

NBER WORKING PAPER SERIES

GEOGRAPHY, SEARCH FRICTIONS AND ENDOGENOUS TRADE COSTS

Giulia Brancaccio  
Myrto Kalouptsidi  
Theodore Papageorgiou

Working Paper 23581  
<http://www.nber.org/papers/w23581>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
July 2017, Revised October 2018

We are thankful to Pol Antras, Costas Arkolakis, Aris Christou, Manolis Galenianos, Paul Grieco, Ken Hendricks, Fabian Lange, Robin Lee, Marc Melitz, Guido Menzio, Ariel Pakes, Nicola Rosaia and Jim Tybout for their many helpful comments. We also gratefully acknowledge financial assistance from the Griswold Center for Economic Policy Studies at Princeton University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Giulia Brancaccio, Myrto Kalouptsidi, and Theodore Papageorgiou. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Geography, Search Frictions and Endogenous Trade Costs  
Giulia Brancaccio, Myrto Kalouptsidi, and Theodore Papageorgiou  
NBER Working Paper No. 23581  
July 2017, Revised October 2018  
JEL No. F1,F14,L0,L91,R4,R41

### **ABSTRACT**

In this paper we study the role of the transportation sector in world trade. We build a spatial model that centers on the interaction of the market for (oceanic) transportation services and the market for world trade in goods. The model delivers equilibrium trade flows, as well as equilibrium trade costs (shipping prices). Using detailed data on vessel movements and shipping prices, we document novel facts about shipping patterns; we then flexibly estimate our model. We use this setup to demonstrate that the transportation sector (i) implies that net exporters (importers) face higher (lower) trade costs leading to misallocation of productive activities across countries; (ii) creates network effects in trade costs; and (iii) dampens the impact of shocks on trade flows. These three mechanisms reveal a new role for geography in international trade that was previously concealed by the common assumption of exogenous trade costs. Finally, we illustrate how our setup can be used for policy analysis by evaluating the impact of future and existing infrastructure projects (e.g. Northwest Passage, Panama Canal).

Giulia Brancaccio  
Fisher Hall  
Princeton University  
Princeton NJ 08544  
giulia@princeton.edu

Theodore Papageorgiou  
Department of Economics  
McGill University  
855 Sherbrooke St West, Room 519  
Montreal, QC H3A 2T7, Canada  
theodore.papageorgiou@mcgill.ca

Myrto Kalouptsidi  
Department of Economics  
Harvard University  
Littauer 124  
Cambridge, MA 02138  
and NBER  
myrto@fas.harvard.edu

# 1 Introduction

Whether by sea, land or air, the entirety of trade in goods is carried out by the transportation sector. With world trade at full steam- the sum of world exports and imports is now more than 50% of the global value of production- the transportation sector has become central in everyday life. Yet, little is known about how the market for transportation services interacts with the market for world trade in goods.

In this paper, we study how this interaction shapes trade flows, trade costs, the propagation of shocks and the allocation of productive activities across countries. As we demonstrate, the transportation sector (i) implies that net exporters (importers) face higher (lower) trade costs leading to misallocation of productive activities across countries; (ii) creates network effects in trade costs; and (iii) dampens the impact of shocks on trade flows. These three mechanisms reveal a new role for geography in international trade that was previously shrouded by the common assumption of exogenous trade costs.

The transportation sector includes several different segments which can be split into two categories: those that operate on fixed itineraries, much like buses, and those that operate on flexible routes, much like taxis. Containerships, airplanes and trains primarily belong to the first group, while trucks, gas and oil tankers, and dry bulk ships to the second. Here we focus on oceanic shipping and in particular, dry bulk shipping; 80% of world trade volume is carried by ships and dry bulkers carry about half of that.<sup>1</sup> Dry bulk ships are the main mode of transportation for commodities, such as grain, ore, and coal. They are often termed the “taxis of the oceans,” as an exporter has to find an available vessel and hire it for a specific voyage, with prices set in the spot market. Despite the operational differences of the various transport modes, we argue that the core economic mechanisms hold for most, if not all of them.

We leverage detailed micro-data on vessel movements, as well as rich data on contracts between exporters and shipowners to uncover some novel facts. First, satellite data of ships’ movements reveal that most countries are either large net importers or large net exporters and that, related to this, at any point in time a staggering 42% of ships are traveling without cargo (termed “ballast”). This natural trade imbalance is a key driver of trade costs. Indeed, transportation prices are largely asymmetric and depend on the destination’s trade imbalance: all else equal, the prospect of having to ballast after offloading leads to higher prices. For instance, shipping from Australia to China is 30% more expensive than the reverse: as China mostly imports raw materials, ships arriving there have limited opportunities to reload. This phenomenon is pervasive in most, if not all modes of transportation: trucks, trains, container air and

---

<sup>1</sup>Source: International Chamber of Shipping and UNCTAD. Seaborne trade accounts for about 70% of trade in terms of value.

ocean shipping, all exhibit similar price asymmetries that correlate with trade imbalances (the direction of the imbalance, however, may be the opposite). In fact, the US-China trade deficit in manufacturing has incentivized US exports of low value cargo, such as scrap or hay, to fill up the empty backhauls.<sup>2</sup>

Second, we compute the elasticity of trade with respect to shipping costs, by regressing trade by country pair on the corresponding shipping price. To do so, we employ a novel instrument inspired by the insight that the shipping price exporters face depends on how attractive their destination is to the ship in terms of future loading opportunities, as discussed above. The estimated trade elasticity indicates that the transport sector has a substantial impact on world trade, especially given the large fluctuations in shipping prices.

Third, a number of features of dry bulk shipping, such as information frictions and port infrastructure, can hinder the matching of ships and exporters. We propose a novel test to argue that these frictions lead to unrealized potential trade. The test is based on weather shocks at sea that exogenously shift ship arrivals at port: in a frictionless world, in regions with more ships than exporters, the change in the number of ships does not affect matches, since the short side of the market is always matched. Here, instead, matches are indeed affected by weather shocks. In addition, the law of one price does not hold: shipping prices exhibit substantial dispersion within a time-origin-destination triplet, also consistent with frictions. Finally, at any given time, in most countries there are simultaneous arrivals of empty ships that load and departures of empty ships, even though ships are homogeneous. This also suggests wastefulness.

Inspired by these facts, we build a spatial model that centers on the interaction of the market for transport and the market for world trade in goods, that is in the spirit of the search and matching literature. The globe is split into a number of regions that trade with each other. Geography enters the model both through regions' location in space, as well as their natural inheritance in commodities of different value. In each region, available ships and exporters participate in a random matching process. When matched with an exporter, ships transport the exporter's cargo to its destination for a negotiated price, and restart there. Ships that do not get matched decide whether to wait at their current location

---

<sup>2</sup>In 2005, about 60% of the containers sent via ships from Asia to North America came back empty (Drewry Consultants), and those "that did come back full were often transported at a steep discount for lack of demand (...) Shippers are so eager to fill their vessels for the return voyage to East Asia that they accept many types of unprofitable cargo, like bales of hay." Similarly, "airlines had become so eager to put something in their cargo holds on the inbound journey to China that rates go as low as 30 to 40 cents a kilogram, compared with \$3 to \$3.50 a kilogram leaving China [...] Very bluntly speaking, they're flying in empty and flying out full." (The International Herald Tribune, 01/30/2006)

In trucking, "[b]ackhaul practices are extremely important in explaining differences in prices; i.e. the truck companies are compensating their expenses on the empty backhaul in the first leg of the trip." (The World Bank, 2012). Even Uhaul moving trucks are priced similarly, "Rent a moving truck from Las Vegas to San Jose and you'll pay about \$100. In the opposite direction, the same truck will cost you 16 times that. (...) What accounts for the difference? (...) With so many people leaving the Bay Area, there are not enough rental trucks to go around." (SFgate, 02/15/2018).

or ballast elsewhere to search there. Exporters that get matched have their cargo delivered and collect its revenue, while exporters that do not get matched wait at port. Finally, a large number of potential exporters decide whether and where to export, thus replenishing the exporter pool seeking transportation.

We derive the equilibrium trade costs (shipping prices), as well as an expression for the equilibrium bilateral trade flows, that is reminiscent of a gravity equation. As ships are forward-looking, trade costs depend on the attractiveness of both the origin and the destination, a rich object that captures a region’s location, freight values, matching probabilities, as well as its neighbors’ attractiveness. This insight applies beyond dry bulk. Although other transport modes require different modeling assumptions regarding their operational practices, in equilibrium, prices are formed by the optimizing behavior of forward-looking transport agents and unavoidably depend, as above, on the attractiveness of origins and destinations, as well as that of their neighbors.<sup>3</sup>

Next, we estimate the model using the collected data. We first estimate the matching function, which gives the number of matches as a function of the number of agents searching on each side of the market. A sizable literature has estimated matching functions in different contexts (e.g. labor markets, taxicabs).<sup>4</sup> Here, we adopt a novel approach to flexibly recover both the matching function, as well as searching exporters, which, unlike ships and matches, are not observed. Our approach draws from the literature on nonparametric identification (Matzkin, 2003) and, to our knowledge, we are the first to apply it to matching function estimation. Our approach extends the literature in two dimensions. First, we do not take a stance on the presence of search frictions. When one side of the market (in this case exporters) is unobserved or mismeasured, it is difficult to discern whether search frictions are present. Second, we avoid parametric restrictions on the matching function; this is important, since in frictional markets, the shape of the underlying matching function is directly linked to welfare.

We then estimate the remaining primitives including ship costs, the values of exporters’ cargo and exporter entry costs. In particular, we recover ship sailing and port costs from the optimal ballast choice probabilities, via Maximum Likelihood, following the dynamic discrete choice literature (Rust, 1987). Then, we obtain exporter valuations directly from observed prices. Finally, we use trade flows to recover exporter costs by destination.

---

<sup>3</sup>For instance, a model for container shipping (or other modes operating on fixed itineraries), would likely feature a small number of shipping firms that decide on their itineraries, as well as route pricing. The itinerary decision involves similar trade-offs as the ballast choice here; e.g. what network of countries can be serviced or how to manage backhaul trips. Prices similarly depend on the number and valuations of exporters, ship supply and crucially, the entire route serviced including the backhaul trips.

<sup>4</sup>For instance, in labor markets, data on unemployed workers, vacancies and matches delivers the underlying matching function. In the market for taxi rides, one observes taxis and their rides, but not hailing passengers; in recent work, Buchholz (2018) and Frechette et al. (2018) have used a parametric assumption on the matching function, to recover the passengers.

Why is it important to account for the transport sector to study international trade in goods? We illustrate the role of endogenous trade costs through three experiments.

First, we compare our setup to one with “iceberg” trade costs that are exogenous and depend only on distance and the cargo’s value, as is the case in canonical trade models. Strikingly, we find that the transport sector acts as a smoothing factor that reduces world trade imbalances. Indeed, under endogenous trade costs net exporters (importers) export less (more) than under exogenous trade costs, leading to a misallocation of productive activities from (efficient) net exporters to (inefficient) net importers. This happens because of ships’ equilibrium behavior and in particular the strength of their bargaining position at different regions. Net exporters offer loading opportunities to ships, thus allowing them to command high prices, which in turn restrains their exports. The converse holds for net importers. This argument extends to a country’s neighbors: a net exporter close to other net exporters offers even more options to ships and prices are even higher, which inhibits the neighborhood’s exports.

Second, we illustrate that the transport sector dampens the impact of shocks on trade flows, by considering a fuel cost shock. A decline in the fuel cost has a direct and an indirect effect. The direct effect is straightforward: as costs fall, shipping prices also fall and thus exports rise. The novel indirect effect is that a decline in fuel costs improves a ship’s bargaining position, as it makes ballasting less costly. This dampens the original decline in prices and increase in exporting. Indeed, the overall increase in world trade would be 40% higher if ships were not allowed to optimally adjust their behavior, in response to a 10% decline in fuel costs.

Third, we explore the spatial propagation of a macro shock: a slow-down in China. Besides the direct effect to countries whose exports rely heavily on the Chinese economy, the optimal reallocation of ships over space differentially filters the shock in neighboring vs. distant regions. Because of the slow-down, fewer ships offload in China, reducing ship supply in the region. Although this impacts negatively China’s own exports by raising prices, it benefits distant countries, such as Brazil, because ships reallocate there.

Finally, we consider the role of maritime infrastructure on world trade, as an illustration of how our setup can be used for policy evaluation. To do so, we examine the opening of the Northwest Passage: the melting of the arctic ice would reduce the travel distance between Northeast America/Northern Europe and the Far East. Although the shock is local, it has global effects: as Northeast America becomes a more attractive ballasting choice, ships have a stronger bargaining position and demand higher prices, pushing exports down everywhere else. Moreover, we consider the impact of four natural and man-made passages: the Panama Canal, the Suez Canal, Gibraltar and the Malacca Straits and show that all passages tend to

substantially increase world trade.

**Related Literature** We relate to three broad strands of literature: (i) trade and geography; (ii) search and matching; (iii) industry dynamics.

First, our paper endogenizes trade costs and so it naturally relates to the large literature in international trade studying the importance of trade costs in explaining trade flows between countries (e.g. Anderson and Van Wincoop, 2003, Eaton and Kortum, 2002). In much of the literature, trade costs are treated as exogenous and follow the iceberg formulation of Samuelson (1954). Here, we consider what happens to trade flows when the equilibrium of the transport market is taken into account, so that transport prices (an important component of trade costs, at least as large or larger than tariffs; Hummels, 2007) are determined in equilibrium, jointly with trade flows. Moreover, related to some of our empirical findings, Waugh (2010) has argued that asymmetric trade costs are necessary to explain some empirical regularities regarding trade flows across rich and poor countries.

We also contribute to a literature that has considered the role and features of the (container) shipping industry; e.g. Hummels and Skiba (2004) explore the relationship between product prices at different destinations and shipping costs; Hummels et al. (2009) explore market power in container shipping; Ishikawa and Tarui (2015) theoretically investigate the impact of “backhaul” and its interaction with industrial policy; Cosar and Demir (2018) and Holmes and Singer (2018) study container usage; Asturias (2018) explores the impact of the number of shipping firms on transport prices and trade; Wong (2018) incorporates container shipping prices featuring a “round-trip” effect in a trade model. Finally, recent work has explored the matching of importers and exporters under frictions (Eaton et al., 2016, and Krolkowski and McCallum, 2018).

Our paper is also related to both old and new work on the role of geography in international trade (e.g. Krugman, 1991, Head and Mayer, 2004, Allen and Arkolakis, 2014), as well as the impact of transportation infrastructure and networks (e.g. Donaldson, 2012, Allen and Arkolakis, 2016, Donaldson and Hornbeck, 2016, Fajgelbaum and Schaal, 2017). We extend this literature by demonstrating that the transport sector reveals a new role for geography through three novel mechanisms (it misallocates productive activities, creates network effects and dampens the impact of shocks).

Second, our paper relates to the search and matching literature (see Rogerson et al., 2005 for a survey). Our model is a search model in the spirit of the seminal work of Mortensen and Pissarides (1994), where firms and workers (randomly) meet subject to search frictions and Nash bargain over a wage. An important

addition in our case is the spatial nature of our setup: there are several interconnected markets at which agents (ships) can search. Such a spatial search model was first proposed by Lagos (2000) (and analyzed empirically in Lagos, 2003) in the context of taxi cabs. We borrow heavily from his model; the key difference is that prices are set in equilibrium, while in the taxi market prices are exogenously set by regulation. This is crucial, as the role of endogenous trade costs is at the core of our paper.<sup>5</sup> Finally, as discussed above, our paper also contributes to the literature on matching function estimation (see Petrongolo and Pissarides (2001) for a survey).

Third, we relate to the literature on industry dynamics (e.g. Hopenhayn, 1992, Ericson and Pakes, 1995). Consistent with this research agenda, we study the long-run industry equilibrium properties, in our case the spatial distribution of ships and exporters. Moreover, our empirical methodology borrows from the literature on the estimation of dynamic setups (e.g. Rust, 1987, Bajari et al., 2007, Pakes et al., 2007; applications include Ryan, 2012 and Collard-Wexler, 2013). Buchholz (2018) and Frechette et al. (2018) also explore dynamic decisions in the context of taxi cabs' search and shift choices respectively. Finally, Kalouptsidi (2014) has also looked at the shipping industry, albeit at the entry decisions of shipowners and the resulting investment cycles in new ships, while Kalouptsidi (2018) focuses on industrial policy in the Chinese shipbuilding industry.

The rest of the paper is structured as follows: Section 2 provides a description of the industry and the data used. Section 3 presents the facts. Section 4 describes the model. Section 5 lays out our empirical strategy, while Section 6 presents the estimation results. Section 7 demonstrates the importance of endogenous trade costs, while Section 8 assesses the role of maritime infrastructure projects. Section 9 concludes. The Appendix contains additional tables and figures, proofs to our propositions, as well as further data and estimation details.

## 2 Industry and Data Description

### 2.1 Dry Bulk Shipping

Dry bulk shipping involves vessels designed to carry a homogeneous unpacked dry cargo, for individual shippers on non-scheduled routes. Bulk carriers operate much like taxi cabs: a specific cargo is transported

---

<sup>5</sup>There are several other differences between our setup and that of Lagos (2000) and Lagos (2003): we model also the demand side (exporters/passengers); we allow for the potential of frictions in each region, while Lagos (2000) assumes that matching is frictionless locally; we allow for several sources of heterogeneity in different regions (travel/port costs, distances, matching rates); we allow trade to be imbalanced, while Lagos (2000) relies on taxi flows in and out of each location to be equal- this distinction is also crucial.



individually by a specific ship, for a trip between a single origin and a single destination. Dry bulk shipping involves mostly commodities, such as iron ore, steel, coal, bauxite, phosphates, but also grain, sugar, chemicals, lumber and wood chips; it accounts for about half of total seaborne trade in tons (UNCTAD) and 45% of the total world fleet, which includes also containerships and oil tankers.<sup>6,7</sup>

There are four categories of dry bulk carriers based on size: Handysize (10,000–40,000 DWT), Handymax (40,000–60,000 DWT), Panamax (60,000–100,000 DWT) and Capesize (larger than 100,000 DWT). The industry is unconcentrated, consisting of a large number of small shipowning firms (Kalouptsidei, 2014): the maximum fleet share is around 4%, while the firm size distribution features a large tail of small shipowners. Moreover, shipping services are largely perceived as homogeneous. In his lifetime, a shipowner will contract with hundreds of different exporters, carry a multitude of different products and visit numerous countries; we discuss this further in Section 3.3.

Trips are realized through individual contracts: shipowners have vessels for hire, exporters have cargo to transport and brokers put the deal together. Ships carry at most one freight at a time: the exporter fills up the hired ship with his cargo. In this paper, we focus on spot contracts and in particular the so-called “trip-charters”, in which the shipowner is paid in a per day rate.<sup>8</sup> The exporter who hires the ship is responsible for the trip costs (e.g. fueling), while the shipowner incurs the remaining ship costs (e.g. crew, maintenance, repairs).

## 2.2 Data

We combine a number of different datasets. First, we employ a dataset of dry bulk shipping contracts, from 2010 to 2016, collected by Clarksons Research. An observation is a transaction between a shipowner and a charterer for a specific trip. We observe the vessel, the charterer, the contract signing date, the loading and unloading dates, the price in dollars per day, as well as some information on the origin and destination.

Second, we use satellite AIS (Automatic Identification System) data from exactEarth Ltd (henceforth EE) for the ships in the Clarksons dataset between July 2010 and March 2016. AIS transceivers on the ships automatically broadcast information, such as their position (longitude and latitude), speed, and level

---

<sup>6</sup>As already mentioned, bulk ships are different from containerships, which carry cargo (mostly manufactured goods) from many different cargo owners in container boxes, along fixed itineraries according to a timetable. It is not technologically possible to substitute bulk with container shipping.

<sup>7</sup>It is not straightforward to obtain information on the share of world trade value carried by bulkers. However, mining, agricultural products, chemicals and iron/steel jointly account for about 30% of total trade value (WTO, 2015).

<sup>8</sup>Trip-charters are the most common type of contract. Long-term contracts (“time-charters”), however, do exist: on average, about 10% of contracts signed are long-term. Interestingly, though, ships in long-term contracts, are often “relet” in a series of spot contracts, suggesting that arbitrage is possible.

of draft (the vertical distance between the waterline and the bottom of the ship’s hull), at regular intervals of at most six minutes. The draft is a crucial variable, as it allows us to determine whether a ship is loaded or not at any point in time.

We also use the ERA-Interim archive, from the European Centre for Medium-Range Weather Forecasts (CMWF), to collect global data on daily sea weather. This allows us to construct weekly data on the wind speed (in each direction) on a 0.75° grid across all oceans. Finally, we employ several time series from Clarksons on e.g. the total fleet and fuel prices, as well as country-level imports/exports, production and commodity prices from other sources (e.g. Comtrade, IEA).

	Mean	St. Dev	Median	Min	Max
Contract price per day (thousand US dollars)	13.9	8.6	12	3.2	43
Contract trip price (thousand US dollars)	291	304	178	30.5	1,367
Contracts per ship	2.9	2.2	2	1	10
Trip duration (weeks)	2.89	1.36	2.95	0.5	5.44
Days between contract signing and loading date	6.39	7.12	5	0	30
Prob of ship staying at port conditional on not signing a contract	0.77	0.12	0.76	0.59	0.95

**Table 1:** Summary statistics. The contract price per day is reported by Clarksons. To create the price per trip, we multiply price per day with the average number of days required to perform the trip. Contracts per ship counts the number of contracts observed for each ship in the Clarksons dataset. To proxy for trip duration, we compute the nautical distance in miles and divide it by the average speed observed in the EE data. The probability of staying at port is calculated from the EE data by computing the frequency at which waiting ships that did not find a contract in a given week remain at port instead of ballasting elsewhere. We have 12,007 observations of shipping contracts and 393,058 ship-week observations at which the ship decides to either ballast someplace or stay at its current location.

