference approach, we find that country pairs whose fastest connection passed through the Panama Canal prior to the expansion traded 10% more after the expansion compared to other country pairs.

Finally, we develop a spatial model of trade to quantify the general equilibrium effects of the Panama Canal expansion. In contrast to standard trade models, goods are passing through a shipping network when traveling from a source to a destination location. The model allows for economies of scale in shipping, so that larger ships on a given route may potentially lead to lower average transport costs. We build on the work of Allen and Arkolakis (2020), but while their application is on urban economics, our focus is on international trade. Using the model to quantify the effects of the Panama Canal expansion on global trade, we assess the welfare impact of the canal expansion. The increase in world real income were orders of magnitude greater than the construction costs, and while the building costs were borne by Panama alone, the gains per capita were shared by many countries. Furthermore, trade costs declined substantially, both for location-pairs directly connected by the canal (e.g., Panama Pacific to Atlantic side), and also for locations indirectly connected (e.g. Long Beach to Hamburg via intermediate ports).

We contrast these counterfactual results to a similar model without a shipping network, i.e. where goods travel directly from source i to destination j. The decline in trade costs is greater, and more smoothly distributed, in the network model compared to the no-network

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

model, as goods in the network model may pass through segments (e.g., the canal) where transport costs declined. Therefore, the network model also generates greater increases in trade and real income compared to the no-network model. In sum, we conclude that the network structure of shipping is of first order importance to assess the impact of changes in transport costs on trade and welfare.

Our paper is closely related 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. (2017) 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. (2020) study the role of shipping hubs for global trade and welfare. While our paper focuses on the role of the global shipping network and use the Panama Canal expansion as a natural experiment, their paper instead estimates a quantitative model to assess the role of hubs in transportation.

Our paper also aims to contribute 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., 2018 and Ducruet et al., 2019). Finally, this paper is related 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 (2009) studies the relationship between trade and the closing and opening of the Suez Canal.

The rest of the paper is structured as follow. Section 2 documents the satellite data and the construction of the shipping network and presents salient features of the network. Section 3 analyzes the global impact of the Panama Canal expansion on trade. Section 4 presents a spatial model of trade capturing the network features of the global shipping network, while Section 5 uses the model to quantify the general equilibrium effects of the canal expansion and compare them to those that would arise without a network environment. Section 6 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 status[3], and current draught (i.e., by how much a ship is 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 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 dead weight tonnage (i.e. the maximum tonnes of goods that a ship can carry). Combining the two with a ship's current draught, we can back out how much cargo a ship was carrying using formulas from the marine traffic literature (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 3 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 4 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 three 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

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

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. Only ships with deadweight tons>15,800 and trips with non-zero duration are used. Summary statistics include only routes taken by at least 5 ships and only routes between ports that appear both as arrival and departure ports.

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 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.[4] Trade between these countries accounts for only 54 percent of world trade. Therefore, a large share of global trade does not travel on direct routes, but on routes with multiple hops.

[4]The share of directly connected pairs is impacted by the restrictions we have imposed on the sample. If we include connections with less than five ships and sailings by very small ships, the share of directly connected country pairs is still only 11%.

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. Only ships with deadweight tons>15,800 and trips with non-zero duration are used. Summary statistics include only routes taken by at least 5 ships and only routes between ports that appear both as arrival and departure ports.

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. To do so, we need to identify 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.

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. This section documents our methodology to calculate routes and shows descriptive statistics on those routes.

Based on the schedule of departure and arrival times of all container ships in our dataset (inferred from the time stamps indicating arrival at and departure from a given port), we compute the optimal path in terms of travel time from any port i to any port j at a any start time h during the year 2016 using a simple algorithm described in Appendix C.[5] Among the set of optimal paths connecting two ports at different points in time, we select the route (the sequence of intermediate ports) that is used most frequently.[6]

Figure 1 visualizes the fastest routes for U.S. exports to all other countries based on our calculation.[7] 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.

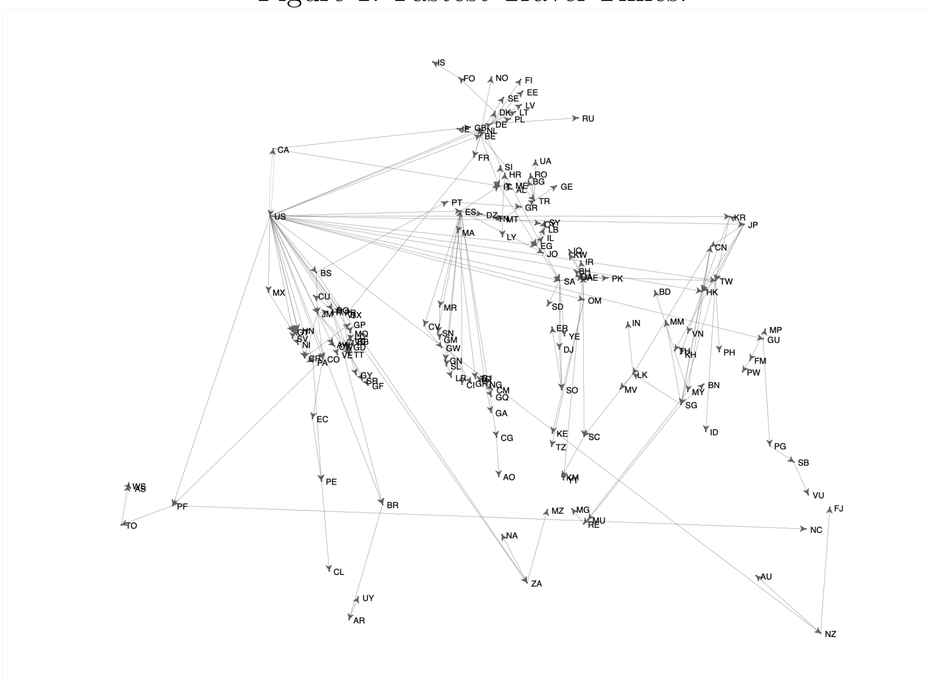
To understand further the role of shipping hubs and indirect routes in the shipping network, we examine the number of hops on the fastest shipping routes between all ports

[5]The algorithm finds all optimal paths for a predetermined maximum number of intermediate stops. For computational reasons, we set the maximum number of intermediate ports equal to 15.

[6]In 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.

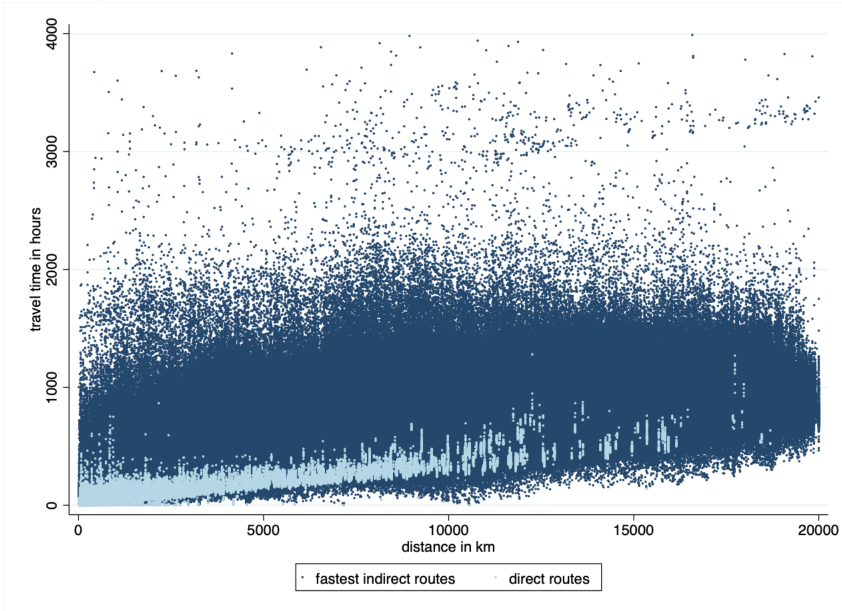
[7]The figure displays only one route per country pair, namely the fastest one among the routes connecting all U.S. ports to the port(s) in the partner country.

Figure 1: Fastest Travel Times.



Note: The figure plots the fastest routes from the U.S. to other countries. All computations are based on observed arrival and departure times of container ships with deadweight tonnes>15,800 in 2016. Routes with less than 5 ships are dropped. The plotted route is the fastest one among the routes connecting any U.S. port and any port in the destination country.

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.

in the network. Figure 3 shows the frequency of hops after aggregating ports by country.[8] Most country pairs are connected by routes involving at leastone to four hops.[9]

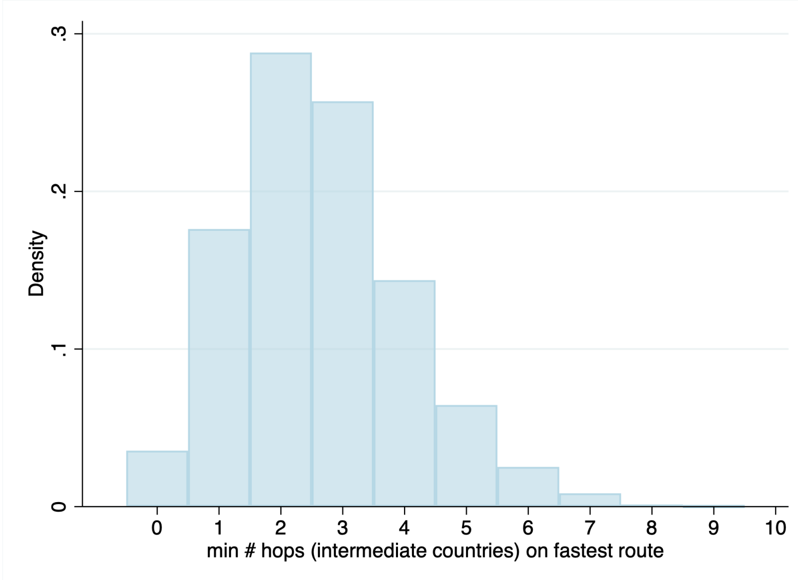
Our calculation of routes relies on the assumption that the fastest route will be the cost minimizing route. Appendix Section D.1 provides empirical evidence on the correlation between freight costs and travel time that supports this assumption. To verify that our calculated fastest routes captures the 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.

Having established the global shipping network and thereby the routes connecting all trading partners, we move on to analyze the role of the network for the propagation of local shocks.

[8]For countries with multiple ports, we use the minimum number of hops across multiple connections to the partner country.

[9]Figure 3 shows that the share of country pairs with zero-hop routes is lower than the 6% percent of directly connected country pairs reported above. This is due to the fact that in some cases the route that was optimal in terms of travel time *most often* in 2016 was an indirect one.

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 Impact of the Panama Canal Expansion on Global Trade

Guided by the stylized facts on the shipping network presented in Section 2, we now investigate how a shock to a particular link in the network not only affects trade between ports/countries on either side of the link, but also trade between any port/country that is using that link indirectly. We use the Panama Canal expansion in 2016 as a natural experiment.

3.1 The Panama Canal Expansion

The Panama Canal opened in 1914 and is one of the important links of worldwide maritime trade. The motivation for the 2016 expansion was twofold. First, because of the rapid increase in global trade, the Panama Canal had started to reach its capacity constraint. Second, container ships were getting larger, and many of them were too big to use the canal. By the turn of the millenium, the Panama Canal was already a bottleneck for American and Asian-American-East coast trade. According to Wilson and Ho (2018), only 41 percent of container ships and 52 percent of dry bulk ships would have been able to pass through the original canal.[10] Therefore, in 2006, the Panama Authority decided to expand the canal by

[10] Wilson and Ho (2018) provide a comprehensive case study of the Panama Canal. They calculated the cited figures based on numbers from *Fairplay* in 2015.

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.[11]

The Panama Canal Authority initially announced that the Canal expansion would be completed by August 2014 to coincide with the 100th anniversary of the opening of the Panama Canal. 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 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 Neopanamax ships, which carry more than twice as much cargo as the older Panamax ships.[12] 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). From June to December 2016, the share of Neopanamax ships passing through the canal increased from 0 to 15 percent. In 2017, the canal container tonnage increased by 22%.[13]

Overall, the Panama Canal expansion serves as an ideal natural experiment for our study for several reasons. First, the canal is one of the most important hubs in the global shipping network; as such the expansion is likely to have a large aggregate impact. Second, the old canal continued to operate both during and after the construction period, facilitating clean identification of the impact of the expansion on global trade and container traffic. Third, uncertainty about the exact opening date of the expanded canal suggests that anticipation effects around the time of opening was limited.

A potential concern is that, although the opening date was subject to uncertainty, knowledge about a future expanded canal might have encouraged the building of larger ships and investment in port capacity. These general equilibrium effects are not identified in this paper. However, based on a detailed analysis of the development of shipbuilding we find no evidence of an escalation in the building of Neopanamax ships from the announcement of expansion of the canal and onwards. The building of such ships was relatively stable between the announcement and completion year. Appendix Section E provides details on the analysis and a more extensive discussion of the container shipping market.

3.2 Empirical Strategy

Combining the AIS based network data with COMTRADE data on bilateral world trade, we investigate how the Panama Canal expansion affected global trade. We do so by employing a simple differences-in-differences analysis:

---

[11] https://web.archive.org/web/20110721055325/http://www.acp.gob.pa/eng/plan/documentos/propuesta/acp-expansion-proposal.pdf

[12] For a detailed description of the expansion project, see e.g. https://www.nationalgeographic.com/news/2014/8/140815-panama-canal-culebra-cut-lake-gatun-focus/

[13] http://www.pancanal.com/common/maritime/advisories/2017/a-02-2017.pdf and https://www.moodys.com/research/Moodys-Upgrades-Panama-Canal-Authority-to-A1-Outlook-stable--PR_396338

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

where y_{ikt} is log of imports from country i to country j in quarter t.[14] The variable $Post_t$ is a dummy that takes on the value one if the date is after June 2016, and zero otherwise. Exposure to the Panama Canal expansion is captured by the variable $PanExposure_{ij}$, which takes on a value between zero and one. Z_{ikt} refers to 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.[15] Hence, we allow for trends in trade that may differ according to observed geographical characteristics and for trends among pairs relying differentially on Neopanamax ships prior to the expansion. We also include a large set of fixed effects: source country-time δ_{it} and destination country-time δ_{kt} fixed effects will control for trends in overall exporting and importing for each country, while source-destination country fixed effects δ_{ik} control for time-invariant country pair characteristics.

Panama Canal exposure. The exposure measure is constructed as follows: We define exposure at the route level equal to one if the route passes the Panama Canal and zero otherwise.[16] For country pairs with multiple ports we average over the exposure of all port-to-port connections using the source and destination port size as weights, where port size is measured in terms of total incoming (outgoing) tonnes in 2016. We also explore alternative ways of inferring Panama Canal exposure in Appendix Section I.

Table 3 presents summary statistics for the Panama Canal exposure measure in 2016. There are 3,623 country pairs (14% of 25,025 pairs with positive trade flows) which are connected by a fastest route passing the Panama Canal. The value shipped between these countries accounts for 12% of global trade. The table shows that 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 4 shows the share of imports passing through the Panama Canal by country, and illustrates the importance of the Panama Canal as a shipping route for the Americas.

The exposure variable is calculated based on the observed container traffic data *prior* to the opening of the expanded canal. The exposure variable is relatively stable over time. We calculate the same variable for the post period (i.e., second half of 2016), and find that the correlation between the pre- and post-period exposure measure is 0.95. Furthermore, 95.3% of port pairs experience zero change in exposure between the pre and post period, 2.5% (2.2%) increase (decrease) their exposure. Hence, we find no large or systematic changes towards a higher exposure in the post period.

The methodology relies on the fastest time algorithm correctly predicting port-to-port connections that actually use the canal. This is more likely to be true if alternative routes

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

[15] A Neopanamax ship is the term that characterizes the maximum ship size for crossing the expanded canal.