**Summary statistics** Our final dataset involves 5,398 ships, which corresponds to about half the world fleet, between 2012 and 2016.<sup>9</sup> We end up with 12,007 shipping contracts, for which we know the price, as well as the exact origin and destination (see Appendix A for our data matching procedure).<sup>10</sup> As shown in Table 1, the average price is 14,000 dollars per day (or 290,000 dollars for the entire trip), with substantial variation. Trips last on average 2.9 weeks. Contracts are signed close to the loading date, on average six days before. The most popular loaded trips are from Australia, Brazil and Northwest America to China, while the most popular ballast trip is from China to Australia (5.7% of ballast trips). We have 393,058 ship-week observations at which the ship decides to either ballast someplace or stay at

<sup>9</sup>We drop the first two years (until May 2012) of vessel movement data, as satellites are still launched at that time and the geographic coverage is more limited.

<sup>10</sup>The Clarksons contracts somewhat oversample the intermediate size categories (Handymax and Panamax) and younger ships.

its current location. Ships that do not sign a contract, remain in their current location with probability 77%, while the remaining ships ballast elsewhere. Finally, Clarksons reports the product carried in about 20% of the sample. The main commodity categories are grain (29%), ores (21%), coal (25%), steel (8%) and chemicals/fertilizers (6%).

### 3 Facts

In this section, we present some novel facts about the transport sector and trade: we first discuss the implications of trade imbalances (section 3.1); then, we quantify the impact of transport costs on world trade (section 3.2); finally, we provide evidence of frictions (section 3.3). Throughout the paper, we most often split ports into 15 geographical regions, depicted in Figure 10 in Appendix B.<sup>11</sup>

#### 3.1 Trade Imbalances

World trade in commodities is greatly imbalanced. Indeed, most countries are either large net importers or large net exporters. This is shown in Figure 1, which plots the difference between the number of ships departing loaded and the number of ships arriving loaded, over the sum of the two. Australia, Brazil and Northwest America are big net exporters, whereas China and India are big net importers.

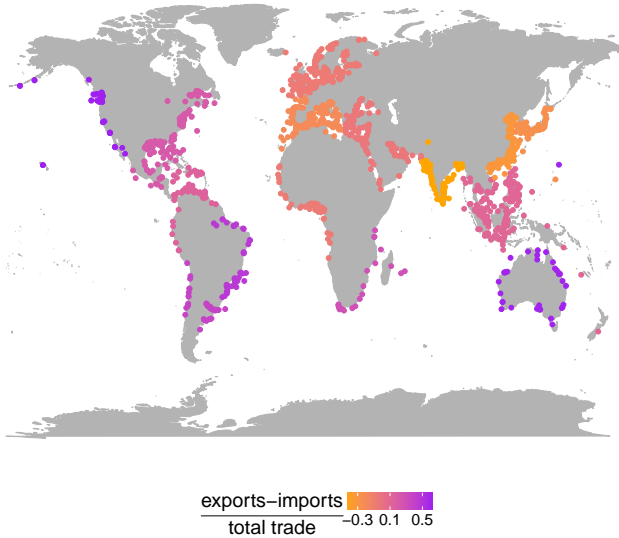
The fact that global trade features such large imbalances is not as surprising; to a large extent, this is due to the different natural inheritance of countries. For instance, Australia, Brazil and Northwest America are rich in minerals, grain, coal, etc. At the same time, growing developing countries require imports of raw materials to achieve industrial expansion and infrastructure building. For instance, in recent years, Chinese growth has relied on massive imports of raw materials. As a result, commodities flow out of producers such as Australia and Brazil, towards China and India.

As a consequence of the imbalanced nature of international trade, ships spend much of their time traveling ballast, i.e. without cargo. Indeed, we find that 42% of a ship's traveled miles are ballast, so that a ship is traveling empty close to half the time.<sup>12</sup> Finally, the trade imbalance is a key driver of the trade costs that exporters face. First, a quick inspection of the data reveals that there are large asymmetries in trade costs across space: for instance, a trip from China to Australia costs on average 7,500 dollars per

---

<sup>11</sup>The trade-off is that we need a large number of observations per region, while allowing for sufficient geographical detail. To determine the regions, we employ a clustering algorithm that minimizes the within-region distance between ports. The regions are: West Coast of North America, East Coast of North America, Central America, West Coast of South America, East Coast of South America, West Africa, Mediterranean, North Europe, South Africa, Middle East, India, Southeast Asia, China, Australia, Japan-Korea. We ignore intra-regional trips and entirely drop these observations.

<sup>12</sup>This percentage is lower for smaller ships— it is 32% for Handysize, 41% for Handymax, 45% for Panamax and 49% for Capesize.



**Figure 1:** Trade Imbalances. Difference between exports (ships leaving loaded) and imports (ships arriving loaded) over total trade (all ships). A positive (negative) ratio indicates that a country is a net exporter (importer); a ratio close to zero implies balanced trade.

day, while a trip from Australia to China costs on average 10,000 dollars per day.<sup>13</sup> In fact, most trips exhibit substantial asymmetry: the average ratio of the price from  $i$  to  $j$  over the price from  $j$  to  $i$  (highest over lowest), is 1.6 and can be as high as 4.1.

We further investigate the determinants of trade costs by considering how shipping prices are associated with the attractiveness of the destination, such as its demand for shipping. Indeed, ships may demand a premium to travel towards a destination with low exports (e.g. China), to compensate for the difficulty of finding a new cargo originating from that destination. As shown in Column III of Table 2, shipping to a destination where the probability of a ballast trip afterwards is ten percentage points higher, costs 2.3% more on average. Similarly, a 10% increase in the average distance traveled ballast after the destination, is associated with a 1.7% increase in prices.

### 3.2 Trade Elasticity

Do shipping prices have an impact on world trade? In this section we address this question in the context of bulk shipping. Ideally, we would like to regress bilateral trade flows on shipping prices, i.e.,

$$\log Q_t^{i \rightarrow j} = \beta_0 + \beta_1 \log \tau_t^{i \rightarrow j} + \epsilon_{ijt}$$

<sup>13</sup>This price asymmetry has been documented also in container shipping; see e.g. Wong (2018) and references therein.

	I	II	III
	log(price per day)		
Probability of ballast		0.234**	0.556**
		(0.030)	(0.081)
Avg duration of ballast trip (log)		0.166**	0.065**
		(0.014)	(0.032)
Coal			0.088**
			(0.045)
Fertilizer			0.245**
			(0.051)
Grain			0.131**
			(0.048)
Ore			0.124**
			(0.045)
Steel			0.135**
			(0.049)
Constant	10.284**	9.127**	8.915**
	(0.103)	(0.099)	(0.408)
Destination FE	Yes	No	No
Origin FE	Yes	Yes	Yes
Ship type FE	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes
Obs	11,014	11,011	1,662
R <sup>2</sup>	0.694	0.674	0.664

\*\*  $p < 0.05$ , \*  $p < 0.1$

**Table 2:** Shipping price regressions. The dependent variable is the logged price per day in USD. The independent variables include combinations of: the average frequency of ballast traveling after the contract's destination (Probability of ballast), the average logged duration (in days) of the ballast trip after the contract's destination, as well as ship type, origin, destination and quarter FEs. The product is reported in only 20% of the sample, so the regression in column III has substantially fewer observations. The omitted product category is cement.

where  $Q_t^{i \rightarrow j}$  is the total trade value from country  $i$  to country  $j$  (in bulk commodities) at time period (month)  $t$  and  $\tau_t^{i \rightarrow j}$  is the shipping price from  $i$  to  $j$  at  $t$ . Naturally, this regression is going to lead to biased estimates, as prices are likely correlated with the error,  $\epsilon_{ijt}$ . Thus, an instrument is required.

The instrument we leverage is inspired by the insight that, as discussed above, the attractiveness of an exporter’s destination impacts the shipping price it faces. Consider the trade flow from  $i$  to  $j$ ,  $Q_t^{i \rightarrow j}$ ; the instrument we use for the shipping price  $\tau_t^{i \rightarrow j}$  consists of the tariffs levied on commodity exports *from* the destination  $j$ . For example, the price to ship goods from Indonesia to China is instrumented using the tariffs on raw materials on routes starting from China. These tariffs do not directly affect the flows from Indonesia to China. However, they affect the value of a ship unloading in China. Indeed, tariffs on  $j$ ’s exports lead to a reduction in shipments from  $j$ , thus dampening the demand for shipping services in  $j$  and making  $j$  a less attractive destination for ships. Therefore, when negotiating a price to ship goods from  $i$  to  $j$ , a ship demands a higher price in order to compensate for its reduced opportunities upon arrival at  $j$ .<sup>14</sup> Similarly, we also use the tariffs levied on commodities imported at the exporter’s origin,  $i$ , as an instrument for the price  $\tau_t^{i \rightarrow j}$ . These tariffs reduce  $i$ ’s imports, leading to lower ship supply in origin  $i$ , and, thus, higher shipping prices to export from  $i$  to  $j$ .

Table 3 presents both stages of the two-stage least squares. Both the first and second stage regressions are run in differences to control for any fixed, route-specific characteristics; we also control for GDP, tariffs on the route considered, as well as tariffs on all goods other than commodities (all in differences). The first stage regresses per-day shipping prices from  $i$  to  $j$  both on the tariffs levied on exports from  $j$  to its first and second biggest trading partners (tariff  $j \rightarrow (1)$  and tariff  $j \rightarrow (2)$ ), as well as on tariffs on  $i$ ’s imports from its first and second biggest trading partners (tariff  $(1) \rightarrow i$  and tariff  $(2) \rightarrow i$ ).<sup>15</sup> The instruments are jointly significant at the 1% level, and the signs are as expected: higher tariffs tend to increase shipping prices. The first stage results are interesting per se, as they showcase that shipping prices between any two countries are affected by shipping conditions on other routes, creating inter-dependencies and network effects in trade costs; this mechanism, which is formalized in our model and is central to the paper, is quantitatively important.

The second stage of the IV regression produces a trade elasticity of 1.02 with respect to shipping prices.

---

<sup>14</sup>This instrument is valid as it should not impact directly  $Q_t^{i \rightarrow j}$ . Recall that here we focus only on raw materials, hence the supply chain should not be a concern (e.g. the instrument would be problematic if  $j$  imports steel and exports cars and we considered tariffs on cars). Moreover, we control directly for the tariffs from  $i$  to  $j$  and the overall level of tariffs on all goods other than commodities.

<sup>15</sup>We obtain yearly country-level trade flows from Comtrade and tariffs from the World Bank (WITS) and we focus only on bulk commodities; yearly average shipping prices come from our Clarksons dataset. The results are robust if we add country fixed effects, or if we use the weighted average of tariffs instead.

In other words, a 1% increase in shipping prices leads to a 1.02% decline in trade flows. This elasticity indicates that the transport sector has a substantial impact on world trade, especially given the large observed fluctuations in shipping prices (for instance, shipping prices experienced an 8-fold increase in the late 2000s, see Kalouptsidi, 2014).

A few recent papers have estimated the same elasticity for the case of container shipping. Asturias (2018), who uses population as an instrument, finds the elasticity to be about 5. Wong (2018), who uses the round-trip effect as an instrument (in particular, for route  $i, j$  she uses a Bartik-style instrument to proxy for the predicted trade volume on route  $j, i$ ) finds an elasticity of about 3. It is also worth comparing the elasticity of trade with respect to shipping prices to that with respect to tariffs, which is estimated to be between 1.5 and 5 on average (e.g. Simonovska and Waugh, 2014, Caliendo and Parro, 2015, Brandt et al., 2017 and Arkolakis et al., 2018). The two elasticities are comparable, with our estimate overall somewhat lower. Recall, however, that we use total rather than waterborne trade value; therefore our estimate should be considered a lower bound of the trade elasticity with respect to shipping prices, as it ignores substitution towards other modes of transportation.<sup>16</sup>

### 3.3 Search Frictions

A number of features of dry bulk shipping, such as information frictions and port infrastructure, can hinder the matching of ships and exporters. In this section we argue that these frictions indeed lead to unrealized potential trade. Consider a geographical region, such as a country or a set of neighboring countries, where there are  $s$  ships available to pick up cargo and  $e$  exporters searching for a ship to transport their cargo. We define search frictions by the inequality:

$$m < \min \{s, e\} \tag{1}$$

where  $m$  is the number of matched ships and exporters. In other words, under frictions there is potential trade that remains unrealized; in contrast, in a frictionless world, the entire short side of the market gets matched, so that  $m = \min \{s, e\}$ . When (1) holds, matches are often modeled via a matching function,  $m = m(s, e)$ , as is done extensively in the labor literature. As Petrongolo and Pissarides (2001) note, “[...] the matching function [...] enables the modeling of frictions [...] with a minimum of added complexity. Frictions derive from information imperfections about potential trading partners, heterogeneities, the

---

<sup>16</sup>For instance, if we exclude EU countries, which can easily substitute from oceanic shipping to land shipping, the elasticity increases to -3.

	$\Delta \log (\tau_t^{i \rightarrow j})$	$\Delta \log (Q_t^{i \rightarrow j})$
	First Stage	IV
$\Delta \log (\tau_t^{i \rightarrow j})$		-1.02** (0.425)
$\Delta \log (\text{tariff}_t^{j \rightarrow (1)})$	0.070* (0.040)	
$\Delta \log (\text{tariff}_t^{j \rightarrow (2)})$	0.135** (0.027)	
$\Delta \log (\text{tariff}_t^{(1) \rightarrow i})$	0.152 (0.096)	
$\Delta \log (\text{tariff}_t^{(2) \rightarrow i})$	-0.034 (0.082)	
$\Delta \log (\text{tariff}_t^{i \rightarrow j})$	0.123** (0.058)	-0.326** (0.109)
Constant	-0.225** (0.021)	-2.173** (0.647)
	GDP of $i$ and $j$	
Controls (changes of)	tariff on $i$ 's import (non-commodities) tariff on $j$ 's export (non-commodities)	
Obs	470	470
R <sup>2</sup>	0.143	-
F-stat	7.04	

\*\*  $p < 0.05$ , \*  $p < 0.1$

**Table 3:** Elasticity of trade with respect to shipping prices. Data on yearly bilateral country-level trade value and tariffs are obtained from the World Bank (WITS) for the period 2010-2016. We focus on trade and tariffs for bulk commodities. To construct tariffs, we consider the minimum between the most favored nation tariff and preferential rates, if applicable, and consider a weighted average across commodities. Shipping prices are calculated from the per-day prices in Clarksons contracts, averaged at the year and country-pair level. We group countries in EU-27 and exclude countries without no access to sea.



absence of perfect insurance markets, slow mobility, congestion from large numbers, and other similar factors.”

We present three facts consistent with frictions, as defined by (1). These facts are inspired by labor markets, where search frictions are generally thought to be present. In particular, we (i) provide a direct test for inequality (1); (ii) we document wastefulness in ship loadings; (iii) we document substantial price dispersion.

First, we provide a simple **test for search frictions**. If we observed all variables  $s, e, m$ , it would be straightforward to test (1); this is essentially what is done in the labor literature, where the co-existence of unemployed workers and vacant firms is interpreted as evidence of frictions. However in our setup, as discussed at length in Section 5.1, we observe  $m$  (i.e. ships leaving loaded) and  $s$ , but not  $e$ ; we thus need to adopt a different approach.

Assume it is known that there are more ships than exporters, i.e.  $\min(s, e) = e$ . We begin with this assumption, because our sample period is one of low shipping demand and severe ship oversupply due to high ship investment between 2005 and 2008 (see Kalouptsidei, 2014, 2018). If there are no search frictions, so that  $m = \min(s, e) = e$ , small exogenous changes in the number of ships should not affect the number of matches. In contrast, if there are search frictions, an exogenous change in the number of ships changes the number of matches, through the matching function  $m = m(s, e)$ . We thus test for search frictions by using ocean weather conditions (unpredictable wind at sea), which affect travel times, to explore whether exogenously changing the number of ships in regions with a lot more ships than exporters affects the realized number of matches.<sup>17</sup> Since we do not observe exporters directly, to select periods in which there are more ships than exporters, for each region we consider weeks when there are at least twice as many ships waiting in port as matches. Table 4 presents the results. We find that indeed matches are affected by weather conditions in all regions, consistent with the presence of search frictions.

Second, we document **simultaneous arrivals** of empty ships that then load **and departures of empty ships**. Indeed, the first two panels of Figure 2 display the weekly number of ships that arrive empty and load, as well as the number of ships that leave empty, in two net exporting countries: Norway and Chile. In Norway, several ships arrive empty and load, but almost no ship departs empty. In Chile, however, the picture is quite different: it frequently happens that an empty ship arrives and picks up cargo, while at the same time another ship departs empty. This is suggestive of wastefulness in Chile:

---

<sup>17</sup>In particular, we divide the sea surrounding each region into zones; for each zone we use information on the wind speed at different distances from the coast and in different directions. To obtain the unpredictable component of weather we run a VAR regression of these weather indicators. The results are robust to the lag structure, as well as estimating jointly for neighboring zones.

	N	Joint Significance	$\frac{s}{m}$
North America West Coast	193	0	2.706
North America East Coast	200	0.013	3.172
Central America	199	0	3.451
South America West Coast	198	0	2.913
South America East Coast	200	0	4.083
West Africa	200	0	5.862
Mediterranean	200	0	4.244
North Europe	200	0	3.577
South Africa	200	0.01	2.862
Middle East	200	0.001	3.86
India	200	0.12	8.58
South East Asia	200	0.005	3.334
China	200	0	6.194
Australia	187	0.008	2.457
Japan-Korea	200	0.003	5.311

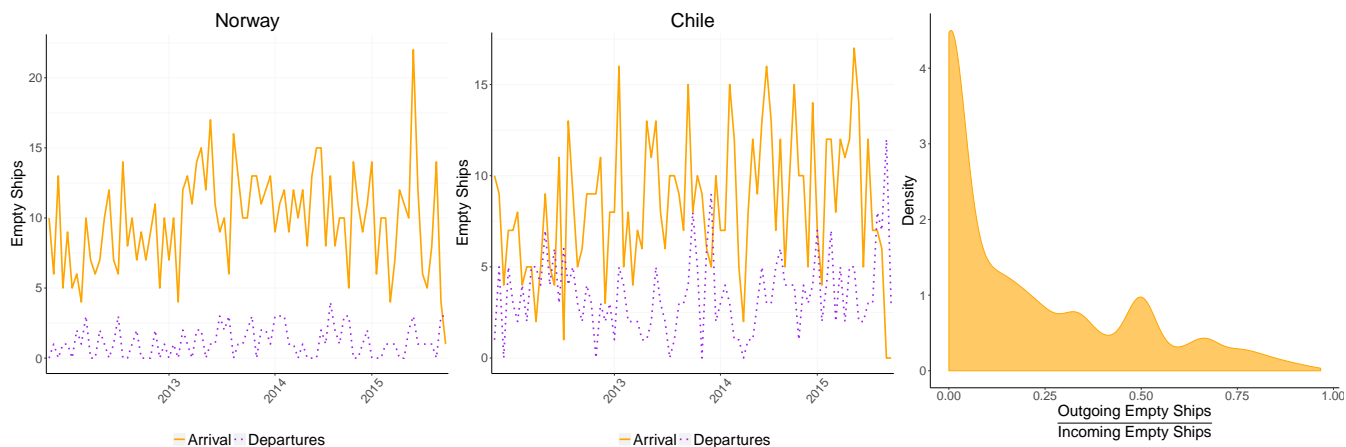
**Table 4:** Test for search frictions. Regressions of the number of matches in each region on the unpredictable component of weather conditions in the surrounding seas. For each region we use weeks in which there are at least twice as many ships as matches. The first column reports the number of observations; the second column joint significance; and the third column the average ratio between matches and ships in each region during these weeks. To proxy for the unpredictable component of weather, we divide the sea surrounding each region into 8 different zones (Northeast, Southeast, Southwest and Northwest both within 1,500 miles of the coast and between 1,500 and 2,500 miles from the coast), and we use the speed of the horizontal (E/W) and vertical (N/S) component of wind in each zone to proxy for weather conditions. We run a VAR regression of these weather variables on their lag component and season fixed effects and use the residuals, together with their squared term, as independent variables in the regression.

why does the ship that depart empty, not pick up the cargo, instead of having another ship arrive from elsewhere to pick it up?

This pattern is observed in many countries. Indeed, the third panel of Figure 2 depicts the histogram of the bi-weekly ratios of outgoing empty ships over incoming empty and loading ships for net exporting countries. In the absence of frictions, one would expect this ratio to be close to zero. However, we see that the average ratio is well above zero. Moreover, this pattern is quite robust in a number of dimensions.<sup>18</sup>

Third, again inspired by the labor literature, we investigate **dispersion in prices**. In markets with no frictions, the law of one price holds, so that there is a single price for the same service. This does not hold in labor markets, where there is large wage dispersion among workers who are observationally identical.