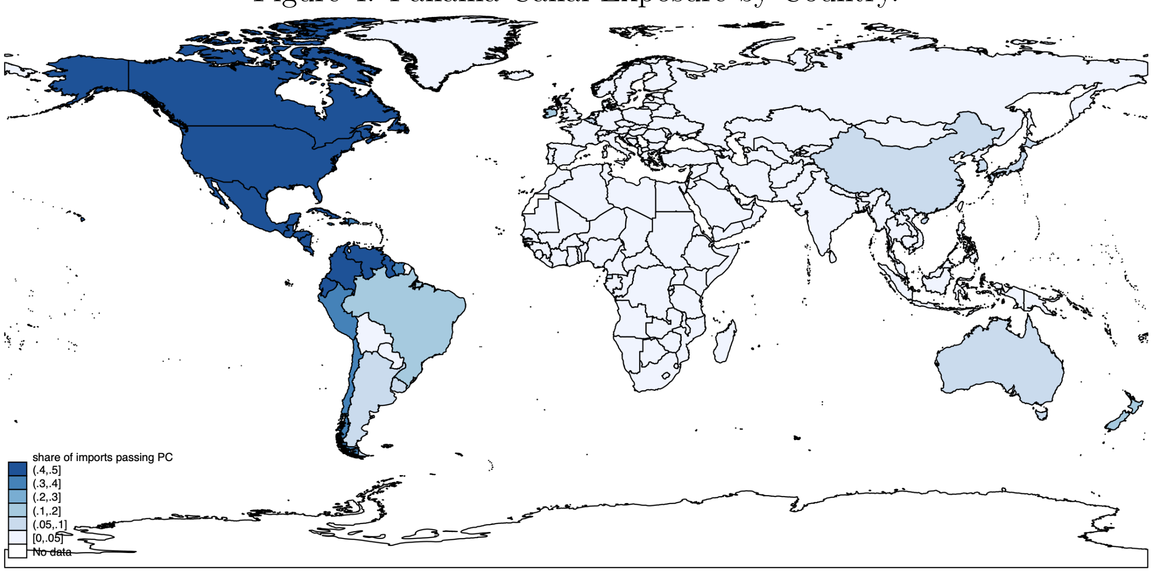
[16] In cases where the most frequent route is not unique, we average over the binary exposure measure of these routes.

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

Country pairs with exposure		Global trade exposed		Importers with exposure	
(1)	(2)	(3)	(4)	(5)	(6)
# pairs	% of total	value in trn $	% of total	# importers	% 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 column (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 4: Panama Canal Exposure by Country.



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

would take much longer time. To address this, we perform the following thought experiment. 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. The result of the experiment is illustrated in Figure 12 in the Appendix. On average, travel time increases by 14 days (67%) if the canal cannot be used. For 97.6% of the affected paths, travel time increases by 3 days or more. This suggests that other factors that may affect the choice of alternative routes are a secondary concern and, furthermore, that it is unlikely that other cost factors would change a "treated" country pair to an "untreated" one, or the other way around.

.

3.3 Empirical Results

We estimate the empirical specification in equation (1) using quarterly COMTRADE trade data for the period 2013Q1 to 2019Q4 as the dependent variable. Table 12 in the Appendix Section G summarizes the estimation sample.[17] Estimation results are reported in Table 4. Columns (1)-(2) report results for the baseline specification without and with controls, respectively. In both cases, we find that bilateral trade between country pairs whose fastest route passes the Panama Canal, increased by around 10 percent after the expansion.

Heterogeneity. We also explore whether the treatment effect is heterogeneous across country pairs. One hypothesis is that country pairs with fewer hops along the route will have a greater treatment effect than country pairs with many hops. For example, if the expansion reduces shipping costs due to the adoption of larger ships, then the cost savings in percent will be higher on routes with fewer hops. In Column (3) in Table 4 the main regressor is interacted with an indicator variable for whether the number of hops between i and j is below or above the median number of hops. The estimation results support the hypothesis; the treatment effect is 30% higher for country pairs with below median number of hops as compared to to those with above median hops.

Pre-trends. Figure 5 shows the estimated coefficient β by quarter. We find that the quarterly treatment effects in the post period are positive though not individually significant, but the sum of them, which corresponds to the regression is in Column (2) in Table 4, is strongly significant. Importantly, the figure illustrates the absence of pre-trends, indicating that the identifying assumption holds.

3.4 Robustness

To check the robustness of our results, we re-estimate the specification from column (2) of Table 4 with a modified set of controls, for a shorter time span, for a balanced panel, for monthly data and with alternative Panama Exposure measures. The results are reported in Table 5. Overall, we find that the results are relatively insensitive to various perturbations of the data, underscoring the robustness of the baseline results.

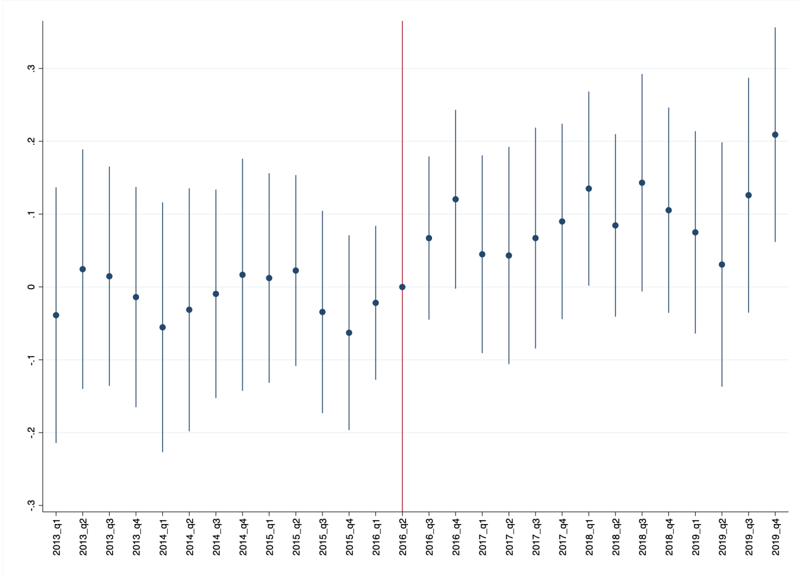
[17] Our estimation sample covers about 82% of global imports reported to COMTRADE. The missing 18% are due to countries not reporting trade data to Comtrade on a monthly basis (which are aggregated to the quarterly level).

Table 4: The impact of the Panama Canal expansion on trade

	(1)	(2)	(3)
$Post_t \times PanExposure_{ij}$	.105***	.108***	
	(.038)	(.040)	
$\times[\#hops \leq med]$			.126**
			(.052)
$\times[\#hops > med]$			.098**
			(.044)
Controls	No	Yes	Yes
FEs	ij,it,jt	ij,it,jt	ij,it,jt
Observations	199,177	199,177	199,177
Exporters/Importers	140/105	140/105	140/105
adj. R^2	.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) 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 $Post_t$. The triple interaction term in column 4 is an indicator variable for whether the number of hops between i and j is below of above the median number for the treated group. Standard errors are clustered by i, j. Significance levels: $*p < 0.1, **p < 0.05, ***p < 0.01$.

Figure 5: Panama Canal Exposure Coefficient by Quarter.



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

Column (1) shows that the estimated magnitude is not driven by the disproportionately large effect in Q4 of 2019 visible in Figure 5. Column (2) document that limiting the sample to the 38 importing countries that reported monthly trade flows in every month in our sample period does not change our results. Therefore our findings are not driven by countries entering or exiting the sample or by an increase inreporting activity by particular countries. Next, we estimate a specification based on monthly data. This produces a slightly smaller coefficient estimate (.082; see Column (3)). This is consistent with measurement error in monthly flows due to lagged reporting, which is smoothed out by aggregating to quarters.

Finally, we construct two alternative Panama exposure measures using all optimal paths found by our algorithm, instead of relying on the most frequent routes. In Column (5) the Panama exposure measure at the port-to-port level is the simple average across the binary exposure measure of all paths that were optimal at least once in the first half of 2016. In Column (4), the paths are weighted by the amount of time during which they were optimal during the first half of 2016. Aggregation from the port-to-port level to the country-pair level is done in a similar way as above.