<sup>18</sup>This figure is robust to alternative market definitions, time periods and ship types. Capesize vessels exhibit somewhat larger mass towards zero, consistent with the somewhat higher concentration of ships and charterers, as well as the ships' ability to approach fewer ports. The figure is also similar if done by port rather than country. To control for repairs we remove stops longer than 6 weeks. Finally, we only consider as "ships arriving empty" the ships arriving empty and sailing full towards another region, and we consider as "ships leaving empty" ships sailing empty toward a different country; so movements to nearby ports are excluded. This definition also implies that refueling cannot explain the fact either- though there are very small differences in fuel prices across space anyway (less than 10%).



**Figure 2:** Simultaneous arrivals and departures of empty ships: The first two panels depict the flow of ships arriving empty and loading, and ships leaving empty in two-week intervals in Norway and Chile. The last panel shows the histogram of the ratio of outgoing empty over incoming empty and loading ships across all net exporting countries.

This observation has generated a substantial and influential literature on frictional wage inequality, i.e. wage inequality that is driven by search frictions.<sup>19</sup> As we already saw in Table 2 there is substantial price dispersion in shipping contracts. More specifically, at best we can account for about 70% of price variation, controlling for ship size, as well as quarter, origin and destination fixed effects. Moreover, the coefficient of variation of prices within a given quarter, origin and destination triplet is about 30% (23%) on average (median). In the most popular trip, from Australia to China, the *weekly* coefficient of variation is on average 34% and ranges from 15% to 65% across weeks. In addition, it is worth noting that the type of product carried affects the price paid and overall more valuable goods lead to higher contracted prices, as shown in Table 2. In the absence of frictions, if there are more ships than exporters, as is the case during our sample period, we would expect prices to be bid down to the ships’ opportunity cost.<sup>20</sup> In contrast, in markets with frictions and bilateral bargaining, as shown formally in the model of Section 4, the buyer’s valuation affects the price he pays and exporters with higher valuations pay more.

As in labor markets, a multitude of factors can lead to frictions (i.e. unrealized matches) in shipping. First, the decentralized and unconcentrated nature of the market and the mere existence of brokers, suggest that information frictions are present.<sup>21</sup> Port infrastructure, congestion or capacity constraints may also hinder matching. In addition, regulations may impose special ship requirements (e.g. flags, environmental

<sup>19</sup>See for instance Burdett and Mortensen (1998), Postel-Vinay and Robin (2002), Mortensen (2003) and references therein.

<sup>20</sup>In a frictionless market with more ships than freights and homogeneous ships, in equilibrium the price from an origin to a destination would be such that ships are indifferent between transporting the cargo and staying unmatched.

<sup>21</sup>The meeting process involves a disperse network of brokers; oftentimes more than two brokers intervene to close a deal, suggesting that the ship’s and the exporter’s brokers do not always find each other, and that an “intermediate broker” was necessary to bring the two together (Panayides, 2016). In interviews, brokers claimed to receive 5,000-7,000 emails per day; sorting through these emails is reminiscent of an unemployed worker sorting through hundreds of vacancy postings.

rules). Finally, heterogeneities, such as long-run relations or special cargo requirements in ship investments may result in unrealized trade.

While in labor markets, as some researchers have argued, observed or unobserved heterogeneity may partly explain the documented facts, in shipping heterogeneity is much more limited. Indeed, the data suggests that ship heterogeneity alone is not a prominent explanation for search frictions. Ships do not specialize neither geographically, nor in terms of products: the majority of ships deliver cargo to 13 out of 15 regions and carry at least 2 of the 3 main products (coal, ore and grain). Moreover, neither shipowner characteristics, nor shipowner fixed effects have any explanatory power in price regressions, as shown in Table 9 in Appendix B, while ballast decisions of ships in the same region are concentrated around the same options.<sup>22</sup> Nonetheless, despite the anecdotal and descriptive evidence presented, it is not possible to reject that heterogeneity can also play a role in the market.

## 4 Model

In this section, motivated by the above findings, we introduce a spatial model that centers on the interaction between the market for transport and the market for world trade in goods. In each period, the timing is as follows: In each region, available ships and exporters participate in a random matching process. Ships that get matched transport their exporter’s cargo to its destination for a negotiated price, and restart there. Ships that do not get matched decide whether to wait at their current location or ballast elsewhere and search there. Exporters that get matched have their cargo delivered and collect their revenue. Exporters that do not get matched wait at port. Finally, a large number of potential exporters decide whether and where to export, thus replenishing the exporter pool seeking transportation the following period.

We first lay out the model’s setup; we then present the agents’ value functions and derive the equilibrium objects of interest: trade costs (shipping prices) and trade flows (gravity equation). We close the section with a detailed discussion of our main assumptions.

### 4.1 Environment

Time is discrete. There are  $I$  locations/regions,  $i \in \{1, 2, \dots, I\}$ . There are two types of agents, exporters and ships. Both are risk neutral and have discount factor  $\beta$ .

---

<sup>22</sup>If heterogeneity were an important driver of ships’ ballasting decisions, we would expect ships to choose diverse destinations from a given origin. Yet we find that ballast choices are similar across ships (the  $CR_3$  measure for the chosen destinations is higher than 70% in most regions). Moreover, home-ports are not an important consideration for shipowners, as the crew flies to their home country every 6-8 months.

At each location  $i$  and period  $t$ , there are  $e_{it}$  exporters/freights that need to be delivered to another location. An exporter obtains revenue (or valuation),  $r$ , from shipping the good. Every period, at each location  $i$ ,  $\mathcal{E}_i$  potential exporters decide whether and where to export. If they decide to export, they pay production and export costs,  $\kappa_{ij}$  and draw their revenue  $r$ , from a distribution  $F_{ij}^r$  with mean  $\bar{r}_{ij}$ .

There are  $S$  homogeneous ships in the world.<sup>23</sup> In every period, a ship is either at port in some region  $i$ , or it is traveling loaded or ballast, from some location  $i$  to some location  $j$ . A ship at port in location  $i$  incurs a per period waiting cost  $c_i^w$ , while a ship sailing from  $i$  to  $j$  incurs a per period sailing cost  $c_{ij}^s$ . The duration of a trip between region  $i$  and region  $j$  is stochastic: a traveling ship arrives at  $j$  in the current period with probability  $d_{ij}$ , so that the average duration of the trip is  $1/d_{ij}$ .<sup>24</sup>

Freights can only be delivered to their destination by ships and each ship can carry (at most) one freight. Following the search and matching literature, we model new matches every period,  $m_{it}$ , using a matching function, whereby the number of matches at time  $t$  in region  $i$  is

$$m_{it} = m_i(s_{it}, e_{it})$$

where  $s_{it}$  is the number of unmatched ships in region  $i$ .  $m_i(s_{it}, e_{it})$  is increasing in both arguments. Let  $\lambda_{it}$  denote the probability that an unmatched ship in location  $i$  meets an exporter;  $\lambda_{it} = m_{it}/s_{it}$ . Similarly, let  $\lambda_{it}^e$  denote the probability with which an unmatched exporter meets a ship;  $\lambda_{it}^e = m_{it}/e_{it}$ .<sup>25</sup> As discussed in Section 3.3, the matching function captures the implications of frictional trading, in a parsimonious fashion. In other words, we do not explicitly model the meeting technology between exporters and ships, which “would introduce intractable complexities”; instead, the matching function captures the several realities of the market discussed earlier, “without explicit reference to the source of the friction” (Petrongolo and Pissarides, 2001).

When a ship and an exporter meet, they either agree on a price to be paid by the exporter to the ship or they both revert to their outside options. The outside option of the exporter is to remain unmatched and wait for another ship, while the outside option of the ship is to either remain unmatched in the current region or to ballast elsewhere. The surplus of the match over the parties’ outside options is split via the price-setting mechanism. The price,  $\tau_{ijr}$ , paid to the ship delivering a freight of valuation  $r$  from region  $i$  to destination  $j$ , is determined by generalized Nash bargaining, with  $\gamma \in (0, 1)$  denoting the exporter’s

<sup>23</sup>We follow Kalouptsi (2014) and assume constant returns to scale so that a shipowner is a ship.

<sup>24</sup>It is straightforward to have deterministic trip durations instead. Our specification, however, preserves tractability and allows for some variability e.g. due to weather shocks, without affecting the steady state properties of the model.

<sup>25</sup>Note that in the frictionless case, if  $s_{it} > e_{it}$ ,  $\lambda_{it}^e = 1$  and  $\lambda_{it} = m_{it}/s_{it} = e_{it}/s_{it}$ .

bargaining power. The price is paid upfront and the ship commits to begin its voyage immediately to  $j$ .

Ships that remain unmatched decide whether to remain in their current region or ballast elsewhere subject to i.i.d. logit shocks. Exporters that remain unmatched survive with probability  $\delta > 0$  and wait in their current region.

## 4.2 Equilibrium

The state variable of a ship in region  $i$  includes its current location  $i$ , as well as the state  $(s_t, e_t)$ , where  $e_t = [e_{1t}, \dots, e_{It}]$  denotes the distribution of exporters over all regions, and  $s_t$  is an  $I \times I$  matrix including the ships that are traveling from  $i$  to  $j$ ,  $s_{ijt}$ , as well as the ships at port  $s_{it}$ ,  $i, j = 1, \dots, I$ . The state variable of an exporter in  $i$  includes his location  $i$ , valuation  $r$  and destination  $j$ , as well as  $(s_t, e_t)$ . In this paper, we consider the steady state of our industry model, following the tradition of Hopenhayn (1992). More specifically, agents view the spatial distribution of ships and exporters,  $(s_t, e_t)$ , as fixed and make decisions based on its steady-state value.

**Ships** Let  $V_{ij}$  denote the value of a ship that starts the period traveling from  $i$  to  $j$  (empty or loaded),  $V_i$  the value of a ship that starts the period at port in location  $i$ , and  $U_i$  the value of a ship that remained unmatched at  $i$  at the end of the period (we suppress the dependence on the steady state values  $(s, e)$ ). Then:

$$V_{ij} = -c_{ij}^s + d_{ij}\beta V_j + (1 - d_{ij})\beta V_{ij} \quad (2)$$

In words, the ship that is traveling from  $i$  to  $j$ , pays the per period cost of sailing  $c_{ij}^s$ ; with probability  $d_{ij}$  it arrives at its destination  $j$ , where it begins unmatched with value  $V_j$ ; with the remaining probability  $1 - d_{ij}$  the ship does not arrive and keeps traveling.

A ship that starts the period in region  $i$  obtains:

$$V_i = -c_i^w + \lambda_i E_{j,r} (\tau_{ijr} + V_{ij}) + (1 - \lambda_i)U_i \quad (3)$$

In words, the ship pays the per period port wait cost  $c_i^w$ ; it gets matched with probability  $\lambda_i$ , in which case it receives the agreed upon price,  $\tau_{ijr}$ , and begins traveling. The ship takes expectation over the type of exporter it meets, i.e. its revenue and destination. With the remaining probability,  $1 - \lambda_i$ , the ship does not find an exporter and gets the value of being unmatched  $U_i$ .

If the ship remains unmatched, it faces the choice of either staying at  $i$  or ballasting to another region;

in the latter case, the ship can choose among all possible destinations. The unmatched ship's value function is:

$$U_i(\epsilon) = \max \left\{ \beta V_i + \sigma \epsilon_i, \max_{j \neq i} V_{ij} + \sigma \epsilon_j \right\} \quad (4)$$

where the shocks  $\epsilon \in \mathbb{R}^I$  are drawn from a type I Extreme Value (Gumbel) distribution, with standard deviation  $\sigma$ . In words, if the ship stays in its current region  $i$ , it obtains value  $V_i$ ; otherwise the ship chooses another region and begins its trip there.

Let  $P_{ii}$  denote the probability that a ship in location  $i$  chooses to remain there, and  $P_{ij}$  the probability it chooses to ballast to  $j$ . We have:

$$P_{ii} = \frac{\exp(\beta V_i / \sigma)}{\exp(\beta V_i / \sigma) + \sum_{l \neq i} \exp(V_{il} / \sigma)} \quad (5)$$

and

$$P_{ij} = \frac{\exp(V_{ij} / \sigma)}{\exp(\beta V_i / \sigma) + \sum_{l \neq i} \exp(V_{il} / \sigma)}. \quad (6)$$

**Exporters** We now turn to the value functions of exporters; we begin with existing exporters and then consider exporter entry. An exporter that is matched in location  $i$  receives his revenue,  $r$  and pays the agreed price,  $\tau_{ijr}$  for a total payoff of  $r - \tau_{ijr}$ . The value of an exporter that remains unmatched,  $U_{ijr}^e$ , is therefore given by

$$U_{ijr}^e = \beta \delta \left[ \lambda_i^e (r - \tau_{ijr}) + (1 - \lambda_i^e) U_{ijr}^e \right] \quad (7)$$

In words, the exporter receives no payoff in the period and survives with probability  $\delta$ ; if so, the following period with probability  $\lambda_i^e$  he gets matched and receives  $r - \tau_{ijr}$ , while with the remaining probability  $1 - \lambda_i^e$  he remains unmatched again.

Each potential entrant, makes a discrete choice between destinations, as well as not exporting, also subject to i.i.d. shocks  $\epsilon^e \in \mathbb{R}^I$ , distributed according to a type I Extreme Value (Gumbel) distribution. Therefore, a potential entrant solves:

$$\max \left\{ \epsilon_0^e, \max_{j \neq i} \{ E_r U_{ijr}^e - \kappa_{ij} + \epsilon_j^e \} \right\}$$

where we denote by 0 the option of not exporting and normalize the payoff in that case to zero.

Potential exporters' behavior is given by the choice probabilities:

$$P_{ij}^e \equiv \frac{\exp(U_{ij}^e - \kappa_{ij})}{1 + \sum_{l \neq i} \exp(U_{il}^e - \kappa_{il})} \quad (8)$$

and

$$P_{i0}^e \equiv \frac{1}{1 + \sum_{l \neq i} \exp(U_{il}^e - \kappa_{il})} \quad (9)$$

where  $U_{ij}^e \equiv E_r U_{ijr}^e$ . Therefore, the number of entrant exporters in  $i$  equals  $\mathcal{E}_i (1 - P_{i0}^e)$ .

**Trade Costs (Shipping Prices)** As discussed above, the rents generated by a match between an exporter and a ship, are split via Nash bargaining. This implies the surplus sharing condition:

$$\gamma [(\tau_{ijr} + V_{ij}) - U_i] = (1 - \gamma) [(r - \tau_{ijr}) - U_{ijr}^e] \quad (10)$$

where  $U_i \equiv E_\epsilon U_i(\epsilon)$ . We use this condition to solve out for the equilibrium price  $\tau_{ijr}$ , in the following lemma:

**Lemma 1.** *The agreed upon price between a ship and an exporter with valuation  $r$  and destination  $j$  in location  $i$  is given by:*

$$\tau_{ijr} = (1 - \mu_i) (U_i - V_{ij}) + \mu_i r \quad (11)$$

where  $\mu_i = (1 - \gamma) (1 - \beta\delta) / (1 - \beta\delta (1 - \gamma\lambda_i^e))$ .

*Proof.* Substitute  $U_{ijr}^e$  in (10). □

In other words, the price is a convex combination of the exporter's revenue,  $r$ , and the difference between the ship's value of transporting the freight,  $V_{ij}$ , and its outside option,  $U_i$ . Consistent with the evidence in Table 2 exporters that have a higher value,  $r$ , pay higher prices. As discussed in Section 3.3, this is true because when there are search frictions, the law of one price no longer holds.

Crucially, the price depends on ships' equilibrium behavior through the value of traveling from  $i$  to  $j$ ,  $V_{ij}$  (which in turn depends on  $V_j$ ), as well as the value of the "outside option,"  $U_i$ . These objects are very rich, as they capture the attractiveness of both the origin  $i$ , as well as the destination  $j$ , which consists of numerous features. For instance, destinations that are unappealing to ships because there are few exporters relative to ships and the probability of ballasting afterwards is high, command higher prices (consistent with the evidence presented in Table 2). The same holds for destinations that are further away



(low  $d_{ij}$ ), have low value exporters or severe search frictions. Moreover,  $V_j$  controls for conditions at all possible ballast destinations from  $j$ , as well as for conditions at all possible export destinations from  $j$ , revealing the importance of network effects. Similarly,  $U_i$  controls for the attractiveness of the origin (e.g. exporter revenues, nearby ballast opportunities, matching probability). As a result, the price between  $i$  and  $j$  depends on all countries, rather than just  $i$  and  $j$ .

**Trade Flows** From equation (8) total flows from  $i$  to  $j$  equal

$$\mathcal{E}_i P_{ij}^e = \mathcal{E}_i \frac{\exp(U_{ij}^e - \kappa_{ij})}{1 + \sum_{l \neq i} \exp(U_{il}^e - \kappa_{il})} = \mathcal{E}_i \frac{\exp(\alpha_i (\bar{r}_{ij} - \tau_{ij}) - \kappa_{ij})}{1 + \sum_{l \neq i} \exp(\alpha_i (\bar{r}_{il} - \tau_{il}) - \kappa_{il})}$$

where  $\bar{r}_{ij}$  is the average revenue from exporting from  $i$  to  $j$ ,  $\tau_{ij} \equiv E_r \tau_{ijr}$  and  $\alpha_i = \beta \delta \lambda_i^e / (1 - \beta \delta (1 - \lambda_i^e))$ . To obtain this expression, we solve for  $U_{ijr}^e$  from (7) to obtain  $U_{ijr}^e = \alpha_i (r - \tau_{ijr})$ .

This equation is a “gravity equation”; it delivers the trade flow (in quantity rather than value) from  $i$  to  $j$  as a function of two components. First, the primitives  $\{\lambda_i^e, \bar{r}_{ij}, \kappa_{ij}, \mathcal{E}_i\}$  not just for  $i$  and  $j$  but for all regions; this is reminiscent of the analysis in Anderson and Van Wincoop (2003) who show that the gravity equation in a trade model needs to include a country’s overall trade disposition.

Second, it is a function of the endogenous trade costs,  $\tau_{ij}$ , for all  $j$ , which are the key addition here. In this model, trade costs introduce network effects between countries: indeed,  $\tau_{ij}$  depends on all locations both through the outside option of the ship at the origin  $i$ ,  $U_i$ , as well as the ballast and export opportunities from the destination  $j$ , captured by  $V_j$ . Overall, any change in the primitives affects trade flows both directly, but also indirectly through its impact on trade costs. We illustrate the importance of this mechanism in Section 7.

**Steady State Equilibrium** A steady state equilibrium,  $(s^*, e^*)$ , is a distribution of ships and exporters over locations, that satisfies the following conditions:

- (i) Ships optimal behavior,  $P_{ij}$  follows (5) and (6)
- (ii) Potential exporters behavior,  $P_{ij}^e$ , follows (8) and (9)
- (iii) Prices are determined by Nash bargaining, according to (11)
- (iv) Ships and exporters satisfy the steady state equations (established in the proof of Proposition 1 below):

$$s_i^* = \sum_j P_{ji} (s_j^* - m_j(s_j^*, e_j^*)) + \sum_{j \neq i} \frac{P_{ji}^e}{1 - P_{i0}^e} m_j(s_j^*, e_j^*) \quad (12)$$

$$e_i^* = \delta (e_i^* - m_i(s_i^*, e_i^*)) + \mathcal{E}_i (1 - P_{i0}^e) \quad (13)$$

$$s_{ij}^* = \frac{1}{d_{ij}} \left( P_{ij}(s_i^* - m_i(s_i^*, e_i^*)) + \frac{P_{ij}^e}{1 - P_{i0}^e} m_i(s_i^*, e_i^*) \right)$$

**Proposition 1.** *Suppose that the matching function is continuous,  $\epsilon$  and  $e^e$  have full support,  $\mathcal{E}_i$  and  $S$  are finite and  $e_i \leq \mathcal{E}_i/(1 - \delta)$ . Then, an equilibrium exists.*

*Proof.* See Appendix C. □

### 4.3 Discussion

We close this section with a discussion on several of our assumptions and some caveats.

We begin our discussion with the **matching process**. In our model the matching function is local, so that an exporter meets a ship only if they are in the same region, much like taxis and passengers. This is a modeling assumption, as there is no technological or other constraint that prevents an exporter from meeting and matching with a ship in another region. Nonetheless, there are economic disincentives that make distant matching unlikely, suggesting this is a reasonable approximation.

More specifically, practitioners explain that contracts tend to be signed with ships that are nearby, by arguing that “a ship is not a train” and it cannot promise exact arrival times far in advance due to weather conditions and port congestion. These delays are costly for exporters. Moreover, ships that are already in the region of the exporter have a distinct cost advantage over ships in other regions, since they do not need to incur the additional cost of sailing empty to the exporter’s region. Given that ships are in oversupply during our time period, exporters are not willing to pay (and wait) in order to contract with ships that are far away.<sup>26</sup> Reassuringly, the data support the local matching function assumption. For instance, about 20% of the contracts specify different signing and loading regions. Furthermore, as shown in Table 1, contracts are signed just 6 days on average prior to the loading date. In addition, the satellite data reveals that ships enter the region within 12 days of loading, which is well before the signing date.