Table 5: Robustness

Robustness check:	2019Q4 dropped	Balanced sample	Monthly data	PanExposure: all paths	
				Weighted avg.	Simple avg.
$Post_t \times PanExposure_{ij}$	.102**	.097*	.082**	.103**	.098**
	(.040)	(.054)	(.040)	(.048)	(.048)
Observations	193,450	86,576	600,884	199,177	199,177
Exporters/Importers	140/105	133/36	140/107	140/105	140/105
adj. R^2	.937	.947	.900	.937	.937

Note: Dependent variable is the log of imports from country i to country j in quarter t over the period 2013Q1 − 2019Q4 in columns (2,4,5). In column (1) the last quarter of 2019 is dropped. Column (3) is based on monthly data for the full sample period. Column (2) is restricted to set of pairs for which trade flows exist in every quarter. All columns include ij, it and jt fixed effects as well as controls. The control variables are: an FTA indicator and geographical variables (distance, contiguity and common language) 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 $Post_t$. Columns (4) and (5) are based on a $PanExposure$ measure computed as a weighted (column (4)) and simple (column (5)) 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 a 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. Standard errors are clustered by i and j in columns (1,2,4,5). In column (3) where the number of importers is very low, standard errors are clustered by ij. Significance levels: $*p < 0.1, **p < 0.05, ***p < 0.01$.

4 A Quantitative Trade Model with Optimal Routing

So far, the empirical analyses have provided evidence of the impact of the Panama Canal expansion on global trade. This section introduces a parsimonious quantitative model of world container traffic and trade to quantify the general equilibrium and welfare effects of the Panama Canal expansion. In contrast to standard trade models, goods are passing through a shipping network when departing from an origin and arriving in adestination port. Agents choose the optimal route endogenously in order to minimize transport costs. The model also allows for economics of scale in shipping, so that larger ships on a given route may potentially lead to lower average transport costs. We build on the work of Allen and Arkolakis (2020) (henceforth, AA2020): 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, but where individuals are immobile across countries.

While the model allows for endogenous routes and reallocation of shipping across ports, we have chosen to abstract from endogenous investment in ships and ports. There are two main reasons for this. First, as documented in Section 3.1 and Appendix E, the first-order effect of the canal expansion was that more and larger ships could transit. Second, we do not have a clean identification strategy for estimating these additional margins of adjustment.[18]

After presenting the economic framework, we quantify the effect of the expansion in

[18]

Section 5. The quantification will allow us to assess the welfare impact of the canal expansion and to assess the importance of ship size (economics of scale in transportation) in generating those welfare gains. Finally, we contrast our results with a model with no shipping network. This allows us to assess the role of the shipping network for trade and welfare.

4.1 Model Setup

Consider a world with N locations indexed by i and j, each is 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 $\Pi_{k=1}^{K} t_{r_{k-1},r_k}$. 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.[19] Furthermore, it buys us tractability in terms of producing analytical expressions for many key objects of interest.

4.2 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:

$$Y_i = \sum_j X_{ij} \qquad E_i = \sum_j X_{ji}, \tag{2}$$

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

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

where

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

[19]See e.g. https://www.cma-cgm.com/ebusiness/schedules/routing-finder.

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

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

and P_j is the consumer price index:

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

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

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 $\mathbf{A} \equiv \left[t_{ij}^{-\theta}\right]$, AA2020 shows that transport costs τ_{ij} can be rewritten as

$$\tau_{ij} = b_{ij}^{-1/\theta}, \tag{7}$$

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

$$\Pi_i^{-\theta} = \frac{E_i}{P_i^{-\theta}} + \sum_j t_{ij}^{-\theta} \Pi_i^{-\theta} \tag{8}$$

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

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, Eaton and Kortum (2008), to solve for a counterfactual equilibrium.

Trade and Traffic

We end this subsection by characterizing traffic flows according to the model. We define traffic as the the total value of all cargo passing though a segment (k, l) in the network. It can be shown that, in equilibrium, the value of traffic between k and l is

$$\Xi_{kl} = t_{kl}^{-\theta} P_k^{-\theta} \Pi_l^{-\theta}. \tag{10}$$

Furthermore, there is a simple mapping between trade and traffic.[20] One can express equilibrium trade flows as:

$$X_{ij} = c_{ij}^X Y_i E_j \tag{11}$$

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

5 The General Equilibrium Effects of the Panama Canal Expansion

We now turn to applying the model to quantify the impact of the Panama Canal expansion. The empirical strategy is as follows. First, we estimate the impact of the expansion on the following margins: container ship size, the frequency of ships and ship capacity utilization. These steps will rely on the MarineTraffic traffic data between all port pairs, before and after the expansion. Using the model, we can then back out the reduction in transport costs caused by the canal expansion. Third, we perform a counterfactual simulation of the model, where we use (i) the estimated change in transport costs and (ii) the MarineTraffic data, to investigate the welfare impact of the expansion. Fourth, we estimate our reduced-form model from Section 3.2 on the simulated data coming from the counterfactual. This enables us to assess the importance of one specific mechanism, ship size, in generating the growth in trade that we estimated in Section 3.3. It also serves as a validation of the quantitative model. Fifth, and finally, we compare the counterfactual results based on the network model with a counterfactual from a model without a shipping network, i.e. where goods are shipped directly from source i to destination j. This helps us isolate and understand the importance of the shipping network in generating the main quantitative results on trade and welfare.

5.1 The Impact on the Margins of Shipping

The first step of the analysis is to determine the impact of the expansion on three margins of shipping: (i) ship size, (ii) frequency (the number of ships) and (iii) ship utilization, i.e. the percentage of used ship capacity. We proceed by creating a dataset of all three variables from the MarineTraffic data, for every segment (port-pair kl) and for the 1st and 2nd half of 2016. The construction of these variables is described in detail in Appendix Section B. We estimate the following regression

$$\Delta \ln y_{kl} = \alpha_0 + \beta PanamaCanal_{kl} + D_k + D_l + \epsilon_{kl}, \tag{12}$$

[20] See AA2020 for detailed derivations.

Table 6: The Margins of Shipping: Results

Dependent variable	$\Delta \ln ShipSize_{kl}$			$\Delta \ln Frequency_{kl}$			$\Delta \ln Utilization_{kl}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$PanamaCanal_{kl}$	.22***	.14**	.12*	.01	.02	.02	.05	.02	.02
	(.05)	(.06)	(.06)	(.12)	(.14)	(.15)	(.04)	(.05)	(.05)
Controls	No	Yes	No	No	Yes	No	No	Yes	No
Source/destination FE	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

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 ($ShipSize_{kl} \times Frequency_{kl}$). $PanamaCanal_{kl}$ 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: $*p < 0.1, **p < 0.05, ***p < 0.01$.

where Δ refers to the change between the 1st and 2nd half of 2016 (recall, the expanded canal opened 26 June 2016), y_{kl} is one of the three outcome variables described above and $PanamaCanal_{kl}$ takes the value one if the segment kl is using the canal and zero otherwise. The variables D_k and D_l are origin and destination port fixed effects, respectively.

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 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.

By construction, the volume of traffic on a segment kl, Ξ^V_{kl}, is the product of margins (i)-(iii):

$$\Xi^V_{kl} = ShipSize_{kl} \times Frequency_{kl} \times Utilization_{kl}.$$

Since margins (ii)-(iii) are both economically and statistically insignificant, our results suggest that the impact of Panama Canal expansion on trade traffic is mainly through allowing bigger ships to operate on relevant links.

5.2 Transport Costs

In the next step, we use the results above to infer the change in transport costs due to the canal expansion. We start by assuming that transport costs t_{kl} are a log-linear function of ship size and other factors, such as travel time. Specifically, we assume that

$$t_{kl} = ShipSize_{kl}^{-\delta} f_{kl}(.), \tag{13}$$

where $f_{kl}(.)$ is a flexible function of all other factors (e.g., travel time). 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 (13) yields

$$\Delta \ln \Xi_{kl} = \theta\delta \Delta \ln ShipSize_{kl} - \theta \Delta \ln \Pi_l - \theta \Delta \ln P_k - \theta \Delta \ln f_{kl}(.). \tag{14}$$

where $\varepsilon_{kl} \equiv -\theta \Delta \ln \Pi_l - \theta \Delta \ln P_k - \theta \Delta \ln f(.)$.

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.