Also related to the matching process, we assume that exporter valuations are sufficiently high so that in equilibrium, when a ship and an exporter meet, they always agree to form a match. It is easy to see that for every origin-destination pair, there exists a threshold of exporter value, below which the match surplus becomes negative and meetings do not result in matches. In this case, the price that a

---

<sup>26</sup>We formalize this using the model estimates produced in Section 6. In particular, we use our model to examine whether an exporter would benefit from searching in multiple regions. We find that when we allow an exporter to search in the “best” other region, in addition to the loading region, exporters always prefer to match with a ship in the loading region. The large majority of matches (>80%) still take place in the loading region. In addition, the price paid and the exporter value function are virtually unchanged.

ship demands to stop searching is too high for the low value exporters to pay and the match becomes unprofitable. Thus, the support of the distribution of revenues,  $F_{ij}^r$ , needs to be bounded below by this threshold. This assumption is reasonable in our context, since exporter revenue is an order of magnitude higher than transport prices. Using the model estimates produced in Section 6, we find that the surplus from matching is fairly high; on average it equals 371,270\$, and it remains high even if we focus on the lowest priced commodity, namely coal (the average surplus in this case is \$249,150, well above zero). This is not surprising: to get a sense of the magnitudes, the average value of a coal cargo is \$3,864,250, while its average shipping price is \$297,720.

Next, it is important to discuss how the model’s main outputs would change in the **absence of search frictions**; i.e. if the matching function were assumed to be  $m_i = \min\{s_i, e_i\}$ . In a world without search frictions and more ships than exporters, the shipping price is given by  $\tau_{ij} = U_i - V_{ij}$ , therefore the properties of endogenous trade costs (dependence on distance, origin, destination and entire network of countries) are independent of the presence of search frictions and *still hold*.<sup>27</sup> The paper’s main implications presented in Section 7 are thus not driven by search frictions, but rather by the endogeneity of trade costs.

We now turn to the **steady state** assumption of our industry model, which follows the tradition of Hopenhayn (1992) and a large body of other work in macro, IO, trade, etc. In particular, we consider the steady state of our dynamic system, where agents’ strategies and value functions depend only on their own state and the long-run average aggregate state, which is constant. This assumption renders the problem more tractable and allows us to derive simple(r) expressions for prices and trade flows.

It is worth connecting this assumption to the empirical exercise coming up. Naturally, the data exhibits variation over time. Strictly speaking, our empirical approach relies on the assumption that agents “play against the steady state”; i.e. their strategies rely on their own state and the long-run industry average.<sup>28</sup> That said, the steady state assumption is not unreasonable for the data at hand, which covers a period that is uniformly characterized by ship oversupply and relatively low demand for shipping services without any major shocks. Moreover, given the short-lived nature of the ships’ ballasting decisions it does not feel unreasonable that they would ignore aggregate long-run shocks when making these weekly choices; in addition, transition dynamics to a new steady state should be short given that a ship can travel to most ports in the world in a month.<sup>29</sup>

---

<sup>27</sup>Since ships are on the “long side” of the market, the price has to be such that ships are indifferent between loading and going to destination  $j$  and remaining unmatched, i.e. it must be that  $\tau_{ijr} + V_{ij} = U_i$ . In this case the price does not depend on the exporter’s valuation  $r$  (law of one price). Recall that this contrasts with the results of Table 2 where the shipping price is higher for higher revenue products.

<sup>28</sup>This is reminiscent of the interpretation of the Oblivious Equilibrium (OE) concept of Weintraub et al. (2008).

<sup>29</sup>To test whether ships respond to transitory shocks, we check if weather shocks affect ships’ ballast decisions and find

In this paper, we do not model **ship entry and exit**; exit is overall very small, while due to long construction lags in shipbuilding (two to six years), the fleet is fixed in the short run; see Kalouptsi (2014, 2018). Our counterfactual results therefore hold in the short/medium run.<sup>30,31</sup> We also do not model speed choice.

Finally, in the trade literature, trade costs include transport costs, tariffs and other barriers. Here we focus on microfounding and endogenizing transport costs. In our setup, other trade barriers are included in  $\kappa_{ij}$ , the cost of entry into exporting. In addition, we do not consider the determination of commodity prices; in other words, we take exporter valuations  $r$  to be exogenous. Determining this object in equilibrium within our setup is an interesting avenue for future research.

## 5 Empirical Strategy

In this section we lay out the empirical strategy followed to estimate the model of Section 4. Our empirical exercise consists of two distinct components: (i) estimation of the matching function and the searching exporters; (ii) estimation of ship travel and wait costs, exporter valuations and costs. We describe the empirical strategy for each component here, and present all results in Section 6.

### 5.1 Matching Function Estimation

A sizable literature has estimated matching functions in several different contexts (e.g. labor markets, marriage markets, taxicabs). For instance, in labor markets, one can use data on unemployed workers, job vacancies and matches to recover the underlying matching function. In the market for taxi rides, one observes taxis and their rides, but not hailing passengers; in recent work, Buchholz (2018), and Frechette et al. (2018) have used such data, coupled with a “parametric” assumption on the matching function to recover the hailing passengers.<sup>32</sup> Similar to the taxi market, we observe ships and matches, but not searching exporters. Here, we adopt a novel approach to simultaneously recover both exporters, as well as a nonparametric matching function.

---

that they do not.

<sup>30</sup>It is also worth noting that risk neutrality is a natural starting point for our analysis. In our setup, agents make risky decisions at a weekly frequency, so it may be reasonable to expect that over the course of the year risk washes out. Moreover, in our data we find that neither the variance of prices, nor of trip durations affect ships’ ballast decisions.

<sup>31</sup>It is straightforward to include a ship free entry condition in the model in order to consider longer-run counterfactuals. However in this case, we would need to take a stand on shipowners’ expectations of future demand.

<sup>32</sup>Buchholz (2018) assumes an “urn-ball” matching function. Frechette et al. (2018) construct a numerical simulation of taxi drivers that randomly meet passengers over a grid that resembles Manhattan; this spatial simulation essentially corresponds to the matching function, and can be inverted to recover hailing passengers.

Our approach extends the literature in two dimensions. First, we do not take a stance on the presence and magnitude of search frictions in the industry. Consider the case of no search frictions, so that the matching function is,  $m_{it} = \min(s_{it}, e_{it})$  and all potential matches are realized. In contrast, if there are search frictions, we have  $m_{it} = m_i(s_{it}, e_{it}) \leq \min(s_{it}, e_{it})$ , so that some potential matches are not realized. If one side of the market is unobserved (here  $e_{it}$ ) or mismeasured it is not straightforward to differentiate these two cases. The test presented in Section 3.3 serves to distinguish the two cases.

Second, we avoid imposing parametric restrictions on the matching function. The literature has imposed functional forms, most commonly the Cobb-Douglas; but parametric restrictions are directly linked to welfare and so can be overly restrictive. For instance, it has been shown in a wide class of labor market models, that the condition for constrained efficiency depends crucially on the elasticity of the matching function with respect to the search input (Hosios, 1990). In much of the matching function estimation literature this elasticity has been restricted to be constant.

Suppose we have a sample  $\{s_{it}, m_{it}\}_{t=0}^T$  for each market  $i$ . The unknowns of interest are the  $I$  matching functions  $m_i(\cdot)$  and the exporters  $e_{it}$ , all  $i, t$ ; henceforth, we suppress the  $i$  subscript to ease notation. Our approach relies on the literature on nonparametric identification (Matzkin, 2003) and nonseparable instrumental variable techniques (e.g. Imbens and Newey, 2009).

We assume that (i) the matching function  $m(s, e)$  is continuous and strictly increasing in  $e$ ; (ii) the matching function exhibits constant returns to scale (CRS), so that  $m(as, ae) = am(s, e)$  for all  $a > 0$  and there is a known point  $\{\bar{m}, \bar{s}, \bar{e}\}$ , such that  $\bar{m} = m(\bar{s}, \bar{e})$ ; (iii) the random variables  $s$  and  $e$  are independent. Assumption (i) is natural, as more exporters lead to more matches, all else equal. Assumption (ii) is a restriction that guarantees identification of both sets of unknowns and is discussed further below. Assumption (iii) is made for expositional purposes and is relaxed later on.

Let  $F_{m|s}$  denote the distribution of matches conditional on ships, and  $F_e$  the distribution of exporters,  $e$ . Then at a given point  $\{s_t, e_t, m_t\}$  we have:

$$\begin{aligned}
 F_{m|s=s_t}(m_t|s=s_t) &= \Pr(m(s, e) \leq m_t | s=s_t) \\
 \text{monotonicity} &= \Pr(e \leq m^{-1}(s, m_t) | s=s_t) \\
 \text{independence} &= \Pr(e \leq m^{-1}(s_t, m_t)) \\
 &= F_e(e_t)
 \end{aligned} \tag{14}$$

The equation  $F_e(e_t) = F_{m|s=s_t}(m_t|s = s_t)$  forms the basis for identification and estimation. Indeed note that this equation, along with the CRS assumption, allows us to recover the distribution  $F_e(e)$ , for all  $e$ : using the known point  $\{\bar{m}, \bar{s}, \bar{e}\}$  and letting  $a = e/\bar{e}$ , for all  $e$ ,

$$\begin{aligned} F_e(a\bar{e}) &= F_{m|s=a\bar{s}}(m(a\bar{s}, a\bar{e})|s = a\bar{s}) \\ &= F_{m|s=a\bar{s}}(a\bar{m}|s = a\bar{s}) \end{aligned} \tag{15}$$

We use (15) and vary  $a$  to trace out  $\hat{F}_e(e)$ , relying on a kernel density estimator for the conditional distribution  $\hat{F}_{m|s=a\bar{s}}(a\bar{m}|s = a\bar{s})$ .<sup>33</sup> Once the distribution  $\hat{F}_e$  is recovered, we obtain the number of exporters  $e_t$  from

$$e_t = \hat{F}_e^{-1}\left(\hat{F}_{m|s=s_t}(m_t|s = s_t)\right),$$

and the matching function at *any* point  $(s, e)$  from

$$m(s, e) = \hat{F}_{m|s}^{-1}(F_e(e)).$$

We choose the known point,  $\{\bar{m}, \bar{s}, \bar{e}\}$ , to be of the form  $1 = m(\bar{s}, 1)$ , so that one exporter is always matched when there are  $\bar{s}$  ships. We set  $\bar{s}$  iteratively, to be the lowest value such that  $m_t \leq e_t$ , for all  $t$ , thus obtaining a conservative bound on search frictions.

The intuition behind the identification argument is as follows: the observed correlation between  $s$  and  $m$  informs us on  $\partial m(s, e)/\partial s$ , since the sensitivity of matches to changes in ships is observed and  $s$  is independent of  $e$  by assumption; then, due to homogeneity, this derivative also delivers the derivative  $\partial m(s, e)/\partial e$ ; and once these derivatives are known, integration leads to the matching function, which can be inverted to provide the number of exporters.

The CRS assumption is a reasonable starting point. In the labor literature, the majority of matching function estimates find support for constant returns to scale; as Petrongolo and Pissarides (2001) point out “divergences from constant returns are only mild and rare”. Nonetheless, to explore the robustness of our findings, in Section 6.1 we consider an alternative approach that does not make an assumption on the returns to scale, but instead relies on a restriction on the distribution  $F_e$  (Poisson).

<sup>33</sup>For instance, one can use a simple frequency estimator:

$$F_{m|s=s_t}(m_t|s = s_t) = \Pr(m \leq m_t|s = s_t) = \frac{\Pr(m \leq m_t, s = s_t)}{\Pr(s = s_t)} = \frac{\#1\{(m \leq m_t, s = s_t)\}}{\#1\{(s = s_t)\}}$$

where  $1\{\cdot\}$  denotes the indicator function and  $\#$  denotes the number of times. In practice, we use a Gaussian kernel density estimator.

Finally, as mentioned above, independence of ships and exporters is not a natural assumption in our setting. To relax it, we employ the literature on nonlinear IV techniques (e.g. Imbens and Newey, 2009, while for an application similar to ours see Bajari and Benkard, 2005). In particular, assume that an instrument  $z$  exists such that  $s = h(z, \eta)$ , with  $z$  independent of  $e, \eta$ . Under this formulation the endogeneity is driven by the correlation between  $\eta$  and  $e$  and, therefore,  $s$  is independent of  $e$ , conditional on  $\eta$ .

The approach now has two steps. In the first step, we recover  $\eta$  using the relationship  $s = h(z, \eta)$ ; in practice we regress flexibly the number of ships  $s$  on the instrument,  $z$ , and set  $\eta$  equal to the residual. In the second step, we employ (14) conditioning on both  $s$  (as before) and  $\eta$ :

$$F_{m|s=s_t, \eta} (m_t | s = s_t, \eta) = F_{e|\eta} (e_t | \eta)$$

Similarly to above we recover the unknowns of interest  $e$  and  $m(\cdot)$ , by integrating both sides over  $\eta$ .

In this case,  $z$  consists of unpredictable sea weather shocks that shift the arrival of ships at a port without affecting the number of exporters; this is the same instrument used in our frictions test presented in Section 3.3.

## 5.2 Ship Costs and Exporter Revenue

We now turn to the ship cost parameters,  $\{c_{ij}^s, c_i^w, \sigma\}$ , for all  $i, j$ , as well as the exporter revenues  $r \sim F_{ij}^r$  and production and export costs  $\kappa_{ij}$ , for all  $i, j$ . The estimation amounts to essentially matching the observed ballast decisions of ships to the model's predicted choice probabilities  $P_{ij}$ ; the observed prices  $\tau_{ijr}$ , to the equilibrium Nash-bargained prices; and the observed trade flows (loaded trips) to the model's predicted equilibrium flows  $P_{ij}^e$ . Intuitively, ships' observed choices, conditional on observed prices, deliver the ship costs. Then, in equilibrium, prices inform us on the exporter's revenues. Finally, potential exporters make their entry decisions taking into account the (expected) prices they will face. From these decisions, we are able to back out the production and export cost.

**Ship Costs** Consider first the ship sailing costs,  $c_{ij}^s$ , wait costs,  $c_i^w$ , and the standard deviation of the logit shocks,  $\sigma$ . These parameters determine ships' optimal ballast choice probabilities, (5) and (6), given prices, through the value functions  $V_{ij}$ . Thus, we can estimate them via Maximum Likelihood using the observed ship choices. In particular, we use a nested fixed point algorithm to solve for the ship value functions at every guess of the parameter values, compute the predicted choice probabilities and then

calculate the likelihood, as in Rust (1987). We provide the details of the approach in Appendix D, where we also prove that our value functions are well-defined using a contraction argument.

As is always the case in dynamic discrete choice, not all parameters are identified and some restriction needs to be imposed. Here, we have  $I^2 + 1$  parameters and  $I^2 - I$  choice probabilities, so we require  $I + 1$  restrictions; we show this formally, borrowing from the analysis of Kalouptsi et al. (2016) in Appendix E. The additional restrictions amount to using observed fuel prices to determine  $c_{ij}^s$  for some  $i, j$ ; see Section 6.2. Note also that the observed prices pin down the scale of payoffs (in dollars) and allow the identification of  $\sigma$ .<sup>34</sup>

**Exporter revenues and entry costs** We next use data on shipping prices to estimate exporter revenues. Consider the equilibrium price (11) solved with respect to the exporter’s revenue  $r$ :

$$r = \frac{1 - \beta\delta(1 - \gamma\lambda_i^e)}{(1 - \gamma)(1 - \beta\delta)} \tau_{ijr} - \frac{\gamma(1 - \beta\delta(1 - \lambda_i^e))}{(1 - \gamma)(1 - \beta\delta)} (U_i - V_{ij}) \quad (16)$$

Note that the only unknowns in this expression are the revenue  $r$  and the bargaining coefficient  $\gamma$ , and that we have as many equations as the observed shipping prices. Indeed,  $\lambda_i^e$  is known from the previous step that estimates the matching function and searching exporters ( $\lambda_i^e$  is simply the average ratio of matches to exporters),  $\beta$  and  $\delta$  are calibrated, while  $U_i$  and  $V_{ij}$  are known once  $\{c_{ij}^s, c_i^w, \sigma\}$  have been estimated. Note that the valuation  $r$  is different for each matched exporter, while the bargaining coefficient parameter is a scalar, that is identical for all markets.

We first estimate the bargaining weight,  $\gamma$  by averaging equation (16) across regions and products; this average relationship links the average shipping price, to the average exporter revenue ( $\bar{r} \equiv \sum_{ij} \bar{r}_{ij}/I^2$ ). We calibrate the average exporter revenue across regions,  $\bar{r}$ , to be equal to the average value of total trade in commodities (obtained from external trade data from Index Mundi) and solve (16) for  $\gamma$ .<sup>35</sup>

Given this estimate for the bargaining weight,  $\gamma$ , we recover each exporter valuation  $r$  point-wise from

---

<sup>34</sup>Unlike commonly used discrete choice models where only choices are observed and both the scale and level of utility need to be normalized, here we also observe a component of payoffs (prices in dollars) which allows us to relax the restriction on the scale of utility and identify  $\sigma$ .

<sup>35</sup>In particular, we solve (16) with respect to  $\gamma$  and average over  $i, j, r$  so that:

$$\gamma = \frac{(1 - \beta\delta)(\bar{r} - \bar{\tau})}{\beta\delta E_{ijr} \lambda_i^e \tau_{ijr} + (1 - \beta\delta)\bar{r} - E_{ij} (1 - \beta\delta(1 - \lambda_i^e)) (U_i - V_{ij})}$$

where  $\bar{\tau}$  is the average observed price. To obtain  $\bar{r}$ , we first collect the average price of the five most common commodities (ore, coal, grain, steel, and fertilizers) from Index Mundi, and multiply it by the average tonnage carried by a bulk carrier (this is equal to the average vessel size times its utilization rate which equals about 65%). We then set  $\bar{r}$  as their weighted average based on each commodity’s frequency in shipping contracts; we find  $\bar{r}$  to equal 7 million US dollars. This approach of estimating the Nash bargaining weight is inspired by the empirical literature on oligopoly bargaining (e.g. Crawford and Yurukoglu, 2012, Ho and Lee, 2017).



(16) and obtain their distribution,  $F_{ij}^r$ , nonparametrically. Note that valuations are drawn from an origin-destination specific distribution, which allows for arbitrary correlation between a cargo’s valuation and destination.

We finally estimate the exporter entry costs,  $\kappa_{ij}$ , which capture both the cost of production, as well as any export costs beyond shipping prices. These are estimated from the exporters’ entry decisions, given by the choice probabilities,  $P_{ij}^e$ , defined in (8) and (9). Indeed, we recover  $\kappa_{ij}$  from the following equation (Berry, 1994):

$$\ln P_{ij}^e - \ln P_{i0}^e = U_{ij}^e - \kappa_{ij}. \quad (17)$$

The only unknown in this equation is  $\kappa_{ij}$ . Recall that  $U_{ij}^e = \alpha_i(\bar{r}_{ij} - \tau_{ij})$  and that both  $\alpha_i = \beta\delta\lambda_i^e/(1 - \beta\delta(1 - \lambda_i^e))$  and the mean revenue  $\bar{r}_{ij}$  have been estimated, while  $\tau_{ij}$  is observed. Turning to the left hand side of (17), note that from the satellite data we have information on the proportions of loaded trips from  $i$  to  $j$ , i.e.  $P_{ij}^e/(1 - P_{i0}^e)$ . However, to obtain  $\kappa_{ij}$  we need both  $P_{ij}^e$ , as well as the fraction of potential exporters who choose *not* to export,  $P_{i0}^e$ , or equivalently, the number of potential exporters  $\mathcal{E}_i$ . We set  $\mathcal{E}_i$  equal to the total production of the relevant commodities for each region  $i$ , thus assuming that a region’s total production serves as an upper bound to the region’s exports.<sup>36</sup>

## 6 Results

In this section we present the results from our empirical analysis. We calibrate the discount factor to  $\beta = 0.995$  and the exporter survival rate to  $\delta = 0.99$ . In our baseline estimation we ignore the different ship sizes, but our estimation results are similar when we consider Panamax alone or Handymax alone (these are the categories with sufficient data). Moreover, results are robust when we estimate the model separately by season.

---

<sup>36</sup>More precisely, let  $n_{it}$  denote the number of entrant exporters in period  $t$  and region  $i$ . Then, the number of exporters transitions as follows (see equation (18) in Appendix C):

$$e_{it+1} = \delta(e_{it} - m_{it}) + n_{it}$$

so that in steady state

$$n_i = (1 - \delta)e_i + \delta m_i$$

or

$$n_i = \mathcal{E}_i(1 - P_{i0}^e) = (1 - \delta)e_i + \delta m_i$$

If  $\mathcal{E}_i$  is known, we can solve this equation for  $P_{i0}^e$  since the right-hand-side is known. To determine  $\mathcal{E}_i$ , we collect annual country-level production data for grain (FAO), coal (EIA), iron ore (US Geological Survey), fertilizer (FAO) and steel (World Steel Association). To transform the production tons into a number of potential freights (i.e. shipments that fit in our bulk vessels), we first scale the production to adjust for the coverage of our data (we observe about half of the total fleet) and then divide by the average “active” ship size, taking into account a ship’s utilization rate and the fact that a ship operates on average 340 days per year (due to maintenance, repairs, etc)

## 6.1 Matching Function

We now present the estimates for the exporters and the matching function, obtained as described in Section 5.1.<sup>37</sup> The matching function is estimated separately for each region  $i$ .