Using the finding above that the canal expansion led to larger ships, but not higher frequency and capacity utilization, $\Delta \ln \Xi_{kl}^V = \Delta \ln \Xi_{kl} = \Delta \ln ShipSize_{kl}$ for Panama Canal port pairs kl, we immediately infer that $\theta\delta = 1$. In other words, since ship size is the only margin of adjustment from the canal expansion, the change in the volume of traffic is the same as the change in ship size.[21] The estimate of $\theta\delta$ then allows us to infer the change in transport costs from equation (13), given knowledge about the trade elasticity θ. Specifically, a change in ship size, holding everything else constant, will lead to the following change in transport costs for the canal segments kl:

$$\hat{t}_{kl} = e^{-\delta\beta}, \tag{15}$$

where β refers to the point estimate from the regression of $\Delta \ln ShipSize_{kl}$ on $PanamaCanal_{kl}$ from equation (12) above. For non-canal segments $k'l'$, we have $\hat{t}_{k'l'} = 1$. In our baseline specification we use the value $\theta = 8$, which leads to a change in transport costs $\hat{t}_{kl} \approx 0.98$ for Panama Canal segments kl.

5.3 The Welfare Effect of the Panama Canal Expansion

We have provided evidence above that the expanded canal led to bigger ships and lower transport costs. In this section, we ask, given our estimated change in transport costs, what is the welfare effect of the Panama Canal expansion?

The general equilibrium of the model can be written in changes, using the "exact hat algebra" approach from Dekle, Eaton and Kortum (2008). Appendix H.1 shows that the system of equations can be simplified as:

[21] An alternative approach is to estimate equation (14) and using the Panama canal indicator as an instrument for the change in ship size. This also produces a coefficient estimate close to one.

Table 7: Data and Parameters.

Variable	Description	Value	Source
θ	Trade elasticity	8	Previous literature
$\hat{t}_{kl}$	Canal transport costs, change	.98	Estimated, Sections 5.1 and 5.2
α	The value per dwt of traffic	USD 1,734	Calibrated, Appendix H.3
Ξ_{kl}^V	Initial traffic flows (volume)		MarineTraffic, 1st half 2016
Y_i	Initial Expenditure		Eora Global Supply Chain Database, 2015; World Development Indicators; INSEE

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

$$\hat{P}_i^{-\theta} = \frac{Y_i}{Y_i + \sum_j \Xi_{ji}} \frac{\hat{\Pi}_i^{-\theta/(\theta+1)}}{\hat{\Pi}_i^{-\theta}} + \sum_j \left(\frac{\Xi_{ji}}{Y_i + \sum_j \Xi_{ji}} \right) \hat{t}_{ji}^{-\theta} \hat{P}_j^{-\theta}. \tag{17}$$

The data and parameters required to solve this system are modest: We need data on initial (i.e. 1st half of 2016) traffic Ξ_{kl}, initial expenditure Y_i, the estimated change in transport costs $\hat{t}_{kl}$ and the elasticity θ, see Table 7. We set $\theta = 8$, which is consistent with previous estimates in the literature (e.g., Eaton and Kortum, 2002). As described above, we assume that $\Xi_{kl} = \alpha \Xi_{kl}^V$. The procedure to calibrate the value of α is described in Appendix H.3.[22]

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. Recall that we assume labor immobility across locations. This is appropriate for countries that only have one port, but is less ideal for multi-port countries. In our sample, 46 percent of countries have only one port, suggesting that the labor immobility assumption is a reasonable approximation (and better than assuming the opposite, perfect labor mobility).

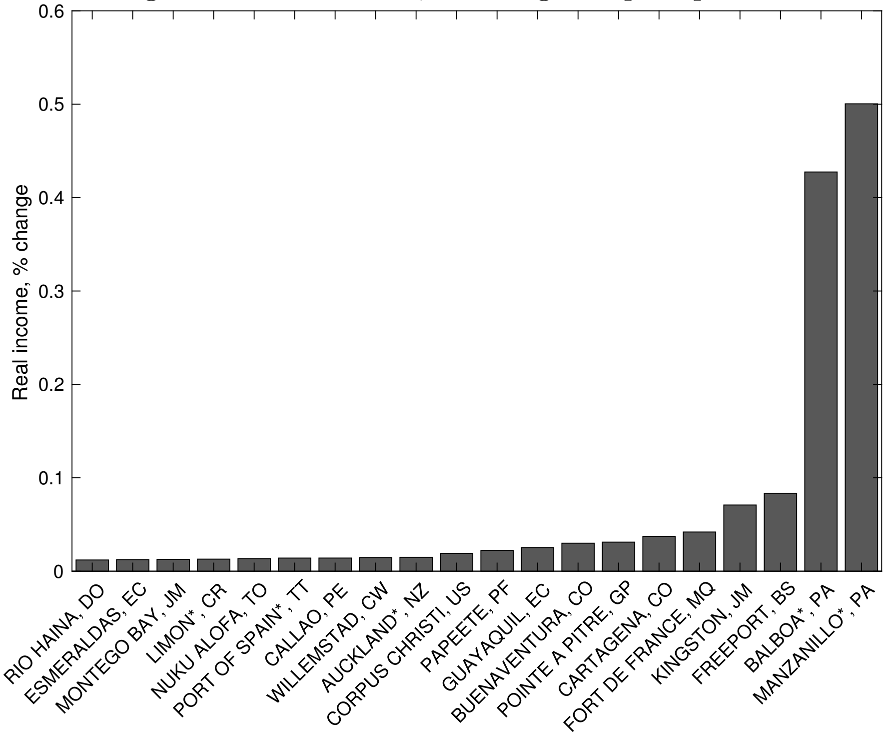
Figure 6 shows the change in real income, w_i/P_i, for the top 20 ports in our dataset. Not surprisingly, the ports closest to the canal are gaining the most. However, we also observe ports further away, such as in Colombia, Ecuador, Peru and the Caribbean, that obtain large welfare gains from the canal expansion. The weighted average of the real wage change across all countries is 0.001%, or 128 billion USD, measured in 2015 prices. The gains from the canal expansion is much higher than its costs, which was estimated 5.25 billion USD in 2006.

5.4 Traffic, Trade and Trade Costs

Next, we explore the impact of the canal expansion on traffic, trade and trade costs. Using the equilibrium objects $\hat{P}_k$ and $\hat{\Pi}_l$ found above, the change in traffic is simply $\hat{\Xi}_{kl} = \hat{t}_{kl}^{-\theta} \hat{P}_k^{-\theta} \hat{\Pi}_l^{-\theta}$. Furthermore, there is a mapping between traffic and trade, see equation (11). Given the

[22] The 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.

Figure 6: Real income, % change. Top 20 ports.



initial and counterfactual matrix of traffic, Ξ_{kl} and Ξ'_{kl}, we can then use equation (11) to infer initial and counterfactual trade flows X_{ij} and X'_{ij}. One can then back out the change in bilateral trade costs, $\hat{\tau}_{ij}$ from the gravity equation (3):

$$\hat{\tau}_{ij}^{-\theta} = \frac{\hat{\Pi}_i^{-\theta}}{\hat{Y}_i} \frac{\hat{P}_j^{-\theta}}{\hat{E}_j} \hat{X}_{ij}.$$

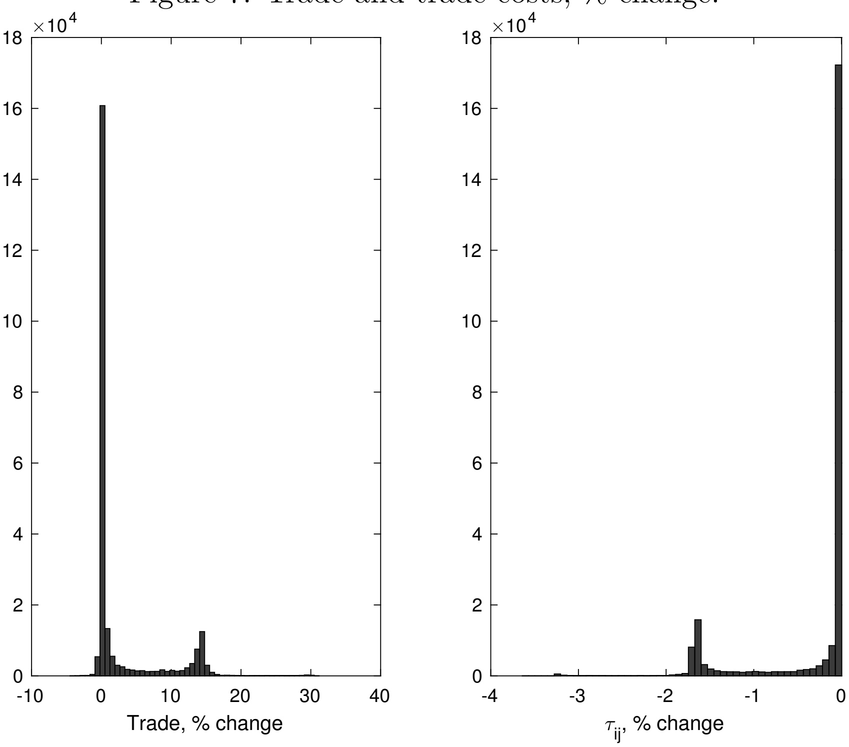
Figure 7 shows the histograms of the % change in bilateral trade costs τ_{ij} and bilateral trade X_{ij}. Trade costs decline by up to 4 percent and trade increases by up to 31 percent in the aftermath of the Panama Canal expansion. In addition to the directly exposed port pairs, such as the Pacific and Atlantic Panama ports, indirectly connected port pairs, such as between Ecuador and the U.S. East Coast, get large declines in trade costs. The average change in trade costs and trade across source-destination ports are -0.32 and 2.68 percent, respectively.