First, we present the results from the first stage regression of the number of ships on the unpredictable component of weather in surrounding seas for all regions. The results, shown in Table 5, indicate that ocean wind has a significant impact. This suggests that weather indeed affects trip duration and therefore weather shocks exogenously shift the supply of ships at port.<sup>38</sup>

	Joint Significance	F-stat
North America West Coast	0	5.82
North America East Coast	0.03	2.22
Central America	0	4.78
South America West Coast	0.01	3.62
South America East Coast	0	5.78
West Africa	0	4.77
Mediterranean	0	4.85
Baltic States	0.03	2.4
South Africa	0.03	2.54
Middle East	0	3.09
India	0	3.48
South East Asia	0	7.87
China	0	5.66
Australia	0	4.54
Japan-Korea	0	4.85

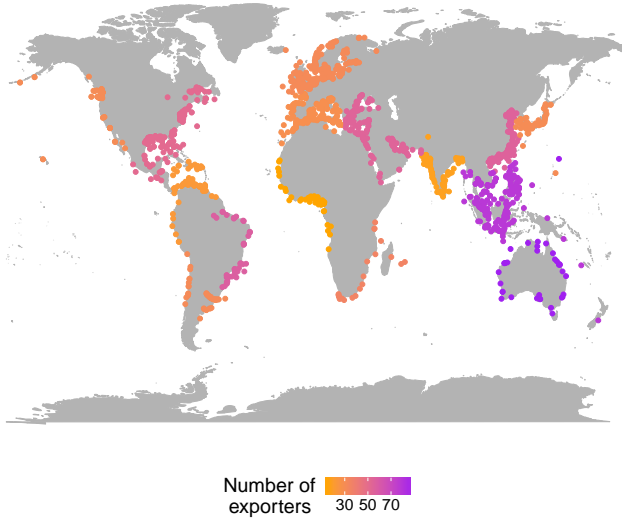
**Table 5:** First Stage, Matching Function Estimation. Regressions of the number of ships in each region on the unpredictable component of weather conditions in the surrounding seas. The first column reports the joint significance of the instruments and the second column the F-statistic. For the construction of the instrument, see Table 4.

Figure 3 presents the weekly average number of exporters in each region. Not surprisingly perhaps, exporters are concentrated in Australia, the East Coast of North and South America and Southeast Asia, which are all rich in raw materials. India, Africa and Central America have the fewest exporters.

The matching function has reasonable properties. Exporters have substantially higher chances of finding a match than ships, consistent with our sample period of high ship supply and low demand. The

<sup>37</sup>The results are fairly robust to the choice of the bandwidth parameters, though some sensitivity is to be expected. In addition, the data exhibits sufficient variation over time to perform the nonparametric analysis.

<sup>38</sup>This is consistent with contracting practices, where the ship promises a range of fuel consumption based on different weather conditions. Moreover, regressing the log of trip duration on wind speed, both upon arrival at the destination and upon departure at the origin confirms that the weather strongly affects trip duration; indeed, trips are 11% longer on average under bad weather conditions.



**Figure 3:** Average weekly number of estimated exporters.

matching rate for ships (exporters) declines as the market gets crowded with ships (exporters). Moreover, we find that regional heterogeneity is important: imposing a single matching function across regions leads to severely biased estimated exporters and matching function elasticities. To illustrate the importance of the nonparametric approach, we compare our estimated matching function to a typical parametric specification: the Cobb-Douglas. This specification, which is common in the literature, imposes that the elasticities of the matching function are constant, a property that can be restrictive. Figure 11 in Appendix B plots our estimates for the elasticity of matches with respect to the number of ships at port,  $\epsilon_s(e, s) = \frac{\partial m_i(e, s)}{\partial s} \frac{s}{m}$ , across regions. Our nonparametric matching function strongly rejects that these elasticities are constant: on average the range of these elasticities is 33% of the elasticity at the average level of ships and exporters in each region. In addition, we estimate a Cobb-Douglas specification for the matching function and find that the exporters recovered under this, more restrictive, assumption are different both in magnitude (the average absolute error is 63%) and in the relative ranking of regions, compared to the nonparametric case.<sup>39</sup>

To measure the extent of search frictions in different regions, we compute the average percentage of weekly “unrealized” matches; i.e.  $(\min\{s_i, e_i\} - m_i) / \min\{s_i, e_i\}$ . The results are plotted in the left panel of Figure 4 along with confidence intervals constructed from 200 bootstrap samples, and reveal that

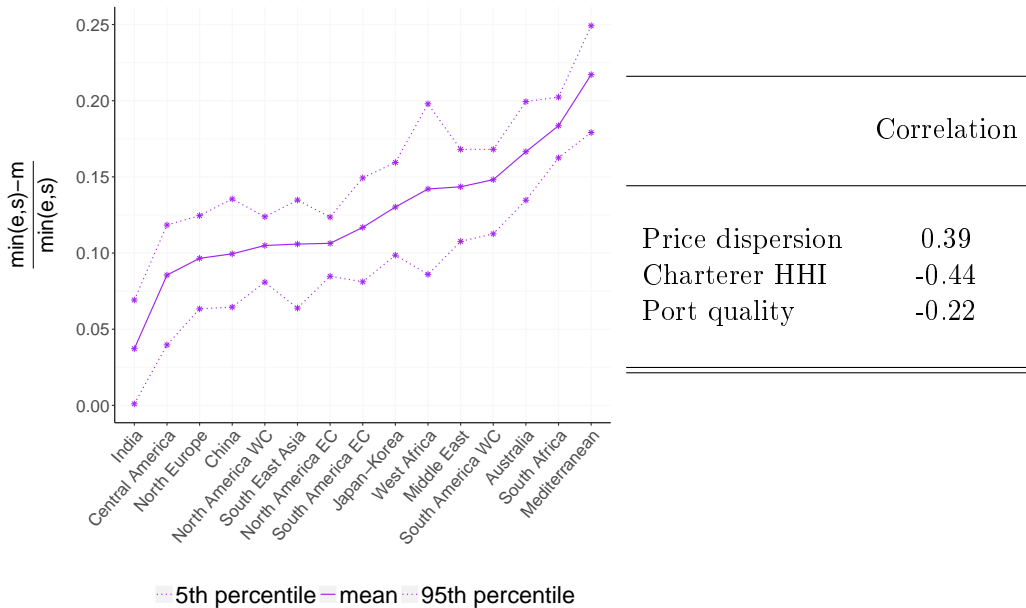
<sup>39</sup>The Cobb-Douglas specification,  $m_{it} = A_i s_{it}^{\alpha_i} e_{it}^{1-\alpha_i}$ , is not straightforward to estimate, as  $e_{it}$  is not observed. We rewrite it as,

$$\log(m_{it}) = \log(A_i) + (1 - \alpha_i) \log e_{it} + \alpha_i \log s_{it} = \alpha_{i0} + \epsilon_{it} + \alpha_i \log s_{it}$$

and we recover  $\alpha_i$  using an instrument (the unpredictable component of weather) for  $s_{it}$ .

search frictions are heterogeneous over space and may be sizable, with up to 20% of potential matches “unrealized” weekly in regions like South Africa and parts of South America and Europe. On average, 12.7% of potential matches are “unrealized”.<sup>40</sup> We can also reject that the matching function equals the minimum for all regions.

Moreover, as shown in the right panel of Figure 4, the estimated search frictions are positively correlated with the observed within-region price dispersion, another indicator of search frictions. We also find that the frictions are negatively correlated with the Herfindahl Index of charterers in a region; this suggests that when the clientele is disperse, frictions are higher. Moreover, they are negatively correlated with the index of quality of port infrastructure issued by Index Mundi, so that regions whose ports have a low quality score, suffer from higher matching frictions. Finally, when we estimate the matching function separately for Capesize (biggest size) and Handysize (smallest size) vessels, we find that for Capesize, where the market is thinner, the ratios of “unrealized” matches are lower.



**Figure 4:** Search Frictions. The left panel shows the average weekly share of unrealized matches, with confidence intervals from 200 bootstrap samples. The right panel displays the correlation between the average weekly share of unrealized matches and several variables of interest. Price dispersion and the Herfindahl Index of charterers are calculated from the Clarksons contract data. Price dispersion is computed from the residuals of a regression of prices on ship characteristics, origin, destination and time fixed effects. Port quality is obtained from the index of quality of port infrastructure issued by Index Mundi.

We close this section by discussing the robustness of our results to the CRS assumption. To do so, instead of assuming CRS, we impose a parametric assumption on the distribution of exporters,  $F_e$ . In

<sup>40</sup>It is worth noting that this does not imply that in the absence of search frictions we would have 12.7% more matches, as we would need to take into account the optimal response of ships and exporters. This is simply a measure of the severity of search frictions in different regions.

particular, we assume that exporters are distributed Poisson, as we can then interpret the number of freights  $e_t$  as the number of arrivals of exporters at port every week. We estimate the number of exporters and the matching function following an analogous strategy, based on Matzkin (2003).<sup>41</sup> The results, shown in Figure 12 in Appendix B, are robust to this alternative restriction. Moreover, we find that the implied degree of homogeneity of the matching function under the Poisson distributed exporters is roughly equal to one, even though we do not assume CRS.

## 6.2 Ship Costs and Exporter Valuations

In our baseline specification, we construct seven groups for the sailing cost  $c_{ij}^s$ , roughly based on the continent and coast of the origin; and we estimate port wait costs  $c_i^w$ , for all  $i$ .<sup>42</sup> Note that  $c_{ij}^s$  is the *per week* sailing cost from  $i$  to  $j$  and its major component is the cost of fuel. Following the discussion on identification in Section 5.2, we set this cost for one of the groups (for trips originating from the East Coast of North and South America) equal to the average weekly fuel price (40,000 US dollars). Moreover, since the fuel cost is paid by the exporter when the ship is loaded, we add it to the observed prices.

The first two columns of Table 7 report the results.<sup>43</sup> Sailing costs are fairly homogeneous. Port wait costs are more heterogeneous and large, ranging between 90,000 and 290,000 US dollars per week. Consistent with industry narratives, waiting at port is costly, both due to direct port and security fees, as well as the rapid depreciation of the ship's machinery and electronics and antifouling costs caused by the accumulation of microorganisms during immobility. Ports in the Americas are the most expensive, while ports in China, India, and Southeast Asia are the cheapest. The standard deviation of the logit shocks,  $\sigma$ , is estimated at about 16,000 US dollars, roughly 5% of the average trip price including the fuel cost payment. This suggests that the logit shocks do not account for a large part of utility or ballast decisions. As shown in Figure 13 in the Appendix, the model's fit is very good, as our predicted choice probabilities are very close to the observed ones.<sup>44</sup>

<sup>41</sup>Recall our basic equation,  $F_{m|s}(m|s) = F_e(e)$ . If  $F_e(e)$  is Poisson with parameter  $\rho$ , we have:

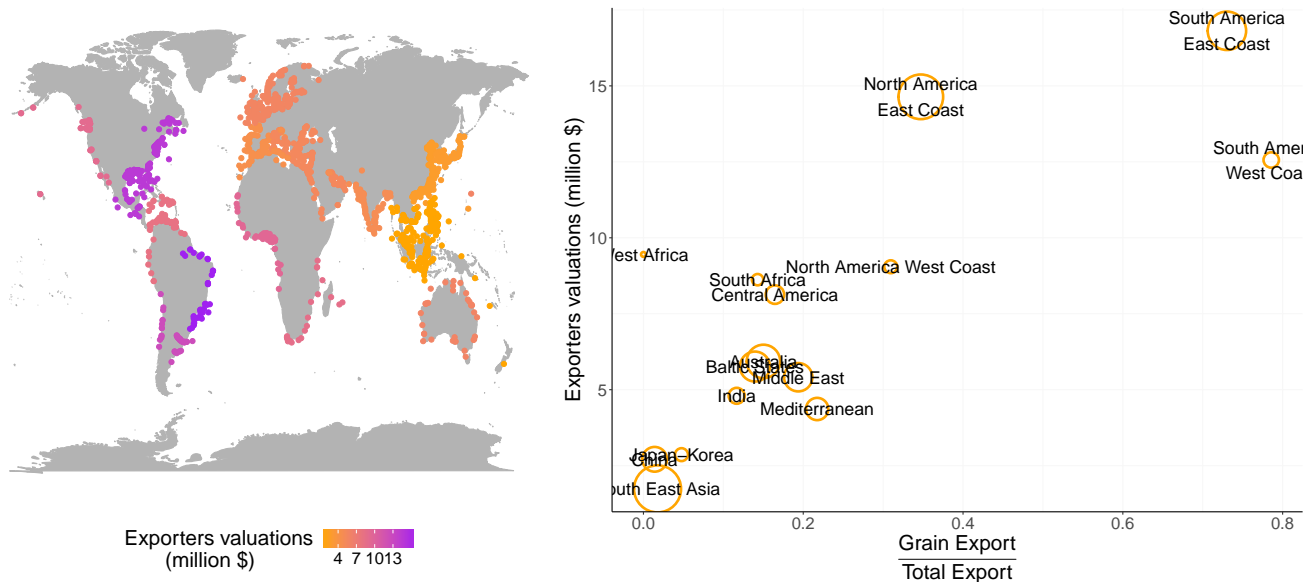
$$F_{m|s}(m|s) = F_e(e) = \exp(-\rho) \sum_{k=1}^e \frac{\rho^k}{k!}$$

If  $\rho$  were known, we could solve this equation for  $e_{it}$ , all  $i, t$ , since the right-hand-side is known. We determine  $\rho$  iteratively by requiring again that the inequalities  $m_{it} \leq e_{it}$  always hold.

<sup>42</sup>The seven groups are: (i) Central America, West Coast Americas; (ii) East Coast Americas; (iii) West and South Africa; (iv) Mediterranean, Middle East and North Europe; (v) India; (vi) Australia and Southeast Asia; (vii) China, Japan and Korea.

<sup>43</sup>The standard errors are computed from 200 bootstrap samples with the resampling done at the ship level. We combine these bootstrap samples with those of the matching function to incorporate the error from the matching function estimation.

<sup>44</sup>As our data comes from a period of historically low shipping prices, our estimated value functions are negative. This is partly due to the fact that we are not modeling ships' expectations, so the value function does not take into that under



**Figure 5:** Exporter valuations. The left panel of the figure plots the estimates for the average exporters’ valuation across regions. The right panel correlates the average exporters’ valuation with the share of exports in grain (source, Comtrade). The size of the circle proxies the number of observations.

In the left panel of Figure 5 we plot the average exporter revenues across origins, while the third column of Table 7 reports the estimates. We estimate the bargaining coefficient to equal  $\gamma = 0.3$ . There is substantial heterogeneity in exporter revenues across space. South and North America have the highest revenues, while China, Japan, and Southeast Asia have the lowest. This ranking is reasonable, as for instance, Brazil exports grain which is expensive, whereas Southeast Asia exports mostly coal, which is one of the cheapest commodities. We generalize this example by focusing on grain, the most expensive frequently shipped commodity. In particular, using data from Comtrade, we explore whether regions that have a high share of grain exports tend to have higher estimated revenues. The results, shown in the right panel of Figure 5, reveal that indeed there is a positive correlation between the two. Of course, there may be other factors determining the valuation of an exporter such as inventory costs, just in time production, etc. On average, the price  $\tau_{ij}$  is equal to about 5% of the mean valuation  $\bar{r}_{ij}$ , consistent with other estimates in the literature (e.g. UNCTAD, 2015, Hummels et al., 2009). Finally, the estimated exporter costs,  $\kappa_{ij}$ , exhibit substantial heterogeneity across destinations from a given origin, as well as across origins. On average  $\kappa_{ij}$  is the same order of magnitude as the average valuation  $\bar{r}_{ij}$ , reminiscent of a free entry condition into exporting. Moreover, we find that exporter costs are lower between an origin  $i$  and a destination  $j$  if the same language is spoken at  $i$  and  $j$ , which is reasonable since  $\kappa_{ij}$  includes both

---

mean-reverting demand for seaborne trade prices will eventually go up (see Kalouptsi, 2014). If we compute the equilibrium under higher exporter revenues that lead to prices closer to the ones observed before 2010, the ship value function indeed becomes positive.

production costs, as well as other exporting costs, as discussed in Section 4.3.

## 7 The Role of Endogenous Trade Costs

In this section, we illustrate the importance of endogenous trade costs. In particular, we demonstrate that the transportation sector (1) implies that net exporters (importers) face higher (lower) trade costs leading to misallocation of productive activities across countries; (2) creates network effects in trade costs; and (3) dampens the impact of shocks on trade flows. We illustrate these mechanisms, by studying how our setup compares to one with exogenous “iceberg” trade costs, the impact of a fuel cost shock and the spatial propagation of a macro shock (a Chinese slow-down). These mechanisms also shape the policy analysis considered in the next section.

### Exogenous Trade Costs

We begin by showcasing how endogenous trade costs lead to misallocation (mechanism 1) and network effects (mechanism 2) by comparing our setup to one with exogenous trade costs. Typical trade models assume that trade costs are “iceberg”, so that a percentage of the traded good’s value is lost in transportation; this percentage is often a function of the distance between the origin and destination (Samuelson, 1954). In our setup, this amounts to assuming that the price paid for transportation from origin  $i$  to destination  $j$  is a function of only the distance,  $1/d_{ij}$ , and the exporter valuation  $\bar{r}_{ij}$ . We thus construct a measure of exogenous trade costs by flexibly regressing the observed shipping prices between  $i$  and  $j$  on distance and  $\bar{r}_{ij}$ . This ensures that the overall level of shipping prices is the same. We then consider the trade patterns under these (counterfactual) trade costs, using the exporters’ choices, (8) and (9).

Strikingly, we find that in this world of exogenous iceberg trade costs, trade is substantially less balanced: trade imbalances are 24% higher on average. Indeed, net exporters experience a 9.4% increase in their exports, (and up to 24% for Australia), while net importers experience a 4.7% decline in their exports (and up to 22% for India). In other words, the transportation sector acts as a smoothing factor for world trade.

The core mechanism is related to the ship’s equilibrium behavior and in particular the strength of its bargaining position. Consider a ship that is located near a large net exporter, such as Brazil. As loading chances are high, the ship’s bargaining position upon meeting an exporter (i.e. its outside option) is strong and the ship can extract a high price, which in turn, tends to restrain Brazilian exports compared

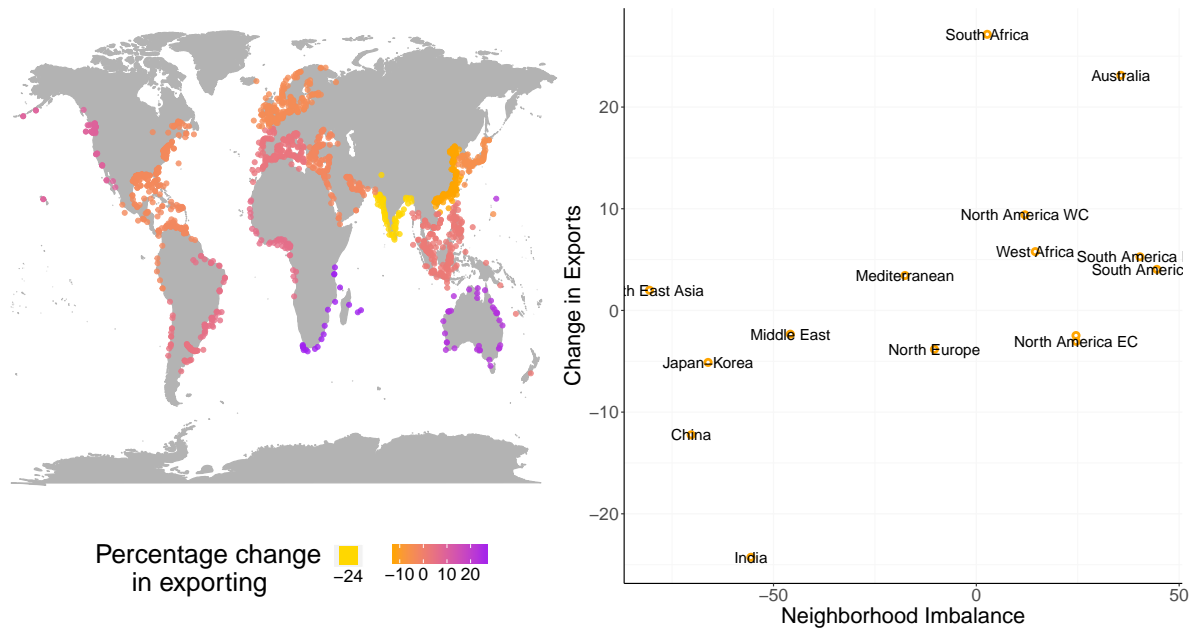
	Port Costs $c^w$	Sailing Costs $c^s$	Exporters Valuations $\bar{r}_i$	Preference Shock $\sigma$
North America West Coast	227.65 (8.77)	46.75 (0.36)	9,047.58 (497.1)	
North America East Coast	272.3 (4.31)	- -	14,639.72 (261.11)	
Central America	175.41 (5.06)	46.75 (0.36)	8,129.37 (400.22)	
South America West Coast	265.55 (7.77)	46.75 (0.36)	12,561.15 (425.65)	
South America East Coast	292.5 (5.23)	- -	16,813.61 (455.06)	
West Africa	145.3 (4.84)	47.65 (0.33)	9,452.09 (659.51)	
Mediterranean	121.89 (3)	46.16 (0.28)	4,364.56 (269.17)	
Baltic States	122.48 (1.71)	46.16 (0.28)	5,761.4 (202.04)	
South Africa	220.11 (7.28)	47.65 (0.33)	8,621.94 (420.55)	
Middle East	118.45 (2.14)	46.16 (0.28)	5,409.67 (252.23)	
India	97.23 (1.8)	45.93 (0.28)	4,800.29 (366.23)	
South East Asia	93.14 (1.02)	40.99 (0.28)	1,734.75 (81.99)	
China	91.07 (0.98)	40.89 (0.25)	2,708.65 (160.27)	
Australia	193.29 (2.85)	40.99 (0.28)	5,929.6 (160.19)	
Japan-Korea	100.41 (1.9)	40.89 (0.25)	2,863.02 (214.54)	
				16.53 (0.1070)

**Table 7:** Ship costs and exporter valuation estimates. All the estimates are in 1,000 USD. Standard errors computed from 200 bootstrap samples. The sailing cost for the East Coast of North and South America is set equal to the weekly fuel cost at 40,000 US dollars.

to a world with exogenous trade costs.<sup>45</sup> In contrast, consider a ship located near a net importer, such as India. The ship is unlikely to reload there. This lack of options puts the ship in a weak bargaining position and forces it to accept lower prices, which in turn increases Indian exports compared to the case

<sup>45</sup>In addition, the likely destinations are net importers where ships' value are low. This further increases prices faced by next exporters like Brazil and dampens exports and trade imbalances.