5.5 Mechanisms and Model Fit

According to the model, the canal expansion caused trade to grow by $\hat{X}_{ij}$. This effect is entirely driven by the finding that the expanded canal caused bigger ships to pass through the canal. In order to evaluate the performance of the model, we estimate a version of the reduced-form equation (1) from Section 3.2, but we replace real data with the simulated data from the model (i.e. the left hand side variable).

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 3.3.

Figure 7: Trade and trade costs, % change.



Interestingly, the treatment effect using simulated data is very close to the treatment effect using real data. This indicates that the main margin of adjustment based on the quantitative analysis, namely ship size, is responsible for the growth in trade that we estimated in Section 3.3. 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 the change in transport costs $\hat{t}_{kl}$ and the parameters δ and β. 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.

A slight disconnect between the reduced form analysis in Section 3 and the model is that the treatment (using the canal) in Section 3 was not inferred directly from the model. We therefore also estimate the reduced form regression, but replace the treatment variable $PanExposure_{ij}$ with a Panama Canal exposure variable according to the model. Appendix Section I provides details about the methodology and Table 13 presents the results. The results are fully in line with our baseline results and are, if anything stronger and more precise.

5.6 The Importance of the Shipping Network

Throughout the paper, we have emphasized the role of the shipping network for trade and welfare. We end this section by comparing the counterfactual results reported above for the network model with counterfactual results from a model without a shipping network, i.e. where goods are shipped directly from source i to destination j.

To fix ideas, consider a world with three locations, i, j and k. In a canonical, no-

Table 8: Reduced-form regression: Simulated data.

Dependent variable: $\ln \hat{X}_{ij}$ (simulated)	Port-pair	Country-pair
$PanExposure_{ij}$	.083***	.085***
	(.002)	(.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: Dep. var. is the relative change in exports from port i to port j implied by the model, $\ln \hat{X}_{ij}$. All columns in include i and j fixed effects. Standard errors clustered by $i$and j. Significance levels: $*p < 0.1, **p < 0.05, ***p < 0.01$.

network trade model, a change in transport costs between i and j, $\hat{t}_{ij}$ can only affect k *indirectly* through general equilibrium effects. In our model, however, k may also be affected by a *network effect* as goods shipped from i to k might pass though j. By comparing the counterfactual results from the network model with the no-network counterfactual results, we can isolate the network effects and quantify their relative importance on global trade.

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 4, we solve the model in changes, as before. Appendix J shows that the system of equations is

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

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

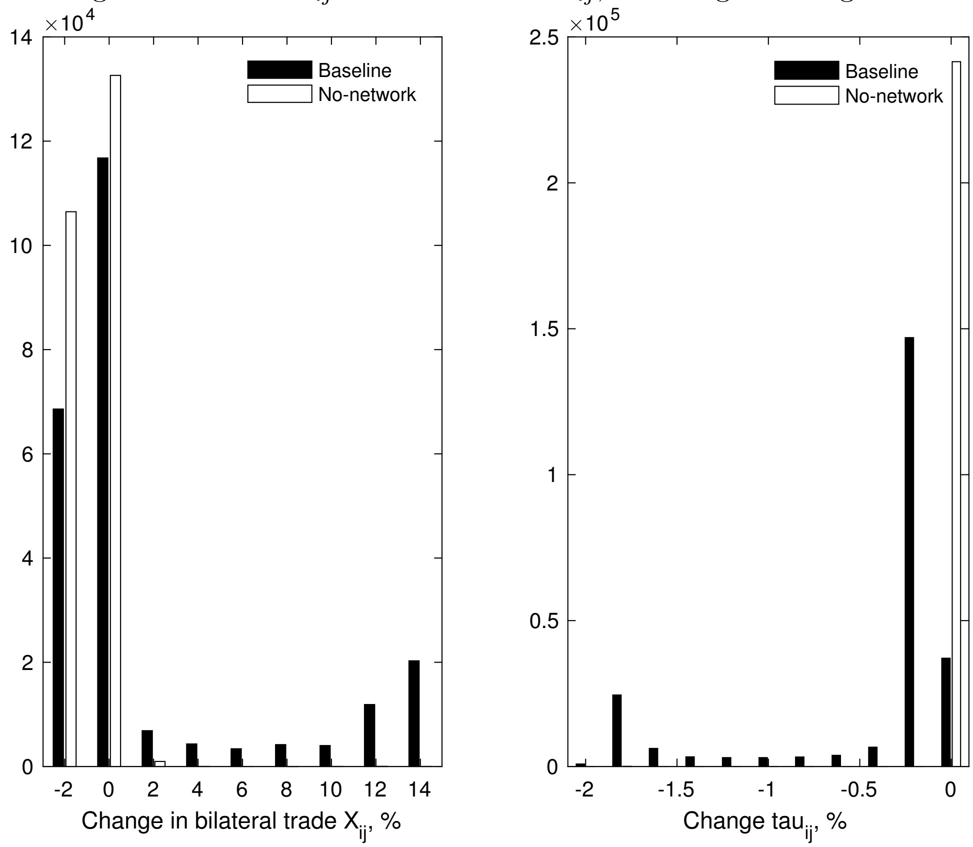
After imposing trade balance, $\hat{E}_j = \hat{Y}_j$, and using the fact that $\hat{Y}_i = \hat{\Pi}_i^{-\theta/(\theta+1)}$, we can solve this system of equations given data on initial trade flows X_{ij}, expenditure E_i, output Y_i, as well as the change in trade costs, $\hat{\tau}_{ij}$, and the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

As in the counterfactual based on the network model, we set $\hat{\tau}_{ij} = 0.98$ for port-pairs directly using the canal, and $\hat{\tau}_{ij} = 1$ otherwise.[23] 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.

Figure 8 presents the changes in trade X_{ij} and trade costs τ_{ij} due to the Panama canal expansion implied by the network and no-network model, respectively. Unsurprisingly, the change in τ_{ij} in the no-network model is 0 for the vast majority of port-pairs, and -2 percent

[23]When goods are shipped directly from i to j, directed affected segments = directly affected port-pairs.

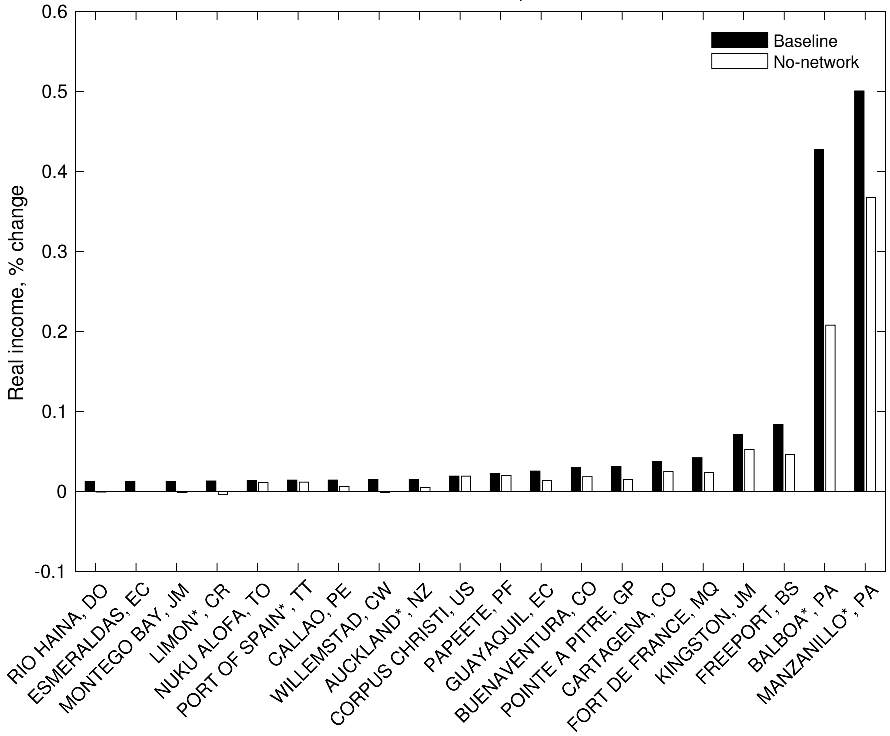
Figure 8: Trade X_{ij} and trade costs τ_{ij}, % change. Histogram.