**Figure 6:** Exogenous Trade Costs: the left panel presents the change in exports. The right panel correlates the change in exports under exogenous trade costs and a region’s centrality, defined as the distance-weighted sum of the imbalance of all other regions, i.e.  $\sum_j (1/d_{ij}) (\text{exports}_j - \text{imports}_j)$ .

of exogenous trade costs. The left panel of Figure 6 presents the change in each region’s exporting under exogenous trade costs and reveals that indeed net exporters experience a disproportionately large increase in their exports.

This implies that in a world of endogenous trade costs, there is a misallocation of productive activities, which are reallocated from (efficient) net exporters to (inefficient) net importers. In our setup this corresponds to production decreasing in high  $\bar{r}$ /low  $\kappa$  countries which face higher trade costs all else equal and increasing in low  $\bar{r}$ /high  $\kappa$  ones, which face lower trade costs. Indeed, total net value of trade increases by 10% under exogenous trade costs. In addition, we find that this argument extends to a region’s neighborhood; this is because trade costs depend on the entire network of neighboring countries. Indeed, a net exporter close to other net exporters offers even more options to ships and prices are even higher, which inhibits the neighborhood’s exports. Hence, a country’s own trade imbalance, as well as the imbalance of its neighborhood, are crucial factors determining its trade disposition. To demonstrate these neighborhood effects, we consider a centrality measure for each region that consists of the weighted sum of trade imbalance in all regions, where the weights are given by the distance (i.e.  $\sum_j (1/d_{ij}) (\text{exports}_j - \text{imports}_j)$ ). The right panel of Figure 6 correlates the change in exports to this centrality measure and shows that the overall imbalance of a neighborhood matters: regions whose neighborhood is overall a net-exporting one, offer high outside options to ships, which pushes prices up and thus exports down (and vice versa).

## A Shock to Fuel Costs

Here we explore how world trade reacts to a fuel cost shock. This shock directly affects the main (variable) cost of transportation,  $c^s$ , and as a result also changes ship optimal behavior. This exercise illustrates how the transport sector dampens the impact of the shock (mechanism 3), while at the same time reallocating production across countries (mechanism 1).<sup>46,47</sup>

Consider a 10% decrease in  $c^s$ . The left panel of Figure 7 presents the resulting change in world exports, while the right panel presents the impact of the shock on several outcomes of interest. A decline in  $c^s$  has a direct and an indirect effect. The direct effect is straightforward: as costs fall, shipping prices also fall and thus exports rise.<sup>48</sup> Indeed, we see that exports increase everywhere, on average by 4.4%. The indirect effect is that the decline in sailing costs lowers the cost of ballasting. This implies that the ships' outside option,  $U$ , is now higher, which leads, all else equal, to an increase in prices that dampens the direct effect (mechanism 3). This is intuitive: reduced sailing costs imply that ships are less "tied" to their current region (and indeed ballast miles increase by 17%), and, as their bargaining position is stronger, they negotiate higher prices. The dampening is substantial: as shown in the second column in the right panel of Figure 7, if ships were not allowed to optimally adjust their behavior, the increase in trade would have been 41% higher.

In addition, as fuel costs decline and the importance of the transport sector is reduced, misallocation declines. Indeed, as shown in the right panel of Figure 7, net exporters experience an increase in exports of about 5%, while net importers experience an increase of about 3% (mechanism 1). This is largely driven by the indirect effect: ships ending a trip at net importing regions, are now less likely to wait there given the lower sailing cost; their outside option is higher and they can command higher prices that further reduce exports from these regions. Indeed, if ships were not allowed to optimally adjust their behavior, the increase in the exports of net importing regions would have been more than twice as high.

---

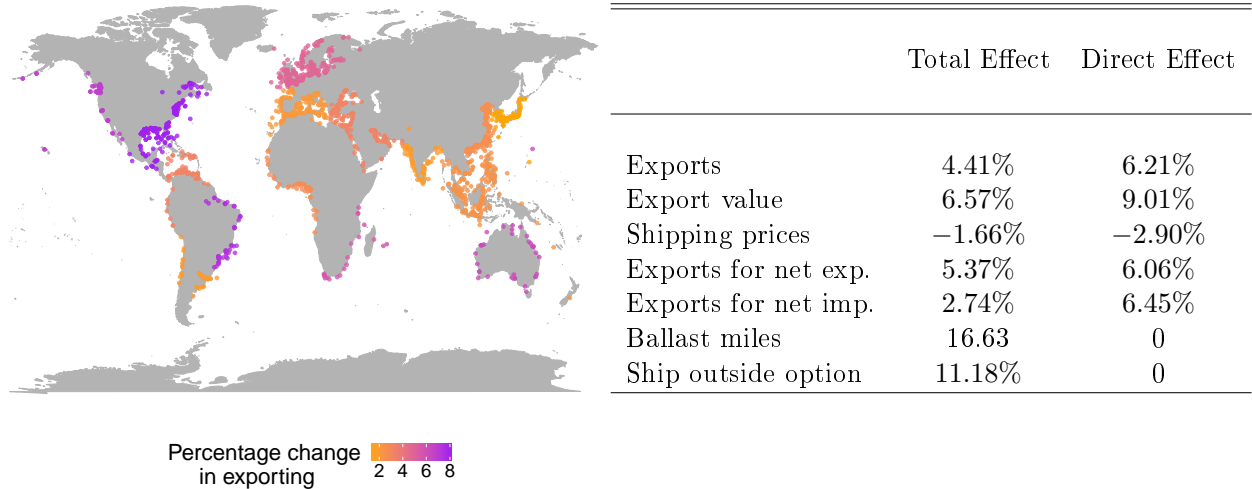
<sup>46</sup>For a more detailed analysis of the impact of fuel costs and ship fuel efficiency see Brancaccio et al. (2018).

<sup>47</sup>In this and all remaining counterfactuals, we compute the steady state spatial equilibrium distribution of ships and exporters. In Appendix F, we provide the computational algorithm employed. The use of nonparametric techniques in the estimation of the matching function may require substantial extrapolation in the counterfactuals; reassuringly, we find that the counterfactual matches and ships are always strictly within the range of our data.

<sup>48</sup>Formally, there is a direct increase in the surplus of all matches, since now a match between a ship and a freight is more valuable. Using the ship and freight value functions, the match surplus is given by

$$S_{ijr} = r - \frac{c_{ij}^s}{1 - \beta(1 - d_{ij})} + \frac{d_{ij}\beta}{1 - \beta(1 - d_{ij})}V_j - U_i - U_{ijr}^e.$$

A decline in  $c_{ij}^s$ , holding everything else constant, directly increases  $S_{ijr}$ . All else equal, this reduces export prices,  $\tau$ , which in turn increases the value of an unmatched exporter,  $U_{ijr}^e$ , and thus induces more entry into the export market.



**Figure 7:** Fuel Cost Shock: impact of a 10% decline in  $c^s$ . The left panel presents the change in exports. In the right panel, the first column presents the total effect of the shock, while the second column presents the direct effect that does not allow ships to optimally adjust their behavior.

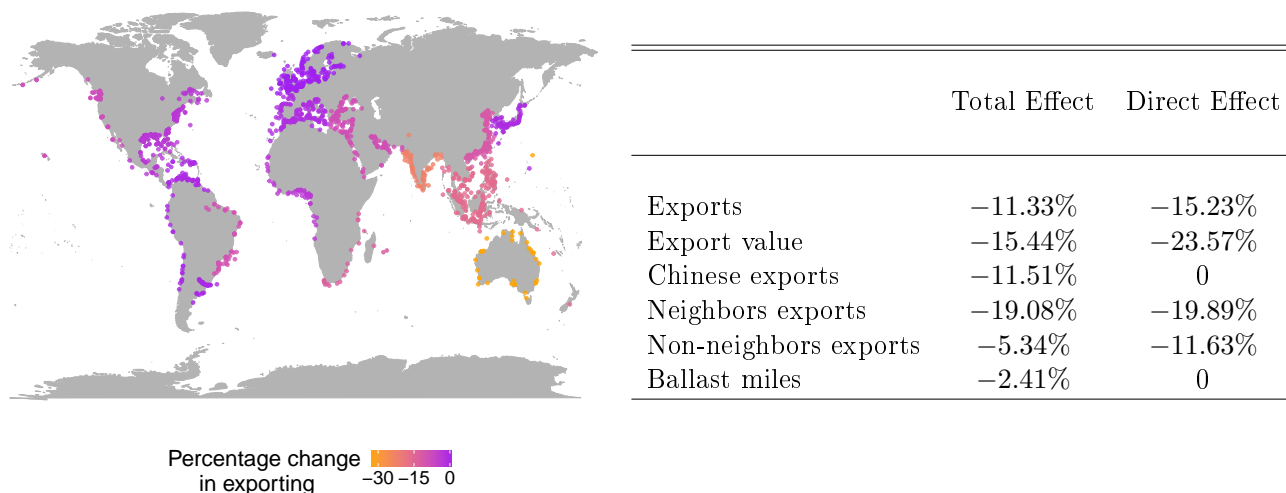
## Chinese Slow-down

Finally, we explore the spatial propagation of a macro shock: a slow-down in China. This experiment showcases how the transport sector creates network effects (mechanism 2), while again misallocating resources (mechanism 1).

We consider a reduction in the revenue of freights going to China,  $\bar{r}_{i,china}$ , by 10%. The left panel of Figure 8 plots the change in exports, while the right panel collects some statistics. We begin by looking at China itself. Chinese exports decline by 11%, even though they are not directly affected by the change in  $\bar{r}_{i,china}$ : the entirety of this decline is driven by the transport sector. Indeed, endogenous trade costs create a complementarity between imports and exports: the high Chinese imports, led to a large number of ships ending their trip in China and looking for a freight there, which in turn reduced trade costs for Chinese exporters (mechanism 1). Therefore, when imports decline, fewer ships end up in China and Chinese exporters are hurt.

Next, note that, as China is a large importer that trades with multiple countries, world exports naturally decline. Indeed, China's large trading partners, such as Australia, Indonesia and Brazil, experience a substantial decline in their exports; total exports decline by 11%. However, in addition to this direct effect, the optimal reallocation of ships over space differentially filters the shock in neighboring vs. distant regions (mechanism 2). Even though distant countries, such as Brazil, also experience a decline in exports, they benefit from the reallocation of ships across space. In particular, prior to the shock, a large fraction of

the fleet was located in the Southeast Pacific region, with ships traveling between China and its neighbors, mainly Australia and Indonesia. Following the shock, these ships reallocate to other parts of the world, pushing up exports there all else equal and dampening the overall decline from the direct effect. To see this, the right panel of Figure 8 shows that if ships were not allowed to optimally adjust their behavior, the decline in exports for distant countries would have been more than double (e.g. Brazilian exports would have fallen by 21% rather than 9%). This underscores the importance of being close to a large net importer like China: exporting countries in that “pocket” of the world, gained not just by directly exporting to China, but also indirectly from the increased supply of ships in that region.<sup>49</sup>



**Figure 8:** Chinese Slow-down: impact of a 10% decline of the revenue from exporting to China. The left panel presents the change in exports. In the right panel, the first column presents the total effect of the slow-down, while the second column presents the direct effect that does not allow ships to optimally adjust their behavior. China’s neighbors include Japan/S.Korea, Australia, Southeast Asia and India. China’s non-neighbors include all other regions.

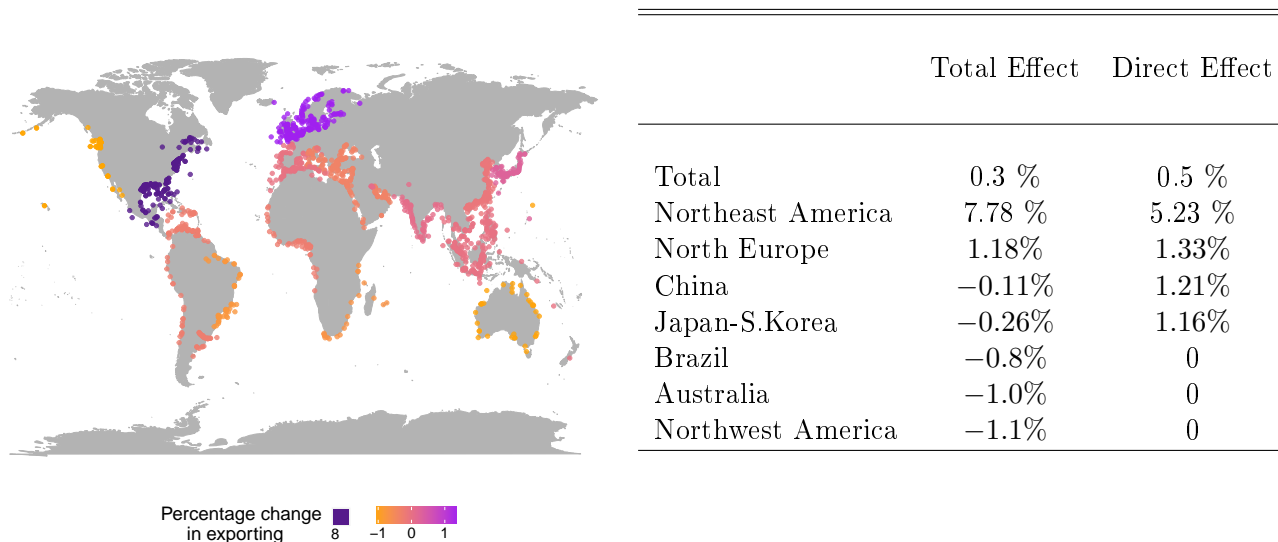
## 8 The Role of Maritime Infrastructure

How much do large maritime infrastructure projects contribute to world trade? We use our estimated setup to address this question by first evaluating a future project, the Northwest Passage. We then examine the impact of four natural and man-made passages: the Panama Canal, the Suez Canal, Gibraltar and the Malacca Straits.

The Northwest Passage is a sea route connecting the northern Atlantic and Pacific Oceans through the Arctic Ocean, along the northern coast of North America. This route is not easily navigable due to

<sup>49</sup>China’s neighbors do not see their exports decline even further, because when Chinese imports decline, ships’ outside options fall everywhere, reducing shipping prices all else equal and dampening the decline in exports (mechanism 3).

Arctic sea ice; with global warming and ice thinning, it is likely that the passage will soon be open for shipping. The opening of the Northwest passage would reduce the distance between East Coast of North America and the Far East, as well as Northern Europe and the Far East.



**Figure 9:** Northwest Passage: The left panel presents the change in exports under the opening of the Northwest passage. In the right panel, the first column presents the total effect of the opening, while the second column the direct effect that does not allow ships to optimally adjust their behavior.

To simulate the impact of this new route, we reduce the nautical distance between Northeast America and Northern Europe to and from China/Japan/S.Korea by 30%.<sup>50</sup> The left panel of Figure 9 presents the resulting change in exports by region, while the right panel collects some statistics.

Not surprisingly, Northeast America sees its exports increase by 8%, while Northern Europe by 1.2%. Interestingly, exports from China and Japan/S.Korea are only marginally affected (0.3% for Japan, and -0.1% for China). On one hand, the import-export complementarity pushes exports up; but on the other hand ballasting is now less costly for ships: when in China or Japan ships can now ballast to the East Coast of North America more cheaply. The ships' higher outside option tends to increase prices and decrease exporting (mechanism 3).

Figure 9 reveals that other countries, not directly affected by the opening of the Northwest Passage, experience changes in their trade. This illustrates how network effects lead to the propagation of local shocks (mechanism 2). For instance, Brazil, Northwest America and Australia see their exports fall by 1%, even though none of these countries are directly affected by this passage. Indeed, ships that used to ballast to these regions now choose to ballast to Northeast America, that experiences a 17% increase in

<sup>50</sup>We calculate this change in travel distance via the Northwest Passage from Ostreng et al. (2013).

ships ballasting there.

	Change in Exports	Max	Min	Most Affected
Suez	-3.33%	3.88%	-25.95%	Middle East
Panama	-4.01%	1.31%	-31.76%	Northeast America
Gibraltar	-6.56%	3.65%	-48.86%	Mediterranean
Malacca	-1.78%	1.61%	-11.21%	Southeast Asia

**Table 8:** Impact of Panama, Suez, Gibraltar and Malacca on world expots. The real and counterfactual maritime routes among regions are calculated using the Dijkstra’s Algorithm. These are combined with ships’ average speed to compute trip duration with and without the corresponding passage.

We also quantify the impact of four existing passages (Panama, Suez, Gibraltar, Malacca) by considering the change in world trade in their absence. These passages reduce nautical distance and thus the duration of specific trips. Table 8 presents the results. All passages substantially increase world trade. Removing the Panama Canal leads a decline in world trade of 4%, but up to 32% in the Northeast America. Removing the Suez Canal reduces trade by 3% and up to 26% in the Middle East. Gibraltar seems to be the most critical one, as removing it would reduce world trade by close to 7% and up to 49% in the Mediterranean. Finally, removing the Malacca Straits would reduce trade by about almost 2% and up to 11% in Southeast Asia.

## 9 Conclusion

In this paper we focus on the importance of the transportation sector in world trade. We build a spatial model where both trade flows and trade costs are equilibrium objects. Different experiments showcase that the transportation sector unveils a new role for geography through three mechanisms: it misallocates productive activities, creates network effects and dampens the impact of shocks. While our empirical implementation focuses on bulk shipping, similar mechanisms are present in most, if not all, modes of transportation. We also demonstrate our setup’s potential to be used for policy evaluation by considering the quantitative impact of maritime infrastructure projects, such as the opening of the Northwest Passage. It is straightforward to use our setup in other counterfactuals, such as tariffs, trade wars, environmental regulations and port infrastructure. Finally, embedding our setup within a general equilibrium framework that endogenizes product prices is an exciting avenue for future research.

# Appendix

## A Construction of Ship Travel Histories and Searching Ships

Here, we describe the construction of ships’ travel histories. The first task is to identify stops that ships make using the EE data. A stop is defined as an interval of at least 24 hours, during which (i) the average speed of the ship is below 5 mph (the sailing speed is between 15 and 20 mph) and, (ii) the ship is located within 250 miles from the coast. A trip is the travel between two stops.

The second task is to identify whether a trip is loaded or ballast. To do so, we use the ship’s draft: high draft indicates that a larger portion of the hull is submerged and therefore the ship is loaded. The distribution of draft for a given vessel is roughly bimodal, since as described in Section 2, a hired ship is usually fully loaded. Therefore, we can assign a “high” and a “low” draft level for each ship and consider a trip loaded if the draft is high (in practice, the low draft is equal to 70% of the high draft). As not all satellite signals contain the draft information, we consider a trip ballast (loaded) if we observe a signal of low (high) draft during the period that the ship is sailing. If we have no draft information during the sailing time, we consider the draft at adjacent stops. Finally, we exclude stops longer than six weeks, as such stops may be related to maintenance or repairs.

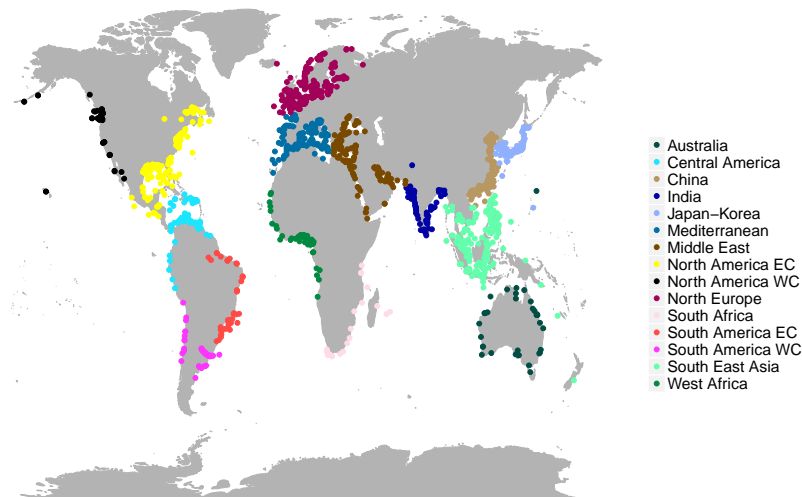
The third and final task is to refine the origin and destination information provided in the Clarksons contracts. Although the majority of Clarksons contracts provide some information on the origin and destination of the trip, this is often vague (e.g. “Far East”, “Japan-S. Korea-Singapore”), especially in the destinations. We use the EE data to refine the contracted trips’ origins and destinations by matching each Clarksons contract to the identified stop in EE that is closest in time and, when possible, location. In particular, we use the loading date annotated on each contract to find a stop in the ship’s movement history that corresponds to the beginning of the contract. For destinations where information in Clarksons is noisy we search the ship’s history for a stop that we can classify as the end of the contract. In particular, we consider all stops within a three month window (duration of the longest trip) since the beginning of the contract. Among these stops we eliminate all those that (i) are in the same country in which the ship loaded the cargo and (ii) are in Panama, South Africa, Gibraltar or at the Suez canal and in which the draft of arrival is the same as the draft of departure (to exclude cases in which the ship is waiting to pass through a strait or a canal). To select the end of the contract among the remaining options we consider the following possibilities:

1. If the contract reports a destination country and if there are stops in this country, select the first of these stops as the end of the trip;
2. If the destination country is “Japan-S. Korea-Singapore”, and if there are stops in either Japan, China, Korea, Taiwan or Singapore, we select the first among these as the end of the trip;
3. If the contract does not report a destination country and there are stops in which the ship arrives full and leaves empty, we select the first of these as the end of the trip.

We check the performance of the algorithm by comparing the duration of some frequent trips, with distances found online (at <https://sea-distances.org>), and find that durations are well matched.

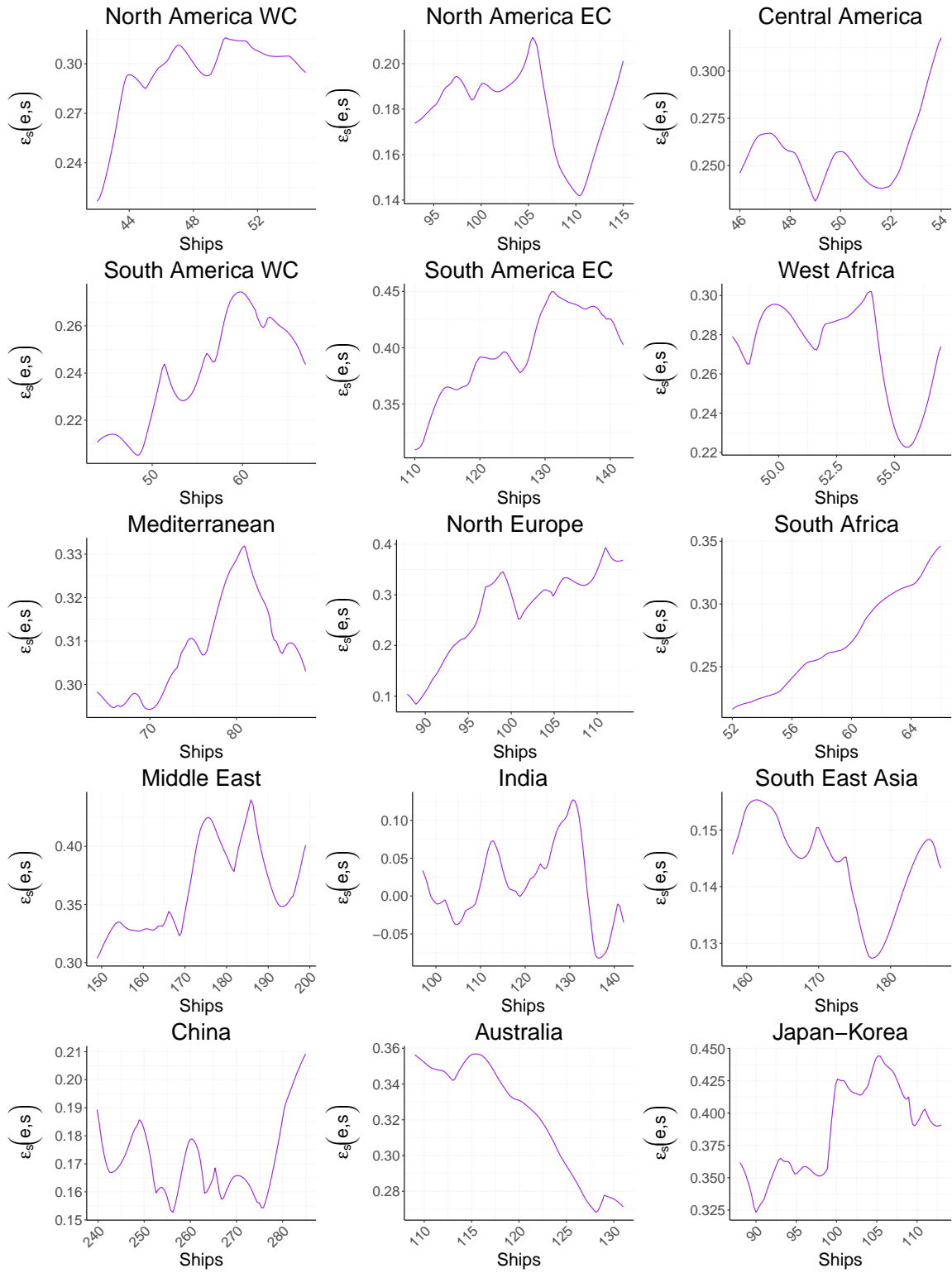
Next, we turn to the construction of searching ships  $s_t = [s_{t1}, \dots, s_{tI}]$  and matches  $m_t = [m_{t1}, \dots, m_{tI}]$ , where  $s_{it}$  denotes the number of ships in region  $i$  and week  $t$  that are available to transport a cargo and  $m_{it}$  the realized matches in region  $i$  and week  $t$ . To construct  $s_{it}$  we consider all ships that ended a trip (loaded or ballast) in region  $i$  and week  $t - 1$ . We exclude the first week post arrival in the region to account for loading/unloading times (on average (un)loading takes 3-4 days but the variance is large; removing one week will tend to underestimate port wait times). To construct  $m_{it}$ , we consider the number of ships that began a loaded trip from region  $i$  in week  $t$ .

## B Additional Figures and Tables

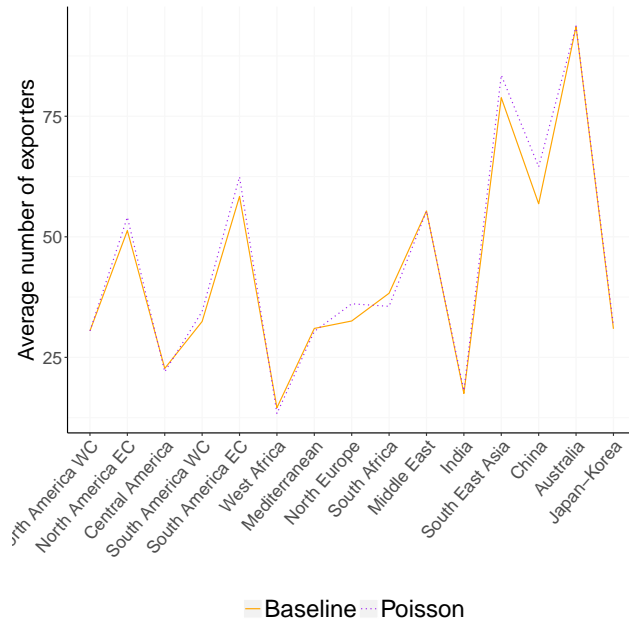


**Figure 10:** Definition of regions. Each color depicts one of the 15 geographical regions.

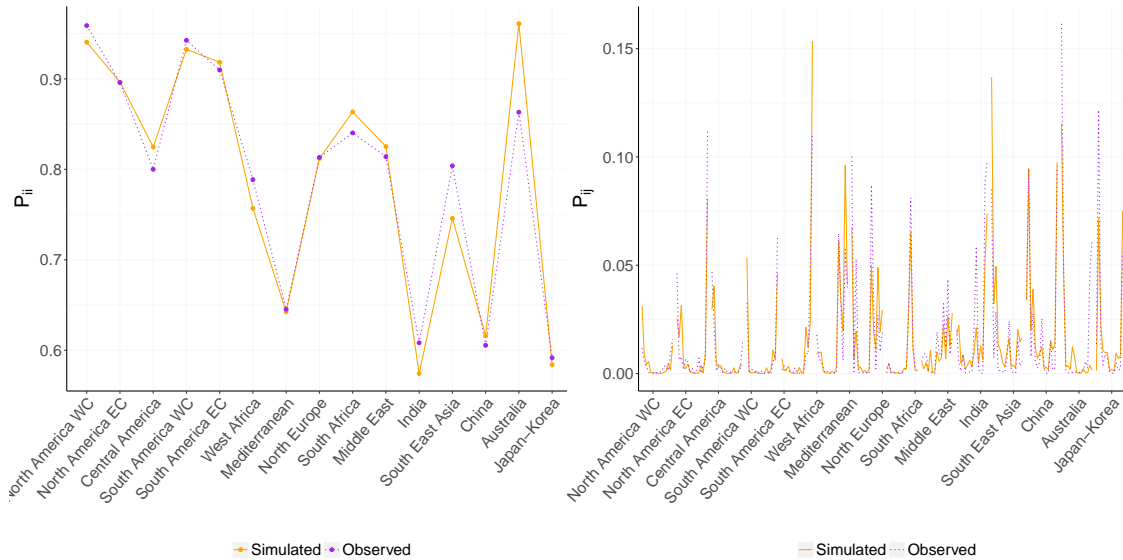




**Figure 11:** Elasticity of the matching function with respect to ships in each region. The range of ships is between the 30th and the 70th percentile of the distribution of ships in the region.



**Figure 12:** Recovered exporters in our baseline specification and under a Poisson distributional assumption.



**Figure 13:** Ballast discrete choice model fit. The left panel depicts the observed and predicted probabilities of staying at port ( $P_{ii}$ ) for all regions  $i$ . The right panel depicts the observed and predicted probabilities of ballasting ( $P_{ij}$ ) to all regions  $i \neq j$ .

	log(price per day)			
	I	II	III	IV
$I \{ \text{orig.} = \text{home country} \}$			0.004 (0.019)	
$I \{ \text{dest.} = \text{home country} \}$			-0.012 (0.015)	
log (Number Employees)				0.008 (0.007)
log (Operating Revenues)				0.003 (0.005)
Time FE	Qtr×Yr	Qtr×Yr	Qtr×Yr	Qtr×Yr
Shipowner FE	No	Yes	No	No
Ship characteristics	Yes	Yes	Yes	Yes
Region FE	Orig. & Dest.	Orig. & Dest.	Orig. & Dest.	Orig. & Dest.
Observations	7,263	7,263	7,973	7,973
Adj. R <sup>2</sup>	0.530	0.540	0.537	0.537

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 9:** Regression of shipping prices on shipowner characteristics and fixed effects. Shipping prices, ships' characteristics (age and size), and the identity of the shipowner are obtained from Clarksons. Information on shipowner characteristics is obtained from ORBIS. In particular, we match the shipowners in Clarksons to ORBIS; we do so for two reasons: (i) ORBIS allows us to have reliable firm identities, as shipowners may appear under different names in the contract data; (ii) ORBIS reports additional firm characteristics (e.g. number of employees, revenue, headquarters). Here we identify the shipowner with the global ultimate owner (GUO); results are robust to controlling for the identity of the domestic owner (DUO) and the shipowner as reported in Clarksons. Finally, the data used span the period 2010-2016.

## C Proof of Proposition 1

We first derive equations (12) and (13). Consider the model's state transitions. Exporters in region  $i$  at time  $t$  transition as follows:

$$e_{it+1} = \delta (e_{it} - m_i(s_{it}, e_{it})) + \mathcal{E}_i (1 - P_{i0}^e) \quad (18)$$

with  $\mathcal{E}_i(1 - P_{i0}^e)$  the (endogenous) flow of new freights. Ships at location  $i$  transition as follows:

$$s_{it+1} = (s_{it} - m_i(s_{it}, e_{it})) P_{ii} + \sum_{j \neq i} d_{ji} s_{jit} \quad (19)$$

In words, out of  $s_{it}$  ships,  $m_{it}$  ships get matched and leave  $i$ , while out of the ships that did not find a match, fraction  $P_{ii}$  chooses to remain at  $i$  rather than ballast away; moreover, out of the ships traveling towards  $i$ , fraction  $d_{ji}$  arrive. Finally, ships that are traveling from  $i$  to  $j$ ,  $s_{ijt}$  evolve as follows:

$$s_{ijt+1} = (1 - d_{ij}) s_{ijt} + P_{ij} (s_{it} - m_i(s_{it}, e_{it})) + \frac{P_{ij}^e}{1 - P_{i0}^e} m_i(s_{it}, e_{it}) \quad (20)$$

In words, fraction  $d_{ij}$  of the traveling ships arrive, fraction  $P_{ij}$  of ships that remained unmatched in location  $i$  chose to ballast to  $j$  and finally,  $P_{ij}^e / (1 - P_{i0}^e)$  of ships matched in  $i$  depart loaded to  $j$ .

Suppose  $s_{ijt}, e_{it}$  approach  $s_{ij}, e_i$  as  $t \rightarrow \infty$ . Then (19) becomes:

$$s_i = (s_i - m_i(s_i, e_i)) P_{ii} + \sum_{j \neq i} d_{ji} s_{ji} \quad (21)$$

while for ships traveling from  $j$  to  $i$ , (20) becomes:

$$s_{ji} = (1 - d_{ji}) s_{ji} + P_{ji} s_j + \left( \frac{P_{ji}^e}{1 - P_{i0}^e} - P_{ji} \right) m_j(s_j, e_j) \quad (22)$$

or

$$d_{ji} s_{ji} = P_{ji} s_j + \left( \frac{P_{ji}^e}{1 - P_{i0}^e} - P_{ji} \right) m_j = P_{ji} (s_j - m_j) + \frac{P_{ji}^e}{1 - P_{i0}^e} m_j$$

where  $m_i = m_i(s_i, e_i)$ . Summing this with respect to  $j \neq i$  we obtain:

$$\sum_{j \neq i} d_{ji} s_{ji} = \sum_{j \neq i} P_{ji} (s_j - m_j) + \sum_{j \neq i} \frac{P_{ji}^e}{1 - P_{i0}^e} m_j$$

and replacing in (21) we get (12).

Equation (13) is a direct consequence of (18).

The steady state equations (12) and (13) have a fixed point over a properly defined subset of  $\mathbb{R}^{2I}$ , by the Leray-Schauder-Tychonoff theorem (Bertsekas and Tsitsiklis, 2015) which states that if  $X$  is a non-empty, convex and compact subset of  $\mathbb{R}^{2I}$  and  $h : X \rightarrow X$  is continuous, then  $h$  has a fixed point. Indeed,

let  $h : \mathbb{R}^{2I} \rightarrow \mathbb{R}^{2I}$ ,  $h = (h^s, h^e)$  with:

$$h_i^s(s, e) = \sum_{j=1}^I P_{ji}(s, e) (s_j - m_j(s_j, e_j)) + \sum_{j \neq i} \frac{P_{ji}^e}{1 - P_{i0}^e} m_j(s, e)$$

$$h_i^e(s, e) = \delta(e_i - m_i(s_i, e_i)) + \mathcal{E}_i \sum_{j \neq 0, i} P_{ij}^e(s, e)$$

for  $i = 1, \dots, I$ . Let  $X = \prod_{i=1}^I [0, \mathcal{E}_i/(1 - \delta)] \times \Delta s$ , where  $\Delta s = \{s_i \geq 0 : \sum_{i=1}^I s_i \leq S\}$ .  $X$  is nonempty, convex and compact, while  $h$  is continuous on  $X$ . We assume that the matching function is such that  $\lambda, \lambda^e$  are zero at the origin and continuous. It remains to show that  $F(X) \subseteq X$ . Let  $(s, e) \in X$ . Then,  $e_i \leq \mathcal{E}_i/(1 - \delta)$  and  $\sum_{i=1}^I s_i \leq S$ . Now,

$$h_i^s(s, e) = \sum_{j=1}^I P_{ji}(s, e) (s_j - \lambda_j(s_j, e_j) s_j) + \sum_{j \neq i} \frac{P_{ji}^e}{1 - P_{i0}^e} \lambda_j(s, e) s_j$$

or

$$h_i^s(s, e) = \sum_{j=1}^I s_j \left[ P_{ji}(s, e) (1 - \lambda_j(s_j, e_j)) + \frac{P_{ji}^e}{1 - P_{i0}^e} \lambda_j(s, e) \right]$$

where let  $P_{ii}^e = 0$  (no inter-region trips). Summing over  $i$  gives:

$$\sum_{i=1}^I h_i^s(s, e) = \sum_{j=1}^I s_j \left[ \sum_{i=1}^I P_{ji}(s, e) (1 - \lambda_j(s_j, e_j)) + \sum_{i=1}^I \frac{P_{ji}^e}{1 - P_{i0}^e} \lambda_j(s, e) \right]$$

or

$$\sum_{i=1}^I h_i^s(s, e) = \sum_{j=1}^I s_j [1 - \lambda_j(s_j, e_j) + \lambda_j(s, e)] \leq S$$

Hence  $h_i^s(s, e) \in \Delta s$ .

Finally, consider  $h^e$ ; since  $m_i \geq 0$ , we have

$$h_i^e \leq \delta e_i + \mathcal{E}_i \sum_{j \neq 0, i} P_{ij}^e(s, e) \leq \delta e_i + \mathcal{E}_i \leq \delta \frac{\mathcal{E}_i}{1 - \delta} + \mathcal{E}_i = \frac{\mathcal{E}_i}{1 - \delta}$$

Hence  $h_i^e(s, e) \in [0, \mathcal{E}_i/(1 - \delta)]$ .

## D Estimation of Ship Costs

Since our model features a number of inter-related value functions  $(V, U)$ , it does not fall strictly into the standard Bellman formulation. Hence, we provide Lemma 2, which proves that our problem is characterized by a contraction map and thus the value functions are well defined.

**Lemma 2.** *For each value of the parameter vector  $\theta \equiv \{c_{ij}^s, c_i^w, \sigma\}$  all  $i, j$ , the map  $T_\theta : \mathbb{R}^I \rightarrow \mathbb{R}^I$ ,  $V \rightarrow T_\theta(V)$  with,*

$$T_\theta(V)_i = -c_i^w + \lambda_i \sum_{j \neq i} G_{ij} \tau_{ij} + \lambda_i \sum_{j \neq i} G_{ij} \left[ -\frac{c_{ij}^s}{1 - \beta(1 - d_{ij})} + \beta d_{ij} \frac{V_j}{1 - \beta(1 - d_{ij})} \right] + (1 - \lambda_i) U_i(\theta, V)$$

where  $\tau_{ij} = E_r \tau_{ijr}$  is the mean price from  $i$  to  $j$  and  $G_{ij} = \frac{P_{ji}^e}{1 - P_{i0}^e}$ , is a contraction and  $V(\theta)$  is the unique fixed point.

*Proof.* Fix  $\theta$ . Let  $\phi_{ij} = \frac{1}{1 - \beta(1 - d_{ij})}$ . The map  $T_\theta(V)$  is differentiable with respect to  $V \in \mathbb{R}^I$  with Jacobian:

$$\frac{\partial T_\theta(V)}{\partial V} = \beta (DG + (I - D)P) \odot Z \quad (23)$$

where  $D$  is a diagonal matrix with  $\lambda_i$  it's  $i$  diagonal entry;  $P$  is the matrix of choice probabilities,  $G$  is the matrix of matched trips,  $Z$  is an  $L \times L$  matrix whose  $(i, j)$  element is  $\phi_{ij} d_{ij}$  and  $\odot$  denotes the pointwise product. We next drop  $\theta$  for notational simplicity; the  $(i, j)$  entry of  $\frac{\partial T}{\partial V}$  is

$$\left( \frac{\partial T}{\partial V} \right)_{ij} = 1 \{i = j\} - \beta \lambda_i G_{ij} d_{ij} \phi_{ij} - (1 - \lambda_i) \frac{\partial U_i}{\partial V_j}$$

Now,

$$\frac{\partial U_i}{\partial V_j} = \frac{1}{e^{\frac{\beta V_i}{\sigma}} + \sum_k e^{\frac{V_{ik}}{\sigma}}} e^{\frac{V_{ij}}{\sigma}} \frac{\partial V_{ij}}{\partial V_j} = \beta P_{ij} d_{ij} \phi_{ij}$$

and thus

$$\left( \frac{\partial T}{\partial V} \right)_{ij} = 1 \{i = j\} - \beta (\lambda_i G_{ij} + (1 - \lambda_i) P_{ij}) d_{ij} \phi_{ij}$$

which in matrix form becomes (23) (as a convention set  $d_{ii} = 1$ ). Let  $H = (DG + (I - D)P) \odot Z$ . Take  $\|H\| = \max_i \sum_j |H_{ij}|$ . Note that  $G, P$  are stochastic matrices and the diagonal matrix  $D$  is positive with entries smaller than 1. Thus  $DG + (I - D)P$  is stochastic. It is also true that  $0 < d_{ij} \phi_{ij} \leq 1$ . Thus,

$$\sum_j |H_{ij}| = \sum_j (\lambda_i G_{ij} + (1 - \lambda_i) P_{ij}) d_{ij} \phi_{ij} \leq \sum_j (\lambda_i G_{ij} + (1 - \lambda_i) P_{ij}) \leq 1$$

and therefore  $\|H\| \leq 1$ . We deduce that  $\|\frac{\partial T_\theta(V)}{\partial V}\| \leq \beta < 1$ . □

In brief, our estimation algorithm proceeds in the following steps:

1. Guess an initial set of parameters  $\{c_{ij}^s, c_i^w, \sigma\}$ .
2. Solve for the ship value functions via a fixed point. Set an initial value  $V^0$ . Then at each iteration  $l$  and until convergence:

- (a) Solve for  $V_{ij}^l$  from:

$$V_{ij}^l = \frac{-c_{ij}^s + d_{ij}\beta V_j^l}{1 - \beta(1 - d_{ij})}$$

- (b) Update  $U^l$  from:

$$U_i^l = \sigma \log \left( \exp \frac{\beta V_i^l}{\sigma} + \sum_{j \neq i} \exp \frac{V_{ij}^l}{\sigma} \right) + \sigma \gamma^{euler}$$

where  $\gamma^{euler}$  is the Euler constant.<sup>51</sup>

- (c) Update  $V_i^{l+1}$  from:

$$V_i^{l+1} = -c_i^w + \lambda_i E_{j,r} \tau_{ijr} + \lambda_i \sum_{j \neq i} \frac{P_{ij}^e}{1 - P_{i0}^e} V_{ij}^l + (1 - \lambda_i) U_i^l$$

where we use the actual average prices from  $i$  to  $j$ , i.e.,  $E_{j,r} \tau_{ijr} = \sum_{j \neq i} \frac{P_{ij}^e}{1 - P_{i0}^e} \tau_{ij}$ . Note that  $\lambda_i$  is known (it is simply the average ratio  $\frac{1}{T} \sum m_{it}/s_{it}$ ). Similarly,  $\frac{P_{ij}^e}{1 - P_{i0}^e}$ , the probability that an exporter ships from  $i$  to  $j$  (conditional on exporting), is obtained directly from the observed trade flows (see Section 5.2).