for the very few port-pairs using the Panama Canal directly. In the network counterfactual, however, the change in τ_{ij} is much more smoothly distributed between -2 and 0%, because many location-pairs are using the canal indirectly. For example, the network model suggests that trade costs between Hamburg and Long Beach (a port-pair which according to our data does not have a direct connection) would decline by 1-2%, while the no-network trade model suggests a zero decline. The heterogeneous and widespread changes in trade costs, $\hat{\tau}_{ij}$, also translate into heterogeneous and widespread changes in trade flows. In the no-network model, trade creation in the aftermath of the canal expansion is limited to very few port-pairs. While in the network model, many port-pairs are indirectly connected via the canal, and therefore experience an increase in trade after the canal expansion. Based on the network model, trade between Hamburg and Long Beach would increase by 14 percent, whereas the no-network model predicts an almost zero percent increase in trade.

Figure 9 reproduces the plot for the change in real income for our network model, but now adds bars for the predicted increase in real income according to the no-network model. Not surprisingly, the network model predicts higher gains from the canal expansion compared to the model with only direct connections, mirroring that trade costs are declining more, and for more location-pairs, in the network model than in the no-network model. For the top 20 locations, the real income gains are 69 percent higher in the network model as compared to the no-network model. In sum, we conclude that the network structure of shipping is of first order importance to assess the impact of changes in transport costs on trade and welfare.

Figure 9: Real income w_i/P_i, % change.



6 Concluding remarks

In this paper we exploit novel satellite data on all port calls made by container ships world wide 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 natural experiment, we show that the 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 counterfactual analyses we find that the Panama Canal expansion produced sizable gains in terms of reduced trade costs, increased trade and higher real income, and that a standard trade model typically will underestimate the widespread gains arising from a local shock to the transport infrastructure.

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 (2020). The welfare effects of transportation infrastructure improvements. Technical report, NBER Working Paper No. 25487.

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.

Brancaccio, G., M. Kalouptsidi, and T. Papageorgiou (2017). Geography, search frictions and endogenous trade costs. Technical report, NBER Working Paper No. 23581, forthcoming in Econometrica.

Brooks, L., N. Gendron-Carrier, and G. Rua (2018). The local impact of containerization. Technical report, manuscript.

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 (2019). All aboard: The aggregate effects of port development. Technical report, manuscript.

Feyrer, J. (2009). Distance, trade, and income - the 1967 to 1975 closing of the suez canal as a natural experiment. Technical report, NBER Working Paper No. 15557.

Ganapati, S., W. F. Wong, and O. Ziv (2020). Entrepot: Hubs, scale, and trade costs. Technical report.

Levinson, M. (2006). *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger*. Princeton University Press.

Limao, N. and A. J. Venables (2001). Infrastructure, geographical disadvantage, transport costs and trade. *The World Bank Economic Review 15*, 451–479.

Maurer, S. and F. Rauch (2019). Economic geography aspects of the panama canal. Technical report.

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.

Rua, G. (2014). Diffusion of containerization. Technical report, Finance and Economics Discussion Series 88.

UN (2013). *International Merchandise Trade Statistics Compilers Manual - Revision 1 (IMTS 2010-CM).*

UNCTAD (2015). *Review of Maritime Transport.*

UNCTAD (2016). *Review of Maritime Transport.*

Wilmsmeier, G. and J. Hoffmann (2008). Liner shipping connectivity and port infrastructure as determinants of freight rates in the caribbean. *Maritime Economics and Logistics 10*, 130–151.

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. 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 volumn 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 × 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 $H_B = 0.55 H_S$, 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 build 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.[24] 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

$$EffectiveDWT = dwt * (H_A - H_B) / (H_S - H_B) . \tag{20}$$

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. (2017) 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.

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

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. Only ships with deadweight tons>15,800 and trips with non-zero duration are used. Summary statistics include only routes taken by at least 5 ships and only routes between ports that appear both as arrival and departure ports.

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. The algorithm works as follows. Every time a ship leaves i to anywhere, we compute all possible paths to j through the network of connections available at that point in time. To limit the computational burden, we consider only paths involving up to 15 intermediate ports. From the set of possible paths, we drop all those that are dominated by others, i.e. 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. The result of the algorithm is a set of paths between i and j that are optimal in terms of travel time at some starting time h in 2016.

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 ($lnHours_{ij}$), geodetic distance ($lnDist_{ij}$), total dwt of ships traveling to US port j from country i in 2016 ($lnDWT$), total number of ships traveling to US port i from country j in 2016 ($lnShips_{ij}$) and average ship size based on the latter two variables ($lnAvgDWT_{ij}$). 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)
$lnHours_{ij}$	.004***		.005***	.005***	.004***	.004***
	(.000)		(.001)	(.001)	(.001)	(.001)
$lnDist_{ij}$		.005*	-.001	-.001	.002	-.001
		(.001)	(.001)	(.001)	(.001)	(.001)
$lnDWT$				-.000		
				(.000)		
$lnShips_{ij}$					-.000	.000
					(.001)	(.001)
$lnAvgDWT_{ij}$					-.005***	-.005***
					(.000)	(.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: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$.

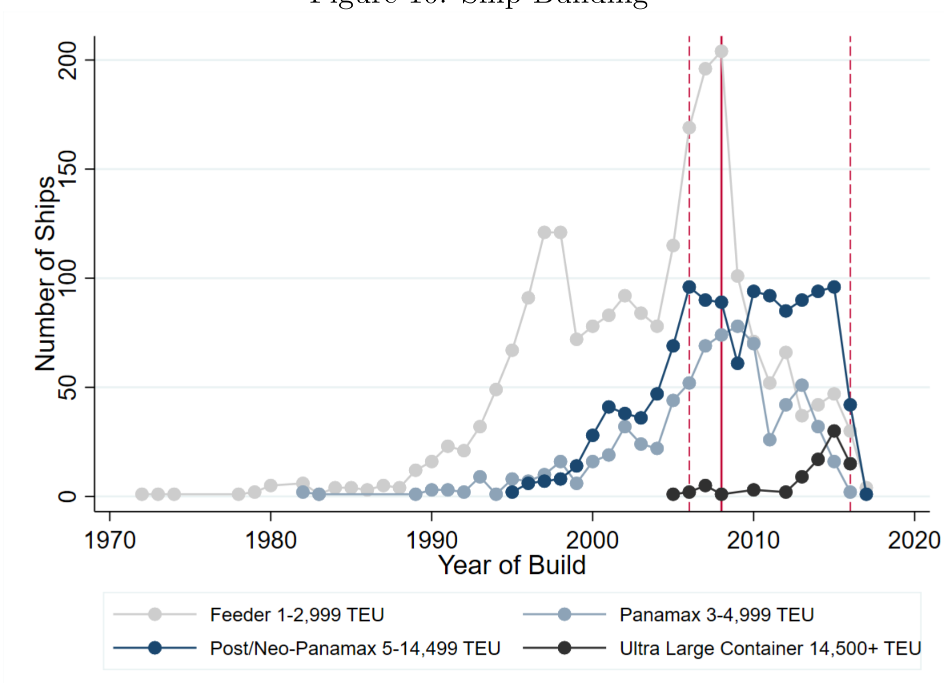
D.2 Evidence on Actual Routes of Chinese Trade

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

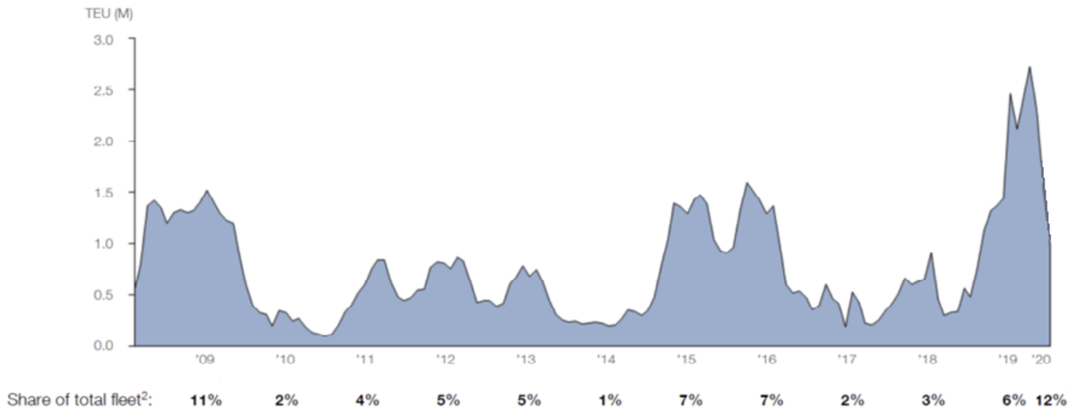
We have compiled data on the size distribution of all container ships over time. Specifically, using data from Clarksons, we can look the ship size distribution *by year of construction*. Figure 10 plots the number of ships by different size bins and by construction year. Interestingly, the number of 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 11), 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 11.[25]. 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 10: Ship Building



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

Figure 11: Idle Containership Capacity 2009-2020



F Panama Canal Exposure

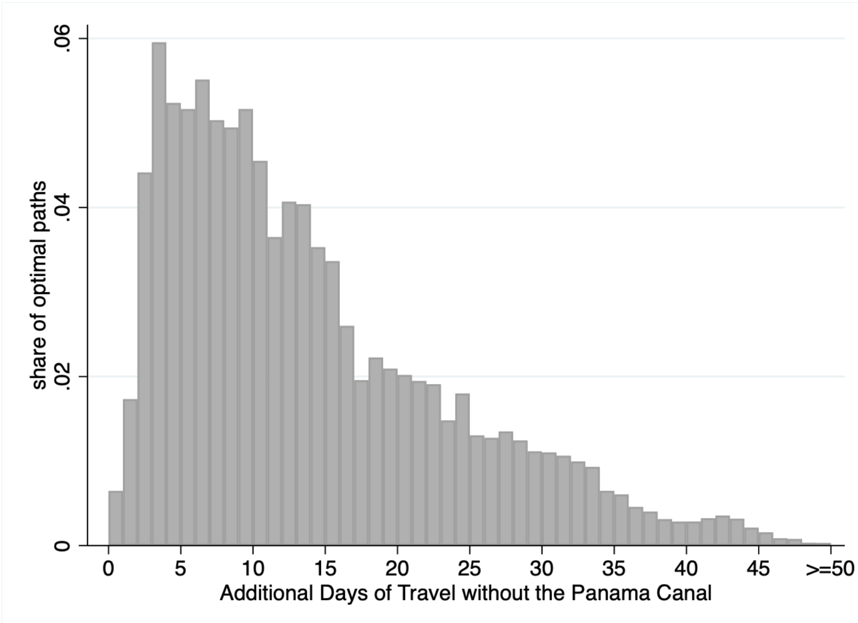
F.1 Summary Statistics

Table 11: Summary statistics on Panama Canal exposure

Rank	Importer	Share of total imports passing PC	Share in world imports	Exporter	Share of total exports passing PC	Share in world 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 12: 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.

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 between Jan 8-10, 2021. We aggregate monthly observations to quarters and keep only quarters were 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)).

G.2 Estimation Data: Summary Statistics

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. 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 Theoretical 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:

$$\left[\Pi_i^{-\theta}\right] = [1 - A]^{-1}\left[\frac{E_i}{P_i^{-\theta}}\right]$$

$$[1 - A]\left[\Pi_i^{-\theta}\right] = \left[\frac{E_i}{P_i^{-\theta}}\right]$$

$$\left[\Pi_i^{-\theta}\right] - A\left[\Pi_i^{-\theta}\right] = \left[\frac{E_i}{P_i^{-\theta}}\right]$$

$$\Pi_i^{-\theta} - \sum_j t_{ij}^{-\theta}\Pi_i^{-\theta} = \frac{E_i}{P_i^{-\theta}}$$

$$\Pi_i^{-\theta} = \frac{E_i}{P_i^{-\theta}} + \sum_j t_{ij}^{-\theta}\Pi_i^{-\theta}.$$

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

$$E_i = \sum_j X_{ji}$$

$$E_i = \frac{E_i}{P_i^{-\theta}}\sum_j \tau_{ji}^{-\theta}\frac{Y_j}{\Pi_j^{-\theta}}$$

$$P_i^{-\theta} = \sum_j \tau_{ji}^{-\theta}\frac{Y_j}{\Pi_j^{-\theta}}$$

$$\left[P_i^{-\theta}\right] = [1 - A']^{-1}\left[\frac{Y_i}{\Pi_i^{-\theta}}\right]$$

$$[1 - A']\left[P_i^{-\theta}\right] = \left[\frac{Y_i}{\Pi_i^{-\theta}}\right]$$

$$P_i^{-\theta} - \sum_j t_{ji}^{-\theta}P_j^{-\theta} = \frac{Y_i}{\Pi_i^{-\theta}}$$

$$P_i^{-\theta} = \frac{Y_i}{\Pi_i^{-\theta}} + \sum_j t_{ji}^{-\theta}P_j^{-\theta}.$$

Expressed in changes, the two equilibrium conditions become

$$\hat{\Pi}_i^{-\theta} = \frac{E_i}{E_i + \sum_j \Xi_{ij}}\frac{\hat{E}_i}{\hat{P}_i^{-\theta}} + \sum_j\left(\frac{\Xi_{ij}}{E_i + \sum_j \Xi_{ij}}\right)\hat{t}_{ij}^{-\theta}\hat{\Pi}_j^{-\theta}$$

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

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

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

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

H.2 Algorithm for solving the equilibrium in changes

The system of equations (21)-(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 $\hat{Y}^W$.

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 shows the sensitivity of the baseline results in Section 3.3 when using a different measure of Panama Canal exposure.

Table 13: Regressions with Panama exposure derived from the model

Exposure measure:	Travel time		Travel time & Ship size	
	continuous	$\pi_{ij}^{PA} > .3$	continuous	$\pi_{ij}^{PA} > .3$
	(1)	(2)	(3)	(4)
$Post_t \times PanExposure_{ij}$	.151**	.086**	.130**	.094**
	(.065)	(.042)	(.066)	(.045)
Observations	192,810	192,810	192,810	192,810
Exporters/Importers	138/102	138/102	138/102	138/102
adj. R^2	.936	.936	.936	.936

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) interacted with $Post_t$, and the share of deadweight tonnes traveling on Neopanamax ships on the route connecting i and j interacted with $Post_t$. Standard errors are clustered by i and j. Significance levels: $*p < 0.1, **p < 0.05, ***p < 0.01$.

We parameterize transport costs t_{kl} as

$$t_{kl} = \left(\frac{TravelTime_{kl}}{ShipSize_{kl}}\right)^{\delta}, \tag{23}$$

where $TravelTime_{kl}$ and $ShipSize_{kl}$ refer to average travel time and ship size across all trips on a link kl. Using the values of δ and θ from Section 5, 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 AA2020):

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

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

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

Table 13 repeats the regression analysis from Section 3.3 when replacing the baseline exposure measure with π_{ij}^{PA}.[26] Column (3) uses the π_{ij}^{PA} as the measure of exposure, while column (4) mimics the baseline regression by instead creating an indicator variable equal to one when $\pi_{ij}^{PA} > .3$. Columns (1) and (2) repeat the analysis, but instead of using t_{kl} from equation (23), we use the simpler version $t_{kl} = TravelTime^{\delta}$.

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).

[26] Similar to the baseline analysis, we aggregate π_{ij}^{PA} from port- to country-pair using port size as weights.

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

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