3. Form the likelihood using the choice probabilities:

$$\mathcal{L} = \sum_i \sum_j \sum_k \sum_t y_{ijkt} \log P_{ij}(c_{ij}^s, c_i^w, \sigma) = \sum_i \sum_j \log P_{ij}(c_{ij}^s, c_i^w, \sigma)^{n_{ij}}$$

where  $y_{ijkt}$  is an indicator equal to 1 if ship  $k$  chose to go from  $i$  to  $j$  in week  $t$ ,  $n_{ij}$  is the number of observations (ship-weeks) that we observe a ship in  $i$  choosing  $j$ , and  $P_{ij}(c_{ij}^s, c_i^w, \sigma)$  are given by (5) and (6).

---

<sup>51</sup>This formula for the ex ante value function  $U_i \equiv E_\epsilon U_i(\epsilon)$  is the closed form expression for the expectation of the maximum over multiple choices, and is obtained by integrating  $U_i(\epsilon)$  over the distribution of  $\epsilon$ .

## E Identification of Ship Port and Sailing Costs

**Proposition 2.** *Given the choice probabilities  $P_{ij}(\theta)$ , the parameters  $\theta = \{\frac{c_{ij}^s}{\sigma}, \frac{c_i^w}{\sigma}, \frac{1}{\sigma}\}$  satisfy a  $(I^2 - I) \times (I^2 + 1)$  linear system of equations of full rank  $I^2 - I$ . Hence,  $I + 1$  additional restrictions are required for identification.*

*Proof.* Let  $\phi_{ij} = \frac{1}{1 - \beta(1 - d_{ij})}$ . The Hotz and Miller (1993) inversion states:

$$\sigma \log \frac{P_{ij}}{P_{ii}} = V_{ij}(\theta) - \beta V_i(\theta)$$

Substituting from (2)-(3) we obtain:

$$\sigma \log \frac{P_{ij}}{P_{ii}} = -\phi_{ij} c_{ij}^s + \beta d_{ij} \phi_{ij} V_j(\theta) - \beta V_i(\theta) \quad (24)$$

It also holds that (see Kalouptsi et al., 2016):

$$\log P_{ij} = \frac{V_{ij}}{\sigma} - \frac{U_i}{\sigma} + \gamma^{euler}$$

or:

$$\sigma \log P_{ij} = -\phi_{ij} c_{ij}^s + \beta d_{ij} \phi_{ij} V_j(\theta) - U_i + \sigma \gamma^{euler} \quad (25)$$

and

$$\sigma \log P_{ii} = \beta V_i(\theta) - U_i + \sigma \gamma^{euler} \quad (26)$$

Now, replace  $V_{ij}$  from (25) into the definition of  $V$ , (3) to get:

$$V_i(\theta) = -c_i^w + \lambda_i \tau_i + \sigma \lambda_i \sum_{j \neq i} G_{ij} \log P_{ij} - \sigma \lambda_i \gamma^{euler} + U_i$$

where  $G_{ij} = \frac{P_{ji}^e}{1 - P_{i0}^e}$  and  $\tau_i \equiv E_{jr} \tau_{ijr} = \sum_{j \neq i} G_{ij} \tau_{ij}$ . Substitute  $U_i$  from (26):

$$V_i(\theta) = -\frac{1}{1 - \beta} c_i^w + \frac{\sigma}{1 - \beta} \left( (1 - \lambda_i) \gamma^{euler} + \lambda_i \sum_{j \neq i} G_{ij} \log P_{ij} - \log P_{ii} \right) + \frac{1}{1 - \beta} \lambda_i \tau_i$$

so that given the CCP's,  $V_i$  is an affine function of  $c_i^w$  and  $\sigma$ . Next, we replace this into the Hotz and



Miller (1993) inversion (24) to obtain:

$$c_{ij}^s = \frac{\beta}{\phi_{ij}(1-\beta)} c_i^w - \frac{\beta}{1-\beta} d_{ij} c_j^w +$$

$$+\sigma \left( \frac{\beta}{1-\beta} \left( d_{ij} \left[ (1-\lambda_j) \gamma^{euler} + \lambda_j \sum_{l \neq j} G_{jl} \log P_{jl} - \log P_{jj} \right] - \frac{1}{\phi_{ij}} \left[ (1-\lambda_i) \gamma^{euler} + \lambda_i \sum_{l \neq i} G_{il} \log P_{il} - \log P_{ii} \right] \right) \right)$$

$$- \frac{\sigma}{\phi_{ij}} \log \frac{P_{ij}}{P_{ii}} + \frac{\beta}{1-\beta} d_{ij} \lambda_j \tau_j - \frac{\beta}{(1-\beta)\phi_{ij}} \lambda_i \tau_i$$

Note that

$$\frac{1}{\phi_{ij}(1-\beta)} = \frac{1-\beta(1-d_{ij})}{1-\beta} = 1 + \frac{\beta d_{ij}}{1-\beta}$$

and set  $\rho_{ij} = \frac{\beta d_{ij}}{1-\beta}$ , then  $\frac{1}{(1-\beta)\phi_{ij}} = 1 + \rho_{ij}$ .

We divide by  $\sigma$ :

$$\frac{c_{ij}^s}{\sigma} = (1 + \rho_{ij}) \frac{c_i^w}{\sigma} - \rho_{ij} \frac{c_j^w}{\sigma} - [\beta(1 + \rho_{ij}) \lambda_i \tau_i - \rho_{ij} \lambda_j \tau_j] \frac{1}{\sigma} +$$

$$+\rho_{ij} \left[ (1-\lambda_j) \gamma^{euler} + \lambda_j \sum_{l \neq j} G_{jl} \log P_{jl} - \log P_{jj} \right] - \beta(1+\rho_{ij}) \left[ (1-\lambda_i) \gamma^{euler} + \lambda_i \sum_{i \neq j} G_{il} \log P_{il} - \log P_{ii} \right]$$

$$- \frac{1}{\phi_{ij}} \log \frac{P_{ij}}{P_{ii}}$$

This is a linear system of full rank in the parameters  $\{\frac{c_{ij}^s}{\sigma}, \frac{c_i^w}{\sigma}, \frac{1}{\sigma}\}$ , since  $\frac{c_{ij}^s}{\sigma}$  can be expressed with respect to  $\{\frac{c_i^w}{\sigma}, \frac{1}{\sigma}\}$ .  $\square$

## F Algorithm for computing the steady state equilibrium

Here, we describe the algorithm employed to compute the steady state of our model to obtain the counterfactuals of Sections 7 and 8.

1. Make an initial guess for  $\{s_i^0, e_i^0, V_i^0\}$  all  $i$ .
2. At each iteration  $l$ , inherit  $\{s_i^l, e_i^l, V_i^l\}$  all  $i$

(a) Update the ship's and exporter's optimal policies by repeating the following steps  $K$  times.<sup>52</sup>

---

<sup>52</sup>K is chosen to accelerate convergence in the spirit of standard modified policy iteration methods.

i. Solve for  $V_{ij}^{l+1}$  from:

$$V_{ij}^{l+1} = \frac{-c_{ij}^s + d_{ij}\beta V_j^l}{1 - \beta(1 - d_{ij})}$$

ii. Update  $U_i^{l+1}$  from:

$$U_i^{l+1} = \sigma \log \left( \exp \frac{\beta V_i^l}{\sigma} + \sum_{j \neq i} \exp \frac{V_{ij}^l}{\sigma} \right) + \sigma \gamma^{euler}$$

iii. Compute the equilibrium prices using

$$\tau_{ijr}^l = \frac{\gamma \left( 1 - \beta \delta \left( 1 - \lambda_i^{e,l} \right) \right)}{1 - \beta \delta \left( 1 - \gamma \lambda_i^{e,l} \right)} \left( U_i^{l+1} - V_{ij}^{l+1} \right) + \frac{(1 - \gamma)(1 - \beta \delta)}{1 - \beta \delta \left( 1 - \gamma \lambda_i^{e,l} \right)} \bar{r}_{ij}$$

iv. Update  $P_{ij}^e$ :

$$P_{ij}^{e,l+1} = \frac{\exp \left( \frac{\beta \delta \lambda_i^e (\bar{r}_{ij} - \tau_{ij}^l)}{1 - \beta \delta (1 - \lambda_i^{e,l})} - \kappa_{ij} \right)}{1 + \sum_{l \neq i} \exp \left( \frac{\beta \delta \lambda_i^e (\bar{r}_{ij} - \tau_{ij}^l)}{1 - \beta \delta (1 - \lambda_i^{e,l})} - \kappa_{il} \right)}$$

v. Update  $V_i^l$ :

$$V_i^{l+1} = -c_i^w + \lambda_i E_{jr} \tau_{ijr} + \lambda_i \sum_{j \neq i} \left( \frac{P_{ij}^{e,l+1}}{1 - P_{i0}^{e,l+1}} \right) V_j^{l+1} + (1 - \lambda_i) U_i^{l+1}$$

vi. Obtain the ships ballast choices  $P_{ij}^{l+1}$ , all  $i, j$ .

3. Update to  $\{\bar{s}^{l+1}, \bar{e}^{l+1}\}$  from:

$$\bar{e}_i^{l+1} = \delta_i \left( e_i^l - m_i^l \right) + \mathcal{E}_i \left( 1 - P_{i0}^{e,l+1} \right)$$

and

$$\bar{s}_i^{l+1} = \sum_j P_{ji}^{l+1} \left( s_j^l - m_j^l \right) + \sum_j \frac{P_{ji}^{e,l+1}}{1 - P_{j0}^{e,l+1}} m_j^l$$

4. If  $\|\bar{s}^{l+1} - s^l\| < \epsilon$ ,  $\|\bar{e}^{l+1} - e^l\| < \epsilon$  and  $\|V^{l+1} - V^l\| < \epsilon$ , stop, otherwise update freights and ships

as follows:

$$\begin{aligned}s^{l+1} &= \alpha s^l + (1 - \alpha) \tilde{s}^{l+1} \\ e^{l+1} &= \alpha e^l + (1 - \alpha) \tilde{e}^{l+1},\end{aligned}$$

where  $\alpha$  is a smoothing parameter.

## References

- ALLEN, T. AND C. ARKOLAKIS (2014): “Trade and the Topography of the Spatial Economy,” *Quarterly Journal of Economics*, 129, 1085–1140.
- (2016): “The Welfare Effects of Transportation Infrastructure Improvements,” *mimeo*, Yale University.
- ANDERSON, J. E. AND E. VAN WINCOOP (2003): “Gravity with Gravitas: A Solution to the Border Puzzle,” *American Economic Review*, 93, 170–192.
- ARKOLAKIS, C., N. RAMONDO, A. RODRIGUEZ-CLARE, AND S. YEAPLE (2018): “Innovation and Production in the Global Economy,” *American Economic Review*, 108, 2128–2173.
- ASTURIAS, J. (2018): “Endogenous Transportation Costs,” *mimeo*, School of Foreign Service in Qatar, Georgetown University.
- BAJARI, P. AND C. L. BENKARD (2005): “Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach,” *Journal of Political Economy*, 113, 1239–1276.
- BAJARI, P., C. L. BENKARD, AND J. LEVIN (2007): “Estimating Dynamic Models of Imperfect Competition,” *Econometrica*, 75, 1331–1370.
- BERRY, S. (1994): “Estimating Discrete-Choice Models of Product Differentiation,” *RAND Journal of Economics*, 25, 242–262.
- BERTSEKAS, D. P. AND J. N. TSITSIKLIS (2015): *Parallel and Distributed Computation: Numerical Methods*, Athena Scientific.
- BRANCACCIO, G., M. KALOUPSIDIS, AND T. PAPAGEORGIOU (2018): “The Impact of Oil Prices on World Trade,” *mimeo*, Harvard University.
- BRANDT, L., J. V. BIESEBROECK, L. WANG, AND Y. ZHANG (2017): “WTO Accession and Performance of Chinese Manufacturing Firms,” *American Economic Review*, 107, 2784–2820.

- BUCHHOLZ, N. (2018): "Spatial equilibrium, search frictions and efficient regulation in the taxi industry," *mimeo, Princeton University*.
- BURDETT, K. AND D. T. MORTENSEN (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39, 257–273.
- CALIENDO, L. AND F. PARRO (2015): "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 82, 1–44.
- COLLARD-WEXLER, A. (2013): "Demand Fluctuations in the Ready-Mix Concrete Industry," *Econometrica*, 81, 1003–1037.
- COSAR, A. K. AND B. DEMIR (2018): "Shipping inside the Box: Containerization and Trade," *Journal of International Economics*, 114, 331–345.
- CRAWFORD, G. S. AND A. YURUKOGLU (2012): "The Welfare Effects of Bundling in Multichannel Television Markets," *American Economic Review*, 102, 643–685.
- DONALDSON, D. (2012): "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," *forthcoming, American Economic Review*.
- DONALDSON, D. AND R. HORNBECK (2016): "Railroads and American Economic Growth: A "Market Access" Approach," *Quarterly Journal of Economics*, 131, 799–858.
- EATON, J., D. JINKINS, J. TYBOUT, AND D. XU (2016): "Two-sided Search in International Markets," *mimeo, Penn State University*.
- EATON, J. AND S. KORTUM (2002): "Technology, Geography and Trade," *Econometrica*, 70, 1741–1779.
- ERICSON, R. AND A. PAKES (1995): "Markov-Perfect Industry Dynamics: A Framework for Empirical Work," *The Review of Economic Studies*, 62, 53–82.
- FAJGELBAUM, P. D. AND E. SCHAAL (2017): "Optimal Transport Networks in Spatial Equilibrium," *NBER working paper*, 23200.
- FRECHETTE, G. R., A. LIZZERI, AND T. SALZ (2018): "Frictions in a Competitive, Regulated Market Evidence from Taxis," *mimeo, New York University*.
- HEAD, K. AND T. MAYER (2004): "The Empirics of Agglomeration and Trade," in *Handbook of Regional and Urban Economics*, ed. by J. V. Henderson and J. F. Thisse, Elsevier, vol. 4.
- HO, K. AND R. S. LEE (2017): "Insurer Competition in Health Care Markets," *Econometrica*, 85, 379–417.
- HOLMES, T. J. AND E. SINGER (2018): "Indivisibilities in Distribution," *NBER working paper*, 24525.

- HOPENHAYN, H. A. (1992): "Entry, Exit, and Firm Dynamics in Long Run Equilibrium," *Econometrica*, 60, 1127–1150.
- HOSIOS, A. J. (1990): "On the Efficiency of Matching and Related Models of Search and Unemployment," *The Review of Economic Studies*, 57, 279–298.
- HOTZ, V. J. AND R. A. MILLER (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.
- HUMMELS, D. (2007): "Transportation Costs and International Trade in the Second Era of Globalization," *The Journal of Economic Perspectives*, 21, 131–154.
- HUMMELS, D., V. LUGOVSKYY, AND S. SKIBA (2009): "The Trade Reducing Effects of Market Power in International Shipping," *Journal of Development Economics*, 89, 84–97.
- HUMMELS, D. AND S. SKIBA (2004): "Shipping the Good Apples Out? An Empirical Confirmation of the Alchian-Allen Conjecture," *Journal of Political Economy*, 112, 1384–1402.
- IMBENS, G. W. AND W. K. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512.
- ISHIKAWA, J. AND N. TARUI (2015): "Backfiring with Backhaul Problems: Trade and Industrial Policies with Endogenous Transport Costs," *mimeo, University of Hawaii*.
- KALOUPTSIDI, M. (2014): "Time to Build and Fluctuations in Bulk Shipping," *American Economic Review*, 104, 564–608.
- (2018): "Detection and Impact of Industrial Subsidies, the Case of Chinese Shipbuilding," *Review of Economic Studies*, 85, 1111–1158.
- KALOUPTSIDI, M., P. T. SCOTT, AND E. SOUZA-RODRIGUES (2016): "Identification of Counterfactuals in Dynamic Discrete Choice Models," *NBER working paper*, 21527.
- KROLIKOWSKI, P. M. AND A. H. MCCALLUM (2018): "The Role of Trade Costs in the Surge of Trade Imbalances," *Federal Reserve Bank of Cleveland, Working Paper no. 16-35R*.
- KRUGMAN, P. (1991): "Increasing Returns and Economic Geography," *The Journal of Political Economy*, 99, 483–499.
- LAGOS, R. (2000): "An Alternative Approach to Search Frictions," *The Journal of Political Economy*, 108, 851–873.

- (2003): “An Analysis of the Market for Taxicab Rides in New York City,” *International Economic Review*, 44, 423–434.
- MATZKIN, R. L. (2003): “Nonparametric Estimation of Nonadditive Random Functions,” *Econometrica*, 71, 1339–1375.
- MORTENSEN, D. T. (2003): *Wage Dispersion: Why Are Similar Workers Paid Differently?*, Cambridge, MA: MIT Press.
- MORTENSEN, D. T. AND C. A. PISSARIDES (1994): “Job Creation and Job Destruction in the Theory of Unemployment,” *The Review of Economic Studies*, 61, 397–415.
- OSTRENG, W., K. M. EGER, B. FLOISTAD, A. JORGENSEN-DAHL, L. LOTHE, M. MEJLAENDER-LARSEN, AND T. WERGELAND (2013): *Shipping in Arctic Waters: A Comparison of the Northeast, Northwest and Trans Polar Passages*, Berlin, Heidelberg: Springer.
- PAKES, A., M. OSTROVSKY, AND S. T. BERRY (2007): “Simple Estimators for the Parameters of Discrete Dynamics Games,” *The RAND Journal of Economics*, 38, 373–399.
- PANAYIDES, P. M. (2016): *Principles of Chartering*, CreateSpace Independent Publishing Platform, 2nd ed.
- PETRONGOLO, B. AND C. A. PISSARIDES (2001): “Looking Into the Black Box: A Survey of the Matching Function,” *Journal of Economic Literature*, 39, 390–431.
- POSTEL-VINAY, F. AND J.-M. ROBIN (2002): “Equilibrium Wage Dispersion with Worker and Employer Heterogeneity,” *Econometrica*, 70, 2295–2350.
- ROGERSON, R., R. SHIMER, AND R. WRIGHT (2005): “Search-Theoretic Models of the Labor Market: A Survey,” *Journal of Economic Literature*, 43, 959–988.
- RUST, J. (1987): “Optimal replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher,” *Econometrica*, 55, 999–1033.
- RYAN, S. P. (2012): “The Costs of Environmental Regulation in a Concentrated Industry,” *Econometrica*, 80, 1019–1062.
- SAMUELSON, P. A. (1954): “The Transfer Problem and Transport Costs, II: Analysis of Effects of Trade Impediments,” *The Economic Journal*, 64, 264–289.
- SIMONOVSKA, I. AND M. E. WAUGH (2014): “The Elasticity of Trade: Estimates and Evidence,” *Journal of International Economics*, 92, 34–50.

- THE WORLD BANK (2012): “Road Freight in Central America: Five Explanations for High Costs of Service Provision,” Tech. rep., The World Bank, Washington, DC.
- UNCTAD (2015): “Review of Maritime Transport,” Tech. rep., United Nations Conference on Trade and Development, Geneva, Switzerland.
- WAUGH, M. (2010): “International Trade and Income Differences,” *American Economic Review*, 100, 2093–2124.
- WEINTRAUB, G. Y., C. L. BENKARD, AND B. V. ROY (2008): “Markov Perfect Industry Dynamics with Many Firms,” *Econometrica*, 76, 1375–1411.
- WONG, W. F. (2018): “The Round Trip Effect: Endogenous Transport Costs and International Trade,” *mimeo, University of Oregon*.
- WTO (2015): “International Trade Statistics,” Tech. rep., World Trade Organization, Geneva, Switzerland.